

Ubiquitous Computing Support for Skills Assessment in Medical School

Kenneth P. Fishkin¹, Sunny Consolvo¹, Jennifer Rode¹, Brian Ross², Ian Smith¹,
Karen Souter²

¹Intel Research Seattle, 1100 NE 45th St, 6th Floor
Seattle, Washington, 98105 USA
{kenneth.p.fishkin, sunny.consolvo, jennifer.a.rode,
ian.e.smith}@intel.com

²Department of Anesthesiology, University of Washington Medical Center
Seattle, Washington, 98195 USA
{bkross, kjsouter}@u.washington.edu

Abstract. This paper describes the research agenda and value proposition for using ubiquitous computing (UbiComp) to support skills assessment and training within a medical school. While the need for automated (or semi-automated) support for skills assessment is increasing, and the technology to allow it is coming within reach, this gap has not yet been bridged. In this paper we report on our attempts to bridge that gap, utilizing recent advances in UbiComp hardware and software, and highlighting the need for further advances. This project is one of the foci of ongoing research at Intel Research Seattle around creating usable, robust, practical UbiComp deployments.

1 Introduction

Ubiquitous computing offers a compelling vision of unobtrusive, proactive aids to our daily life, varying with and appropriate to our location and activities. However, for historical reasons, UbiComp has focused almost exclusively on a narrow subset of this – first-world offices and homes. Now that UbiComp technology is maturing, we wish to test the vision and its applicability in other settings, *e.g.*, settings where desktop computers are inadequate. We wish to explore UbiComp technology in a *non*-office work environment, that of the hospital. As Bardram [2] mentions, hospitals are environments which are particularly well-suited for non-desktop UbiComp, and fundamentally different from typical office work: the extreme physical mobility, the lack of privacy, the high degree of interaction with the physical world, the difficulty of typing (hands often being occupied, wearing gloves, or both), etc. However, with the exception of Bardram’s pioneering work [2,3,4], little has been done on applying UbiComp to this domain. As part of our lab’s effort to investigate portable, powerful, usable UbiComp for human activity inferencing, we are attempting to take on this challenge. We are focusing on a subset of the huge hospital work domain, namely the “skills assessment” training of medical school students and residents. In skills assessment, the student is required to perform some task, often in a simulation lab,

under the watchful eye of teacher(s). Often the task is not specified directly (“do procedure <X>”), but rather indirectly (“symptoms <Y> are present – what do you do?”). After the student has exercised some set of skills in an attempt to solve the problem, her skill level is assessed by the teacher (s), and feedback given.

Our research agenda focuses on Ubicomp technology to aid in the assessment and training of medical school students and residents learning new procedures, which in the long term, could be used to proactively assist practitioners, whatever their skill level, through a procedure.

Of all the parts of the hospital domain where Ubicomp could be brought to bear, why skills assessment? We chose it for these reasons:

- *Value to the medical school.* The United States has recently reduced the number of hours a resident may work/train per week to 80 [1] (workloads of 100+ hours were previously common). Medical schools are trying to find ways to increase the efficiency of education, such that they reach the same skill level as before in 20% less time. Ubicomp solutions can help address this need. For example, many hospitals already have simulation centers for skills assessment and training, centers which may have cost hundreds of thousands of euros to create. However, they often sit empty. For example, at the University of Washington, the anesthesiology simulation room is on average only used 3-6 hours per week, because of the severely limited resource of instructor time. If a Ubicomp solution could reduce the number of instructors needed to run training simulations, students and residents might be able to spend more time practicing their skills and honing their craft.
- *Value to the licensing boards.* There is a drive, especially in the United States, towards a more objective and quantitative skills assessment. Most skills assessment before a certifying board is currently done via the verbal “say-so” of teachers. This is largely because at present, there is no objective data which measures performance. Introducing such objective data and reporting could provide great value to the licensing boards.
- *Value for the future.* There is a world-wide drive towards tele-medicine and virtual surgery. This requires both a greater availability of simulation facilities, and a greater ability for remote assessment. Ubicomp deployments which could evaluate a student (or group of students) in one location, while the grader is not physically present, will be of increasing value.

