Title
PeopleTones: Exploring Peripheral Cues in the Wild Using Mobile Phones

Authors
Li, Kevin A
Sohn, Timothy Y
Huang, Steven
et al.

Publication Date
2007-03-26

Peer reviewed

PeopleTones: Exploring Peripheral Cues in the Wild Using Mobile Phones

Kevin A. Li, Timothy Y. Sohn, Steven Huang, and William G. Griswold

Computer Science and Engineering
University of California, San Diego, La Jolla, CA, USA
{k2li, tsohn, sbhuang, wgg}@cs.ucsd.edu

Abstract. A principal challenge in ubiquitous computing has been identifying a platform and mechanism that can achieve both affordable sensing and unobtrusive notification. In this paper we investigate the use of peripheral cueing mechanisms on commodity mobile phones. Key constraints are the undesirability of using headsets for delivering cues and the limited functionality of commodity vibration actuators. To explore these issues we designed PeopleTones, which conveys buddy proximity via peripheral cues. We addressed the unobtrusiveness of cues by using short sounds, a novel algorithm to translate the sounds to corresponding vibrotactile cues, and context-based adaptation. PeopleTones was deployed to three groups of friends (17 people) for 2 weeks. Underlying the results is a theme of personal control for peripheral cues. When a person could choose familiar sounds with positive associations – typically music – they both understood the cues better and felt they were less obtrusive.

1 Introduction

A common vision for ubiquitous computing is a context-aware infrastructure that can simplify and enrich our lives by helping us with tasks that might otherwise be out of our reach. Location-based services, for example, can detect the proximity of friends that are just out of sight or unnoticed. Realizing such a vision depends upon a ubiquitously available mechanism for detecting such occurrences and an unobtrusive, privacy-aware mechanism for making us aware of them.

In this paper, we investigate the adaptation of two potentially complementary technologies in support of these goals. One, ambient peripheral cues have the potential to unobtrusively communicate information to a person engaged in another activity. With an advanced context-aware computing infrastructure, it is desirable that a language of cues be employed to concisely and conveniently communicate a variety of facts to a person. Two, mobile phones are personal devices that possess both a number of sensors (e.g., microphone, camera, and GSM radio) and proactive actuators (e.g., speaker and vibration actuator). Analysts estimate that there are over 2.5 billion mobile phone subscribers worldwide [28], making phones an ideal platform for ubiquitous computing applications.

Employing these technologies in real-world settings faces challenges. Peripheral cues have been investigated primarily in the controlled office setting. Past work has shown audio to be an effective means for cue delivery, and vibrations from a mobile phone's vibration actuator are likely to be socially etiquette friendly. Could peripheral cues in the wild be designed using the same principles prior research had used in office settings? A mechanism for rich audible cues can achieve a language of cues, but the constant use of headphones or similar devices to ensure delivery and reduce obtrusiveness is untenable. Having vibrotactile cues that correspond to known audible cues is one approach to mitigating this issue. However, the commodity vibrotactile actuators found on mobile phones today only have a binary on/off setting, severely limiting their communication abilities.

We hypothesize that keeping audible cues short can improve their unobtrusiveness. To provide corresponding vibrotactile cues, we introduce an offline digital signal processing (DSP) technique that captures the essence of audio cues. These patterns were realized on the mobile phone's limited vibrotactile actuator using a novel algorithm based on the observation that pulsing the motor can create a range of amplitudes. Finally, we hypothesized that detecting ambient noise using the phone's microphone could help determine the best mechanism for cue delivery.

To explore these mechanisms, we designed PeopleTones, a system for conveying buddy proximity via peripheral cues that are uniquely assigned to buddies. Deploying such a context-aware application on mobile phones presented additional challenges that could not be overlooked as part of a naturalistic study of peripheral cues in the wild. The commodity sensors on mobile phones are notoriously imprecise and uncalibrated, in our case complicating accurate proximity detection. We adapted a simple 802.11b-based proximity algorithm [15] to achieve GSM-based proximity detection that does not need absolute location.

PeopleTones was deployed to 17 users for 2 weeks. Three groups of friends were recruited to test three different cue-to-information mappings. One group's members were mapped to various sounds in a nature ecology. In the second group, participants selected music clips *others* would hear when they themselves became nearby. In the third group, participants selected music clips that *they* would hear when their friends were nearby.

Underlying the results is a general theme of *personal control for peripheral cues*. Using ambient noise volume as a mechanism for detecting social situations was explored, but most users opted to enforce explicit control of their cue delivery modality. Users who selected the cues they heard found the system more useful and were also better at identifying our vibrotactile patterns. Additionally, since they were self-selected cues, they often wanted the cues to be longer in contrast to the short, nature cues that are often cited in ambient information systems. Although haptics has long been heralded as a promising ambient delivery mechanism, sound is still the preferred medium, possibly because of higher fidelity.

2 Prior Work

Peripheral cues have been used in office settings for many years. Audio Aura presented a system that plays auditory cues for nearby people while in an office environment [19]. The Live Wire [29] and ambientROOM [14] are also examples of office setting peripheral displays that use sounds as cues to convey information in the background. While many other systems that use peripheral cues have been proposed [8,11,14,19,23,29], these systems have not been deployed or evaluated in a mobile context. The success of peripheral cues in home and office environments suggests that they may be useful in the wild on mobile phones.

Mobile phones have become the dominant platform for deploying context-aware systems because of their ubiquitous availability. As just two examples, we have seen mobile phone applications of location-sharing [25] and reminders [26]. All of these applications use commodity mobile phones, coupled with a form of context, and evaluated the application in the wild with users. These studies often reveal uses of these applications that would not have been found in their laboratory prototypes.

