Contextualized Mobile Support for Learning by Doing in the Real World
Natalie Linnell, Ray Bareiss, Martin Griss
Carnegie Mellon Silicon Valley
{natalie.linnell, ray.bareiss, martin.griss}@sv.cmu.edu

Abstract
This research addresses the use of mobile devices with both embedded and external sensors to provide contextualized help, advice, and remediation to learners engaged in real-world learn-by-doing tasks. This work is situated within the context of learning a complex procedure, in particular emergency responders learning to conduct urban search and rescue operations. Research issues include the design and delivery of contextualized performance support and the inferring of learner actions and intentions from sensor data to ensure that the right support is delivered just in time, as it is relevant to what the learner is doing.

Introduction
Learning by doing in which learners engage in an authentic task – receiving help, advice, and remediation just in time as it’s relevant to what they are doing – is generally regarded as the most effective way to gain “knowledge to be used” in contrast to “facts to be memorized” [6]. Current eLearning systems provide learning-by-doing experiences via simulations in which a facsimile of a task is created in the closed world of a computer program; the learner’s progress can be tracked, all possible learner actions can be recognized, and thus contextually-relevant help, advice, and remediation can be provided. Many tasks, however, are better mastered in the open, real world. In particular, when learning complex tasks such as urban search and rescue, car engine repair, building construction, or environmental sampling, interaction with the physical world is essential for mastery. In the past, providing contextualized performance support in this environment was too difficult because it was generally not possible to know what the learner was doing in the real world without requiring explicit input. Advances in mobile technology include the incorporation of rich sensor packages, which can enable reasonable inferences about the learner’s needs, potentially making computer-supported, contextualized learning by doing possible for a range of real-world tasks. That said, much work remains to be done on two fronts: The first is interpretation of rich sensor data to derive context, that is inferring actions from position, motion, services accessed, and other sensor data. The second is establishing the mapping between actions in a given context and useful support for the learner, which requires inferring learner intentions from actions.

We are developing an eLearning system on the Android platform as an experimental testbed to research tracking the behavior of learners as they
complete a complex real-world task, and to research the most effective ways to provide contextualized help and advice during the completion of such tasks. This paper reports on preliminary work on this project. This research addresses two key problems in the delivery of contextualized learning via mobile devices:

- Appropriate recognition of actions in the open, real world in contrast to the closed world of a computer-based simulation
- Appropriate learning and interaction design for supporting the learning of complex tasks, using mobile devices to provide performance support as the learner interacts with objects in the world - rather than with the eLearning system per se.

We are conducting this research in the context of learning a complex urban search and rescue procedure, which requires the use of diverse tools, protective equipment, and other resources. This is a task in which recognition of actions and inference of learner intentions is critical to successfully providing contextualized help, advice, and remediation and in which appropriate learning and interaction design are essential to ensure that such support aids rather than distracts the learner.

**Related Work**

Other researchers in the ubiquitous learning community have discussed the utility of a range of sensors in establishing context (e.g., [33,14,4,31]), but interestingly, implemented learning applications tend to use only location, and sometimes proximity to tagged objects, as simplified proxies for rich sensor-determined context. The majority of these previously developed solutions focus on learning tasks that tend to be somewhat unfocused (e.g., learning about plants in a botanic garden [4,19] or paintings in a gallery [23,9]) and are similar in spirit to a number of GPS-based tour guide projects (e.g., [32,1,10]). A few projects do attempt to teach more structured tasks (e.g., performing single-crystal X-ray diffraction [15] or assembling a computer [12]); these too rely on location, proximity to tagged objects, and occasionally, learner input to determine what guidance to provide.

In parallel, researchers in the ubiquitous computing community have been researching the recognition of actions from sensors (e.g., [20,2,24,27,8]). Their approaches use a mix of techniques drawn from statistical machine learning and natural language processing to create recognition models and to apply them to segmented sensor data. Various inference techniques are used to integrate data from different sensors for higher-level recognition (e.g., [25]). Commercial sensor products are also imminent in this space (e.g., [18,13]).

