Designing Easily Learnable Eyes-free Interaction

Kevin A. Li
Computer Science and Engineering
University of California, San Diego
k2li@cs.ucsd.edu

ABSTRACT
Environmental factors in mobile scenarios often prevent the user from being able to interact with a device via visual interfaces. Tactile feedback offers a promising eyes-free alternative. However, haptic research typically focuses on increasing the bandwidth of the tactile communication channel. This results in complex patterns that lack mnemonic properties, making them difficult to learn. I propose the way to enable easily learnable eyes-free interaction is to create tactile feedback based on stimuli users already have semantic associations with. In my work, I examine how music, human touch and speech can be mapped to the tactile channel and how this can be done in a way that exploits pre-learned tactile feedback.

ACM Classification Keywords: H5.2 [Information interfaces and presentation]: User Interfaces, Haptic I/O; B 4.2 Input Output devices.

General Terms: Design, Human Factors

Keywords: eyes-free, haptics, vibrotactile, user interfaces

INTRODUCTION

Mobile phone interfaces are typically designed based on PC design, where the device has the user's undivided attention. This assumption no longer holds in mobile scenarios where a user's attention may be divided between multiple activities. Visual interfaces are unusable in situations such as driving where the user's visual attention is required. In other scenarios, as in the middle of a meeting, attending to a device can be socially unacceptable. Haptic feedback has been proposed as a private form of cue delivery [6], useful for conveying information to the user in an eyes-free manner [2]. Work in this area has typically focused on increasing the bandwidth of the tactile channel [3]. This can result in complex vibrotactile sequences that are difficult to learn because they lack mnemonic properties.

I propose the way to enable easily learnable eyes-free interaction is to leverage pre-learned associations users already have from their interaction with auditory and tactile stimuli from the physical world. The stimuli I plan to examine are human-human touch, music, and speech (Figure 1).

UIST'08, October 19-22, 2008, Monterey, California, USA
ACM 978-1-59593-975-3/08/10.

APPROACH AND METHODOLOGY

People communicate with the world around them in a number of eyes-free scenarios in the *absence of technology*. Although taste and smell can provide valuable information, when the visual channel is unavailable, it is not clear what their computational equivalents are. Instead, I focus on how people interact with the world via their senses of touch and hearing. People receive tactile feedback either from touching an object or another person. Auditory feedback comes in the form of music, speech or other sound effects. People already have semantic associations with many of these stimuli from their experiences with them.

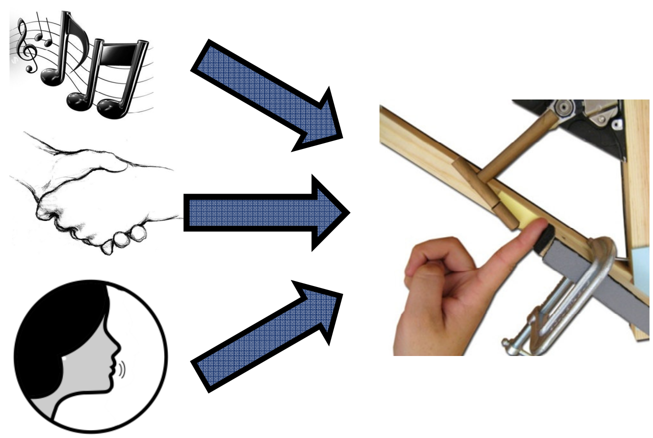

Figure 1. We can create feedback with pre-learned meaning by mapping familiar stimuli such as music, human touch and speech to tactile feedback.

The key question to be answered is how to design eyes-free interaction in a way that maps the pre-learned stimuli to the tactile channel in an appropriate way. Laboratory setting user studies with both quantitative and qualitative measures will be used to evaluate whether the information has been mapped appropriately. I also plan to evaluate how users will use these eyes-free technologies in their day-to-day lives with long term field studies.

RELATED WORK

Traditionally, haptics research has focused on generating a large number of differentiable haptic icons. Brown and Brewster's work with multidimensional *Tactons* is representative of this approach [3]. By varying rhythm, roughness and spatial location, they demonstrated a variety of *tactons* could be generated. However, a complex sequence of vibrotactile patterns can be difficult to learn. Geldard proposed a vibrotactile encoding of the English alphabet [5]. Users were able to obtain 90% accuracy at a speed of

38wpm but this required 65 hours of training. In contrast, Chang's ComTouch explored what could be done without learning (i.e., training) and how users would send vibrotactile cues to each other when used in conjunction with an audio channel [4]. Since users were given no training, the resulting tactile cues they generated were fairly simplistic, consisting of binary or simple numbering schemes. It remains to be studied what can be done with minimal training (e.g. < 5min.) and haptic icon design with a focus on learnability.