Vibrotactile cues have been proposed for a variety of uses such as for conveying information in a non-visual channel. Geldard's Vibratese language proposed a vibrotactile encoding of the English alphabet, using spatially separated vibrotactile actuators to receive messages [6,7]. ComTouch explored the design space of how users might communication with minimal learning, using a simple touch-based mechanism for generating Morse-code like pulses [5].

Human response to tactile stimuli is about 5 times faster than vision, but tactile perception cannot be utilized without a high fidelity delivery channel [2]. Mobile prototypes have been built to explore how information can be conveyed through a high fidelity channel using piezoelectric materials[17,22]. These works suggest tactile interfaces might be used for ambient information delivery. The use of vibrotactile actuators, such as those found in commodity mobile phones, have been explored less, possibly due to their lower fidelity mechanical nature. Tactons pulse actuators similar to those found in mobile phones to generate distinct pulses which have been shown to be effective for alerting users to message type as well as urgency [3,4].

One of the API limitations of commodity vibrotactile actuators is that they typically only provide on/off functionality. VibeTonz technology developed by Immersion Corp supports richer, more complex vibrotactile pattern generation, but utilizes specialized hardware that is currently only available on a handful of commercially available handsets [13].

3 Peripheral Cues in the Wild

The question, for this study, is what constitutes a truly peripheral cue for the mobile context, not only sitting in the periphery of the intended recipient, but for all in the vicinity? There are at least three components to peripherality in this regard:

- First and foremost, the cue should not invade the periphery.
- Second, when the cue is perceived, it should not be seen as inappropriate in any way, most notably by those for whom the cue is not intended – a matter of

etiquette. This is difficult, if only because different people have different value systems, and there is no telling who is nearby when a cue is played.
- Third, because the periphery is constantly shifting with one's attention, perhaps as demanded by other changes in the environment (e.g., someone speaks to you, or shifting traffic conditions while driving), the cues, when perceived, should not be distracting – they should not impede shifts in attention or other natural changes to the periphery. In particular, people should not have to think about the cues that they perceive.

We refer to these three properties collectively as *unobtrusiveness*. With these issues in mind, the principal challenge with the use of peripheral cues in the mobile setting is resolving the tension between reliable receipt of cues and unobtrusiveness, without making unrealistic assumptions such as the required use of headsets. Peripheral cues can be overlooked without harm, and as designers we can err on the side of cues being missed. Yet, if we take the lowest common denominator among all possible settings, the only acceptable peripheral cues might rarely be noticed.

We hoped to gain insight on these complex considerations in the course of our study, but we did have some initial hypotheses. One, short audio cues would be less invasive and more polite than long cues. Two, having corresponding vibration cues could be useful both for politeness and increasing chances of being perceived in noisy environments. Three, environment sensing could support the adaptation of the cues being played to ensure consistent maintenance of peripherality and politeness.

3.1 Auditory Cues

Since much past work with peripheral cue systems has used sound cues to deliver information, we followed in suit. Playing sound cues from a mobile phone is natural, but has potentially different requirements than environmental-based systems. Past work has found that short, rich auditory cues that build off of sounds users are accustomed to hearing in their normal lives can provide information to users serendipitously [19]. Soothing nature ecologies have often been used and so we created a set of nature cues of 3-5 seconds in duration. Many mobile phones have the ability to map specific ringtones or music clips to different users on a contact list. While many people use these, the efficacy of mapping sound clips to identity is relatively unexplored. Yet, for a buddy proximity application, music clips seem promising for mapping the identity of a person to an audio cue and so we explored music cues as well.

3.2 Vibrotactile Cues

In many office setting studies of peripheral cues, a headset or other wearable device is often the delivery mechanism used for delivering auditory cues in an etiquette-friendly manner. When delivering peripheral cues in the wild, where the user can be in a variety of social settings, it is unreasonable to require them to wear an additional device for receiving auditory cues. Mobile phones offer the ability to play sounds using their speakers, which can be effective for informal situations, but it is unlikely that this delivery channel will always be socially acceptable. Much like the silent or vibrate-only modes on mobile phones, peripheral cues delivered via these devices must also have a socially etiquette friendly mode.

Motivated by haptics research suggestions, we explored using vibrotactile patterns to convey ambient information on mobile phones with the actuator that commonly ships with these devices. Ideally, there would be a one-to-one mapping of sound cues to vibrotactile patterns, where a user could easily identify a vibrotactile cue and its respective auditory cue. However, generating a variety of distinguishable vibrotactile cues can be difficult on commodity mobile phones, given the limited API; most mobile phones only support the functionality of turning the actuator on or off. With the exception of phones with specialized built-in hardware [13], the API for most phones does not support playing vibrotactile pulses of different amplitudes nor do they provide any low-level functionality to specify the amount of current used to drive these actuators.