**Background on Urban Search and Rescue**

After a disaster such as an earthquake or hurricane, Urban Search and Rescue (US&R) personnel focus on locating and evacuating victims, especially those trapped in collapsed structures. These personnel are generally firefighters who
receive additional training to become certified in US&R. Though no two disaster sites are the same, a general procedure is followed when searching a site and evacuating victims, and teaching this procedure is the focus of our work. In order to provide context for the discussion that follows, we present a brief overview.

Upon arrival, the team drives around the building to perform the initial size up. They then establish a set of staging areas for command, equipment, and medical, being sure to establish these areas in places that are a safe distance from the building and that ensure easy access and exit. At this point an initial plan for the search is agreed upon, and the team is briefed.

The first major phase of the search is the exterior assessment. This involves walking around the building to determine whether it is safe to enter the building, locate possible points of entry, and to assess possible locations of victims. The personnel conducting the assessment produce an exterior map that is used to adapt the existing search plan.

At this point, the team begins the interior search. There are a set of important activities associated with this phase. The US&R personnel must appropriately mark the buildings, floors, hallways, and rooms to indicate that they have been searched, and whether any victims or hazardous materials were located. Victims must be evacuated. The team will continuously use tools such as air quality monitors to assess the safety of the areas they are searching. They will use highly trained search dogs (Figure 1a), listening devices, and cameras to locate victims, and they will breach obstructions, as necessary, to rescue them (Figure 1b). Correct use of tools is an important part of the training (Figure 1c); Figure 1d shows the US&R personnel using hand tools to build a wooden support for a damaged ceiling.

1 A simpler US&R process is learned by Community Emergency Response (CERT) teams. Such learners may benefit even more from just-in-time support.
US&R training typically presents a scenario that requires the US&R personnel to follow this procedure. However, there is a set of important challenges associated with creating and conducting a realistic US&R training scenario. First, there are a small number of locations nationally that realistically simulate a collapsed structure; training must often be conducted in completely intact buildings. Trainees are asked to imagine that a collapse has occurred, but it can be difficult to map the scenario to the environment without the visual and physical indications of obstacles and issues that exist in a real collapsed building. This can make it difficult to treat the scenario as one would a real emergency. In addition, trainees have a tendency to take shortcuts with safety precautions, such as the use of safety gear including breathing masks. Another important challenge, especially in light of recent budget cuts, is a shortage of training personnel to lead training and provide remediation.

These difficulties with current training, together with the inherently procedural nature of US&R activities, make US&R training an excellent candidate for the real-world eLearning approach we propose. In addition to being a good testbed for research in eLearning, creating a high-quality learning experience with the technology would provide tangible benefit to US&R training (such benefit would also be realized by CERT as mentioned above). In particular, it would allow more frequent training than is currently possible, especially for small or under-resourced departments.

Two Complementary Threads of Research

Our research involves two threads of work. The first focuses on designing and prototyping the core learning system, which will provide help, advice, and remediation to emergency responders learning to conduct urban search and rescue. This work builds on previous research on “Teaching Architectures,” discussed in the next section. The second focuses on adapting and integrating existing techniques for inference of learner actions as appropriate for our system (discussed in the section Sensor Fusion).
Applying Teaching Architectures to Urban Search and Rescue Training

The design of this system builds on a large body of previous work at Northwestern University’s Institute for the Learning Sciences (ILS) on “Teaching Architectures” [28,29,30,5,3,16]. Teaching architectures provide general templates for the development of simulation-based eLearning systems to teach specific kinds of task. They were formulated after ILS had developed a large number of eLearning systems, and a set of patterns of “best practices” began to emerge related to the types of tasks being taught. Thus these teaching architectures are grounded in the practice of building and evaluating eLearning systems.