One potential approach to eyes-free interaction is to use auditory feedback. Auditory feedback is easy to interpret; a system could simply read information aloud to a user, allowing her to use it eyes-free. I explored eyes-free interaction with auditory feedback in a project called blindSight [7]. BlindSight is an application that replaces the in-call menu of a mobile phone. Users interact using the phone keypad, without looking at the screen. BlindSight responds with auditory feedback.

While effective, auditory feedback has a number of drawbacks. Manipulating a device can be unsafe when driving or socially unacceptable when in a meeting. Additionally, audio cannot always be heard because of environmental factors such as a busy road, or at a loud concert. The usage of tactile feedback overcomes these limitations of auditory feedback.

The remainder of this discussion centers on mapping tactile and auditory stimuli to tactile feedback. The focus of my work is to achieve a rich tactile language that can be learned quickly by leveraging pre-learned associations users already have.

MAPPING MUSIC TO VIBRATIONS

Music is a promising form of auditory feedback to examine. The strong associations that people have with music make it easy to learn information mappings to music. In particular, buddy identity is often mapped to music, in the form of ringtones. However, an audible ringtone can be unobtrusive in a number of social environments such as meetings. Vibrotactile feedback has often been proposed as a private, subtle cue [6], but no techniques exist for mapping music to vibrotactile sequences. As described earlier, there are a number of situations where an audible cue would be annoying or unacceptable.

To explore the question of how to map audible cues to a vibrotactile sequence, I built an application called PeopleTones. PeopleTones is an application that runs on commodity Windows Mobile phones [8]. A user's friends are each mapped to different music cues. When a buddy is near, the audible cue for that buddy plays. If the ringer is off, a corresponding vibrotactile cue is played. This system consists of a buddy proximity algorithm as well as a sensor noise filtering mechanism. However, as these components are outside the scope of this paper, I omit them, focusing only on the component that maps music to vibrotactile sequences.

Mapping music to vibrotactile sequences that will play on a mobile phone is a non-trivial process. First of all, the vibrotactile actuators on commodity mobile phones only turn on and off. To create different levels of intensity, I use a software technique similar to pulse width modulation. By varying how long the motor is pulsed on for, 10 user differentiable levels of vibration can be generated. However, because of the pulsing nature of this approach, the shortest unit of vibration is 20ms; playing anything shorter than this yields no mechanical response. As a result, this approach is only capable of generating a signal of about 50Hz, three orders of magnitude coarser than the range of what a human can hear (20Hz-22kHz).

The music information retrieval literature suggests beat and lyrics are important for identifying music [1]. However, these do not map well to 50Hz vibrotactile sequences, due to the disparity in fidelity. Instead, I attempt to capture the energy levels of the music. This involves three main steps, as outlined in Figure 2. First, I remove noise using a bandpass filter, keeping only components between 6.6kHz and 17.6kHz. Next, I take a running sum of the absolute value, generating 1 value every 20ms. This keeps the length of the music cue and vibrotactile sequence consistent. Finally, I compose the resulting output with a power function to exaggerate the features. The result is a sequence of vibration intensity values ranging from 0-10, with each value representing a vibrotactile pulse of 20ms.

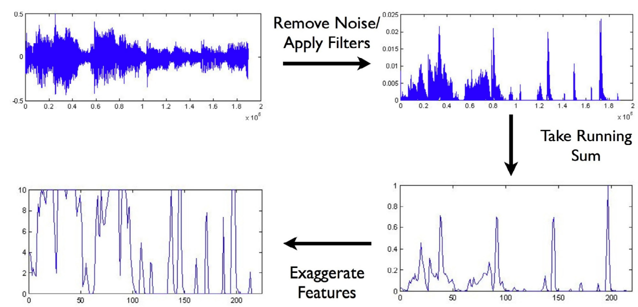

Figure 2. General process for converting music to vibrotactile sequences.

I conducted a field study on 3 groups of friends over 2 weeks. Each group had a different cue condition. The *Nature* group mapped buddy cues to different nature sounds, since these are commonly cited by the peripheral cue literature. In the *MyChoice* group, users selected what they would hear and in the *YourChoice* group, buddies would choose what their friends would hear.

I conducted a pre-study and post-study experiment examining how well participants could map the sounds to the vibrotactile cues I had generated. The purpose of this was twofold. First, I wanted to examine if people were able to map the vibrotactile cues to the associated music cue with no training. Secondly, I wanted to see if there were learning effects that occurred throughout the study regarding the mapping of audio clip to vibrotactile sequence. There were no significant differences between the pre-study and post-study results, suggesting no learning effects. Additionally, many users were consistent in the way they matched vibrotactile patterns to sound, with 7 of 17 participants responding consistently in the pre and post study experiments.

