Sonar-based Measurement of User Presence and Attention

Stephen P. Tarzia[†] Robert P. Dick[‡] Peter A. Dinda[†] Gokhan Memik[†]

ABSTRACT

We describe a technique to detect the presence of computer users. This technique relies on sonar using hardware that already exists on commodity laptop computers and other electronic devices. It leverages the fact that human bodies have a different effect on sound waves than air and other objects. We conducted a user study in which 20 volunteers used a computer equipped with our ultrasonic sonar software. Our results show that it is possible to detect the presence or absence of users with near perfect accuracy after only ten seconds of measurement. We find that this technique can differentiate varied user positions and actions, opening the possibility of future use in estimating attention level.

Author Keywords

Sonar, presence, attention, user study, ultrasonics.

ACM Classification Keywords

H.5.2 Information interfaces and presentation (e.g., HCI): Miscellaneous.; D.4.m Operating Systems: Miscellaneous.

General Terms

Measurement, Human Factors, Security

INTRODUCTION

In ubiquitous computing systems, it is often advantageous for distributed electronic devices to sense the presence of roaming humans, even when they are not directly interacting with the devices. This ability allows such a system to provide services to users only when appropriate. In traditional desktop computing, attention information is also useful; it is already used by Operating System (OS) power management systems to save energy by deactivating the display when the keyboard and mouse are inactive. Security systems prevent unauthorized access by logging out or locking a user's session after a timeout period. In both of these cases, the OS must know whether a user is present and *attentive*, i.e., using the computer system, or absent.

We have identified five different human user *attention states* among which an *ideal* system would distinguish, shown in Table 1. The active state is trivially detectable using input activity. Our ultimate goal is to distinguish the remaining four attention states. In this initial paper, however, our evaluation focuses on specific *activities*, as shown in the table. A given attention state may be as associated with numerous activities.

Related Work

Activity detection is one of the fundamental research problems in ubiquitous systems. Such systems typically attempt to determine a human's current task using data from a variety of sensors such as GPS, RFID, infrared motion detectors, accelerometers, and video cameras. Our work differs from past work in that it uses hardware that is already available in many home and office electronic devices. Essentially, our work evaluates a new kind of environment sensor.

In the OS community, we know of only one existing research project that studies user attention detection: "FaceOff" tackles the fine-grained OS power management problem [2]. It processes images captured by a webcam to detect whether a human is sitting in front of the computer.

Ultrasonic sounds have such high frequency that humans cannot hear them. They have already been used in context-aware computing for several different tasks. Madhavapeddy et al. used ultrasonic and audible sound as a short-range low-bandwidth wireless communication medium [4]. The Cricket localization system by Priyantha et al. uses ultrasonic and radio beacons to allow mobile devices to determine their location within a building [7]. Borriello et al. built another room-level location service similar to Cricket [1]. Peng et al. built a ranging system for pairs of mobile devices that uses audio [6]. In this work we propose using ultrasonic sound for another task: directly sensing the user.

Active Sonar

Sonar systems emit sound "pings" and sense the resulting echoes. Based on the characteristics of the echoes, a rough map of the surrounding physical space can be derived. Sonar is used by animals, such as bats and dolphins, for navigation and hunting. Man-made systems have been invented for fishermen, divers, submarine crews, and robotics. The omnidirectional (unfocused)

[†] Northwestern University, EECS Dept., 2145 Sheridan Rd, Evanston, IL 60208, USA. email: starzia@northwestern.edu
[‡] University of Michigan, EECS Dept., 1301 Beal Ave, Ann Arbor, MI 48109, USA. email: dickrp@eecs.umich.edu

This work was supported in part by the National Science Foundation under awards CNS-0720691 and CNS-0347941. Tarzia is supported by a Dr. John N. Nicholson fellowship.