The idea behind teaching architectures is that, given a type of task and the nature of performance deficits that people typically exhibit in task performance, there is a standard “teaching architecture” which will be effective in addressing those deficits. In this paper, we focus on the teaching architecture for the task of learning a complex procedure, which focuses on teaching a procedure by coaching the learner through a sequence of actions or decisions. To illustrate the breadth of these teaching architectures, we quickly describe two other task types. One example of a type of task is “observe and critique,” where the student observes the behavior of others and must identify and critique specific actions based on applicable laws, regulations, codes of conduct, standard process definitions, and so on. Another is “investigate and decide (i.e., diagnose),” in which the student is confronted with a diagnostic problem and charged with determining what evidence is needed to distinguish between the possible diagnoses, collecting the evidence, coming to the correct diagnostic conclusion, and presenting the evidence that supports the conclusion. The teaching architecture itself consists of a set of strategies for supporting the given type of learning problem that are likely to generalize across potentially very different instantiations of that learning problem. Identifying a teaching architecture allows reuse of these teaching techniques across applications; and when an eLearning system is being developed, it also allows re-use of system components and even code. While the teaching architectures require some customization for a given application domain, experience with using teaching architectures to build eLearning systems has shown a great deal of generalizability in these patterns.

In our work, we build upon the teaching architecture for learning a complex procedure. This architecture is a good fit for the set of fairly procedural tasks that must be undertaken by urban search and rescue first responders. One of the most important best practices for this teaching architecture relates to remediating errors. When training novices to perform procedural tasks, ILS found that it is most effective to constrain people to follow a designated path fairly closely by making them aware of their mistakes quickly. If allowed to stray too far down a faulty path, it can become difficult for the learner to identify the actual point at which the error occurred, and the opportunity to learn from the mistake may be lost. The handling of mistakes is one of the most important aspects of the teaching architecture for learning a complex procedure, since it is at these moments when
we have the best opportunities to alter the learner’s mental model of the procedure. In addition to following these best practices, we can also gain from building on the set of important intervention points identified for a system that implements the “learning a complex procedure” teaching architecture:

• **Introduction to the scenario.** Since complex procedures occur in the real world, an important part of successfully enacting this teaching architecture is embedding the learning experience in a realistic scenario. The learner’s first exposure to the scenario should promote engagement and “suspension of disbelief.”

• **Orientation to learning resources.** After the introduction to the scenario, the system should provide an explanation of the learning environment itself. In particular, if we are developing an eLearning system, this step would explain the support that is available throughout the learning experience. Global support might include explanation of the task context or enumeration of goals/subgoals and the accompanying subprocedures for achieving them.

• **Beginning and during a procedural step.** Support for a given procedural step will include such things as an explanation of preconditions, resources required, and a recommended procedure.

• **A mistake is made.** Remediation of incorrect actions (e.g., the learner did not adequately examine the perimeter of a structure for signs of damage before entering it, or he or she is using a tool incorrectly) is perhaps the most important function of the system. It is important that this remediation is done within the scenario, allowing the learner to see the likely outcome of an incorrect action in the world before the learning system provides remediation. Otherwise, as humans tend to anthropomorphize digital devices [26], they are likely to become annoyed with the system for “criticizing” them. Conversely, remediation provided after an outcome has been observed is likely to be regarded as helpful coaching. ILS found that providing a standard set of questions for the learner to ask after a mistake helped the learner feel that they were being coached, not criticized. Questions included: “What happened?” “Why?” and “What should I do next?”

• **The scenario is completed or a significant point of closure is reached within the scenario.** It’s important to provide closure and opportunities for reflection at the end of the learning experience, or at points where the flow of the task has paused because the learner feels that they have accomplished something meaningful. A few specific examples will make the sorts of actions the system must recognize and remediate more tangible; we also describe how the system provides just-in-time help and advice to responders before they undertake actions.

• **A team of responders should conduct a “size up” by walking around the perimeter of a building and taking structured notes before entering.** If they do not walk around the entire structure, as observed from a GPS track, or do not note key aspects of the situation, the system should remediate these errors by showing the potential outcome of a faulty size-up, providing an overview of an appropriate...
size-up procedure, and prompting the learner to enter the correct information in the size-up form embedded in the system.

• If there is a risk of hazardous materials, the team should possess (and wear) all appropriate HAZMAT gear; the system should remediate such things as lack of appropriate breathing apparatus or perhaps even an inadequate air reserve due to previous use. The system can provide guidance on how to correctly inspect, don, and use HAZMAT gear, perhaps in the form of a short video lesson.