The *MyChoice* group was the only condition with an appreciable amount of cue identification from vibration, demonstrating 25% identification rate from vibration cues alone and 83% identification from both vibration and auditory cues. While limited, this serves as a proof-of-concept for the delivery of ambient information via low fidelity haptic channels. Of the participants in the *MyChoice* group, 75% of them were able to correctly map vibrations to sounds and then to people.

Some participants reported mapping vibrotactile cue directly to their buddy first, and then to the sound, even though that was not the task they were given. This suggests that the semantic association of buddy to music clip does carry over when the music clip is mapped to a vibrotactile pattern. Participants in the *Nature* and *YourChoice* conditions were less successful in mapping vibrotactile patterns to music, possibly because of the larger number of cues or because they were not as familiar with the songs selected for cues. These findings suggest that when mapping music to vibrotactile cues, using music cues users are familiar with promotes both cue identification and the ability to identify corresponding vibrotactile sequences.

HUMAN-HUMAN INTERACTION: TAPPING AND RUBBING

One of the most common forms of tactile feedback people experience is from human-to-human interaction. Since two of the most common forms of this are tapping and rubbing, I examined how taps and rubs could be generated and what types of sensations they would give [8].

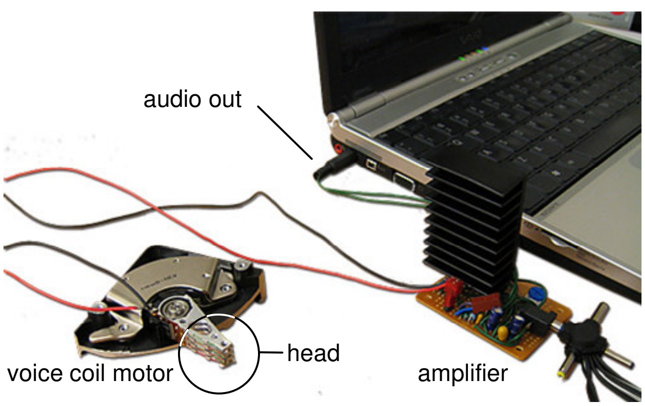

Figure 3. s*oundTouch* prototype.

To generate a variety of measureable, controlled taps and rubs, I built a prototype called *soundTouch* (Figure 3). Input is provided by generating waveforms via a laptop's sound card, which is then amplified by a custom amplifier. *SoundTouch* generates tactile feedback via an arm mounted on a voice coil motor taken from a hard drive. By moving the actuator perpendicular to the contact surface, taps are generated (Figure 4a). Moving the actuator parallel to the contact surface generates rubs (Figure 4b). Tapping and rubbing convey a distinct emotional message, similar to those induced by human-human touch. My findings from two exploratory user studies indicate tapping and rubbing are perceived as being similar to touch interactions exchanged by humans.

In the first study, I examined participants' perceptions of tapping. Because I was particularly interested in how users would describe the sensations they felt, I was very careful to not mention the word "tapping" or any other word that was suggestive of this term. When describing their perceptions, many participants used terminology drawn from human-human interaction. Thirteen of the 16 participants used the word "*tap*" in their descriptions. Additional descriptions included: "getting flicked on the finger", "tickling", "brushing something off", "drumming fingers" and "touch". Twelve participants volunteered that the experience had a human quality to it, often citing that it felt like "getting tapped on the shoulder, but on your finger".

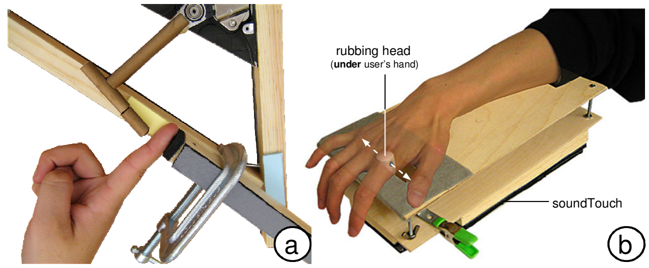

Figure 4. (a) tapping prototype. (b) rubbing prototype.

The second study was similar to the first one, but this one investigated user perceptions of rubbing instead of tapping. Half of the participants felt that the faster rubs felt more natural while the other half thought the slower ones were more natural. Those who cited faster ones being more natural mentioned that it felt more like "sliding your hand across a table" or "dropping a marble through your hands". These participants said that for the slow ones, you could feel the actuator moving against the palm and could tell it was an artificial thing. Participants who said slower was more natural used comments like "I don't come across anything that moves that quickly" to describe their experiences. They also described the sensation as being more like "rubbing your hands together or "playing with a rubber eraser".

The participants consistently described their experiences with terms like "tapping" and "rubbing" and seem to readily relate the experiences to common human-human interactions.