Ubicomp 2009, Sep 30 – Oct 3, 2009, Orlando, Florida, USA.
Copyright 2009 ACM 978-1-60558-431-7/09/09...$10.00

User attention state	Definition	Activity	User-study task
Active	Manipulating the keyboard, mouse, etc	*Typing*	Replicating an on-screen document on a laptop using a word processor
Passively engaged	Reading the computer screen	*Video*	Watching a video on the laptop's display
Disengaged	Sitting in front of the computer, but not facing it	*Phone*	Short multiple-choice telephone survey using telephone next to the laptop
Distant	Moved away from the computer, but is still in the room	*Puzzle*	Pencil-and-paper word-search puzzle on the desk beside the laptop
Absent	User has left the room	*Absent*	After the participant left the room

Table 1. Proposed user attention states and each state's associated user-study task.

and relatively insensitive microphones and speakers built into most laptops are not ideal for building a precise sonar system. However, our expectations for the sonar system are modest; we only need information about the user's activity, not a detailed map of the room.

Audio in the 15 to 20 kHz range can be produced and recorded by a laptop computer but is inaudible to most adults [5]. Thus, by using these audio frequencies and assuming the absence of children and pets that are sensitive to ultrasound, we can program a sonar system that is silent to the user. Our sonar system emits a continuous high frequency (ultrasonic) sine wave and records the resulting echoes using a microphone.

HYPOTHESES

What characteristics of the echoes might vary with user activity? We make the following two conjectures: (1) The user is a close surface that will reflect sound waves emitted from the speaker. (2) The user's presence may affect the amount of reflection and therefore the *intensity* of echoes received by the microphone.

In many scenarios the user is the only moving object near the computer. It might therefore be helpful to listen for signs of movement in echoes; any data related to movement is likely to be related to the physically-active user's behavior. In particular, motion in the environment is likely to introduce additional variance in the echoes since the angles and positions of reflection surfaces will be changing. Thus, the user's presence and activity might affect the *variance* of echo intensity. Our results, presented later, support this claim.

USER STUDY

We conducted a user study to determine how sound echoes vary with changes in user attention state. We were specifically interested in how echo intensities and variances are affected. Our study protocol was reviewed and approved by our university's Institutional Review Board and is described briefly in this section. We recruited twenty paid volunteers from among the graduate students in our department. During the study, participants spent four minutes working on each of four tasks. Each task, plus absence, shown in Table 1 is associated with one of five attention states.

Microphones	*internal:*	Laptop's internal microphone, located near the touchpad
	ST55:	Sterling Audio ST55 large diaphragm FET condenser mic connected through Edirol UA25 USB sound card
	PC:	Inexpensive generic PC microphone connected via Plantronics USB DSP v4 sound card
	webcam:	Built-in microphone on a Logitech Quickcam 3000 pro USB webcam
Speakers	*internal:*	The laptop's internal speakers, located on either side of the keyboard.
	sound-sticks:	Harman Kardon SoundSticks USB speakers that include a subwoofer, left, and right speakers.
	dell:	Dell's standard desktop computer speakers connected via Plantronics USB DSP v4 sound card
	null:	Record without any emitted sound wave

Table 2. Audio hardware used in user study.

A secondary goal of the study was to determine which types of speakers and microphones would be suitable for a computer sonar system. We, therefore, experimented with combinations of four microphones and four speakers. While the users completed the tasks, a 20 kHz sine wave was played, and recordings of the echoes were made. For each task, sixteen recordings were made. The four microphones recorded simultaneously. The four minutes that each participant spent on a task was divided into four one-minute intervals. During each interval a different speaker played the sine wave. In this way, a recording for each combination of microphone and speaker was obtained for each user performing each task. To eliminate temporal biases, the order of tasks completed and speaker activations within those tasks were randomized for each user (except that the "absent" task always occurred last, after the user had left). The total user study duration was twenty minutes: four minutes for each of five tasks.

Experimental Setup

Our experimental setup was modeled after an office environment. The audio equipment and laptop computer were arranged on a large desk and the participant sat in a rolling office chair. The study administrator was seated at an adjacent desk throughout the study. Everything, including the word puzzle clipboard was fastened securely to the desk to ensure consistency between runs. The telephone cord was shortened to force users to remain in front of the laptop while making calls. A Lenovo T61 laptop with a 2.2 GHz Intel T7500 processor and 2 GB RAM was used. The OS was Ubuntu Linux.

Setup details are as follows. We used the audio hardware listed in Table 2. The speaker volumes were set to normal listening levels. We used the right-hand side speakers only, for simplicity. We chose a sonar frequency of 20 kHz because very few people can hear tones at this frequency. Recording and playback audio format was signed 16 bit PCM at 96 kHz sample rate (almost all new laptops support these settings). The first and last five seconds of each recording were discarded leaving a set of fifty-second recordings for analysis.