3.2.1 Generating Rich Vibrotactile Patterns Using Mobile Phone Actuators

To circumvent the constraint on actuator functionality, we present a novel algorithm to generate a wider range of vibrotactile sequences. While a full analysis of the capabilities of this approach is outside the scope of this paper, the basic algorithm for playing a pulse of varying amplitude is presented below for completeness. The general approach for playing a pulse is to repeatedly turn the actuator on for a short period of time, spinning between calls to the function that turns the actuator on. Timing is critical during this process and so the active thread is given the highest priority to avoid inopportune context-switches.

```
end = GetTickCount() + vibeLength;
while(GetTickCount() < end)
{
    playVibrate(len);
    sleep(pause);
}
```

By varying how long the actuator is on for and how much time the thread is slept between calls, different amplitudes can be represented. This approach takes advantage of the non-ideal rampup time for turning an actuator on, effectively turning it off before it reaches its full “on” state. For the purposes of this study, the operating range of the actuator was divided into 10 differentiable amplitude levels, using 20ms as the fundamental pulse length. To generate an amplitude of `1`, values `len=1`, `pause=9` were used. To generate an amplitude of `9`, `len=1`, `pause=1` values were used. A pulse with amplitude of 10 was generated by simply turning the actuator on for the desired pulse length, in this case 20ms. Using this approach, a vibrotactile pattern can be defined as a sequence of pulses of varying amplitude.

3.2.2 Mapping Sounds to Vibrotactile Patterns

With peripheral cues deployed in the wild, a number of situations will arise where sound cues will be socially disruptive and where a quieter cue delivery mechanism is desirable, or where a sound cue might not be heard over ambient noise. As a

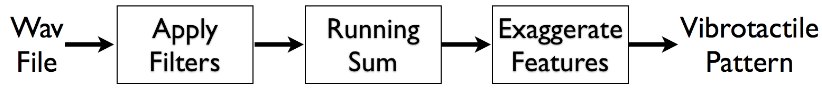


Figure 1. Block diagram showing the process of converting a wav file to a vibrotactile pattern

complement to sound cues, we generated vibrotactile. With an effective mapping in place, information associated with the sound cue would be associated with the felt vibrotactile pattern as well.

Mapping auditory cues to vibrotactile sequences is challenging. On the one hand, there are difficulties associated with trying to map from an auditory system to a tactile one, where different receptors are being used to receive information [2]. This issue becomes further complicated by the significant difference in sample rates. A typical music file is sampled at 44.1kHz whereas our algorithm generates output at a granularity equivalent to a signal sampled at 50Hz, a full three orders of magnitude coarser. To address this gross level of under sampling, we utilized a number of techniques from DSP as part of the encoding process to try to capture the essence of the music clip. We used a semi-automated method for converting a song to its vibrotactile equivalent using Matlab on a desktop PC.

Initially, we considered the beat of the sound by examining lower frequency components of a given sound. This can be effective for certain sound clips, but not all clips have a distinct beat to extract. The lyrics of the song chorus were also thought to be important to characterize, given their use in identifying songs. However, in practice, lyrics are difficult to map to vibrotactile patterns due to the lower fidelity of the commodity vibrotactile actuators. Our initial studies suggest the most effective characterization of sound cues comes from a combination of amplifying certain frequency components and focusing on the louder aspects of the song, while exaggerating the difference between loud and quiet components. The general process can be thought of as trying to create a *humming* sequence for the audio clip. Figure 1 outlines the general steps of this process.

The first step in converting a sound file into a vibrotactile pattern is to remove noise from the original signal. In this context, we consider "noise" to be elements of the sound that are not significant to the vibrotactile encoding of the sound, in addition to the traditional definition of the term. A series of band-pass filters can be used to focus on frequencies in the range between 0.15fs to 0.4fs, effectively reducing noise, where fs is the sampling rate of the original signal. Butterworth Filters are commonly used for band-pass filtering and we use an 8th order implementation for this step (Figure 1 - Apply Filters) [10]. Additionally, we scale different frequency components of the sound signal, emphasizing higher frequency sounds by scaling them with higher coefficients. The next step in the process is to try to characterize the resulting processed signal in a way that preserves the characteristics of the sound file. In some senses, we want to translate a 44.1kHz signal into a 50Hz signal. Clearly a significant amount of information is lost in this translation, but the hope is that the essence of the original content is preserved so that the vibrotactile pattern can still be associated with the original sound. We take a running sum of absolute values post-filtering, creating a sub-sum such that we have 50 samples per second (Figure 1 - Take Running Sum). Finally, certain features of the signal need to be exaggerated in such a way that the characteristics of the sound are emphasized and so we compose it with a power function (Figure 1 - Exaggerate Features). The result is a sequence of values representing a vibrotactile pattern that preserves many of the characteristics of the original sound signal.

3.3 A First Approach to Detecting Ambient Noise

Given the variety of social situations that can arise throughout a person's daily life, it is desirable to have some strategy for detecting ambient noise levels to deliver peripheral cues using the most appropriate channel. While sound cues delivered in the middle of a meeting can be disruptive, a sound cue delivered in the middle of a loud concert would be of fairly low utility.

A simple implementation for detecting ambient sound levels was implemented using the microphone on the mobile phone. To detect ambient noise, the microphone is turned on for 5 seconds. The amplitude over this interval is then averaged and compared to different constants to determine the level of noise in the environment. Constants were obtained through analysis of a number of samples obtained in a quiet office environment and the center of a University during a busy hour. When a cue was triggered to be delivered, ambient noise level was measured using the technique described above. In quiet environments, only the vibrotactile cues were played. In loud environments, both vibrotactile and sound cues were played, with the assumption that a vibrotactile cue would never be socially disruptive in a loud environment. Additionally, vibrotactile cues offer the advantage of being felt in noisy environments when even a very loud sound cue might not be heard.

4 PeopleTones

To explore our hypotheses and mechanisms for peripheral cue delivery in the wild, we developed PeopleTones, an application for informing users of buddy proximity via their mobile phone. To inform the user of buddy proximity, a sound clip and corresponding vibrotactile pattern is associated with each buddy. If the automatic ambient noise detection mode is enabled, it will gauge the situation and deliver a cue in the best possible mode. Alternatively, the user can explicitly specify to have only vibrotactile cues, only audio cues, or both be played by selecting the appropriate phone profile.

