

Usability at the Edges: Bringing the Lab into the Real World and the Real World into the Lab

Valerie Henderson-Summet and James Clawson

Georgia Institute of Technology, GVV Center, Atlanta, GA, USA
{vlh, jamer}@cc.gatech.edu

Abstract. As scientists, we must make ostensibly either-or tradeoffs in order to gather the data needed to properly evaluate a prototype, algorithm, or phenomena. Either we evaluate prototypes in a controlled, laboratory environment, or we evaluate them in an unstructured, real world environment. However, we feel that there are ways to push this seeming boundary between real world and laboratory studies. In this paper, we discuss how to retain some of the control of a laboratory environment while working in the field and how to bring some of the authenticity of the real world into the controlled, laboratory environment.

1 Introduction

The context in which a study takes place impacts the types of questions a researcher can ask as well as the results one can expect. The laboratory and “real world” environments afford researchers the ability to ask and answer different types of questions. Laboratory studies are characterized as taking place in a controlled environment and allow researchers to collect data which are often analyzed using quantitative methods. Some advantages of laboratory studies include a large degree of experimental control over variable manipulation and easy replication. However, laboratory studies can include limited relation to the real world and an unknown level of generalizability of results outside laboratory settings (external validity).

Studies that take place in the “real world,” typically case studies and field studies, are often intensively empirical. They can employ various qualitative research methods and are used to answer questions about the effect that context has on an individual’s usage of technology. Some advantages of real world studies are that the collected data are grounded in natural settings and that the data are often richly detailed. However, real world studies are often very time consuming, the data are not easily analyzed, the results are hard to generalize, and there is a very limited ability to manipulate variables.

We actively pursue a research agenda in which one works at the edges of these two contexts. In our work, we attempt to blur the boundaries of these environments to harness the advantages of each context while minimizing disadvantages. By bringing elements of the laboratory into the real world and elements of the real world into the laboratory, we are able to ask questions that would be much harder to answer if we did not practice at the edges.

2 Educational Games for Deaf Students

Since 2003, we have been working to develop educational games for deaf students in cooperation with a local school for the deaf. One game, CopyCat, is designed to increase a students' receptive abilities in American Sign Language (ASL) by presenting instructions, problems, and other game content via short signed videos. The student's task is to help a character navigate through a virtual world. The player controls the game's protagonist by signing the location of predators such as alligators or spiders. A computer recognizes the child's signing, and if the child signs correctly, the predator vanishes and the character continues on his quest. For a full discussion, see Henderson et al. [3] and Brashear et al. [1].

Study Design Challenges: In developing studies to evaluate the effectiveness of CopyCat, two issues that needed to be addressed were communication challenges and the limitations of our prototype technology.

Due to the language barrier between the children and researchers, interpreters were necessary to facilitate communication. This introduced interference that hindered direct communication. For example, after the children played the game for the first time, they were asked several questions regarding the enjoyment and usability of the game. The students had trouble giving an answer and first looked to the interpreter to see if there was a "correct" answer that they were supposed to deliver.

Although our goal was to test the development of a game built around the concept of a sign language recognition engine, using a recognition engine during early, iterative testing would unnecessarily introduce errors and confounding factors to the study. While introducing the sign recognition system early would be closer to the end-goal system which would occasionally make mistakes (just as speech recognition does), we felt that unanticipated errors would corrupt the study and skew our participants' opinions.

Bringing the Lab into the Real World: Bringing 10-15 deaf children to our lab in downtown Atlanta twice per week for multiple weeks was logistically infeasible. Based on this logistics issue, we decided to evaluate the game at the children's school. However, to ensure we collected the data we needed, we introduced some measures of control into the setting, similar to what we would normally employ in a laboratory setting. We addressed the limitations of our technology by employing a Wizard-of-Oz approach and attempted to eliminate communication challenges by incorporating alternative evaluation metrics to remove communication barriers.

The Wizard-of-Oz setup allowed us to introduce more stability into the testing process. A child would sign a phrase, and the wizard (with assistance from an ASL linguist) would determine the correctness of the signed phrase. The Wizard-of-Oz setup also allowed us to collect video data of the children's signing and automatically label it as correct or incorrect, in much the same way as we would in a laboratory setting.

To combat the language barrier issues, we wanted to allow the children as much uninterrupted time with the game as possible without adult interference.

