NORTHWESTERN
UNIVERSITY

Electrical Engineering and Computer Science Department

Technical Report
NWU-EECS-09-06
April 06, 2009

Prospects for Sonar-based Measurement of User Attentiveness

Stephen P. Tarzia **Robert P. Dick** **Peter A. Dinda** **Gokhan Memik**

Abstract

We describe a technique to determine presence and attention of computer users. This technique relies on sonar, the hardware for which already exists on commodity laptop computers. Determining user presence and attention enables a number of system-level optimizations, e.g., the screen may be dimmed when the user is not paying attention in order to reduce power consumption. The proposed technique relies on 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 distinguish between cases when the user is attentive and inattentive with a statistical confidence of 98.8%. Our experiment is the first to demonstrate that user attentiveness states can be differentiated using sonar. We plan to make our sonar trace gathering and analysis software available.

This project is made possible by support from the National Science Foundation (NSF) via grant CNS-0720691. Stephen P. Tarzia is supported by a Dr. John N. Nicholson fellowship.

Keywords: sonar, attentiveness, presence, power management, empathic systems, user study, ultrasonics

1 Introduction

This section points out the motivation for developing the proposed sonar-based user attention estimation technique, summarizes related work, and provides background information on the fundamental principals on which the technique builds.

1.1 Motivation

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 attentiveness. 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. It is our goal to distinguish between attentiveness and inattentiveness, not merely between the presence and absence of user input. A more reliable indicator of user attentiveness has the potential to improve power management and security subsystems; they could be triggered more quickly and confidently, avoiding deactivating or locking portions of a computer system that are actually in use.

We have identified five different user attentiveness states among which the OS may want to distinguish, shown in Table 1. The active state is trivially detectable using input activity; our goal is to distinguish the remaining four states.

1.2 Related Work

We know of only one existing research project that studies user attentiveness detection. "FaceOff" tackles the fine-grained power management problem [2]. It processes images captured by a webcam to detect whether a human is sitting in front of the computer. This method relies on peripheral hardware (a webcam) and may incur significant energy cost in the image processing. In addition, no experimental evaluation of FaceOff is reported. Other works try sense user emotions [9] and satisfaction levels [3].

Ultrasonics have already been used in context-aware computing for several different tasks. Madhavapeddy et al. used ultrasonics and audible sound as a short-range low-bandwidth wireless communication medium [5, 6]. The Cricket localization system by Priyantha et al. uses ultrasonic and radio beacons to allow mobile devices to determine their location within a building [10]. 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 [8].

1.3 Background on Sonar

Sonar systems emit sound "pings" and listen for the resulting echoes. Based on the characteristics of the echos, 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 [11]. Man-made systems have been invented for fishermen, divers, submarine crews, and robotics. 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 attentiveness state, not a detailed map of the room.

Audio in the 15 to 20 kilohertz range can be produced and recorded by a laptop computer but is inaudible to most adults [7]. Thus, by using these audio frequencies, we can program a sonar system that is silent to the user. Our sonar system continuously emits a high frequency (ultrasonic) sine wave and records the resulting echoes using a microphone.

2 Hypotheses

An open question is: what characteristics of the echoes might vary with attentiveness state? We make the following conjectures:

1. The human user is one of several close surfaces 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 help-

state	*definition*	*user-study realization*
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.

Table 1: Proposed user attentiveness states and our user-study realization of each.

ful to listen for signs of movement in the ultrasonic 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 attentiveness state might affect the *variance* of echo intensity.

3 User Study

We conducted a user study to determine how sound echoes vary with changes in user attentiveness 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 breifly 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 the third column of Table 1 is associated with one of the five attentiveness states.

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 different microphones and four different speakers (hardware details are provided in Section 3.1). 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.

As illustrated in Figure 1, 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.

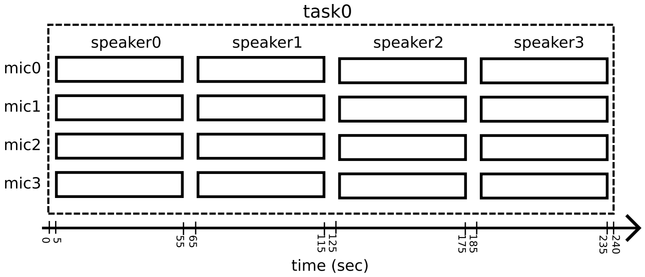

Figure 1: Timeline of all sixteen recordings taken while a user study participant completes a single task.

3.1 Experimental Setup

Our experimental setup is shown in Figure 2. The equipment was 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

Microphones

internal:	The 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:	An inexpensive generic PC microphone connected through a Plantronics USB sound card.
webcam:	The 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.
soundsticks:	Harman Kardon SoundSticks usb speakers that include a subwoofer and two satellites.
dell:	Dell's standard desktop computer speakers connected through a Plantronics USB DSP v4 sound card.
null:	We also record once without any emitted sound wave.