PeopleTones is implemented as a client-server application using standard SOAP web services. The client-side application is written in C#.NET on the Windows Mobile Smartphone platform. The interface is shown in Figure 2. Each phone periodically pushes its GSM cell tower readings to the server, which computes buddy proximities and then notifies the phones of changes. Two substantial technical issues had to be resolved to successfully implement and deploy this application. The first is the accurate, privacy-aware detection of buddy proximity with GSM cell tower sensing, and the second is battery life.

(a) **(b)**

Figure 2. **(a)** PeopleTones interface showing nearby and faraway buddies with corresponding last update times **(b)** Menu showing options for peripheral cue delivery

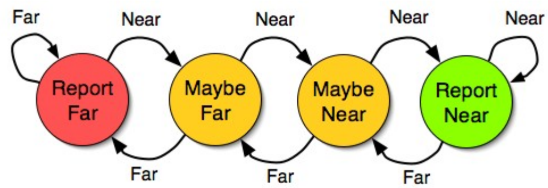
4.1 Proximity Detection

One method of acquiring location on some phones is through the network carrier, but they often do not release the required APIs. Another method is to use a location infrastructure. Place Lab [16], Activecampus [9], and Plazes [21] are examples of systems that offer both absolute and relative positioning schemes. These infrastructures limit the area of sensing to areas with pre-mapped access points. This hinders ubiquitous deployment of our application since we cannot anticipate every geographic area a person may visit, and even if we could, pre-mapping can be costly.

PeopleTones does not need a person's geographic location to find the proximity of nearby buddies. Hence, we used a relative positioning method in the spirit of the Nearme server [15]. Nearme used a variety of metrics for comparing the distance between two wireless measurements, such as Euclidean distance, spearman rank correlation, and the ratio of common access points. In our experiments, we found that computing the ratio of common GSM cell towers between two readings provided the best proximity indicator. (Our i-mate SP3i's are capable of reporting 7 towers.) Figure 3(b) shows the equation for computing the ratio of cell towers given two readings a and b, each consisting of a set of cell tower sector identifiers.

In practice, we found that a ratio of 0.4 or higher provided reliable indication that two phones were within 0.2 miles of each other in most settings. This is the cutoff used by PeopleTones for detecting proximity.

4.2 Sensor Noise and Power Considerations



$$\text{proximity-ratio}(a,b) = \frac{|a \cap b|}{\left(\frac{|a|+|b|}{2}\right)}$$

(a) **(b)**

Figure 3. (a) Two-bit counter for eliminating noise in proximity detection **(b)** Equation to compute ratio of common GSM towers where a and b are two cell tower readings

GSM readings can vary widely from moment to moment in ways unrelated to the phone's proximity to the cell towers in the region, creating the possibility of false proximity detection. Additionally, if buddies hover around a 0.2mi distance from each other for a prolonged period of time, multiple cues might be triggered, creating an annoyance. To mitigate such errors, we took two steps in client-side sensor filtering that address these concerns. First, we utilized a location reporting mechanism whereby a friend's nearby state is updated only after a number of consistent, consecutive readings. Figure 3(a) illustrates the logic used for this approach. Buddies are initially reported as far away. Edge transitions represent a sensor sampling, yielding *near* or *far*. The state of a buddy is only updated to far or near when the states "Report Far" and "Report Near" are reached. This approach is motivated by the 2-bit counters used by branch predictors in computer architecture [20]. Although this approach addresses a number of accuracy issues associated with irregular sensor fluctuations associated with location, it also introduces a notification delay of up to 3 update intervals, the number of consecutive similar required to move from one report state to another. To reduce the worst-case delay, two sampling rates were used. Initially, buddy proximity is sampled every 90 seconds. When the counter moves into a "maybe" state, the sampling rate is increased to 1 sample every 20 seconds, until steady state is reached again, at which point the sampling interval reverts to 90 seconds. Initial data collection suggested a 2-bit counter represented a good tradeoff between filtering noisy readings and delayed proximity notification, while minimizing the delay. To avoid redundant notifications for buddies hovering around the near/far cutoff, the cues for a pair of buddies are delivered at least 90 minutes apart.

The limited power supply from mobile phone batteries requires careful consideration in a continuously running context-aware system. Since sending data over General Packet Radio Service (GPRS) drains more power than utilizing the processor, we attempt to minimize unnecessary use of the network. Our use of two sampling rates helps to some degree, but measures are needed for situations with poor network signal. For one, sending data over a poor link tends to consume more battery power, in part because web service calls are more likely to fail, causing the underlying system to attempt additional calls. Two, a poor link is indicative that there are no cell towers that are strongly suggestive of the phone's relative location, so there is no point in making a report anyway. These *black hole* situations are common in the USA, such as inside buildings with lots of metal or concrete. The PeopleTones client detects these situations by comparing the phone's signal strength to a threshold. To compensate for when clients in these situations do not update, the server retains the last reported reading from the phone along with a timestamp, so that others can still make inferences about proximity for a time, assuming that the non-reporting is due to being in a building.

5 User Study

We deployed PeopleTones to three groups of friends, forming three different test conditions, each for the course of two weeks. The purpose of this study was to gauge the feasibility of deploying peripheral cues from mobile phones and to examine what, if any, design requirements for peripheral cues in the wild might be different from office-based systems.

