An Initial Usability Assessment for Symbolic Haptic Rendering of Music Parameters

Meghan Allen[1], Jennifer Gluck [1], Karon MacLean[1], Erwin Tang[2]

[1] Department of Computer Science
University of British Columbia
2366 Main Mall
Vancouver, BC V6T 1Z4
{meghana, gluck, maclean}@cs.ubc.ca

[2] Department of Mechanical Engineering
University of British Columbia
2054-6250 Applied Science Lane
Vancouver, BC V6T 1Z4
erwin@mech.ubc.ca

ABSTRACT

Current methods of playlist creation and maintenance do not support user needs, especially in a mobile context. Furthermore, they do not scale: studies show that users with large mp3 collections have abandoned the concept of playlists. To remedy the usability problems associated with playlist creation and navigation – in particular, reliance on visual feedback and the absence of rapid content scanning mechanisms – we propose a system that utilizes the haptic channel. A necessary first step in this objective is the creation of a haptic mapping for music. In this paper, we describe an exploratory study addressed at understanding the feasibility, with respect to learnability and usability, of efficient, eyes-free playlist navigation based on symbolic haptic renderings of key song parameters. Users were able to learn haptic mappings for music parameters to usable accuracy with 4 minutes of training. These results indicate promise for the approach and support for continued effort in both improving the rendering scheme and implementing the haptic playlist system.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, haptic i/o, prototyping, user-centered design.*

General Terms

Design, Experimentation, Human Factors.

Keywords

Haptics, physical interfaces, playlist creation, digital music, music classification, force feedback, vibrotactile feedback, mp3.

1. INTRODUCTION

Digital media has exploded in popularity in recent years. The mp3 digital music standard in particular has enjoyed massive adoption in the popular culture, evidenced by rising sales in digital audio hardware. Samsung Electronics Co. projects that global shipments of portable mp3 players will reach 35 to 45 million units in 2005 [8]. With this growing centrality of digital audio comes a new set of problems, particularly in the context of browsing and listening to large music collections.

Traversing a large digital music collection for the purposes of building a playlist can be a time-consuming task [11]. When an unfamiliar song is encountered, the user must listen to a musical passage in order to make the decision of whether to add the song to the playlist. Even in the event that the user is immediately familiar with every song based on filename, exploring the collection via file and menu scrolling is slow and tedious. The problem is compounded by the widespread use of mobile listening devices that offer little or no screen space with which to work: visual methods are infeasible. Often, users resort to creating one large playlist that includes every song in the collection, and then engage the media player in random play mode [13]. This usage requires the user to interact continually with the system, skipping through the list until a desired song is encountered.

We envision a haptic solution to the problem of playlist creation and navigation, whereby a user receives a concise symbolic representation of key song characteristics via a haptic channel and uses this to make listening decisions. While this scheme could also be used with visual encoding in a desktop environment, the visual display would not be suitable for mobile and multitasking contexts. Likewise, a symbolic auditory coding might be difficult to hear in some mobile contexts and compete with the music itself. We propose a one-degree-of-freedom (e.g. haptic knob) solution which allows the user to scroll through a list of songs while simultaneously receiving song information, thereby achieving efficient bidirectional interaction.

A typical use case shows how this mechanism can alleviate the high engagement demands of the current common practice of using a large collection with random selection and play-time song skipping. Our proposed system instead allows users to interact once with the haptic knob prior to the listening session and then listen without interruption. The user scrolls rapidly through the randomly-ordered collection, cognitively primed to haptically search for features in the range of interest – e.g. a particular tempo range or genre. When such a rendering is encountered, the user can quickly select the song by pushing a button; the system adds that song to the playlist then advances to the next. The user continues to scroll and select in this manner until a sufficient

ICMI'05, October 4–6, 2005, Trento, Italy.

number of songs have been selected. An on-the-fly playlist has thus been generated quickly and easily.

In order to be sufficiently usable, the haptic rendering used in such a system will require fast recognition on the part of the user; based on related research with haptic icons (e.g. [7]), we anticipate that 1 second per song is a usable and possible goal. However, a necessary first step is to determine whether this kind of haptic mapping for music can be easy to learn and use, regardless of recognition speed, and gain insight on how to optimize it. This initial feasibility analysis entails two research tasks. We began by establishing the attributes of music considered by users during song selection. We then investigated various ways in which they could be rendered on a haptic device so as to facilitate association of the haptic sensation with the music.

1.1 Approach

Our design approach was guided by the following criteria, which were arrived at through discussion with current digital music users:

1. The haptic mapping is intended to facilitate selection, rather than to aesthetically augment or replace the listening experience itself.
2. The system must be scalable to large collections, and therefore the parameters used in the symbolic representations must be automatically extractable from the song media, in the near future if not now.
3. Initial studies showed that users do not necessarily select songs based on artist and title, and so we do not expect the haptic system to enable the user to uniquely identify a particular song. Rather, the feedback is intended to let a user decide if a song fits generally with the *style* of music to which he is currently in the mood to listen.