• The team should mark a building with a FEMA-standard marking of an appropriate size, using spray paint, to indicate its status. If they do not have a can of spray paint of an appropriate color before beginning the operation, the system should remediate. If they do not mark a building, mark it in an appropriate location (e.g., their mark is not visible from the street), or do not use a standard marking, the system will recognize errors by comparing a learner-input photo of his or her marking with a catalog of FEMA-standard markings for buildings, and will present this error to the user together with how to correct it.

• Responders should use tools correctly. For example, responders use a tool called a TAC stick to detect live electrical current. The TAC stick should be swept over the exterior walls of the building in an arcing motion. If the TAC stick is being used incorrectly, the system can show the learner the real-world implications of not checking for electric current, provide guidance, again perhaps in the form of short video lessons, on the appropriate use of tools.

Design Process

Through extensive experience with purely computer-based eLearning simulations, ILS settled upon a standard set of items to present in a fairly standard layout in the interface of a system to teach the learner to follow a complex procedure. Figure 2 provides a conceptual overview of the elements of the knowledge-level teaching architecture. Rather than starting from scratch in our design process, we decided to begin from this structure, and change it as necessary to accommodate use for activities in the real world, and use on a mobile phone. In the rest of this section, we describe the process we took to translate this structure into something that could be used on a mobile phone, for tasks that occur in the real world.
In the computer-based learning environments developed at ILS, after an introduction to the scenario, the learner is taken through the task step-by-step. At each step, the learner is presented with a screen that contains four elements: presentation of the step in the scenario, the set of available actions in the scenario, access to tools and resources, and performance support in the form of questions that the learner might want to ask. If the learner makes a mistake, they are taken to the remediation screen, where the outcome of their mistake is presented in the scenario together with information to help them understand what happened and how to correct their mistake, before returning to the main flow of the scenario.

When beginning the design process for this project, we took the approach of beginning with a fairly direct translation of this simulation-based layout to the mobile platform, creating a low-fidelity mock-up of what this would look like for a fragment of an urban search and rescue scenario. We then proceeded to remove information and actions that would correspond to interactions with the real world, iterating on the mock-up. The use of a tangible artifact (the mock-up) to inform our thinking greatly aided in the design process.

The design process was heavily influenced by a couple of important factors. First, the learner would be interacting with the technology in a fundamentally different manner, since the majority of the learning activity would now take place in the real world. This meant that the system could not distract from the real-world task, requiring the application to take a much more secondary, or transient, posture in its interactions with the learner. In addition, we were now designing for the small screen of a smart phone. In each domain where a teaching architecture is applied, the architecture must be tuned to work with that domain. Dealing with the small screen of a smart phone exacerbated the necessary tuning; since such a small
amount of information can be presented on each screen, multiple-screen interactions must be created for more complex steps such as filling out a size-up form. This meant that we needed to be much more judicious in presenting information to the learner, since the small screen limits the amount of information can be presented.

While developing this system, we found it useful to be able to parallelize the design of the learning environment with the development of the activity recognition system. Significant advantages of this separation of concerns are that design of the learning system will be learner centered rather than technology centered, and correspondingly that requirements for recognizing particular actions will be driven by the needs of the learner's task rather than by what is possible a priori. In order to achieve this, the system under construction is adopting a “Wizard of Oz” approach [17,7,23] to initial usability and learner testing of the prototype. Wizard of Oz prototyping is a fairly standard user testing technique where certain aspects of system function are simulated by a human experimenter during a user test. This allows designers to test the system before all of the technology has been completed. Since no systems for conducting Wizard of Oz tests on Android phones were available, we built a system that allows us to run a Wizard of Oz experiment from a laptop on an Android phone over a WiFi connection. In our experiments, the learner will interact with the system as if it is completely autonomous, except for the fact that an experimenter will inform the system of learner actions rather than attempting to rely upon autonomous, sensor-based inference of those actions.

Sensor Fusion

The second stream of work is to adapt and integrate the approaches devised by colleagues (especially [8]) and other researchers (e.g., [20,2,24,27,25]) for action recognition from sensor data.

Most of the sensor work that we will be doing will focus on two areas: determining the location of the learner, even when indoors, and inferring learner actions from accelerometers and other sensors.