Group	Group Size (Gender)	Age Range	Makeup	Condition
Nature	5 (F)	19-21	Roommates	Nature Sounds
Your Choice	8 (5F, 3M)	22-26	Friends From Church	*You* choose what I hear
My Choice	4 (1M, 3F)	19-22	Close Friends	*I* choose what I hear

Table 1. Group makeup for the three different groups

5.1 Participants

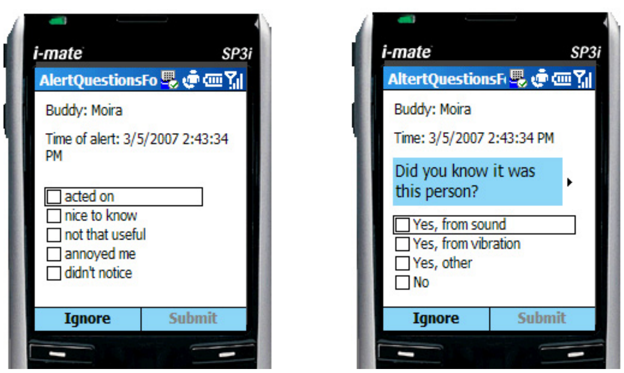
We recruited three groups of friends forming groups of sizes 4, 5 and 8 people. These 17 participants consisted of students and young working professionals, 12 women and 5 men, aged 19-26. Participants were recruited based on interest in a buddy proximity application as well as having physically proximal friends. Participants were given $25 American Express Giftcards as a thank you for their time.

5.2 Methodology

Participants used PeopleTones over the course of 2 weeks. Prior to the study, a pre-study interview was conducted to gather basic demographic information, mobile phone usage habits and general "closeness" to the participant's friends whom were also participating in the study. Additionally, a test was conducted to evaluate whether participants could match the semi-automatically generated vibrotactile patterns to sound clips. A mid-study evaluation was also conducted to make sure there were no problems with the system. Finally, a post-study interview was conducted to reflect on the participant's experience. A test of matching vibrotactile patterns to music cues was again performed to measure learning effects that may have taken place over the course of the study, and to evaluate consistency.

The three groups of friends formed three different conditions for cue-to-information mapping methods. Group *Nature (N)* consisted of 5 friends who were given a set of nature sounds to assign to their friends but for the most part they opted for automatic assignment of cues. Group *Your Choice (YC)* consisted of 8 friends who selected a single sound for themselves, representing the cue that their friends would hear when they were nearby. Group *My Choice (MC)* consisted of 4 friends who each selected the cues that they would hear, when their friends were nearby. Table 1 summarizes these group conditions.

For conditions where participants selected their own cues, they opted to select clips from music. Before the study, participants identified music cues that they would want to use for the study. They were given the option to identify specific parts of the song that they wanted to use. Alternatively participants could select from 2-3 different 3-5 second segments of the song selected by the authors, typically chosen for their mappability to vibrotactile patterns. After participants selected song segments they wanted to use, a corresponding vibrotactile pattern was generated, using the procedure described in section 3.



(a) **(b)**

Figure 4. (a) Post-alert form asking how the user responded to the cue. **(b)** Post-alert form asking if the user could identify the buddy identity.

6 Usage and Self-reported Data

	Acted On	Nice To Know	Didn't Notice	Not That Useful	Ignored	Annoyed
Nature	4%	38%	22%	25%	10%	1%
Your Choice	9%	34%	31%	5%	20%	1%
My Choice	12%	60%	15%	2%	8%	3%

Table 2. Self-reported response to the cue.

	Yes From Sound	Yes From Vibration	Yes Other	Ignored	No
Nature	3%	2%	17%	2%	76%
Your Choice	76%	0%	7%	1%	16%
My Choice	64%	16%	3%	3%	14%

Table 3. Self-reported identification of the cue's information.

To perform data analysis, we performed client-side logging when a cue was triggered. Once a cue was triggered, the participant was presented with a form asking them if they acted on it, if it was nice to know, if it was not useful, or if it was annoying. This post-cue questionnaire was left on the screen so they could later respond if they did not notice the cue. Alternatively, they had the option to ignore the form if they were busy. If they did not choose to ignore the cue, they were presented with a form asking them if they could tell who it was from the sound or the vibration or from other unspecified cues, or if they could not identify who it was. Figure 4 shows the post-cue questionnaire form used to self-report responses to the cues after they were triggered.

A total of 683 cues were sent over the course of 2 weeks, across all conditions with 122 cues in Nature, 466 cues in Your Choice, and 95 cues in My Choice. Using self-reported forms displayed on the mobile phone, the user was queried both for their response to the cue as well as whether they could identify buddy that the cue represented. The breakdown for these post-cue responses is shown in Table 2 and Table 3 respectively.

7 Post-Study Responses and Discussion

In the following section we reflect on our two major research questions and one secondary question: the suitability of peripheral cues as an in-the-wild communication mechanism, the suitability of mobile phones for providing such cues, and the suitability of mobile phones as a platform for context-aware ubiquitous computing. We draw on the observations and data above, as well as from our interviews with the study participants.

7.1 Peripheral Cues are a Viable Communication Mechanism in the Wild

Unobtrusively delivering peripheral cues in the wild, though challenging, can be achieved through a combination of informed cue design and personal control for selecting and controlling peripheral cues.

7.1.1 Designing and Choosing Cues for the Wild: Music and Personal Control

Although office-setting studies have found soothing nature ecologies to be effective for comprehension and unobtrusiveness, cues in the wild should be composed of music, and perhaps repeated.