Our investigation led us to a prototype that utilizes both vibrotactile and force feedback. Frequency and amplitude of force feedback detents convey information about song tempo and energy, respectively, while vibration provides additional song-specific information. Thus, our system produces a unique haptic icon for each song, but representations of similar songs will also tend to share ‘family’-like features. The rationale for this unique mapping strategy versus a simpler one of categorization is fully discussed in Section 3.2.

Because humans are better at making relative haptic discriminations (tactile, proprioceptive or kinesthetic) than at assessing absolute haptic magnitudes (e.g. [14]), our rendering algorithm maps tempo and energy information to the display dimensions of the knob in a way that is perceptually relative, not absolute. Thus with our system, a user may feel one song and then the next and be able to observe, “There are more detents in Song 2 than Song 1, and so I know that Song 2 is faster,” as well as, “There is more detent resistance in Song 1 than Song 2, so I know that Song 1 has higher energy.”

We have not examined in detail the question of whether an absolute versus a relative mapping to, for example, some kind of gestalt of song ‘mood,’ is more appropriate in the anticipated usage where the user is rapidly scanning a large collection; arguments can be made either way. However, our principal goal here is to assess whether a haptic mapping is possible and usable, and to this end, it makes sense to orient the scheme so as to take maximum advantage of human perceptual abilities.

We began our research with a number of exploratory studies. More rigorous prototyping and evaluation of the final prototype followed, aimed at determining how well subjects could map renderings to music. In this manner, our assessment iteratively moved towards an understanding of performance potential for the haptic mapping approach.

2. RELATED WORK

2.1 Haptic Icons

The development of functional haptic icons is a necessary first step in the creation of haptic interfaces. Multiple terms – tactile icons, tactons, hapticons, etc. – have been used to refer to essentially the same concept: the use of tactile messages to communicate information non-visually [6, 9]. MacLean and Enriquez began by developing a technique for optimizing the perceptual spacing of completely abstract sets of vibrotactile icons [9]. Brown et al. encoded cell phone call type and priority into frequency and amplitude of “Tactons,” and referred to these icons as ‘structured’ and ‘abstract’ [6]; however, they do constitute a direct symbolic mapping because they are artifacts carefully designed to convey a specific message rather than arbitrarily memorized sensations. Either way, promising recognition rates in their use of rhythm and texture suggest the basic feasibility of haptic icons.

Work done by Chan et al. on haptic icon design using families and metaphors [7] is an example of symbolic representation. This research indicated that grouping icons into families facilitated learning through cognitive chunking. Steeping the icon renderings in metaphors such as a heartbeat and impatient finger drumming further added to learnability and usability. The deliberate mechanism of supplying users with a metaphorical “story” to help them understand the still-systematic and perceptually optimized mappings may explain the substantially better learning times and performance obtained as compared with [6].

Past work has been largely limited to low-quality vibrotactile stimuli, in part on the principle of recognizing the current state-of-art in commodity devices. However, force feedback enables additional parameters for information transmission and, more importantly, contributes substantially to active exploration and bidirectional control. Recent hardware research developments give promise of a new generation of portable devices which could incorporate both. For these reasons, our work explores the richer context of hardware devices capable of both force-feedback and vibrotactile stimuli.

2.2 Haptics and Music

van Erp and Spapé recently carried out a study involving “tactile melodies:” pieces of music translated directly from the auditory to the vibrotactile domain [17]. Although this research is very similar to the vibrotactile aspect of our prototype, the goals of the tactile melody project are very different from ours. van Erp and Spapé are experimenting with haptic rendering of music as a tool for building haptic icons intended to communicate messages unrelated to the music itself. In contrast, our research goal is to convey song information directly.

D'Groove, a haptic turntable for disk jockeys, demonstrates an inventive blending of haptics and music [4]. Two of the interactive haptic features in particular – the "beats-for-bumps" and "resistance" modes – were inspirational to our prototype design. Though our techniques are similar in this respect, the projects differ widely. D'Groove allows the user to manipulate and modify digital audio. Our system, on the other hand, communicates properties of existing audio.

This idea of incorporating haptics into a media control interface has been broached by Pauws et al. [10], but there, haptic feedback is meant to aid in basic fixturing (location of menu borders, etc), rather than content discovery. In contrast, our work focuses on haptic representation of song content.

2.3 Automatic Classification and Playlist Creation

There has been significant recent headway in automatic classification of digital music, technology which is essential to the eventual success of our scheme. Tzanetakis and Cook [16] have developed a set of algorithms for the automatic genre categorization of audio signals, as well as graphical user interfaces that support genre-based interaction with large audio collections. We believe that there is promise in translating these visually-based methods into a multi-modal, haptic system. Furthermore, their work on automatic feature extraction from mp3 compressed data [15] could prove indispensable if the information could be used as the basis for haptic mapping.

