Hands on Music: Physical Approach to Interaction with Digital Music

Janne Bergman, Jarmo Kauko, Jaakko Keränen
Nokia Research Center
P.O. Box 1000
33721 Tampere, Finland
{janne.bergman, jarmo.kauko, jaakko.keranen}@nokia.com

ABSTRACT
Mobile users listen to large digital music libraries with thousands of songs. Browsing such libraries in mobile contexts is difficult due to constraints of the context and devices. We explore the usage of physical interaction with digital music to overcome these limitations. Our solution is to utilize the physical orientation of a mobile device as a tool for exploring music. We focus on allowing users to manage their music for easier mobile access. We present a novel bimanual interaction method for mapping items from a music library into different orientations of the mobile device. An experiment was conducted to test our prototype, focusing on efficiency and impact on users' ability to recall locations of items in the mapping. The interaction method was found to be significantly faster than using a touch screen without the mobile device. Subjectively users valued the bimanual method in ease of use, efficiency, and pleasantness.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User interfaces; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information

General Terms
Design, Experimentation, Human Factors.

Keywords
Bimanual interaction, graspable interfaces, interaction techniques, mobile music.

1. INTRODUCTION
Today people have personal digital music libraries that may contain thousands of songs, and have access to internet music services that house millions of songs. While portable music players are popular for the consumption of music, managing and exploring a music library typically occurs on a PC. The restrictions inherent in mobile user interfaces make it inconvenient to browse large libraries using a mobile device. In addition, the increased cognitive load and attentional resource fragmentation in mobile contexts pose further challenges [16].

In the past, people have had collections of vinyl records and CDs, which can be organized and manipulated physically. People naturally understand and handle physical items in 3D space with two hands [5, 18]. Nowadays, however, digital music lacks physical manifestation. Our work is motivated by the potential of physicality – whether concrete, spatial interaction is beneficial with digital music.

Previously we developed Gravity Sphere [10] for tackling issues in mobile music exploration and playlist creation. In this paper, we extend Gravity Sphere toward a complete music solution that takes advantage of our innate spatial skills and ability to interact with physical objects. The system integrates a mobile device with a PC in a novel way that allows consistent usage in both mobile and desktop contexts.

The paper is structured as follows. Section 2 takes a look at the theoretical background and solutions in related work. Section 3 summarizes our previous work on Gravity Sphere. In Section 4, we present our concept design, novel bimanual interaction method and prototype implementation of the system. In Section 5, we describe an experiment to compare alternative interaction methods for the system, and discuss the results of the experiment. Section 6 concludes our work and its implications.

2. RELATED WORK
In this section, we first discuss theoretical background of tangible user interfaces. Then we review some earlier solutions that have been developed for or can be applied to managing music collections on mobile devices.

2.1 Theoretical Background
The concept of tangible user interfaces (TUIs) was originally introduced as graspable user interfaces by Fitzmaurice et al. [6]. Fitzmaurice [5] defined a graspable user interface as "a physical handle to a virtual function where the physical handle serves as a dedicated functional manipulator". Our concept is based on general-purpose computing devices rather than dedicated devices or everyday objects. However, it realizes the definition by utilizing a mobile device as a physical handle to a virtual function, and thus partially shares the same design space with graspable and tangible user interfaces.

Sharin et al. [18] examine human spatial skills and present three heuristics for incorporation of spatiality into TUI application

MobileHCI'09, September 15 - 18, 2009, Bonn, Germany.
ACM 978-1-60558-281-8.

design. The first heuristic concerns intuitive spatial mappings to the application task, exploiting spatial abilities and mappings that are known innately or learned early in life. The second heuristic is I/O unification, coupling of action and perception space and embodying a clear representation of application state across all sensory modalities. The third heuristic suggests that exploratory or epistemic activity should be supported in addition to goal-oriented or pragmatic activity, enabling low-cost "trial and error" learning.

Fitzmaurice [5] discusses how graspable user interfaces naturally encourage two-handed interaction. Guiard [7] presented a Kinematic Chain (KC) model for asymmetric bimanual interaction. Many bimanual interface designs have been guided by the three principles of the KC model, as summarized in [3]:

1. Dominant-to-Non-Dominant Spatial Reference: The nondominant hand sets the frame of reference relative to which the dominant hand performs its motions.
2. Asymmetric Scales of Motion: The two hands operate in asymmetric spatial-temporal scales of motion. For instance, when writing on a piece of paper, the motion of the non-dominant hand controlling the position of the paper is of lower temporal and spatial frequency than the writing movements of the dominant hand which nonetheless depends on the non-dominant hand's movement for spatial reference.
3. Precedence of the Non-Dominant Hand: Contribution of the non-dominant hand to a cooperative bimanual task starts earlier than the dominant hand. In the handwriting example, the dominant hand starts writing after the paper has been oriented and positioned by the non-dominant hand.

According to the third principle of the KC model, in asymmetric tasks the movement of the hands is sequential rather than parallel. However, the ability to use both hands doubles the available degrees of freedom and often reduces the amount of physical movement required to perform a task. In addition, Leganchuk et al. propose that appropriately designed bimanual interaction reduces cognitive load by allowing cognitive unit tasks to be combined and planned as a chunk [11]. Bimanual interaction for 3D camera control and object manipulation tasks is studied in [3], providing a reference point to our bimanual design.

Our design is also influenced by previous work in 3D rotation techniques. ArcBall is a mouse-controlled 3D rotation technique based on mathematically rigorous quaternion implementation, resulting in simple and consistent behavior [19]. More recent analysis and comparison of mouse-controlled techniques is presented in [2], showing that ArcBall is still one of the best techniques. It was outperformed only by Two-Axis Valuator technique, which is less attractive for tasks involving rotation around the view vector. Hinckley et al. [9] compare mouse-controlled rotation techniques to physical rotation devices based on magnetic orientation sensors. The study showed that physical rotation devices were 33%–36% faster than ArcBall, without any significant loss of accuracy.

