Sonar-Based Measurement of User Attention

Stephen P. Tarzia
Northwestern University

Robert P. Dick
University of Michigan

Peter A. Dinda, Gokhan Memik
Northwestern University

***Abstract*—We describe a technique to determine presence and attention levels 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 users with near perfect accuracy after only ten seconds of measurement. Our experiment is the first to demonstrate that user attention states can be differentiated using sonar.**

I. Introduction

Several operating system (OS) subsystems are triggered by user inactivity. For example, power management systems save energy by deactivating, or *sleeping*, 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 whether the user is absent. Input activity is often used as an indicator of attention. This works in some cases because it captures engagement between the user and computer. However, engagement is not well-measured: input activity based techniques are unable to distinguish between a truly inattentive user and one who is actively reading the display without using the mouse or keyboard.

We have identified five different human user attention states among which a system may want to distinguish, shown in the table. The active state is trivially detectable using input activity; our goal is to distinguish the remaining four states.

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.

II. Active Sonar

Audio in the 15 to 25 kilohertz range can be produced and recorded by a laptop computer but is inaudible to most adults; these frequencies are called *ultrasonic*. Ultrasonics have already been used for communication [3] and localization [4], [1]. In this work, we describe an ultrasonic sonar software system.

Sonar systems emit sound "pings" and listen for the resulting echoes. Based on the characteristics of the echoes, a rough map of the surrounding physical space can be derived. The omnidirectional (unfocused) 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 attention state, not a detailed map of the room. Our sonar system simply emits a continuous 20 kHz sine wave and records the resulting echoes using a microphone.

We hypothesize that small movements of the human user will change the angle and degree of ping reflection, thus we expect the echo intensity in the recordings to vary if a user is present. We assume that human users are the only nearby moving objects.

To detect user presence we look for changes in the echo's intensity. 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 echoes; 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 simple characterization of the echo's variance that we call *echo delta*, Δ_e. 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.

III. User Study

We conducted a user study to measure how sonar readings Δ_e varied with user attention state. A more detailed description of the study protocol is given in the technical report [5]. We recruited twenty paid volunteers from among the graduate students in our department. During the study, participants spent some time working on each of four tasks. Each task, plus absence, shown in the table is associated with one of five attention states. While the users completed the tasks a 20 kHz sine wave was played, and we captured a fifty-second recording of echoes for each user in each of the attention states. To eliminate temporal biases, task ordering was randomized for each user except that "absent" task always occurred last, after the user had left).

state	*definition*	*user-study task*
Active:	the user is manipulating the keyboard or mouse	Replicating an on-screen document on a laptop using a word processor.
Passively engaged:	the user is reading the computer screen.	Watching a video being played on the laptop's display.
Disengaged:	the user is sitting in front of the computer, but not facing it.	Completing a short multiple-choice telephone survey using a telephone located to the side of the laptop.
Distant:	the user has moved away from the computer, but is still in the room.	Completing a word-search puzzle with pencil and paper on the desk beside the laptop.
Absent:	the user has left the room.	After the participant left the room.

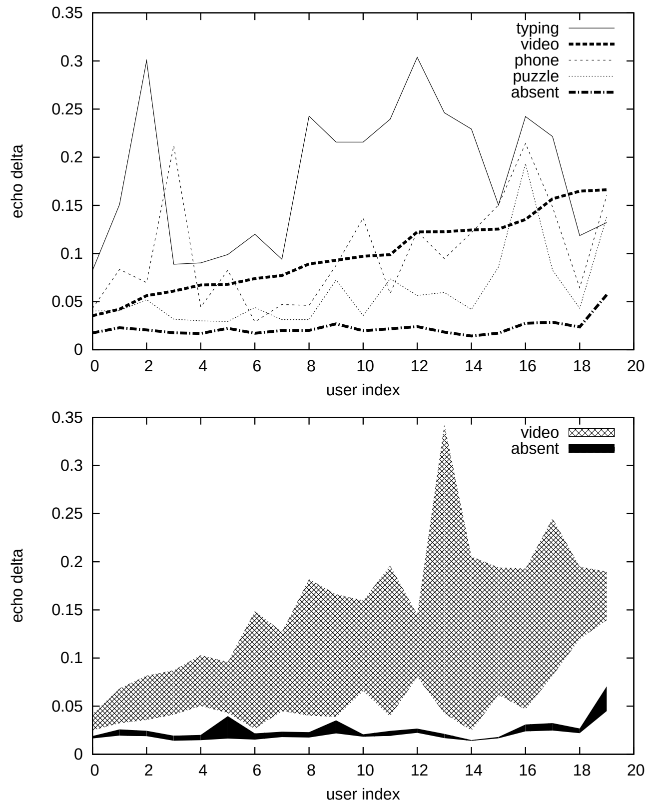

Fig. 1. Echo delta sonar measurements for all 20 users in each of 5 states. Users are sorted by video state data. Note the clear difference between measurements of users in the video and phone states versus the absent state. Top chart shows mean value over 50 seconds, bottom chart shows range of values among five 10-second windows.

IV. 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 present results from only one combination: those obtained using Harmon Kardon SoundSticks USB speakers and a Logitech Quickcam 3000 Pro webcam's microphone. Similar but weaker results were obtained from the other audio hardware combinations.

A comparison of echo delta (Δ_e) among different attention states is quite 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 users cause 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 HID monitoring, the proposed sonar approach makes it possible to differentiate between interactive users, present but non-interactive users, and absent users.

Practically speaking, these results imply that we can build a sonar system that measures Δ_e and classifies the user as present and attentive or absent by simply comparing Δ_e to a threshold. However, precisely identifying which of the attentive states was observed is more difficult since there is some overlap of lines in Figure 1A. For example, it is not clear that distinction between video and phone states is possible, but this was expected since users' behavior and positioning in these states is varied.

V. Conclusion and Future Work

The experimental results support the hypothesis that the user's presence and attention state indeed cause 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 research group is already working on tackling practical issues such as calibration with the goal of on implementing effective sonar-based fine-grained power management.

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

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

[5] S. P. Tarzia, R. P. Dick, P. A. Dinda, and G. Memik. Prospects for sonar-based measurement of user attentiveness. Technical Report NWU-EECS-09-06, Northwestern University, 2009.