The task of speeding music playlist creation has also been a hot research topic. Pauws and Eggen [11] take an AI approach, generating playlists automatically. However, their PATS system requires user interaction during listening to train the system to learn user preferences. We endeavour to create a system that will never require user interaction during listening. More troubling is the catch that an AI system must store a large amount of attribute data for each song. There may be limited space on a portable player, especially on smaller memory flash players. Users would likely be unwilling to sacrifice precious memory to AI systems, limiting the amount of music that can be stored on the device. Conversely, a negligible amount of space (only two extra numerical values per song) is necessary to include the data needed by our system.

Other techniques [3, 5] utilize information visualization methods to aid the user in the process of playlist generation. In the face of mobile music players, such methods are inadequate: Screen space limitations necessitate a non-visual approach.

3. PHASE I: INITIAL EXPLORATION

3.1 Initial User Interviews

We interviewed 7 subjects (3 females, 4 males) in order to (a) explore current user practices for playlist creation and music listening in order to clearly establish the existence and nature of need; and (b) identify the musical dimensions essential to a user when selecting a song to listen to or to add to a playlist, from which we could then derive requirements for a mapping from musical dimensions to haptic sensation.

We employed an open-ended interview strategy to understand how users interact with digital collections and playlists of songs. A few key observations were noted across the study. The most apparent trend was that *mood dictates song selection*. Definition of 'mood' had some minor variation across subjects, but most indicated a meaning that combines the energy of the song (relating to beat, tempo, and instrumentation), the feelings it inspires, and the context in which they are listening. Examples of different contexts include working out, doing homework, or getting ready to go out. Users did not generally tie selection to artist or song title information.

We were surprised to find that very few of our subjects build and use playlists. Some indicated that they had previously used playlists, but that the effort required to keep them updated was too great. Instead, they tend to select a large number of songs, engage the player in random mode, then continually interact with the system, skipping the songs to which they aren't interested in listening. This observation provided the primary impetus for our proposed use scenario.

3.2 Rationale for Unique Mapping Strategy

The rationale behind our decision to use a one-to-one mapping from song to haptic icon, the implementation of which is described below, emerged during conversations with users and discussion of their use strategies. An alternate and potentially simpler approach could be based on binning or categorization; for example, we could have used several classes based on genre, or divided the spectrums of tempo or energy into a number of discrete ranges. We might even have defined several distinct 'moods.' We could then create a set of distinct haptic icons that could be easily learned and associated by users. This lower-cognitive-effort approach would very likely have eased the rapid recognition required by our target application.

We took the more complex approach principally because music is too rich and nuanced to lend itself to such simple categorization. We do not believe that answers to an aesthetic problem such as classification for listening purposes lie in simplifying music so that we can deal with it more efficiently. Instead, we strive to convey the richness of each individual song in such a way that the user can grasp the richness inherent in the listening experience, but in a much shorter time than it takes to listen to the song.

Should the strict categorization approach nevertheless seem of value in some context, it would be pointless to require users to learn a system of haptic renderings and scan them song-by-song. Instead, such a system should automatically sort songs into the pre-set categories; a user would then simply indicate to the system, "Play songs from Category 2."

We did consider another approach, which utilized the idea of themes, defined loosely as some combination of mood and context. Theme examples include 'sad,' 'mellow,' 'upbeat,' 'angry,' 'sophisticated,' and 'background.' The haptic encoding reflected how closely a song adheres to the theme. A song could adhere to various themes over a range of values. This paradigm could be a useful way for users to select songs that correspond to a specific mood. However, this method requires the user to provide a rating for each individual song in the collection, and for every theme. Such a system is not scalable. Conversely, the system we implemented lends itself well to automatic feature extraction. During the development of our prototype, tempo and energy information was computed manually, but it is reasonable to expect one of two future scenarios: (a) the development of reliable algorithms to compute tempo and energy ratings, or (b) that the recording industry will provide tempo and energy information for songs as they are recorded and released into the market. Thus, the proposed prototype is or will become scalable.

3.3 Initial Mapping and Rendering Strategy

Based on our initial interview results we decided on a direct haptic mapping of song characteristics. The attributes we selected were song tempo and energy, as these characteristics were perceived by users to be strongly related to the mood of the music and were specifically mentioned as influencing song selection. Tempo information for each song was obtained using an online beat counter [1] and recorded as the number of beats per minute. In this iteration, energy was a subjective measure determined by the experimenters on a scale of 1-10, where 1 indicates very low and 10 very high energy.