2.2 Previous Solutions

Some previous work has studied new interaction techniques for music applications, but only little in respect to mobile usage. Hiipakka et al. have studied spatial audio interface for creating playlists in mobile contexts [8]. The interface allows navigating typical genre-artist-album-song hierarchies using four keys, and item selection to a playlist is done with a separate key. Eye-free usage is enabled by providing spatial text-to-speech and non-speech audio feedback, along with music previews. However, the tasks are directly adopted from desktop interfaces and require considerable effort and concentration. Mobile aspects were not covered in the evaluation of the system.

Pauws et al. present a similar multimodal approach to playlist creation [17]. A force-feedback trackball was used to navigate a music hierarchy visualized using four rollers. Tactile and audio feedback allowed the system to be used without a visual display. In this study, the system was designed for the home environment and it does not consider mobile context aspects.

In general, any mobile approach to list navigation can be utilized for music navigation and playlist composition. Tilt based interfaces for list navigation are studied in [15]. An absolute mapping of device orientation to the list position is presented and shown to outperform rate based mapping in relatively short lists. Non-visual list navigation based on circular touchpad and spatial audio output is studied in [20]. Segments of the touchpad are directly mapped to list items, allowing smooth and natural learning path from long radial navigation towards short direct selections. Short mechanical click sounds and interruptible audio increase reactivity. In contrast, our system utilizes absolute mapping of 3D orientation, and is used for exploring large music collections rather than making selections in short lists. As in [15] and [20], immediate audio or tactile feedback and the potential learnability of absolute mappings are important factors in our work.

Recently a popular trend has been to utilize algorithmic feature extraction to visualize and interact with music content. Neumayer et al. [14] use a self-organizing map (SOM) to allow creating playlists by drawing shapes on the map. A graphical interface is presented also for mobile usage. Following the same trend, a spherical visualization technique called Globe of Music is presented in [12]. A spherical SOM is used to map music on a sphere based on similarity of features. Our work uses a similar mapping technique to enable non-visual usage in mobile contexts.

Ängeslevä et al. [21] introduce a concept and design strategies for mapping information to body-space, accessed by moving a mobile device to different locations around one's body. As our system, their concept relies on proprioception and vibrotactile feedback to ease cognitive load and minimize need for visual attention. However, we utilize a continuous mapping of orientations instead of distinct locations in body-space.

As music libraries are typically stored on a PC, important features of a portable music device are interoperability and integration to other devices. The idea of using mobile devices in combination with desktop PCs and other interaction devices is investigated in [13]. They utilize a touch screen enabled PDA in the non-dominant hand to support standard GUIs operated with a mouse in the dominant hand. We extend this idea by designing a multi-device system based on Guiard's KC model to better support bimanual interaction.

A tangible music player called MusicCube is presented in [1]. Four predefined playlists are mapped to four sides of a cube. One of the playlists can be selected by placing the corresponding side of the cube upwards. One side is reserved for controls. The system also utilizes shake gestures, text-to-speech song descriptions, and rhythm visualizations using LED lights. Although not designed for exploring entire music libraries, MusicCube demonstrates the potential of TUIs for music consumption.

3. MOBILE MUSIC EXPLORATION

In our previous work, we concentrated in enabling eyes-free exploration of large music collections and creating quick on-the-go playlists in mobile contexts. This chapter describes our mobile music concept, prototype implementation and key findings from the usability evaluations.

3.1 The Concept

Consumer use of photos and music is studied in [4], suggesting that systems should be designed to support serendipitous browsing of content rather than goal-oriented search only. In our previous work in mobile music exploration, we have looked into addressing serendipitous aspects such as sidetracking, opportunity to rediscover forgotten music, and settling for adequate but non-optimal results. Our solution to supporting these factors is based on physical manipulation of the body of a mobile device [10]. Figure 1 illustrates the idea. The music library that is being explored has been mapped onto a virtual 3D surface around the mobile device. An accelerometer within the device is used for determining the direction of gravity in the device's frame of reference. This direction is used as a cursor for locating songs in the library. Since music libraries may contain thousands of songs, there is typically more than one song near the cursor. As the user rotates the device in 3D, brief samples of songs near the cursor are played. Finally, once suitable music has been found, the user can lock the cursor to keep playing music from the current orientation of the device. Locking can be done for instance with a spatial gesture such as shaking the device. This kind of physical manipulation of the device is well suited for non-visual use, as it does not require looking at the device or changing grip on it.

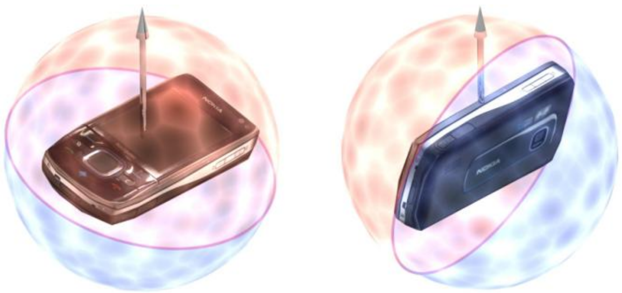

Figure 1. Mobile exploration of a 3D music mapping.

Due to lack of visual modality, tactile feedback is used for providing more information about how the system reacts to user actions. When the device is held steady, the only feedback the user receives is the brief song samples. However, while he is rotating the device, vibration pulses are used for communicating how the orientation is changing. This is comparable to tactile feedback in [15] and audio feedback in [20]. However, our system did not use audio clicks because they would have interfered with the music. In addition, the tactile feedback from the device contributes to action-perception coupling, strengthening I/O unification [18].

The system has a strong random element, which suits the serendipitous nature of the task. Due to its limited resolution and noise from external factors, the interaction technique is not suitable for finding a particular song in the library. Instead, the system is designed for quickly locating a set of songs that fulfill the user's wishes. For example, the user might wish to hear relaxing music while sitting in a bus – a specific song in this case is not important. To achieve this, he takes out his mobile phone, turns it on its left side (he recalls it may contain something suitable), listens to a couple of the song samples while making fine adjustments to the orientation, and then picks the first song that he likes well enough. This task can be accomplished without visual attention, so it can be performed in more demanding contexts, too, for instance while walking.