Table 2: Audio hardware used in user study

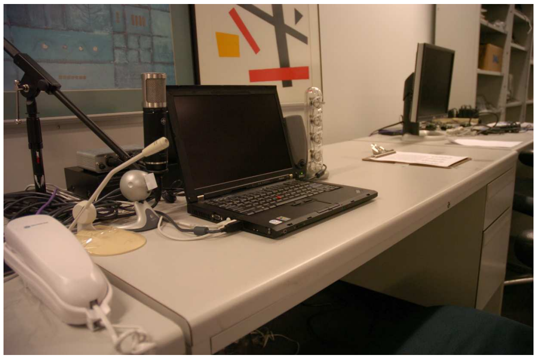

Figure 2: User study setup.

clipboard was fastened securely to the desk to ensure consistency between runs. The telephone cord was shortened to force the user to remain in front of the laptop while using it. A Lenovo T61 laptop with a 2.2 GHz Intel T7500 processor and 2 GB RAM was used.

Table 2 describes the audio hardware used. The speaker volumes were set to normal listening levels. We used the right-hand side speakers only, to prevent speaker interference. 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 (the sound hardware on 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.

3.2 Analysis

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 techniques standard in digital audio signal processing. In this section we describe how echo intensity measurements were calculated from the recordings and what statistical properties of these intensities were used in 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. To approximately measure the echo intensity, we did a 1024-point Fast Fourier Transform (FFT) and recorded the amplitude of the fourier coefficient nearest 20 kHz. This value was squared to get an energy value and then scaled down with a base-10 logarithm (log scales are common in audio measurement).

The statistical property that we use in our results is an average variance of echo intensity within the entire fifty-second recording. We divided each recording into five hundred 100 ms subwindows and calculated the echo intensity in each of these. We then calculated the variance of these echo intensities within each one second window; that is, we calculated the variance within fifty groups of ten. The mean of these fifty variance values was considered the overall echo intensity variance for the recording and this value was used in our results. Thus, each fifty-second recording was characterized by a single statistical measure.

4 Results

Although our experiments included three different speakers and four microphones, for brevity, we fully present only the most promising results: those obtained using the soundstick speaker and webcam microphone.

A comparison of the change in variance among attentiveness states is quite compelling. Figure 3 shows echo variance values for each study participant, in each of the five attentiveness states. We don't see a consistent ordering of the values for each state when moving across users; instead, the lines intersect each other. However, a trend of decreasing values when moving from the active state, through the intermediary states, to the absent state is apparent for individual users. Variance measurements for the absent state were the most consistent; in for every participant, they were the lowest among all states.

We used two statistical hypothesis tests to evaluate the difference in measurements taken in the different pairs of attentiveness states; p-values for a paired t-test are shown in Table 3 and for a sign-test in Table 4. These tables show the statistical significance of the results [4]. Roughly speaking, p-values estimate the probabilities that two sets of values were drawn from the same probability distribution. In other words a p-value gives the probability that the measurement differences we observed were due to entirely to randomness. Both of these tests used paired observations; that is, the difference between the measurements for each user are used. The passively engaged and disengaged (video watching and telephone survey) state-pair had a non-negligible p-value, so they were not clearly distinct. The remaining pairs of states were distinct with confidence greater than 98.8% (p-value less than 0.012 for both tests).

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 5 indicates which hardware combinations gave measurements that provide 95% confidence under both t-test and sign-test; half of the combinations meet this criterion.

Practically speaking, these results imply that we can make the following judgement on subsequently collected data: If we are given two sets of variance measurements taken from a single user while in a given two attentiveness states, then we will be able to say, with high confidence, which set corresponds to which state.

Speaker	Microphone			
	internal	ST55	PC	webcam
internal		X		X
soundsticks	X	X		X
dell		X		X
null				X

Table 5: Speaker and microphone combinations capable of distinguishing between the passively engaged (video watching) and absent states with 95% confidence under both t-test and sign-test.

Processing overhead for sonar is negligible. As an indication, the analysis runtime for a fifty-second sample was only 1.6 seconds on our study laptop, described above. A real-time-processing implementation would add a load of about 3% to one of the CPU cores. Our implementation was not optimized for performance, so these figures are conservative.

5 Conclusion and Future Work

The experimental results support the hypothesis that the user's presence and attentiveness state indeed affect the variance of echo intensity. More generally, we have demonstrated that sonar implemented using commodity computer hardware can measure useful information with low computational burden.