The softer taps were consistently reported as feeling natural. The hardest and fastest taps felt less natural. The fastest taps were frequently described as vibrations. This implies that tapping and vibration are perhaps on a frequency continuum, yet perceptually distinct.

The results from these studies validate that the prototypes I have built are indeed able to create rubbing and tapping sensations. There is some evidence of emotive connotations and further work needs to be done to formalize what emotive aspects exist in this communication medium.

NEXT STEPS

In addition to exploring the emotive aspects of tapping and rubbing, I also plan to pursue a tactile mapping of spoken language. The results of these two components will be incorporated with the other work presented here in a messaging backchannel application.

Mapping English to the tactile channel

One promising area for future work is a vibrotactile messaging backchannel. This could be used to keep in touch with a spouse or significant other throughout the day. Key to realizing this is determining how to map the English language to a vibrotactile channel.

To explore this idea, I ran a pilot study. First, 20 vibrotactile sequences were generated. Then, for each vibrotactile sequence, participants had to select 1 of 5 text phrases that best matched it, in a forced choice response. The text phrases were selected from commonly sent text messages. The same 5 phrases were used for each vibrotactile sequence: "hello?", "goodbye.", "I miss you.", "where are you?" and "are you there?" At least 10 of 16 participants agreed on a single vibrotactile pattern for each of the 5 phrases.

This is a promising finding, suggesting that there are aspects of written text that map to vibrotactile sequences in a pre-learned way. From these results and analysis of the confusion matrix, it appears that vibrotactile amplitude has a relationship with intonation and the number of syllables has a relationship with the number of pulses. Since some of these characteristics have a stronger relationship to speech, I am examining what characteristics of spoken language can be mapped to the tactile channel. Drawing from the linguistics literature, I plan to find a small number (about 5) of distinguishing characteristics of language. I then plan to explore how these might be used to map speech to vibrotactile patterns in a pre-learned way. Using a series of controlled forced-choice response studies with a number of participants, I can identify which of these characteristics are most important, and how to map them.

Messaging backchannel field study

Once I have refined the technique for mapping spoken language to tactile patterns, and formalized the emotive aspects of tapping and rubbing, I will validate them with controlled laboratory experiments. I then plan to combine these components into a messaging backchannel application. The techniques used for mapping music to vibration could act as a form of eyes-free caller-id. Pre-learned vibrotactile messages could also be generated using the results from my work with tactile messages. Finally, these messages could be augmented with emotive aspects with tapping and rubbing.

A field deployment with dyads will be used to examine how close friends and significant others use such a messaging backchannel in their daily lives. Post-study interviews will be used to further examine interesting usage habits.

CONCLUSION

With these elements in place, I will have explored how to map different stimuli to the tactile channel and demonstrated how to generate easily learnable eyes-free interaction. This will allow mobile devices to convey information to users in scenarios where they are currently unable to. Additionally, this will enable a new class of applications that users will be able to use eyes-free without requiring a lengthy training process.

ACKNOWLEDGEMENTS

I would like to thank Bill Griswold, Jim Hollan, and Patrick Baudisch for feedback.

REFERENCES

1. Berenzweig, A., Logan, B., Ellis, D. P. W, and Whitman, B., A large-scale evaluation of acoustic and subjective music similarity measures. *Proc. ISMIR'03*, pp.63-76.
2. Brewster, S. and Brown, L. M. Tactons: structured tactile messages for non-visual information display. *Proc. AUIC'04*, pp.15-23.
3. Brown, L.M., Brewster, S.A., Purchase, H.C. Multidimensional tactons for non-visual information presentation in mobile devices. *Proc. MobileHCI'06*, pp. 231-238.
4. Chang, A. and O'Modhrain, S., Jacob, R., Gunther, E., and Ishii, H. ComTouch: design of a vibrotactile communication device. *Proc. DIS'02,* pp. 312-320.
5. Geldard, F.A. Some neglected Possibilities of Communication, Science 131, 1960, 1583-1588.
6. Hansson, R., Ljungstrand, P., Redstrom, J. Subtle and public notification cues for mobile devices. *Proc. Ubicomp'01,* pp. 240-246.
7. Li, K. A., Baudisch, P., Hinckley, K. BlindSight: eyes-free access to mobile phones. *Proc. CHI'08*, pp.1389-1398.
8. Li, K. A., Sohn, T., Huang, S., Griswold, W.G. PeopleTones: A System for the Detection and Notification of Buddy Proximity on Mobile Phones. *Proc. MobiSys'08*, pp. 160-173.
9. Li, K. A., Baudisch, P., Griswold, W.G., Hollan, J.D. Tapping and rubbing: exploring new dimensions of tactile feedback with voice coil motors. To appear in *Proc. UIST'08*.