The mapping of the songs in the library is done in such a way that similar songs are placed close to each other on the surface. This allows the user to explore the collection in a more coherent manner, as the randomness is reduced in the songs played by the system. There are different ways to determine song similarity, e.g., content based analysis of audio features [12, 14], song metadata, and statistical and user-submitted data from online music stores and communities (such as Last.fm). Once some measure of similarity between songs exists, the songs can be automatically arranged on a sphere surface for instance using a spherical SOM algorithm [12].

3.2 The Prototype

This section briefly presents the design and implementation of our solution for mobile music exploration called Gravity Sphere [10]. Our prototype was implemented in python on a Nokia N95, which has a built-in accelerometer. The music library was stored on a memory card in the phone, containing hundreds of songs.

The application was provided a pre-calculated music mapping that defined 3D coordinates on a unit sphere for each song in the music library. The application continuously read the accelerometer sensor output at a rate of approximately 30Hz, filtered it by averaging the three latest samples. The resulting vector was normalized and used to select the nearest song on the unit sphere.

A playlist of 20 songs most similar to the currently playing song could be created with one of three different actions: shaking the phone, pressing the Select button, or closing the N95's two-way slide. During browsing, the two-way slide was open with the media keys exposed so that closing it would require a short sliding motion.

We experimented with different forms of visual feedback that would help novice users understand the concept. These included textual instructions about what to do with the device, and a visualization of the device's orientation as a cue for the user to see what effect turning the device has. We decided not to use any complex graphical feedback, as one of the main design drivers was eyes-free usage. However, our visual cues were insufficient in helping the users to understand the system without instructions.

The application responded to a change in the orientation of the device after a minimum change of α_{min} degrees. The angle was defined as the difference between the current orientation and the last orientation that triggered a song change. While the user was walking, the accelerometer was subject to a lot of noise due to the user's body movements. To accommodate this we implemented a simple adaptation algorithm that would increase the value of α_{min} to a maximum of 45º whenever large noise was detected and decrease it to a minimum of 4º as the noise reduced. This allowed the user to walk normally and not cause inadvertent song changes due to the motions of walking.

Haptic feedback in the form of vibration pulses was used to indicate when α_{min} threshold was crossed. These provided immediate feedback, unlike song changes, which due to technical limitations suffered from a file loading latency of 0.5 – 0.8 seconds.

Gravity Sphere also kept a history of the six last played songs. Songs in the history were prioritized over other songs in the vicinity of the current orientation vector to allow the user to return to a recently played song sample. This was also intended to help the user to realize that there is a fixed mapping between orientations and corresponding songs.

3.3 Key Findings from Usability Evaluations

Gravity Sphere was developed in two iterations. Formative usability evaluations were performed to gather guidelines for improving the design of the prototype. Evaluations were conducted for both sitting and walking conditions to study differences between them. Users were unfamiliar with the music library and the mapping. The execution and results of the studies are presented in [10].

Users' comments and questionnaire results indicated that manipulating Gravity Sphere felt more natural while walking than while sitting. The users also appreciated the vibration pulses because they could sense when the phone had been turned enough for the system to react.

The orientation of the phone was continuously recorded during the test sessions. In our data analysis, we discovered that in majority of the orientations one of the six sides of the phone was roughly facing upwards. We observed that users tended to perform tasks in terms of these six primary orientations. However, when it came to the exploration task where the user was expected to rotate the phone freely to discover songs, a rigid song mapping that placed each genre of music on its own side of the phone was a deterrent to free exploration. The rigid genre-based mapping was easier for the users to recognize and remember, though.

One of the largest issues with our implementation of Gravity Sphere was that users were not initially familiar with the song mapping, and thus were forced to keep randomly turning the device to make sense out of the song placement. This led to some users treating the system as a shuffle feature. We concluded that a potential remedy would be the possibility for the users to personalize the mapping. The user would then know the location of each style of music and thus use the system more effectively. Personalization of the mapping was also among the most popular improvement ideas suggested by the users themselves. They expressed willingness to use multiple mappings, and commented that trying out different mappings would make the exploration more fun.

4. HANDS ON MUSIC

In this section, we expand the Gravity Sphere concept toward a complete music application. We focus on music management with a PC and personalization of music mappings for mobile usage. We also discuss the bimanual interaction method employed in our prototype and its implementation.

Automatically generated mappings can be applied to any music library. However, to facilitate remembering the mapping and utilizing the orientations efficiently, the system should allow users to define personalized mappings. In those mappings, the orientation of certain genres, albums, artists, or songs is defined manually. In this way, favorite items can be placed in orientations that can be easily recalled and accessed. The rest of the songs in the music collection can be automatically mapped to fill the gaps between the manually placed items.

Suitable granularity of the mapping is likely to depend on the experience of the user with the system. In our formative usability evaluations, we observed that inexperienced users were confused about the complexity of the mapping. A novice user might instead opt to start out with a very simple mapping containing for example only two meaningful orientations: up and down. After becoming more familiar with the system, the user can assign new music into other orientations, thus building up the mental model of the mapping over time. This suggests that editing and augmenting the mapping should be fast and convenient.

4.1 Concept Design

The concept was designed to allow a user to concretely manipulate the placement of songs, albums, artists, and genres in the music mapping. The system consists of two parts, a mobile device and a tablet PC as shown in Figure 2. The tablet has a GUI for browsing the music collection and a 3D representation of the mobile device and the sphere, whose orientation matches that of the mobile device.

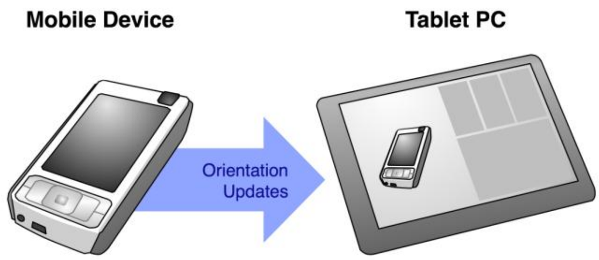

Figure 2. The system is composed of two parts: a mobile device and a tablet PC.

4.1.1 Music Library View

Music library view (right side of the tablet PC display in Figure 3) is used to browse music stored in the tablet PC. The hierarchical navigation model was adopted from Apple iTunes due to its popularity and simplicity. The three topmost lists are used to browse genres, artists and albums. The bottom list is reserved for songs and metadata. Lists can be scrolled by directly dragging on the list or by using the alphabetical scrollbar, adopted from Apple iPod touch.