Feature extraction

Analysis of the recordings was done after the user study was complete. We wrote Python scripts to analyze the 18 GB of WAV files using standard digital audio signal processing techniques. In this section we describe how echo intensities were calculated from the recordings and we describe a feature of these intensities, called *echo delta*, which we used when explaining our results.

To calculate an estimate of the echo intensity, we use a frequency-band filtering approach. We assume that all of the sound energy recorded in the 20 kHz band represents sonar echos; our measurements confirm that ambient noise in that frequency-band was negligible. We use Bartlett's method (with 10 non-overlapping rectangular windows and a 1024-point Fast Fourier Transform (FFT)) to estimate the recording's power spectrum; in each of the ten windows, the amplitude of the Fourier coefficient nearest 20 kHz was squared to get an energy value and then averaged with the other nine values. As is common in audio measurement, we scaled down the results with a base-10 logarithm.

In our results, we use a characterization of the echo's variance that we call *echo delta*. To calculate the echo delta of each recording we first break it into a sequence of 100 ms windows. The echo intensity is calculated for each of these by Bartlett's method, as described above; this gives us a sequence of echo intensity values $e_1 ... e_N$. The echo delta Δ_e is then just the average of absolute differences in that sequence:

$$\Delta_e(e_1 ... e_N) \equiv \frac{1}{N} \sum_{i=1}^{N-1} |e_{i+1} - e_i|$$

Echo delta characterizes echo variances on the time scale of a single echo intensity window, i.e. 100 ms.

RESULTS

We now quantify the effect of user state on sonar measurements in our user study. Although our experiments included three different speakers and four microphones, for brevity, we fully present results from only one combination: those obtained using the soundsticks speaker and webcam microphone.

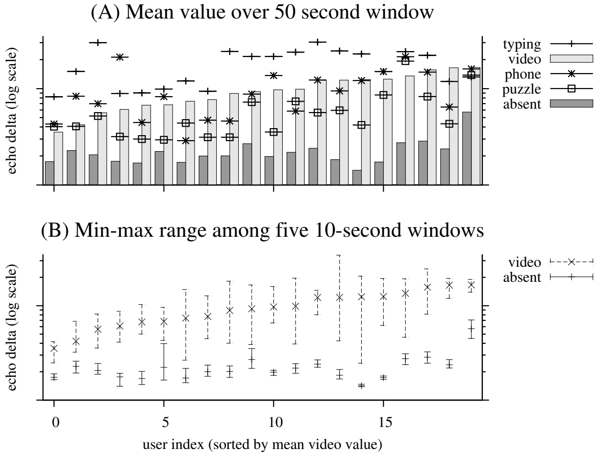

Figure 1. User study echo delta sonar measurements for soundsticks–webcam hardware combination. Note the clear difference between measurements of users during video and phone activities versus the absent state.

A comparison of echo delta (Δ_e) among different activities is compelling. Figure 1A shows Δ_e for each study participant, in each of the five attention states. There is a clear trend of increasing Δ_e when moving from absent to more engaged user states. The exact ordering of the middle states (video and phone in particular) varies between users, but all for all users we observe an increase in Δ_e with their presence in any of the four attention states.

To test the potential responsiveness of our sonar system, we simulated a reduction in the recording time window by splitting each fifty-second recording into five 10-second windows. Figure 1B shows the range of Δ_e values calculated in these smaller windows for a representative pair of states. We can see that, as compared to Figure 1A, the gap between the video and absent states is narrowed, but the two still do not intersect. This demonstrates a tradeoff between time window size and state identification accuracy.

In both plots of Figure 1, there is a clear difference between users who are absent and those who are present but not interacting directly with the machine. Combined with traditional Human Input Device (HID) monitoring, the proposed sonar approach makes it possible to differentiate between interactive users, present but non-interactive users, and absent users.