These song parameters (frequency and energy) were then rendered on the haptic knob (described in Section 4.1.1) by displaying the song tempo as the number of detents per revolution and song energy as detent amplitude.

3.4 Testing and Refinement of Prototype

Between our initial interviews and the evaluation of our final prototype, we carried out 4 additional lightweight user studies: 3 iterative prototype evaluations and 1 follow-up interview with participants from the first two evaluations. The following issues were explored, leading to design refinements that were integrated into the final prototype.

3.4.1 Rendering the Energy Scale

We experimented with two rendering schemes for the energy scale. The first associated high energy with *low* detent amplitude: there was lower resistance when turning the knob for a high-energy song and thus the user could turn the knob very quickly, generating a sense of motion and activity. The second associated high energy with *high* detent amplitude, resulting in a higher resistance (and more user effort) when turning the knob for a high-energy song. Testing showed that the high-energy-to-high-amplitude mapping made more sense to users.

3.4.2 Explaining and Training the Rendering

Because we were trying to gauge the intuitiveness of the rendering, we did not inform users of the rendering scheme during the first two evaluations. Results were rather dismal, and so in the third study, subjects were informed and trained on a set of sample songs and haptic icons prior to testing. The training time was brief (averaging 5-10 minutes), but test performance indicated a marked improvement in user comprehension of the mapping. These results encouraged us to continue to pursue the proposed rendering scheme. We note that the concept of symbolic haptic renderings was completely new to all of our users and interface ‘intuitiveness’ is often strongly related to degree of transfer from a user’s experience. Thus, this need for some introduction in a situation devoid of transfer is not necessarily an indication of ultimate acceptability or learnability.

3.4.3 Improving Intuitiveness: Additional Parameter

We conducted follow-up interviews with subjects from the first two evaluations in order to learn about the mental models and reasoning strategies employed during the test, and hence move towards a more intuitive rendering strategy. We started by revealing the rendering scheme that had been used.

Users responded well to the ideas of conveying information about energy and tempo, but many felt something more was needed to differentiate songs with similar tempo and energy yet different mood.

Wary of adding another subjective attribute, we explored conveying temporal waveform data directly, using a vibrotactile frequency range of the haptic knob. Our final prototype incorporates this additional information into the haptic rendering.

4. PHASE II

We developed a final prototype based on the information collected during Phase I, and tested it more carefully. The goal of this phase was to establish a performance baseline for how well users could match songs to their symbolic haptic renderings, given the mapping approach taken here.

4.1 Final Prototype

The prototype utilizes both vibrotactile and force feedback. Knob detents convey information about song tempo and energy. Similarly to one version of our initial prototype, the number of detents per rotation is a linear function of song tempo. Energy is presented as the amplitude of the detents, where a high energy is associated with high amplitude and thus with high user energy. We also added a layer of temporal song information displayed with vibrotactile feedback, detailed below. The resultant texture is the sensation one would get from putting one’s hand on a very powerful speaker and feeling its lower-frequency vibrations as the song plays.

4.1.1 Hardware System

The hardware device used in this study was a Twiddler [12], a single degree of freedom rotary haptic device consisting of a 50mm diameter machined aluminum knob connected to a DC brush (Pittman 8324) motor with an attached encoder. The Twiddler system is capable of running a host-based haptic loop at a frequency of 1 kHz, connecting to the haptic display through the host’s parallel port. It can display force vibrations without significant attenuation up to about 150 Hz. Figure 1 shows the Twiddler knob, motor, stand, and electronics box.

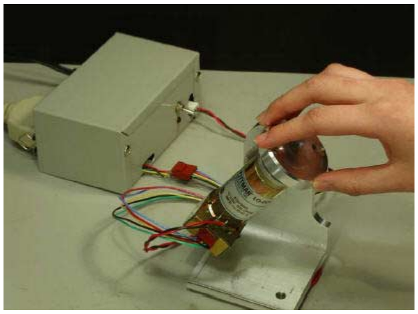

Figure 1. The Twiddler haptic knob [12].

4.1.2 Software System

The rendering implementation can be examined as two separate systems: the force-feedback stimuli and vibrotactile stimuli.

The force feedback system renders information about tempo and energy as knob detents. The detents are modeled as a sine wave. As the angular position of the knob is changed, the forces felt by the user vary as a sinusoid. Song tempo and energy are mapped to the two parameters of the detent sine wave (amplitude and spatial frequency) as follows:

$$\textit{detent spatial frequency} = \textit{tempo value} \div 10 \qquad (1)$$

$$\textit{detent amplitude} = \textit{energy value} \times 500\ \textit{mN} \qquad (2)$$

The detent spatial frequency defines how many detents the user feels per knob revolution. The detent amplitude represents how much force the user must exert to rotate through the peak of the detent. Using Equation 2, low energy values produce detent amplitudes that are too low for distinct detents to be perceived by users. To solve this problem, any calculated amplitude value lower than 1700 mN is automatically brought up to 1700 mN. Figure 2 illustrates these renderings graphically.