We experimented with different methods for adding music items from the library view to the phone view. Drag and drop allows users to select an item and then drag it over the phone. Releasing the item outside the phone cancels the operation. A variation of the drag and drop was tapping an item and then specifying the orientation by tapping on the phone. Double tapping was designed for quickly adding an item to the current orientation. Informal evaluation showed that drag and drop was discovered by the users without any instructions. It is a well-known technique and supports our goal of physically placing music items on the phone. Drag and drop was selected for this experiment, although double tapping technique is likely to be beneficial for expert users, as it requires smaller amount of hand movement.

4.1.2 Phone View

Phone view (left side of the tablet PC display in Figure 3) shows a 3D model of the phone viewed from above. The orientation of the model represents the orientation of the phone in mobile usage. There is a map surface around the phone representing the music mapping. Once items are placed onto the mapping, they appear as dots on the surface. Visual cues superimposed on the surface indicate the different sides of the phone.

4.1.3 Bimanual Interaction Method

Our design goal was to enable fast and convenient creation and modification of music mappings. This would allow users to define personal mappings that are comfortable to use in mobile contexts, without burdening users with a laborious configuration tool. Another design goal was to allow consistent usage in mobile and desktop contexts, minimizing learning requirements and reinforcing users' mental models to help non-visual usage in mobile contexts.

In order to fulfill the design goals, we designed a novel interaction technique based on bimanual usage of an orientation-sensing mobile device and a graphical touch UI. Similarly, to physically manipulating a pincushion, the non-dominant hand rotates the mobile phone while the dominant hand manipulates the music items. The orientation of the 3D model is synchronized with the real phone. Usage of the interaction method is shown in Figure 3, where the user holds the phone in his left hand while using his right hand to drag items from the music library browser onto the 3D surface.

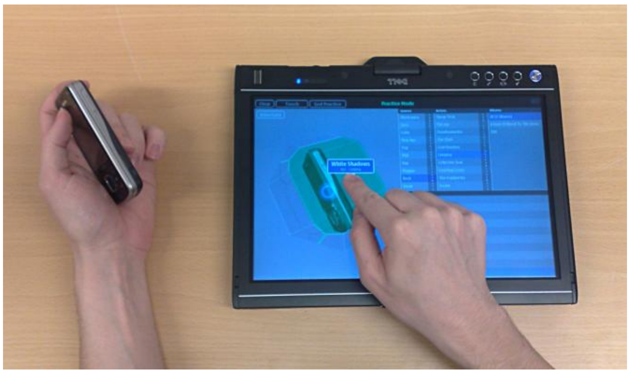

Figure 3. The prototype in use. The user is dragging a song from the browser to the map surface.

The design of the interaction method follows the guidelines of Guiard's KC model [7]. The non-dominant hand sets the orientation as a frame of reference for the dominant hand to manipulate the music items. The dominant hand requires larger and more frequent movements to browse and drag the music items. Finally, the non-dominant hand sets the orientation of the mobile device before any manipulation. However, the orientation can also be changed during the manipulations, for example while dragging a music item.

We also designed the interaction method to take advantage of human spatiality [18]. The first heuristic concerning intuitive spatial mappings is addressed by the personalization feature in general. While mapping music to different orientations is not something that users naturally understand, personalized mappings are likely to be more meaningful and intuitive. I/O unification is addressed by displaying the 3D model of the phone next to the actual phone and synchronizing the orientations. It can be further strengthened by combining haptic, audio and visual modalities to reproduce the feeling of physically manipulating items around the phone. I/O unification was also the reason for selecting touchscreen instead of a mouse. Epistemic activity is supported by immediate feedback and easily reversible actions. In fact, even unintentional movements of the non-dominant hand provide direct feedback on how the system reacts to movements.

4.1.4 Touch Only Usage

To allow comparing our bimanual interaction method with traditional one-handed touch UI, Music Tablet was fully usable with one hand as well. Instead of using the actual mobile device to set the orientation, the 3D model of the phone can be rotated by dragging a finger over it. We did informal experiments with the two most potential rotation techniques, ArcBall and Two-Axis Valuator [2]. With the Two-Axis Valuator, users were confused by the lack of explicit rotation around the view vector. ArcBall technique, on the other hand, was understood without any instructions and was preferred over Two-Axis Valuator. Our current impression is that with well-known commonplace objects such as the mobile phone, users want to control the exact orientation, even if it is not required by the task. The 3D object used in [2] was a symmetric and unknown object, and thus favorable for Two-Axis Valuator.

4.2 Implementation

The prototype system consists of two pieces of hardware: a Dell Latitude XT tablet and a Nokia 6210 Navigator. We chose this particular phone model because it has both an accelerometer and a magnetometer built in, unlike the Nokia N95, which only has an accelerometer. The application was running on the tablet PC, and communicated with the phone over a serial Bluetooth link. The software was written in Python and OpenGL was used for rendering both 2D and 3D graphics.

In this prototype, the phone had a role as an input device for specifying the orientation. Sensor data from the accelerometer and magnetometer sensors was sent 20 times per second to the tablet PC. In the final system, the mobile device executes the same music application that is used in mobile contexts. Therefore, the mobile device should also provide visual, tactile or audio feedback during the interaction.

4.2.1 Sensor Data Filtering and 3D Rotation

The data read from the accelerometer and magnetometer had to be filtered to reduce noise so that the visualization would represent the physical orientation of the phone convincingly, without shaking or excessive lag. We designed a filter to fulfill two goals:

allow the system to quickly respond to fast rotation of the phone, and reduce jitter in visualization when the phone was held steady.