With regards to comprehension, the self-reported usage data shows that groups Your Choice (YC) and My Choice (MC) both demonstrated an 83% comprehension rate, where comprehension is defined to be when the user can identify the buddy from the cue (Table 3). In contrast, the nature group demonstrated a significantly lower rate of 22%. Interestingly, this lower rate did not result in significantly lower usefulness ratings (42%) for the application when compared to the Your Choice group (43%). It may be that the ability to look at the phone after receiving a cue mitigated the negative effects on cue comprehension. This might not carry over to a more comprehensive system that used cues for multiple purposes (e.g., buddies and stock alerts). Many participants cited that they would prefer longer cues since they could be difficult to catch in the dynamic environments of their daily lives. Participant <MC-1> commented: *"sometimes couldn't hear because the song was too short."*

The obtrusiveness of music cues was not a concern. The reasons are somewhat surprising. <MC-3> comments: *"When it went off in [the library] it didn't actually seem to annoy other people too much, they just thought it was just another phone."* This observation points to the fact that mobile phones have become largely invisible and socially accepted, at least for young adults, even in a "quiet zone" like a library. (We reflect more on etiquette concerns in the next section.) Another reason cited for the unobtrusiveness of music cues was the positive feelings generated by the music. <MC-1> comments: *"I would like longer songs so I could hear it and because I like the songs."* The Your Choice group made similar comments, even though they did not pick their music cues. <YC-5> comments that she liked: *"Just hearing the songs. I liked the fact that each person could choose whatever they want for their own identity. Since it was a small group of us, it's kind of fun and it felt like this is a group of us."* Overall, 9 of the 12 music participants volunteered a liking for hearing music. It appears that cues with emotionally positive associations are generally unobtrusive.

Music cues are similar to the ringtones sometimes used for caller ID on mobile phones. However, ringtones in the wild are relatively unexplored. <YC-3> comments: *"It was fun how everyone had a song specific to them. Add's a little bit of personality. I don't use ringtones, that's why it was neat for me. Too much trouble to do on my phone."* The results of this study validate the usefulness of ringtones, at least if the volume is kept down, in being able to successfully convey information about people in a pleasurable way.

The usage of music cues also seems to support learning, with 83% of users in conditions My Choice and Your Choice being able to identify who the cue was for, based on self-reported post-notification questions. This learning effect is also reflected to some extent with vibrotactile patterns. The My Choice group was the only condition with an appreciable amount of cue identification from vibration, demonstrating 16% identification rate in the wild from vibration cues alone. While not overwhelming, this acts as a proof-of-concept for the delivery of ambient information via low fidelity haptic channels. Analysis of before and after vibration studies suggest some users are consistent in the way they match vibrotactile patterns to sound, with 7 participants responding consistently, when comparing their before and after responses. 75% of My Choice was able to correctly map vibrations to sounds and then to people in both the pre-study and post-study vibration-to-sound tasks. Additionally, participants in this group exhibited association of buddy identity with vibrotactile patterns. When presented with the task of matching vibrations to sounds in the post-study interview, participant <MC-3> exclaimed *"Oh that's Cathy!"* when she felt the vibration associated with the music cue associated with Cathy. Participants in the Your Choice condition were less successful in mapping vibrotactile patterns to music in these vibration-to-sound matching tasks, possibly because of the larger number of cues or because of the fact that they were not as familiar with the songs selected for cues. When comparing error rates for a matching vibration-to-sound task from before and after studies, minimal improvements were observed suggesting weak or minimal learning effects.

7.1.2 Personal Control over Cueing Mechanism for Unobtrusiveness

The discussion above suggests that personal selection of cues aids both comprehension and unobtrusiveness. In addition, for many users, explicit control of the notification modes was important. Although personal control has been cited as important in the design of a number of social mobile systems, these concerns typically have to do with privacy [12]. In our study, personal control over the cueing mechanism was a critical element for controlling unobtrusiveness and interruptibility.

Even though an automatic ambient noise level feature was provided, many users opted not to use this mode, not even trying it before dismissing it. In fact, 12 of the 17 participants did not even try the Automatic mode, despite the fact that the phone was put in Automatic mode when given to the participants. When asked about the automatic mode in the post-study questionnaire, <MC-4> commented *"I didn't use it. I was afraid to use it since my professors this quarter are pretty anal. I kept it mostly on vibrate when I was in class, or in normal mode when I wasn't."* For this participant, personal control of the notification mode was important because they feared PeopleTone cues being triggered in the audio mode in a classroom setting where it might be disruptive. Participant <YC-1> expressed a similar concern, saying *"I was afraid that if I was at church, it wouldn't work. It would just backfire on me and I wanted to have been more sure about it."* Like <MC-4>, <YC-1> was afraid of unwanted notifications while at church, another social context where an audio notification would be unacceptable. It should be noted that both <MC-4> and <YC-1> considered social contexts where they expected the notifications to be triggered, in this case as defined by their group's shared interests. In addition to a lack of trust in the application's accuracy for detecting ambient noise in high stakes situations, a number of users also expressed uncertainty as to what the system considered to be loud or quiet environments. <YC-4> said *"I'm not too sure what happened or how loud the environment needed to be. I'd want to determine how reliable the function is before using it and to check how often it looks at the environment."* <YC-4>'s comment suggests a potential solution to this problem is for some type of system feedback whereby through manually user-controlled notification management, the application gains the user's trust, demonstrating that it works well in a variety of environments. This could be done with some type of visual indicator in the application of the type of cue that would be delivered, which the user could check while in different environments. Of course, this solution requires the user's visual attention during the familiarization period.