For the songs we tested, this algorithm resulted in tempos ranging from 4 to 21 detents per revolution (45 to 214 bpm), and detent amplitudes from barely discernible (1700 mN) to very definite and crisp but still comfortable to push through with one-handed operation (5000 mN).

The vibrotactile algorithm renders temporal song waveform data, by vibrating the knob according to the momentary acoustical energy of the song. The vibrations are modeled as a sine wave superimposed on the force feedback (detents) sine wave. While the force felt from the detents varies with user-controlled position, the stimulus associated with playback data varies with time. This second sine wave has a fixed temporal frequency of 30 Hz, but the amplitude varies, dependent on the real-time audio data. A carrying frequency of 30 Hz was chosen because lower frequencies began to feel sluggish while higher frequencies exhibited a jittery feel and were not well matched to the force display bandwidth of our device. In general, the playback rate of the detent sinusoid (dependent on the users' actuation velocity) will be several times lower than 30 Hz, so the vibrations will be perceived as a high frequency component of the overall signal.

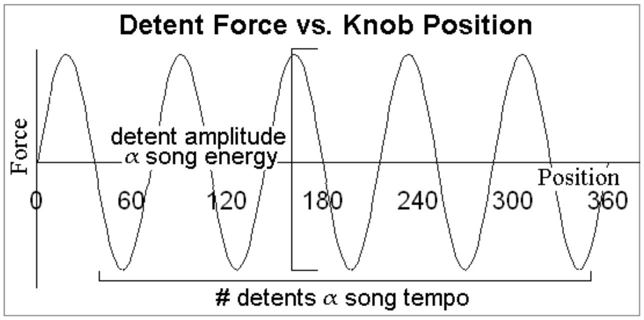

Figure 2: Graphical representation of the force-feedback renderings. 5 detents per knob revolution are shown here, and the height of the sine wave corresponds to the song energy.

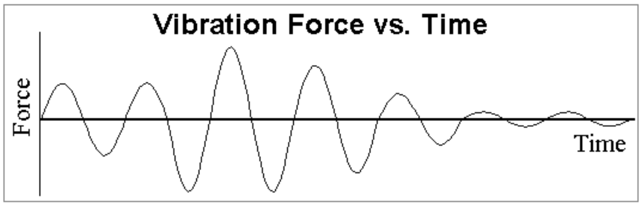

Figure 3: Graphical representation of the time-variant vibrotactile rendering of song data.

The BASS audio library [2] was used to aid in decoding the audio waveform data for all the songs. The audio data for all the songs was represented in the mp3 audio file format. Using a BASS function call, the mp3 file is loaded and played in realtime at 1x speed and zero volume, then sub-sampled at 1 KHz using another BASS function call. This value was then used to continuously modulate the amplitude of the 30 Hz sinusoidal signal. Although we were initially concerned that a 1 KHz modulation of a 30 Hz carrier might result in unstable or jittery renderings, we found the richness of the vibrations quite satisfactory. We leave as an option the possibility of sampling song data at frequencies other than 1000 Hz.

As mentioned before, the forces felt from the vibrations are then superposed on the forces from the detents. This enables users to feel the vibrations regardless of the knob position. Figure 3 shows a graphical explanation of the vibrotactile rendering.

4.2 Experimental Design

We conducted a study to assess the feasibility of our proposed haptic mapping. Our principal goal was to determine how well subjects could associate the haptic icons with the songs they represent, and thus establish an initial performance metric. Secondly, we took the opportunity to explore how the vibrotactile aspect of the rendering contributes to the mapping's usability.

4.2.1 Conditions

It was not relevant to compare basic performance levels using our haptic mapping scheme with, for example, a non-haptic mapping, since we envisage use in contexts where haptic feedback is the only usable feedback source. We were primarily after a first baseline performance assessment. However, we did establish two conditions to aid in establishing the marginal utility of the *vibrotactile* rendering component:

1. **Force Feedback Alone (F)**: Song tempo and energy information were rendered via detent number and amplitude, but without the vibrotactile display.
2. **Force Feedback + Vibrotactile (F+V)**: Song tempo and energy information were rendered via detent number and amplitude, along with vibrotactile rendering of temporal song data.

4.2.2 Procedure

The study required one 1-hour session for which subjects were compensated $10. As an added incentive, subjects were told that test scores would be evaluated and the top 20% of the subjects would receive a bonus prize. Subjects were reminded of this prize during the instruction for each phase. We anticipated that this would motivate subjects to focus on the training and listening tasks enough to ensure collection of relevant results during the testing phase.