The accelerometer and magnetometer in the Nokia 6210 Navigator are three-axis sensors that produce 3D vectors. The vectors were normalized and stored in a fixed-length buffer:

$$\left[\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N\right], \quad (1)$$

where $\vec{x}_1$ is the latest vector read from the sensor, and N is the length of the buffer. First, the system determined whether the phone was currently held steady or was being rotated. This was done by examining the difference between consecutive vectors in the buffer. If the length of the sum of delta vectors was within a certain threshold, the phone was considered steady. We defined stability s as

$$s = S - 2\left\|\sum_{i=1}^{n-1}(\vec{x}_i - \vec{x}_{i+1})\right\|, \quad (2)$$

where S is the empirically determined stability constant (S = 1.2) and n is the stability observation window size (n = 10).

However, steadiness was only judged by examining accelerometer vectors, because of their relatively low noise level. Therefore, the unsteady state was not correctly detected in a scenario where the magnetometer vector changed but the accelerometer vector stayed the same, for instance when rotating the phone carefully on a table surface. In practice, while holding the phone in a hand, any rotation that resulted in a different compass reading would also cause changes in the accelerometer vector, so this was acceptable.

While steady, sensor readings were filtered by taking the arithmetic mean of 30 samples (filter A). While unsteady, an arithmetic mean of the 6 latest samples (filter B) was used. The mean vectors $\vec{m}$ are:

$$\vec{m}_A = \frac{\sum_{i=1}^{30}\vec{x}_i}{30} \qquad \vec{m}_B = \frac{\sum_{i=1}^{6}\vec{x}_i}{6}. \quad (3)$$

To avoid an abrupt change when switching between steady and unsteady states, the final output vector was a weighted arithmetic mean of A and B. Thus the final output vector $\vec{y}$ is:

$$\vec{y} = \begin{cases} \vec{m}_B & \text{if } s \le 0 \\ (1-s)\vec{m}_B + s\vec{m}_A & \text{if } 0 < s < 1 \\ \vec{m}_A & \text{if } s \ge 1 \end{cases}. \quad (4)$$

Finally, the filtered acceleration and magnetometer vectors ($\vec{y}_{acc}, \vec{y}_{mag}$) were used to calculate rotation for the 3D visualization of the mobile device. As the acceleration vector is pointing down due to gravity and the magnetometer is pointing towards the magnetic north, the angle between the vectors depends on geographical location. The coordinate axes for the rotation R were therefore calculated as:

$$\vec{s} = \frac{\vec{y}_{acc} \times \vec{y}_{mag}}{\left\|\vec{y}_{acc} \times \vec{y}_{mag}\right\|}, R = \begin{bmatrix} \vec{s} \times \vec{y}_{acc} \\ \vec{s} \\ \vec{y}_{acc} \end{bmatrix}. \quad (5)$$

In order to reduce the jitter further in the steady state, the orientation was updated only at one-second intervals and linear interpolation was applied to animate the orientation changes smoothly. A more sophisticated filter may have helped to cancel the noise, but we found this simple approach sufficient for our needs. In the unsteady state, the orientation of the visualization was updated on every frame to keep up with fast rotations.

4.2.2 Compass Alignment and Magnetic Interference

The prototype used a manual compass alignment method for ensuring that the 3D model of the phone matched the orientation of the real one. The tablet PC user interface had a button reserved for realigning the phone orientation at any time. Alignment was done by holding the phone display upward with the top pointing forward, and pressing the alignment button on the tablet. We used quaternions to calculate the final rotation:

$$r_{final} = r^*_{align} \times r_{current}, \quad (6)$$

where r^*_{align} is the conjugate of the rotation quaternion during alignment. Ideally, the tablet PC would contain corresponding sensors that could be used to calculate the alignment rotation automatically.

There were magnetic fields around the tablet PC that interfered with the magnetometer in the phone. Because of this, the user had to keep the phone at least 20 cm away from the tablet. However, this did not cause the user's posture to become unnatural for the use case. A gyroscope in the mobile device could be used to compensate the magnetometer interference caused by any local magnetic fields.

5. EXPERIMENT

We conducted an experiment to evaluate the difference between our bimanual interaction method and the alternative method using the touch screen only. We were primarily interested in aspects related to efficiency and memory. In addition, we were interested in subjective measures such as pleasantness and difficulties in physically handling the mobile phone. In this section, we describe the design of the experiment, present the results and discuss their implications.

The experiment was motivated by the potential benefits of physical interaction with digital music. Here we focus on personalization of the music mapping, whereas music exploration was studied in [10]. As discussed in previous sections (3.3 and 4), enabling fast and effortless creation and manipulation of music mappings is essential to support music exploration in mobile contexts. In this experiment, we concentrated on the task of adding new music onto a mapping. The goal was to test the following hypotheses:

- Bimanual interaction is more efficient in terms of time required for placing items onto the mapping.
- Bimanual interaction makes it easier to recall the placement of the items on the mapping after they have been placed.
- Bimanual interaction is subjectively experienced as easier, faster, and more pleasant.
- Using touch UI to set the orientations is subjectively experienced to be more accurate, as it does not suffer from noise or unintentional hand movements.

5.1 Experiment Design

To test the hypotheses we designed two tasks: a speed task, where the goal was to quickly place a set of items onto the mapping, and a memory task, where the user tried to recall the placement of recently added items. Every user performed these tasks using both interaction methods. Quantitative data was automatically gathered on task execution times and the number of remembered items. In addition, subjective data was collected with 5-level Likert questionnaires.

One pilot user and 20 participants were recruited for the experiment. All participants were right-handed, listeners of mobile music and had at least some hands-on experience with pen or touch based user interfaces. The tests were recorded on video. The test was conducted in usability laboratory setting (Figure 4). The user sat by the tablet and held the phone in the left hand, while interacting with the tablet PC with the right hand. Music was played through the speakers in the corners of the table. The test moderator was sitting at the end of the table.

The duration of test sessions was approximately 35 minutes. The participants were unfamiliar with the system and the contents of the music library. In the beginning of the test session, the moderator introduced the concept to the participants. The users were then allowed to familiarize themselves with the system using both interaction methods. Once they were comfortable with the system, they initiated the tasks by clicking a button on the screen. All participants performed the speed tasks before the memory tasks. The order of the interaction methods was counterbalanced between the participants.

Figure 4. Test setup

5.1.1 Speed Task

The speed task was divided in three difficulty levels according to the type of the items to be placed. Both interaction methods were used to perform the task for each difficulty level, resulting in 2x3 subtasks. The difficulty level was always increasing from easy to more difficult. This was because we expected novice users to start with simple mappings, and gradually adding more details while gaining experience with the system.