When we first began envisioning the portion of the system that infers learner actions, we expected to rely primarily upon sensors embedded in the phone itself. However an important factor led us away from this idea: first responders’ extensive use of a large set of tools (including tools for conducting breaches such as saws and jackhammers as in Figure 1, and tools for assessing safety such as air quality monitors and TAC sticks, which detect live electrical current). The training focuses to a large extent on the correct use of these tools. Detecting which tool is in use and whether it is being used correctly using only sensors on the phone is problematic. Thus when trying to detect whether the learner is using the tools correctly, it makes sense to instrument the tools themselves with sensors. In particular, we will rely heavily upon wireless accelerometers that can be attached to the tools, and can transmit data to the phone (e.g., [13]). In addition, we will use Near Field Communication (NFC) tags to detect which tools the learner has selected.
When the learner is outdoors, we can use the GPS on the phone to determine their location. However, GPS will not allow us to track the learner when they are inside a building – and tracking people inside a building is a difficult problem – one without a consensus good solution [22,21,36]. For this reason we are investigating a set of existing techniques for indoor locationing. The most commonly used set of techniques require the building to be consistently covered by WiFi or bluetooth; the system then uses the signal strength from different wireless access points to determine its location [22,21]. Since walls are sometimes transparent to the wireless signal, in order to determine which room the learner is in, the system must be trained to recognize the pattern of signal strength of wireless access points for a given location (this is referred to as the location’s “fingerprint.”). Another technique involves inferring learner movement from the accelerometer in a cell phone and comparing these movements to known maps of the building to determine the learner’s position. This is known as the “dead reckoning” approach, but errors quickly accumulate and make this approach difficult to implement in practice [36]. Finally, it is possible to instrument the building with RFID readers and attach RFIDs to objects we wish to track; however this requires a dense deployment of RFID readers, which are very expensive [35]. Active Badging takes a similar approach, using infrared instead of RFID, and suffers similar drawbacks [34]. All of these techniques require some instrumentation of the space, however since this is a training environment we feel that this is a reasonable requirement.

Based upon our review of the literature and consultations with colleagues, we intend to adopt a hybrid approach using dead reckoning augmented by periodic wireless fingerprinting for errors. We have tentatively chosen to use Bluetooth instead of WiFi as our fingerprinting mechanism because of the significantly lower price of deployment.

We envision the architecture of this system as shown in Figure 3. Computation will be carried out on the mobile device whenever possible, but complex processing of sensor data and processing of information from sensors in the larger environment and from services accessed will be carried out in a cloud-based backend. Much data collection will employ external sensors attached to tools in the environment.
Conclusions

This paper has reported on preliminary work on the design and development of a system for delivering a contextualized ubiquitous learning experience via mobile devices. To date, we have completed the design of the system, developed a rich and authentic scenario for teaching urban search and rescue, developed a Wizard of Oz system via which an experimenter can interact with a program running on a mobile device, and completed fairly comprehensive surveys of activity recognition using phone-based and environmental sensors and of indoor locating. Even at this relatively early stage, we can report on several lessons learned: Basing design on a well-understood teaching architecture has provided not only strong design guidelines for the system but also detailed data requirements for implementing the learning experience. The technique of beginning with a direct translation of a desktop-based design to a mobile interface and iterating on this concrete artifact has proven to be an effective design process. The principle of separation of concerns has enabled us to focus on the design of the learning environment, freeing the design process from consideration of the difficult problems of activity recognition and indoor locationing. This decision has also led us to develop a Wizard of Oz system for mobile devices with potential applicability beyond our project. The next phases of our work will involve implementation of the learning system and usability and learning outcomes testing; we plan to complete these by the end of summer 2012.
Acknowledgments

The authors would like to thank members of the NASA Ames Disaster Assistance and Rescue Team, especially Roger Miller, Bob Dolci, and Erik Rockwell, as well as Lynn Brown of the Mountain View Fire Department, the Central (San Mateo) County and South San Francisco Fire Departments, and members of Carnegie Mellon University’s Disaster Management Initiative. We would also like to thank our graduate research assistant Kristoffer Pantic.
References