Subjects wore noise-canceling headphones to block audible artifacts, as well as to listen to the music during the phases that necessitated listening. The following three phases were carried out for each of the conditions (F and F+V).

> *Training Phase:* We began with an explanation of the rendering scheme and metaphor. Subjects were then presented with the Twiddler knob and a GUI with a set of 20 buttons. Each button was labeled with the artist and title of a song. When a button was pressed, the song began to play while the Twiddler displayed the haptic icon associated with that song, and subjects were encouraged to interact with the knob. Each subject was instructed to spend as much time as necessary listening to the songs and feeling the renderings to ensure that he or she understood how the haptic icons related to the songs.
>
> *Listening Phase:* We attempted to select songs that our pool of subjects would already be familiar with prior to the experiment, based on pre-experiment surveys. This was augmented with an additional song familiarization period, during which users listened to the set of 14 songs which they would need to recall (without audio cues) in the testing phase. Thus, in the listening phase, subjects were instructed to take as long as necessary to familiarize themselves with the songs and their names. Actual subject familiarity with the song set was gauged via a survey at this point.
>
> *Testing Phase:* Subjects were presented with five renderings, as well as artist and title data for five of the songs heard in the previous phase. Subjects were encouraged to feel each rendering as often and for as long as desired, but were not permitted to listen to the songs. Subjects were instructed to match each rendering to one and only one song. The task was executed twice (denoted as Repetition 1 and Repetition 2) for each condition. Each task repetition within one condition used a different set of songs; however, the same two sets of songs were used across conditions. Thus 10 different songs in total were used in the testing phase (2x1x5). These songs were selected from the 14 in the listening phase. Selection was nonrandom because it was necessary to satisfy conditions described below.

This testing phase matching task was devised for its properties of testing the rendering scheme, which is based on relative comparisons. Thus, it is only possible to gauge the ability of the prototype to convey song information if the testing task is designed in a way that allows the user to make comparisons between renderings.

We prohibited the user from listening to the song during the testing phase on the grounds of emulating the context of the target playlist application, in which we assume that users will not listen to songs at the same time as they are interacting with the haptic knob. However, we *do* expect that users will be somewhat familiar with the songs in their own digital music collections. We attempted to simulate this situation as accurately as possible in the experimental design.

Songs used in the testing phase were selected on the basis of tempo and energy values such that there would be diverse representation across the spectrum of values. For each of the two matching tasks, we wanted values widely distributed in the song energy-tempo space, as illustrated in Figure 4. In this manner, it was our aim that the five-song matching task would be representative not of five specific songs, but of five distinct moods. This goal was complicated by the need to also select songs that would be familiar to our subject pool. This is why the same 10 songs were used in both conditions. The alternative – finding 20 songs that would both be familiar to users and satisfy the energy-tempo constraints – seemed infeasible.

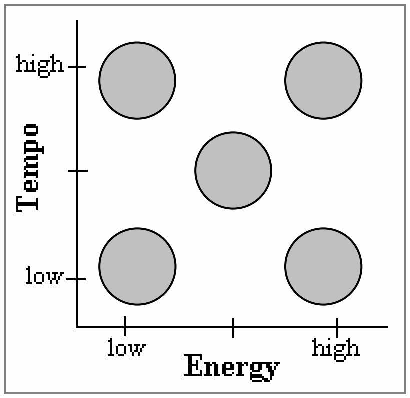

Figure 4: Graphical representation of the song-space. Circles represent the tempo and energy diversity goals for each set of 5 songs used in the matching tasks.

4.2.3 Design and Subjects

A within-subjects design was used to avoid a confound arising from differences in individuals' familiarity with members of the song set. The order of the F and F+V conditions, each administered once per subject, was counterbalanced across subjects.

The 10 subjects (7 male, 3 female) were graduate students in Computer Science at the University of British Columbia, aged 20-40. All were right handed. Subjects' experience with haptics varied: 2 subjects had no previous experience, 5 had limited experience and 3 had extensive experience. Subjects were screened for familiarity with the music used in our test.

4.2.4 Dependent Measures

We measured subjects' scores on the matching task and considered this score as a baseline performance measure for the haptic mapping as implemented here. Subjects were also asked to indicate a confidence rating for each rendering matched. We measured learning effort through the amount of time spent exploring the haptic stimuli during the training phase. We also tracked the amount of time spent on the listening phase, as well as song familiarity via the survey distributed during this phase. In a post-study questionnaire, we collected subject data as well as suggestions for improvement to the prototype.

5. RESULTS AND DISCUSSION

One participant's data was eliminated from the evaluation when test performance indicated insufficient understanding of the test instructions. Results are reported for the remaining 9 participants.