It is important to understand the limits of our claim. We have only motivated and laid the groundwork for useful computer sonar systems. We have not provided an attentiveness-state recognition algorithm. Such an algorithm now seems possible with sonar, but more experimentation and evaluation is needed to develop such tools. Our research group is already working in that direction, with the goal of on implementing an effective sonar-based fine-grained power management daemon. However, given the potential value of sonar-based measurements with commodity computer hardware for power management, security, and other applications we have not yet considered, we thought it useful to share our findings.

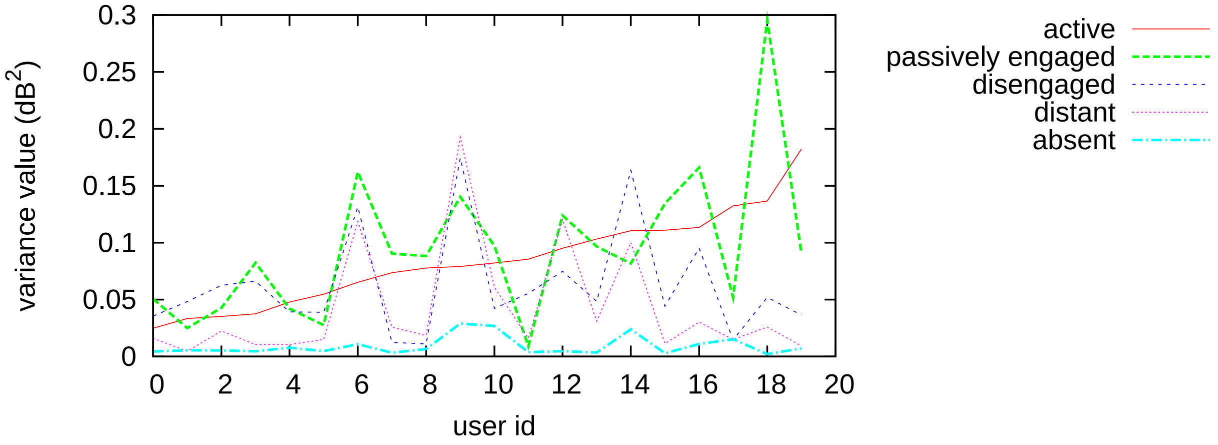

Figure 3: Variance in intensity of echoes for all 20 users in each of 5 states for soundsticks-webcam combination. Users are sorted by active state variance.

passively engaged	disengaged	distant	absent	
0.002	**0.001**	< **0.001**	< **0.001**	active
	0.231	< **0.001**	< **0.001**	passively engaged
		< **0.001**	< **0.001**	disengaged
			0.007	distant

Table 3: T-test p-values for distinguishing between pairs of attentiveness states. Lower values are better. Pairs clearly distinguished by this test are in bold.

passively engaged	disengaged	distant	absent	
0.012	**0.012**	< **0.001**	< **0.001**	active
	0.0414	**0.003**	< **0.001**	passively engaged
		< **0.001**	< **0.001**	disengaged
			< **0.001**	distant

Table 4: Sign-test p-values for distinguishing between pairs of attentiveness states. Lower values are better. All pairs are clearly distinguished by this test.

References

[1] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp. Walrus: wireless acoustic location with room-level resolution using ultrasound. In *MobiSys '05: Proceedings of the 3rd international conference on Mobile systems, applications, and services*, pages 191–203, New York, NY, USA, 2005. ACM.

[2] A. B. Dalton and C. S. Ellis. Sensing user intention and context for energy management. In *HotOS '03: Proceedings of the 9th Workshop on Hot Topics in Operating Systems*, 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 *ExpCS '07: Proceedings of the 2007 workshop on Experimental computer science*, page 10, New York, NY, USA, 2007. ACM.

[4] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design Measurement Simulation and Modeling.* Wiley, 1991.

[5] A. Madhavapeddy, D. Scott, and R. Sharp. Context-aware computing with sound. In *UbiComp '03: In Proceedings of The 5th International Conference on Ubiquitous Computing*, pages 315–332, 2003.

[6] A. Madhavapeddy, R. Sharp, D. Scott, and A. Tse. Audio networking: the forgotten wireless technology. *Pervasive Computing, IEEE*, 4(3):55–60, July-Sept. 2005.

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

[8] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *SenSys '07: Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 1–14, New York, NY, USA, 2007. ACM.

[9] R. W. Picard. Toward machines with emotional intelligence. In G. Matthews, M. Zeidner, and R. D. Roberts, editors, *The Science of Emotional Intelligence: Knowns and Unknowns.* Oxford University Press, 2007.

[10] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The cricket location-support system. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 32–43, New York, NY, USA, 2000. ACM.

[11] J. A. Thomas, C. F. Moss, and M. Vater, editors. *Echolocation in bats and dolphins.* University of Chicago Press, 2004.