The tasks were to place six items on the map surface, one to each side of the phone. Timing was started when the first item was placed, and ended when the 6th item was in place. The items were given one at a time, and users were allowed to freely select an empty side to place them on. The easiest difficulty level involved placing artists, which required browsing the artists list only. The second level was placement of albums, where both artist and album lists had to be used. The last level was the placement of songs, which required browsing in all three lists. The genre list was not used after the first insertion and remaining constant within the tasks. The set of songs, albums, and artists used in the tasks was the same for all users, and each user placed the items in the same order. The number of items in artists list was 13–24, 1–7 in albums list and 2–8 in songs list. As each list view can show 10 items, only artist selection required some scrolling.

Subjective feedback was collected in the form of questionnaires after performing the tasks with each interaction method. The statements in the questionnaire concerned ease of use, speed, accuracy, and pleasantness. The questionnaire related to bimanual interaction method contained an additional statement regarding difficulty of physically handling the phone. After completing the whole speed task, a third questionnaire was used to directly compare the interaction methods.

5.1.2 Memory Task

For the memory task, the users were first asked to pick 12 artists they recognized from the list of artists in the music library. Then the users were instructed to place 6 randomly selected familiar artists onto the mapping using one of the interaction methods. The artists were placed one by one and in random order. After finding and placing all six artists, the GUI was disabled and the users were instructed to show the orientations of each artist with the mobile phone. The orientations were queried again one artist at a time, in random order. The same task was repeated with the 6 remaining familiar artists using the other interaction technique.

5.2 Results

5.2.1 Effect on Speed

A 2 x 3 (Technique X Complexity) repeated measures ANOVA was conducted to analyze the results of the speed test. Each measurement represents the mean item placement time for a sequence of five items. 9 out of 600 (1.5%) data points were excluded as outliers, as they deviated more that three times the standard deviation from the mean. Such extreme outliers can be explained by interruptions in the task execution, e.g., due to software problems.

There was no significant Technique X Complexity interaction ($F_{2,38} = 0.547$, $p = 0.583$), suggesting that the speed difference between the techniques is not substantially influenced by the task difficulty or vice versa. There was a significant main effect for Technique ($F_{1,19} = 16.220$, $p < 0.005$), indicating that the overall item placement time with the bimanual technique ($M = 9.385$, $SD = 2.949$) was 12.5% faster than with the touch only technique ($M = 10.720$, $SD = 2.985$). As expected, there was a significant main effect for Complexity ($F_{2,38} = 89.801$, $p < 0.001$) verifying that the item placement time increases with the task complexity. The mean times are presented in Figure 5.

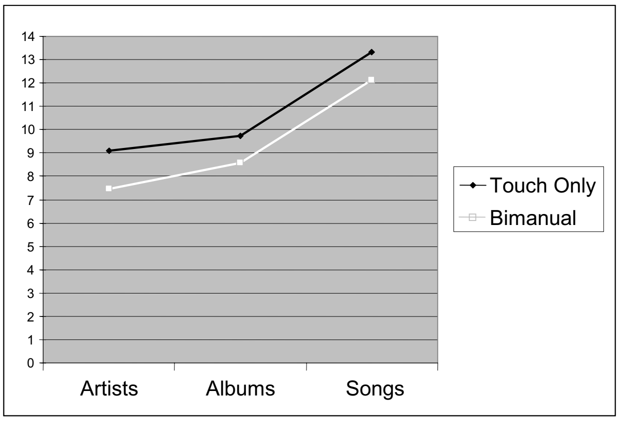

Figure 5. Mean item placement time for each condition.

5.2.2 Effect on Memory

A paired T-test was conducted to analyze the results of the memory test, revealing no significant difference between the techniques ($t_{19} = 0.556$, $p = 0.292$). In fact, the mean number of correct answers with bimanual technique (3.65) was smaller than with the touch only technique (4.0), as presented in Figure 6 a. Number of correct answers with each of the techniques is presented in Figure 6 b, revealing no clear trends towards any direction.

5.2.3 Subjective Measures

For both interaction techniques, the subjective measures for ease of use, speed, accuracy and pleasantness were collected using a five-point Likert questionnaire. In a Wilcoxon signed-rank test, ease of use was found to be significantly higher with bimanual technique ($W = 0$, $N = 7$, $p < 0.005$). Speed and pleasantness were also significantly higher with bimanual technique ($W = 4$, $N = 9$, $p < 0.05$ and $W = 15$, $N = 12$, $p < 0.05$ respectively). However, the accuracy revealed no significant difference ($W = 15.5$, $N = 10$, $p = 0.104$). Contrary to expectation, the mean subjective rating for accuracy was higher for bimanual technique than for touch only technique.

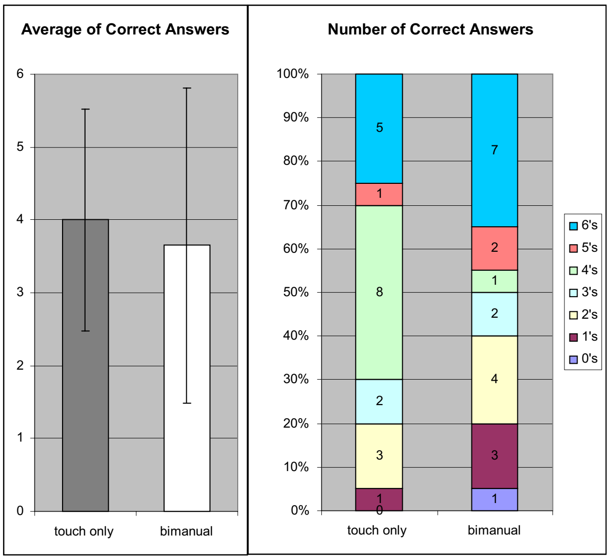

Figure 6. a) Average of correct answers (with standard deviation) and b) distribution of results for both interaction techniques.

Additional question regarding the handling of the phone was overall agreed to be easy. Overall, all questions show positive feedback for both interaction techniques, but moderately in favor of bimanual technique. The questionnaire results are presented in Figure 7.