Table 1: Summary of test score results. Shows percentage of correct matches according to a variety of breakdowns (n=9)

Condition:	F	F+V
Overall (condition only)	60%	66%
Repetition 1	51%	66%
Repetition 2	69%	66%
Condition presented first	58%	70%
Condition presented second	63%	72%
Perfect trials (31% overall)	22%	39%

5.1.1 Execution Times

On average, subjects took 4.02 minutes to complete the training phase of the experiment, 4.47 minutes to complete the listening phase, and 4.33 minutes to complete each matching task.

5.1.2 Results

Overall, subjects matched 65.6% of the haptic icons with the correct songs (n = 9, sd = 16.9%). Subjects performed better when the vibrotactile temporal data was displayed (F+V condition: mean = 66%, sd = 13.0%) as compared to the F condition (mean = 60%, sd = 19.4%), but the difference was not statistically significant. Other breakdowns of the results are shown in Table 1.

Confidence ratings did not reveal any significant differences due to the presence of vibrotactile stimuli. On a 5-point Likert scale where a 5 indicated “very confident,” mean response values were 3.33 and 3.53 for the F and F+V conditions, respectively. Thus, under both conditions, participants were neutral to slightly confident about their performance. However, subjects indicated an overwhelming preference (8 of the 9 participants) for the F+V condition.

5.1.3 Interpretation of Performance Values and Suitability of Test Task

Because of the one-to-one matching structure of the task, a single error in matching meant that the highest possible score for a given trial was capped at 3 out of 5 or 60%. Thus, an average score of 65.6% implies that subjects averaged less than one error per matching task. In fact, subjects often made a single switch-type error, but errors rarely involved more than 2 items; and they made no errors in 31% of the trials. Furthermore, almost all of the errors made involved confusing renderings for songs with similar tempo and energy values. We consider this an encouraging result: it indicates that our rendering scheme was well understood, but our task, as presently posed, may dictate a limit to its resolution.

Recall that the anticipated usage of the system is not to aid users in selecting specific songs, but in choosing music that ‘sounds right’ and fits with a particular mood. Based on this imprecise objective, we feel that the performance values observed were quite reasonable in terms of suggesting potential performance in a task more closely related to the ultimate system goal of choosing tracks based on a desired mood (performance of which would necessarily have been too subjective for an initial usability assessment).

5.1.4 Learning

The mean time spent on the training phase for each of the two conditions was 4.02 minutes. Although we had expected subjects to devote more time to training, we were encouraged that the shorter training time indicated the learnability of the rendering. We did observe improved performance with time and familiarity over the course of the 60-minute study, although this was not a significant effect. We did not observe any cases of performance degradation over time. Note that, with only two trials per subject, noisy data was to be expected in the best of situations.

Subjects spent an average of 4.33 minutes on each matching task, but performance improved after each task. This demonstrates that subjects’ understanding of the rendering improved quickly in a short period.

Performance under both the F and F+V conditions was better when the condition was presented second. Learning of the aspects of the rendering scheme common to both conditions (tempo and energy) were likely the cause of this effect.

5.1.5 Optimal Performance

The best performance was achieved under the F+V condition when presented second. In this condition, subjects correctly matched 72% of the renderings with the corresponding song. Additionally, a greater number of perfect trials were achieved under the F+V condition than under the F condition (7 versus 4). Furthermore, performance under the F+V condition was consistently better than under the F condition. Although none of these results was statistically significant, the consistently higher scores under the F+V condition suggest additional potential of the vibrotactile stimuli, especially in light of overwhelming subject preference for the vibrotactile condition.

5.1.6 Reflection on Experimental Design

The listening-phase questionnaire indicated that subjects were familiar with an average of 11.7 of the 14 songs. This was approximately the level of familiarity we had anticipated when designing our experiment. Still, despite the time spent during the listening phase, comparison of specific performances with post-learning phase familiarity surveys suggested that it was clearly more challenging for subjects to match renderings to a song with which they were not previously familiar. We hypothesize that had we been able to test song familiarity prior to the experiment – or perhaps used songs from each subject’s own music library – and then tailored the content of the matching task to each subject (at substantially increased experiment effort), test scores would have improved. Additional learning for the rendering scheme could also have been helpful.

We also hypothesize that performance might have been higher if subjects had been more formally focused on the training task. An enforced minimum of training time, or else a test to indicate understanding prior to the testing phase, is something to consider for future user testing.

6. CONCLUSIONS AND FUTURE WORK

We have described the development of a haptic rendering scheme for music intended for use in playlist creation and navigation. The work presented is an exploratory first step necessary to the achievement of the ultimate proposed application. Results of user testing demonstrate the potential of a haptic mapping of song parameters. With training and the inclusion of both force-feedback and vibrotactile stimuli, subjects were able to achieve scores as high as 72% on a song-rendering matching task using our rudimentary prototype, and achieved perfect scores 31% of the time.

The results of this initial feasibility study justify further investigation in two areas: improvement of the mapping scheme and rendering, and implementation of the haptic playlist system.