Still, even with a trustworthy system in place, some users had different requirements than those retained by the system. <MC-2> commented *"Sometimes when it's quiet, I don't need it to be quiet, like when I'm at home by myself. I think I felt it one time but I generally like to keep it on [loud]."* While the user trusts the Automatic mode to work as expected, she would prefer the mode that plays both vibrotactile and audio cues, simply because in certain quiet situations, auditory cues will not be irritating to anyone.

Finally, for some participants, different notification modes offered different fidelities of information. <YC-8> commented that *"Did use the vibrations, but didn't work out well. I felt it vibrate, but I could indicate [who it was] better with sound. The sound lets me instantly figure out who it is. With the vibrate, you have to wait 5-10 seconds to figure out who it is."* Although this comment touches on the issue of learning mappings between the different notification modes, <YC-8> expresses a distinct preference for sound cues, crediting their higher fidelity. Similarly, <Nature-1> said *"I wish it had given me louder alerts,"* suggesting that some mechanism for controlling the volume would have been useful for some.

7.1.3 User Information Needs: Peripheral Cues Provide an Overview

When our participants were asked about how physically near friends needed to be to be considered "nearby," many people cited the mode of transportation as being relevant. For people traveling by foot, distances within about 0.5 miles were cited as being "nearby." For people traveling by car, a distance of 2-5 miles was considered nearby.

The proximity algorithm used by PeopleTones, which detected proximity at around 0.2 miles, was accurate enough for user needs. When asked about the accuracy of the system, most participants commented that it detected "near" for a buddy most of the time when the buddy was known to be near, and "far" when far away. Activity detection [27] could conceivably be used to adjust the cutoffs according to one's speed of movement, although the accuracy reported by users suggests that this may not be necessary.

Even though PeopleTones' proximity algorithm was deemed accurate enough, 4 people spontaneously volunteered that they also wanted to know the actual distances of their buddies, and 3 of those wanted the location. *"I'd like if it could tell me exactly how close they actually are,"* said <MC-2>. <YC-5> offered that *"The only thing I didn't like about using this phone was not knowing exactly where that person is."* (PeopleTones's proximity ratio is not suitable for computing exact distances or locations, but ratios significantly above the "near" cutoff could be used to infer "close".) This information need is not surprising, since such information would inform, for example, a decision on whether to call the person to arrange a meeting. This information need not be provided by the cue itself. Schneiderman's Visual-

Information Seeking Mantra "Overview first, zoom and filter, then details-on-demand" [24] suggests that the peripheral cue should be treated as the overview, with additional information displayed in the user interface providing details on demand. Also, because the cue is an *ephemeral* overview, and may not be fully comprehended, the visual interface provides valuable redundancy. For this reason, the PeopleTones user interface used by the participants displayed the near/far state of the buddy and the time of the inference. <MC-4> volunteered *"If I wanted to know if anyone was nearby, I liked how it showed the last time it had checked next to their name."*

7.2 Mobile Phones Are a Viable Platform for Peripheral Cues

As the above results foretell, mobile phones appear to be a viable platform for delivering peripheral cues. Several factors contribute to this. Mobile phones and the sounds they make are socially accepted in many settings, aiding unobtrusiveness. As eminently personal devices they enable personal control, which aids the comprehension of cues, creates the positive associations that permit managing the periphery, and ensures their unobtrusiveness when silence is paramount.

Additionally, with the use of our novel DSP techniques, the commodity vibration actuators found on mobile phones are an adequate channel for etiquette-sensitive situations. For music-based vibrations, people were very good at matching the vibration patterns to their songs. One group, the My Choice group, found the vibrations to be useful in the wild, serving as the delivery mechanism about 16% of the time.

7.3 Mobile Phones Are a Viable Platform for Ubiquitous Context Awareness

Previous studies have suggested that mobile phones are a suitable platform for context awareness for individual phones [26,27]. The results of our study extend these results to phone-phone inference – that is, interpersonal situational inference. Our participants told many stories about how PeopleTones affected their behaviors or dispositions, providing a qualitative picture of the potential power of ubiquitous computing on mobile phones. Here are some typical, selected quotes, at most one per participant:

> *"One time I was walking around with a bra on. I saw <N-3> coming [on the phone] and I thought oh Jake might be coming too. I should put on a shirt."*

> *"One time at the library, I wanted to eat with someone and so I went outside to call someone. The phone vibrated. I just called the person to meet up."*

> *"When I was going to John's birthday, I knew who was there when I pulled up because of the ring tones."*

> *"Whenever I drive to school I found out where <YC-7> works because I always get her alert when I'm driving on Miramar. Oh, so she works around here?"*

> *"I thought it was so neat every time it would ring. It made me really happy. Oh! They're right here, or oh! They're right there."*

"It was cool to see who was home by the time I got home. I could tell if <YC-1> was home when I passed by University. So if we were going to go eat or something I could ask her. Oh she's home, so let's call her and see if she wants to eat."