Similar, but weaker, results were obtained from several other microphone and speaker combinations. For power management, distinguishing between the passively engaged (video watching) and absent states is critical. Table 3 summarizes how well each hardware combinations distinguished between the video and absent states; success varies but does not seem to depend on hardware

Speaker	*Microphone*			
	internal	ST55	PC	webcam
internal	1.28	1.99	1.01	1.74
soundsticks	2.85	3.07	1.40	*4.53*
dell	2.66	7.55	1.86	2.36
null	1.24	1.16	0.97	1.35

Table 3. Ratios of Δ_e measurements for video activity over absent activity averaged across all users. Speaker and microphone combinations with higher ratios are likely capable of better distinction between these two representative activities.

Actual state	*Predicted state*	
	passively engaged	absent
passively engaged	0.9632	0.0368
absent	0.0248	0.9752

Table 4. Confusion matrix for binary presence classifier using 10 s of training and 40 s of test recordings.

cost. It is particularly noteworthy that the webcam's relatively basic microphone was a top performer. We suggest that microphone and speaker positioning and omission of high-frequency noise-filtering circuitry are the most important factors for good sonar performance.

Processing overhead for sonar is negligible. As an indication, the analysis runtime for a fifty-second sample was only 1.6 s on our study laptop. Therefore, a real-time-processing implementation would add a load of about 3% to one of the CPU cores. The energy overhead of activating the audio hardware is negligible compared to display or CPU energy.

STATE CLASSIFIER

Encouraged by the results shown in Figure 1B, we built a binary state classifier to automatically distinguish between passively engaged (video) and absent states. We use a very simple threshold-based classification scheme. Analyzing the training data gives an average echo delta Δ_e for the two states: $\Delta_e^{passive}$ and Δ_e^{absent}. We choose a threshold T between the two values using a weighted geometric mean: $T \equiv (\Delta_e^{passive} * (\Delta_e^{absent})^2)^{1/3}$. To classify the test data, we simply compare its Δ_e to T. If it is greater than or equal to T we classify it as passively engaged, otherwise absent.

Table 4 shows a confusion matrix for the binary state classifier on the user study data. The fifty second recordings from the passively engaged and absent states were broken into ten second windows as in Figure 1B. For each user, one passively engaged window and one absent window were used as training. Classification was repeated using every pair of absent and passively engaged windows as training. False positive and false negative rates were both below 4%. After increasing the length of the training data window to 25 s, classification of the remaining 25 s window became perfect. Note that distinguishing between the four attentive states is much more difficult than the above binary classification. This is evident in the reordering of states among different uses seen in Figure 1A. For example, it is not clear that distinction between video and phone activities is possible, but this was expected since users' behaviors and postures for these activities are varied. Nonetheless, by also monitoring HID events, we can clearly distinguish between three states: active, passively engaged, and absent.

CONCLUSION AND FUTURE WORK

The experimental results support the hypothesis that the user's presence indeed causes changes in echo intensity. More generally, we have demonstrated that sonar implemented using commodity computer hardware can measure useful information with low computational burden. Our user study was performed on a laptop computer and used traditional desktop computing applications. However, any device with a speaker, microphone, and a moderate amount of computational power should be able to use sonar; this includes cellular phones, PDAs, kiosks, and more.

There is still some work to be done in developing a practical sonar attention detection system. Preliminary experiments have shown that our sonar system works on several different laptop models using their built-in hardware. Our research group is already working on implementing effective sonar-based fine-grained power management in the OS.

REFERENCES

1. G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp. WALRUS: wireless acoustic location with room-level resolution using ultrasound. In *Proc. MobiSys '05*, pages 191–203, 2005.

2. A. B. Dalton and C. S. Ellis. Sensing user intention and context for energy management. In *Proc. HotOS '03*, May 2003.

3. P. A. Dinda, G. Memik, R. P. Dick, B. Lin, A. Mallik, A. Gupta, and S. Rossoff. The user in experimental computer systems research. In *Proc. ExpCS '07*, 2007.

4. A. Madhavapeddy, D. Scott, and R. Sharp. Context-aware computing with sound. In *Proc. UbiComp '03*, pages 315–332, 2003.

5. B. C. J. Moore. *An Introduction to Psychology of Hearing.* Emerald Group Publishing, 2003.

6. C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices. In *Proc. SenSys '07*, pages 1–14, 2007.

7. N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket location-support system. In *Proc. MobiCom '00*, pages 32–43, 2000.