5.3 Discussion

5.3.1 Speed Test

As hypothesized, the bimanual interaction method was significantly faster than using the touch only. The users also noticed the effect subjectively, scoring bimanual method significantly faster. After completing the speed task with both methods, 75% of users agreed to the claim.

All input events were logged during the tests to allow further analysis. To provide insight to the speed difference, we estimated the time used for setting the orientation by analyzing the touch

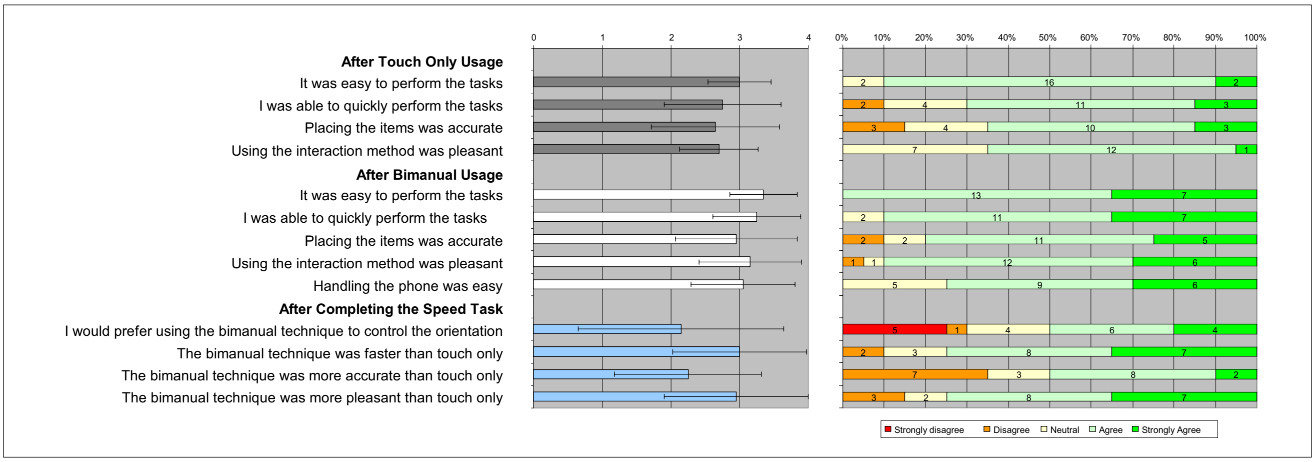

Figure 7. Questionnaire results and standard deviations for the speed task. On the right, distribution of answers.

input events. The mean time for setting the orientation using the ArcBall was 1.645 seconds. Rotation time with the bimanual technique was more difficult to estimate, as the user may rotate the mobile device at any time, and the device is not completely stable even when the user is not intentionally rotating it.

The speed difference in a rotation task between a mouse controlled ArcBall and a 3D orientation input device has been studied in [9], showing that the 3D orientation input device is 33%-36% faster than the ArcBall. Touch-controlled ArcBall can be assumed to be comparable to the mouse controlled version. By using the same percentual speed difference to estimate the rotation time with the bimanual technique, we get a rotation time of 1.05 seconds. Comparing the estimated rotation times suggests that the time difference between the techniques should be approximately 0.6 seconds. However, we found the speed difference to be on average 1.335 seconds, more than double the estimated difference. This confirms that the coordination of two hands in the bimanual technique was not difficult for the users. It also suggests that there were some additional benefits in the bimanual technique, such as less divided attention or cognitive benefits [11]. Further tests would be needed to find out the actual reason for the speed difference. For example, using a gaze tracker would reveal information about the users' visual attention during the tests.

Overall, the subjective results were positive for both interaction methods. This was not surprising, as the users had some previous experience with touch UIs, and the ArcBall method is known to be one of the best mouse-based rotation techniques. Ease of use gained the strongest scoring difference, which is also evident from the answer distributions. With 35% vs. 10% strong positive opinions and 0% vs. 10% neutral opinions in method specific scores, it is clear that the users had no trouble in performing the tasks with the bimanual method. It seems to be more intuitive to control the orientation with a physical rotation method than with ArcBall, regardless of its consistent and simple behavior [19].

Accuracy was the only measure where the method specific score difference was not statistically significant. This is consistent with the results in [9]. However, 80% of the users agreed that the accuracy of bimanual interaction was good.

Clear majority of the users found the bimanual interaction to be more pleasant, with 75% at least agreeing and 35% strongly agreeing. In method specific scores 65% of the users found touch to be pleasant to use, but for bimanual the scoring was given by 90% of the users. However, users did not have a clear consensus on which interaction method was preferred.

5.3.2 Memory

The memory test gave no conclusive results about interaction technique's effect on memory. In fact, it seems that other variables, such as users' memorization schemes, had much larger impact than interaction technique.

We observed a carry-over effect, although it was not statistically significant – users tended to score slightly better using the first interaction method than the second one. This suggests that the test setup failed to eliminate all the non-desirable variables from the test. For example, the second round may have been affected by mixing memorized positions from the first round, even though the set of items was different.

5.3.3 Other Remarks

Our approach to filtering the sensor data proved to be sufficient. In the stable state users found the interaction method accurate, and none of the users complained about the responsiveness during fast rotations. Instead, the largest issue with the sensors was the magnetic interference from the tablet PC, which occasionally required the user to recalibrate the system. Mostly it did not hinder task execution, but was a source of puzzlement during the practice phase. A product quality solution based on more reliable sensors, e.g., gyroscopes, might increase the value of the technique. We also predict that tasks requiring more rotation and less traditional list navigation might benefit more from the interaction technique. In particular, editing an existing mapping by moving and removing items would be an interesting task.

6. CONCLUSIONS

We investigated a physical approach to interaction with digital music to take advantage of our innate spatial and physical skills. Previously we studied the idea of using the physical orientation of a mobile device as a tool for exploring a music library in mobile contexts. We expanded our previous work towards a full music application, allowing users to manage and arrange their music for mobile usage. We concentrated on the task of creating personal mappings between the music and orientations of the device.