8 Conclusion

Peripheral cues delivered via mobile phones are a promising approach to providing ubiquitous, unobtrusive situational awareness. We invented a novel method of mapping sounds to vibration patterns on the simple on/off vibration motors of commodity mobile phones. After conducting a naturalistic study of the use of peripheral cues for supporting buddy proximity awareness, we contributed the following insights:

- Peripheral cues in the wild are better comprehended and less obtrusive if derived from music and are chosen by the intended recipient. Moreover, people have an overriding need to directly control the modality of cue delivery to manage etiquette. Context-adaptive cueing requires support and mechanisms for gaining a person's trust. Peripheral cues can provide a sparse overview of the underlying situation, but the ability to get details on demand is important to users, especially since the cues are ephemeral and sometimes not understood.
- Mobile phones are a suitable platform for delivering peripheral cues, as they are socially accepted in many settings and naturally provide much of the personal control over cue selection and delivery that people desire. Mobile phones' complementary support for context awareness is an added consideration.
- Our method for encoding sounds into vibration patterns on the limited vibration motors of mobile phones produces a representation of sound that is sensible to many, but not all people. Surprisingly, there appear to be minimal learning effects from repeated exposure to the vibration patterns.

As an aside, many interesting and positive behaviors arose out of the use of peripheral cues for buddy proximity, providing a tantalizing peek at what the future of ubiquitous computing might hold.

References

1. Brewster, S. A. and Brown, L. M. 2004. Non-visual information display using tactons. In *CHI '04 Extended Abstracts*.
2. Boff, K.R., Kaufman, L., Thomas, J.P., eds., Handbook of Perception and Human Performance: Sensory Processes and perception, vol. 2, John Wiley & Sons, 1986.
3. Brown, L. M., Brewster, S. A., and Purchase, H. C. 2006. Multidimensional tactons for non-visual information presentation in mobile devices. In *Proc. of* MobileHCI 2006.
4. Brown, L. M. and Kaaresoja, T. 2006. Feel who's talking: using tactons for mobile phone alerts. In *CHI 2006 Extended*.Abstracts.

5. Chang, A., O'Modhrain, S., Jacob, R., Gunther, E., and Ishii, H. 2002. ComTouch: design of a vibrotactile communication device. In Proceedings of DIS 2002.
6. Geldard, F.A., Adventures in tactile literacy. *American Psychologist*, 1957. 12: p.115-124.
7. Geldard, F.A. Some Neglected Possibilities of Communication, *Science*, 131, 1960, 1583-1588.
8. Gilfix, M. and Couch, A. Peep (The Network Auralizer): Monitoring Your Network With Sound. In Proc. of LISA (Conference on Systems Administration) 2000.
9. Griswold, W.G., Shanahan, P., Brown, S.W., Boyer, R.T. ActiveCampus: Experiments in Community-Oriented Ubiquitous Computing. *IEEE Computer* 37(10), 73-80.
10. Haykin, S., Van Veen, B. Signals and Systems. pp. 511-520. John Wiley & Sons, New York, CA. 1999.
11. Ho, J. and Intille, S.S. Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In Proc. of CHI 2005.
12. Iachello, G., Smith., I., Consolvo, S., Abowd, G., Hughes, J., Howard, J., Potter, F., Scott, J., Sohn, T., Hightower, J., and LaMarca, A. Control, Deception, and Communication: Evaluating the Deployment of a Location-Enhanced Messaging Service. In Ubicomp 2005.
13. Immersion VibeTonz. http://www.immersion.com.
14. Ishii, H. and Ullmer, B. Tangible Bits: Towards Seamless Interface Between People, Bits, and Atoms. In Proceedings of CHI 1997, pp. 234-241.
15. Krumm, J. and Hinckley, K. The NearMe Wireless Proximity Server. In *Ubicomp 2004*.
16. LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G., Schilit, B. Place Lab: Device Positioning Using Radio Beacons in the Wild. In *Pervasive 2005.*
17. Luk, J., Pasquero, J., Little, S., MacLean, K., Levesque, V., and Hayward, V. 2006. A role for haptics in mobile interaction: initial design using a handheld tactile display prototype. In *Proceedings of CHI 2006*.
18. Marmasse, N., Schmandt, C., and Spectre, D. WatchMe: communication and awareness between members of a closely-knit group. In Proc. of Ubicomp 2004.
19. Mynatt, E. D., Back, M., Want, R., and Fredrick, R. Audio Aura: Light-Weight Audio Augmented Reality. In Proceedings of UIST '97, pp. 211-212.
20. Patterson, D., Hennessey, J. Computer Organization and Design. pp501-02. Morgan Kaufmann, San Francisco, CA. 1998.
21. Plazes. http://www.plazes.com
22. Poupyrev, I., Maruyama, S., and Rekimoto, J. Ambient Touch: Designing tactile interfaces for handheld devices. Proceedings of UIST 2002.
23. Sawhney, N. and Schmandt, C. Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging. In Proc. of CHI 1999.
24. Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy of information visualizations, *Proc. IEEE Symposium on Visual Languages '96*, IEEE, Los Alamitos, CA (September 1996), 336-343.
25. Smith, I., Consolvo, S., Hightower, J., Hughes, J., Iachello, G., LaMarca, A., Scott, J., Sohn, T., Abowd, G. Social Disclosure of Place: From Location Technology to Communication Practice. In *Proceedings Pervasive 2005.*
26. Sohn, T., Li, K.A., Lee, G., Smith, I., Scott, J., and Griswold W.G. Place-Its: A Study of Location-Based Reminders on Mobile Phones. In Ubicomp 2005.
27. Sohn, T., Varshavsky, A., LaMarca, A., Chen, M.Y., Choudhury, T., Smith, I., Consolvo, S., Hightower, J., Griswold, W.G., and de Lara, E. In Ubicomp 2006.
28. Telecom Consulting (OVUM). http://www.ovum.com/go/content/c,377,66726
29. Weiser, M. and Brown, J.S. Designing Calm Technology. http://www.ubiq.com/hypertext/weiser/calmtech/calmtech.htm, December 1995.