2 Challenges

The hospital environment, though one with promise for Ubicomp, is also one with its own challenges. Here are some of the challenges we have faced and are facing, and their implications on our design and lessons for Ubicomp researchers:

- *A Challenging Locale.* Hospitals are difficult deployment environments. For understandable legal reasons, it is extremely difficult to deploy or test any technology on or even near real patients. In the United States, this hurdle is so high that we have designed our research to omit patients as subjects – our subjects are medical school students, residents, and physicians; a special mannequin plays the role of the patient. Even without involving patients, hospitals are still understandably reticent about allowing novel equipment, particularly equipment emitting electro-magnetic energy, within a hospital building; this was a concern of ours as well. We are therefore conducting our trial deployments in existing simulation rooms, already cleared for such experimentation and far from patients.
- *Challenging subjects.* Our decision to focus on medical students and doctors as our subjects brings its own challenges. Medical students and doctors are *extremely* busy people, who are extremely busy treating (and learning how to better treat) patients. Any minute they spend with us is a minute away from this vital task. We have pared our surveys, user tasks, and user subject population size to the bone to accommodate this.
- *Truly Ubiquitous deployment.* The skills assessment domain is one in which subjects (students) are already performing a procedure as part of their grading and education. This means that the mode of operation of the student must be as close as possible to that of a normal, un-augmented environment – it defeats the purpose of skills assessment if they are not allowed to perform using the same tools, in the same setting, in the same way, as they do normally. Secondly, the deployment must have literally *zero* cognitive overhead while the student is performing their simulation – they are being assessed on a demanding cognitive task already, and all their attention will be focused on that. They have no patience for special instructions, tools, or people that would require them to assist the Ubi-comp system in any way – pressing a special button, holding an instrument just so, etc.
- *Challenging sensing.* We must use the existing tools the students use, supporting their current modes of use – whatever physical forces they exert must not break our system, and whatever tools they use must be observed in their existing size, shape, and material (this has posed significant problems, as we discuss later). Furthermore, for this to be used in grading, accuracy is paramount. Any false negatives or false positives can dramatically reduce the system’s value. Finally, grading is not complete, for more complicated procedures, without 1) the integration of object-based interactions (which tools they used, when), 2) with kinematic interactions (what force they used, the angle at which a tool was held, etc.), 3) with audio interactions (who they spoke to, when, in what tone of voice), and 4) with visual interactions (which monitor they consulted, when).
- *Challenging Machine Learning.* Existing machine learning techniques can be used to classify novice vs. expert users, and even to create a numerical score indicating deviation from the expert model. However, to make this useful to the doctors and students, more is needed: a mecha-

nism must be created to easily explain *why* the system produced the score it did, and said explanation must be understandable by medical professionals with no computer science background.

- *Challenging Assessment.* Medical technique is both art and science. Different doctors will teach different methods for the same procedure in a variety of valid fashions. Further, correctness often varies in the eye of the beholder. Taking into account the variety of valid methods when doing assessment is a key challenge.

3 Current Work

Our deployment is focused on skills assessment of students in the school of Anesthesiology. Of the dozens of procedures that our system could assess, our first version is focusing on a procedure known as “pre-intubation setup”. This is a basic procedure in anesthesiology: we chose it as our initial procedure due to its simplicity and as a first step towards more complicated assessments.

The procedure consists of subjects searching through a cart (Figure 1), taking out some of the objects in it, and then using some of those objects to perform an intubation (e.g., inserting a breathing tube into an airway). The assessment focuses on the “correctness” of the set of instruments they set out for the procedure.



Figure 1: an anesthesiology cart (left) and the contents of a drawer (right). The black circles are RFID tags

Even this simple procedure poses a significant challenge on the sensor front. The cart drawer contains roughly 40 different objects which need to be tracked. These objects are of drastically different sizes and shapes, many quite small and curved, and many are metal or transparent. Detecting when these objects are used via vision is

impractical, given the current state of machine vision, especially as the procedure is heavily visually occluded – the student is often leaning over the tools, surrounded by other people, etc. Another option would have been to use contact sensors, detecting when a tool is lifted from its resting place. This would not have fit with their current usage context, where the instruments are often jumbled and piled in rough arrangements, like a household’s crowded silverware drawer (see Figure 1). Another option would have been to use smart accelerometers that report a wireless signal when they are moved, but those are too large to be placed on many of the instruments.

Given the problems with the aforementioned sensing options, we are currently relying heavily on a different sensing option, *i.e.*, passive RFID tags. These tags are quite small (Figure 2 shows the tags we used, which range in size from a 2 cm diameter circle to a 10cm X 1 cm rectangle), cheap, require no battery, can be stuck onto objects, and transfer a wireless identifier to a nearby transmitter. For more detail, see [5]. To sense the tags, we use a very short-range (roughly 2 cm) RFID antenna integrated into a wearable glove, as was done in [6]. A report from this glove is a good indicator that the person is using a tagged object. There are occasional false positives, *e.g.*, when somebody brushes their hand over a row of tools, but this is usually easily detected. False negatives are much more common, we are overcoming this by “stud-ding” some tools with multiple tags, often of different types, without destroying the existing affordances for the tool.

To minimize the tags’ impact on *how* the students perform the procedure, we tag *all* objects in the drawer, place multiple tags on many of the objects (so as not to suggest *where* or *how* an object should be held or used), and place tags on other uninteresting objects. This is critical, because if the tags provide clues that would bias the students to perform the procedure better or worse than they would have done without the tags, our goal of being able to accurately assess their skill could not be met.