The rendering scheme we have developed was a first step and we do not assume this approach to be the most successful one possible. Further development and experimentation is necessary to explore other possible renderings in order to determine an ideal mapping. Particular attention needs to be paid in the future to the speed of recognition possible with a particular mapping – very likely a compromise with the amount of information that can be encoded. Addressing the most common source of errors in this evaluation, we will explore the effect of discretization of the tempo and energy spectrums in order to make the haptic icons more perceptually distinguishable. A next step in extending the system will be to incorporate automatically generated genre information, such as that of Tzanetakis and Cook [16].

Following iteration on mapping strategy, we plan to move to the next step of implementing an actual playlist system. Aspects requiring careful consideration include the track selection mechanism as well as the hardware device on which the system will be implemented. A crucial consideration will be how to display the occurrence of a scrolling action from one song to the next without interfering with the rendering of the music parameters. Once these research questions have been answered, it will be very exciting to study the capability and applicability of the system.

7. ACKNOWLEDGMENTS

We are grateful for the help we received from Timothy Beamish, Mario Enriquez and Colin Swindells. We would also like to thank subjects who volunteered their time to participate in our initial user studies.

8. REFERENCES

[1] Athletic Directory Beat Counter, web source. Accessed April 2005. http://www.athleticdirectory.com/calculators/beat_counter.htm.

[2] BASS Homepage, web source. Accessed April 2005. http://www.un4seen.com/bass.html.

[3] Bainbridge, D., Cunningham, S.J. and Downie, J.S. Visual collaging of music in a digital library. In proceedings of 5th Int'l Conf. on Music Information Retrieval (ISMIR), (2004).

[4] Beamish, T., MacLean, K. and Fels, S., Manipulating Music: Multimodal Interaction for DJs. In proceedings of ACM SIGCHI (CHI '04), (Vienna, Austria, 2004), 327 - 334.

[5] Boman, B., Pak, J. and Sadi, S., MusicVis: a visual interface for playlist generation, Web source. Accessed April 2005. http://www1.cs.columbia.edu/~paley/spring03/assignments/HWFINAL/bgb10/MusicVis_Poster.pdf.

[6] Brown, L.M., Brewster, S.A. and Purchase, H.C., A First Investigation into the Effectiveness of Tactons. In proceedings of the 1st Joint Eurohaptics Conf. and Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, WHC 2005, (Pisa, Italy, 2005).

[7] Chan, A., MacLean, K.E. and McGrenere, J., Learning and Identifying Haptic Icons under Workload. In proceedings of Word Haptics Conference, IEEE-VR2005, (Pisa, Italy, 2005).

[8] ITFacts.biz. “Samsung predicts MP3 player sales at 35-45 mln units in 2005,” Web source. Accessed April 2005. http://www.itfacts.biz/index.php?id=P2858.

[9] MacLean, K., Enriquez, M., Perceptual Design of Haptic Icons. In proceedings of Eurohaptics, (Dublin, Ireland, 2003), 351-363.

[10] Pauws, S., Bouwhuis, D. and Eggen, B., Programming and enjoying music with your eyes closed. In proceedings of SIGCHI Conf. on Human Factors in Computing Systems, (2000), ACM Press, 376 - 383.

[11] Pauws, S. and Eggen, B. Realization and user evaluation of an automatic playlist generator. Journal of New Music Research, 31 (1), 2002.

[12] Shaver, M.J. The Twiddler: A Haptic Teaching Tool: Low-Cost Communication and Mechanical Design Mechanical Engineering, UBC, Vancouver, 2003.

[13] Stavness, I., Gluck, J., Vilhan, L. and Fels., S., The MUSICtable: a map-based ubiquitous system for social interaction with a digital music collection. To appear in proceedings of Int’l Conf. on Entertainment Computing (ICEC 2005), (Sanda, Japan, 2005).

[14] Tan, H.Z., Durlach, N.I., Beauregard, G.L. and Srinivasan, M.A., Manual discrimination of compliance using active pinch grasp: the roles of force and work cues. Perception & Psychophysics, 57 (4). 495-510.

[15] Tzanetakis, G. and Cook, P., Sound analysis using MPEG compressed audio. In proceedings of Int'l Conf. on Audio, Speech and Signal Processing (ICASSP), (Istanbul, Turkey, 2000).

[16] Tzanetakis, G. and Cook, P., Automatic Musical Genre Classification of Audio Signals. In proceedings of Int'l Symp on Music Information Retrieval (ISMIR 2001), (Bloomington, Indiana, 2001).

[17] van Erp, J.B.F. and Spapé, M.M.A., Distilling the Underlying Dimensions of Tactile Melodies. In proceedings of Eurohaptics, (Dublin, Ireland, 2003), 111-120.