We presented a novel interaction technique based on simultaneous bimanual usage of an orientation-sensing mobile device and a graphical touch UI. A 3D model of the device is presented on the GUI and the orientation of the model is synchronized with the device. Music items can be dragged from a list view onto a map surface around the device using the dominant hand. At the same time, the orientation can be controlled with the device in the non-dominant hand.

The bimanual interaction technique was found to be significantly faster to use than unimanual touchscreen interaction in the task of placing items on different sides of a mobile device. Subjective evaluations were in line with the finding, with significant difference in favor of bimanual interaction method for easiness, speed and pleasantness of use.

We believe that the ability to easily create and manipulate personal music mappings is fundamental for the success of our mobile music exploration concept. In addition, providing a consistent means of interaction in both personalization and mobile exploration tasks is critical, as it reduces learning requirements and helps non-visual usage by reinforcing the user's mental model. To generalize, as devices and user interfaces are becoming more versatile, we believe that seamless interoperability and user experience between different devices will help users to take full advantage of our smart environments. In such environments, interaction devices can form a holistic user interface that is greater than the sum of its parts.

7. REFERENCES

[1] Alonso, M. B. and Keyson, D. V. 2005. MusicCube: making digital music tangible. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. CHI '05. ACM, New York, NY, 1176-1179. DOI= http://doi.acm.org/10.1145/1056808.1056870

[2] Bade, R., Ritter, F., and Preim, B. 2005. Usability Comparison of Mouse-Based Interaction Techniques for Predictable 3d Rotation. In *Smart Graphics. 5th International Symposium*. *Proceedings*, Springer, vol. 3638 of *Lecture Notes in Computer Science*, 138-150.

[3] Balakrishnan, R. and Kurtenbach, G. 1999. Exploring bimanual camera control and object manipulation in 3D graphics interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: the CHI Is the Limit*. CHI '99. ACM, New York, NY, 56-63. DOI= http://doi.acm.org/10.1145/302979.302991

[4] Bentley, F., Metcalf, C., and Harboe, G. 2006. Personal vs. commercial content: the similarities between consumer use of photos and music. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '06. ACM, New York, NY, 667-676. DOI= http://doi.acm.org/10.1145/1124772.1124871

[5] Fitzmaurice, G. W. 1996. *Graspable User Interfaces*. Doctoral Thesis. UMI Order Number: AAINN18871, University of Toronto. Available at http://www.dgp.toronto.edu/~gf/papers/Thesis.gf.final.pdf

[6] Fitzmaurice, G. W., Ishii, H., and Buxton, W. A. 1995. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '95. ACM Press/Addison-Wesley Publishing Co., New York, NY, 442-449. DOI= http://doi.acm.org/10.1145/223904.223964

[7] Guiard, Y. 1987. Asymmetric division of labour in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behaviour*; 19, 486-517.

[8] Hiipakka, J., and Lorho, G. 2003. A Spatial Audio User Interface for Generating Music Playlists. In *Proceedings of the 2003 International Conference on Auditory Display* (ICAD '03), 267-270.

[9] Hinckley, K., Tullio, J., Pausch, R., Proffitt, D., and Kassell, N. 1997. Usability analysis of 3D rotation techniques. In *Proceedings of the 10th Annual ACM Symposium on User interface Software and Technology*. UIST '97. ACM, New York, NY, 1-10. DOI= http://doi.acm.org/10.1145/263407.263408

[10] Keränen, J., Bergman, J., and Kauko, J. 2009. Gravity sphere: gestural audio-tactile interface for mobile music exploration. In *Proceedings of the 27th international Conference on Human Factors in Computing Systems*. CHI '09. ACM, New York, NY, 1531-1534. DOI= http://doi.acm.org/10.1145/1518701.1518935

[11] Leganchuk, A., Zhai, S., and Buxton, W. 1998. Manual and cognitive benefits of two-handed input: an experimental study. *ACM Trans. Comput.-Hum. Interact.* 5, 4 (Dec. 1998), 326-359. DOI= http://doi.acm.org/10.1145/300520.300522

[12] Leitich, S. and Topf, M. 2007. Globe of Music: Music Library Visualization Using GEOSOM, In *Proceedings of the 8th International Conference on Music Information Retrieval* (ISMIR '07), 167–170.

[13] Myers, B. A. 2001. Using handhelds and PCs together. *Commun. ACM* 44, 11 (Nov. 2001), 34-41. DOI= http://doi.acm.org/10.1145/384150.384159

[14] Neumayer, R., Dittenbach, M., Rauber, A. 2005. PlaySOM and PocketSOMPlayer, Alternative Interfaces to Large Music Collections. In *Proceedings of the 6th International Conference on Music Information Retrieval* (ISMIR '05), 618-623.

[15] Oakley, I., and O'Modhrain, S. 2005. Tilt to Scroll: Evaluating a Motion Based Vibrotactile Mobile Interface. In *Proceedings of the 1st Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (WHC'05), 40-49.

[16] Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. ACM, New York, NY, 919-928. DOI= http://doi.acm.org/10.1145/1054972.1055101

[17] Pauws, S., Bouwhuis, D., and Eggen, B. 2000. Programming and enjoying music with your eyes closed. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '00. ACM, New York, NY, 376-383. DOI= http://doi.acm.org/10.1145/332040.332460

[18] Sharlin, E., Watson, B., Kitamura, Y., Kishino, F., and Itoh, Y. 2004. On tangible user interfaces, humans and spatiality. *Personal Ubiquitous Comput.* 8, 5 (Sep. 2004), 338-346. DOI= http://dx.doi.org/10.1007/s00779-004-0296-5

[19] Shoemake, K. 1992. ARCBALL: A user interface for specifying three-dimensional orientation using a mouse. *Proc. of Graphics Interface '92*. pp 151-156.

[20] Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., and Baudisch, P. 2007. Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. ACM, New York, NY, 1395-1404. DOI= http://doi.acm.org/10.1145/1240624.1240836

[21] Ängeslevä, J., O'Modhrain, S., Oakley, I., and Hughes, S. 2003. Body Mnemonics. In Physical Interaction (PI03) - Workshop on Real World User Interfaces, MobileHCI '03