To accomplish this, we included help buttons with instructions in sign language and automated some character movements. This reduced confusion on the part of the children and made them less likely to ask for help from an adult. While removing interpreters from the conversation with the children may seem counterintuitive, it introduced a measure of control to our experiment and facilitated concrete data gathering. Instead of one on one interviews, we transitioned to focus groups. The children drew their own video games using crayons and paper and shared them with the class. This interactive activity proved to be a better source of information than interviewing a child with the help of an interpreter.

3 Mobile Evaluation in the Lab

The laboratory environment is particularly suited for gathering empirical human performance measures. We have focused our efforts on evaluating mobile devices with the goal of better understanding the limits of human performance on these devices. Particularly, we have performed evaluations of mobile text input [2] and output [4]. The laboratory environment enables us to perform tightly controlled quantitative studies but makes it difficult to evaluate the way that people use mobile devices while in the wild. To address this challenge, we attempted to simulate real world conditions in a laboratory environment.

Study Design Challenges: When a user attempts to interact with a device while in-motion traditional HCI evaluation methods need to be reconsidered. When evaluating a desktop application or interface a researcher can often collect measures of cognitive load that accurately reflect a user's perception of how challenging it is to accomplish a task. When a user is in-motion, the user's attention is split between the task of navigating the interface of the mobile device and the task of navigating the environment. As such, in the papers cited above, we have attempted to address two primary and one secondary mobile evaluation challenges in the laboratory: simulating the split attention condition encountered when a user interacts with a device while mobile, accurately approximating physical mobility conditions, and finally accomplishing both of these in a way that does not compromise the external validity of the studies.

Bringing the Real World into the Lab: We needed to simulate two real world conditions which drastically affect mobile device use in our laboratory experiments: split attention and mobility.

In our study of blind mobile text input [2] we attempted to tease out the concept of split attention by testing expert mini-QWERTY keyboard typists' ability to type quickly and accurately in conditions of limited visual feedback. For the purposes of this study, the participants were using a mini-QWERTY keyboard attached to a desktop display. This enabled us to decouple the input from the output. This study had three typing conditions: a control condition in which the participants had full visibility of both the keyboard and the display, a "hands blind" condition in which the participants had full visibility of the display but typed with the keyboard placed under a desk, and a "fully blind"

condition in which the participants typed with their hands under the desk and had limited visibility of the display. Each of these conditions simulate a different real world condition encountered when users employ mobile devices. By varying the amount of visual feedback to the user, we were able to simulate the split-attention mobility conditions successfully while still maintaining control over the testing environment.

In our study of mobile output [4] we took a different approach. In this study we created a path inside our laboratory and asked participants to walk around the path while reading passages of text from a variety of mobile devices and responding to reading comprehension questions. By keeping the subjects in the laboratory, we were able to control confounding environmental conditions (light/screen glare, temperature, wind, etc.) while still approximating real world mobile conditions. Additionally, we are able to collect performance measures such as variations in walking speed and the number of steps participants took off of the path and correlate those measures to the task being performed. By conducting this study in the lab we are able to compare performance between subjects since everyone was operating in exactly the same conditions (no variations in terrain, environmental noise, or atmosphere). As such, we were able to run a tightly controlled study while still introducing real world elements into the laboratory environment.

4 Conclusions

Bringing the laboratory into the real world and the real world into the laboratory affords us measures of control unrealized previously and increases our ability to produce externally valid work. Conducting studies at the edges of the real world and laboratory environments enables us to reap the benefits implicit in different methods while mitigating the inherent disadvantages. Careful consideration of different experimental methods can lead to a successful evaluation which has richer results than either a strict laboratory study or a strict “real world” study.

References

1. H. Brashear, V. Henderson, K.-H. Park, H. Hamilton, S. Lee, and T. Starner. American Sign Language recognition in game development for deaf children. In *Proceedings of Assets '06*, pages 79–86, 2006.
2. J. Clawson, K. Lyons, T. Starner, and E. Clarkson. The impacts of limited visual feedback on mobile text entry for the twiddler and mini-QWERTY keyboards. In *Proceedings of ISWC '05*, pages 170–177, 2005.
3. V. Henderson, S. Lee, H. Brashear, H. Hamilton, T. Starner, and S. Hamilton. Development of an American Sign Language game for deaf children. In *Proceedings of IDC '05*, pages 70–79, June 2005.
4. K. Vadas, N. Patel, K. Lyons, T. Starner, and J. Jacko. Reading on-the-go: a comparison of audio and hand-held displays. In *Proceedings of MobileHCI '06*, pages 219–226, 2006.