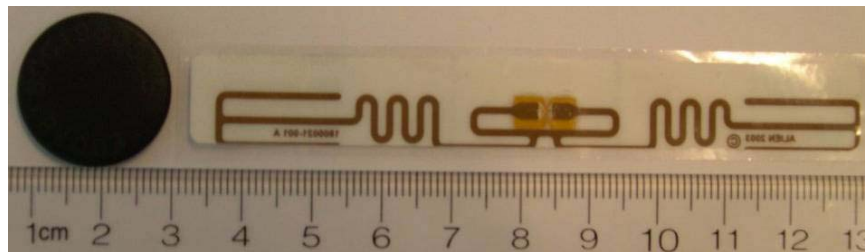


Figure 2: Two of the RFID tags we are using.

The wearable glove used in earlier research [6] had an antenna in the palm, and the computer on the top of the hand. After observing doctors performing procedures in the hospital environment, it became clear that a number of changes had to be made.

1. The computer on the top of the hand was distracting ergonomically and prevented users from wearing latex gloves, as they typically do. In our redesign, the computer is attached to the wrist with wires leading to the glove.
2. The palm antenna wasn’t sufficient, as many instruments (*e.g.* stethoscopes and syringes) are used with the fingertips and never come near the

palm. To detect those interactions, we extended the glove to have two antennae – one located in the palm as before, and one located on the thumb. By using recently developed flexible mylar antennae from SkyeTek, we were able to get a reasonable detection range (2-3 cm), even within this very small, curved form factor.

3. Doctors often used their fingertips for sensing. We had to ensure that the gloves left as much of the fingers free as possible.
4. Doctor's hands come in a variety of shapes and sizes. We had to make all components accommodate this.

Figure 3 shows an example of the resulting wearable in use.



Figure 3: an RFID-detecting pair of gloves in use

During our initial trials, we became aware of an additional variable—a high level of hand strength used, which was not obvious to a third party observer. These forces combined with the users' moist palms and resulted in a high level of false negatives in the trial's later runs, as one of the gloves broke halfway through the session. We are presently designing a next-generation glove to overcome this. This experience demonstrates the need for iterative design and evaluation with the intended users in the intended setting.

We are at the beginning stages of testing the system and its ability to detect and then assess user actions. To date we have tagged 43 objects with 69 tags, and run 7 volunteers from the medical school faculty through the pre-intubation setup 17 times (total) to test the underpinnings of the system. Our initial experience is positive with

respect to the unobtrusiveness of the system – the doctors found the gloves did not interfere with their work practice, nor did the tags. We are beginning to analyze the data; we are averaging 217 sensor reports for each run of the procedure.

5 Future Work and Conclusions

We have outlined our current research into using Ubicomp technology to enable better skills assessment for medical students training to become doctors. This research thrust, which is in its early stages, is already driving research into sensors, wearable computers, and machine learning. As the research matures, we hope to include audio and vision into our multi-modal sensor fusion, to start tracking more complicated procedures (and group procedures), and to eventually deploy the system to more departments. If the project is successful, we hope to have a positive impact on the training and daily life of those learning to become doctors, a difficult but valuable deployment for Ubicomp technology.

Acknowledgements

Thanks to Gretchen Cox, who made the gloves, Adam Rea, who designed and built the RFID wearable system housed in the gloves, and the doctors who volunteered to help test the system.

References

- 1 Accreditation Council for Graduate Medical Education (ACGME). The ACGME's Approach to Limit Resident Duty Hours: The Command Standards and Activities to Promote Adherence. 2004.
- 2 J.E. Bardram, Jakob. Hospitals of the Future – Ubiquitous Computing support for Medical Work in Hospitals. UbiHealth 2003 – the 2nd International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications
- 3 J. E. Bardram. Applications of Context-Aware Computing in Hospital Work – Examples and Design Principles. In *ACM Symposium on Applied Computing (ACM SAC) 2004*.
- 4 J. E. Bardram, R. E. Kjær, and M. Ø . Pedersen. Context-Aware User Authentication – Supporting Proximity-Based Login in Pervasive Computing. *UbiComp 2003*, pages 107–123 , Oct. 2003.
- 5 Fishkin, Kenneth P. ,Jiang, Bing, Philipose, Matthai and Roy, Sumit. I Sense a Disturbance in the Force: Long-range Detection of Interactions with RFID-tagged Objects. *UbiComp 2004*. to appear
- 6 Matthai Philipose, Kenneth P. Fishkin, Mike Perkowitz, Donald J. Patterson, Dirk Hähnel, Dieter Fox and Henry Kautz. Inferring ADLs from Interactions with Objects. *IEEE Pervasive*, to appear.