Shifting contexts in invisible computing environments

David R. McGee
Pacific Northwest National Laboratory
900 Battelle Blvd.
Richland, WA 99352 USA
+1 509 372 4705
dmcgee@pnl.gov

Misha Pavel and Philip R. Cohen
Computer Science and Engineering Department
Oregon Graduate Institute
20000 NW Walker Road
Beaverton, OR 97006
+1 503 748 7806
{pavel, pcohen}@cse.ogi.edu

ABSTRACT
Invisible computing systems are highly context-dependent. Consequently, the influence that language has on contextual interpretation cannot be ignored by such systems. Rather, once language and other forms of human action are perceived by a system, its interpretative processes will of necessity be context-dependent. As an example, we illustrate how people simply and naturally create new contexts for interpretation by creating new names and referring expressions. We then describe Rasa, a mixed-reality system that observes and understands how users in a military command post create such contexts as part of the process of maintaining situational awareness. Rasa augments both the commander's map and the Post-it notes pasted on it, which represent units in the field, with multimodal language, thereby allowing paper-based tools to interact with digital information. Finally, we argue that architectures for such context-aware systems must reduce the inherent ambiguity and uncertainty through fusion and other means.

Keywords
Context-aware computing, invisible computing, mixed reality, multimodal

INTRODUCTION
Future invisible computing platforms will be severely handicapped, if they cannot understand the context of situated use. A large part of context in situations where people live, work, and play is generated by their language. However, current approaches to context-aware computing ignore the interplay of language and context. In so doing, these approaches fail to support human expectations for contextual interpretation during collaborative task execution, whether with other humans or with intelligent, context-aware machines.

In addition to graphical user interfaces, small, mobile, or ubiquitous systems will increasingly employ more perceptual interfaces, which depend on natural human modalities, such as voice, sight, touch, etc. Even for humans, the derivation of the intended meaning from such inputs is ambiguous and uncertain, dependent on context for proper interpretation. Once human-computer interfaces incorporate these types of inputs, simplistic notions of context become insufficient, because we rely heavily on context to refine the range of potential interpretations.

This essay points out that, in the everyday execution of tasks, not only does language influence the context in which we interpret each other's actions, even going as far as to create new context, but prior linguistic context influences our choice of actions. In particular, people use language to transform everyday objects into stand-ins for more complex, abstract, or less tangible objects, thereby extending or "augmenting" their context and the range of potential meanings. Systems designed to be aware of context must recognize this transformation as well as other potential context shifts that language makes possible.

While studying military command posts, we observed the importance of this human ability to extend the meaning of objects and symbols. Based on these observations, we built Rasa, a system that supports the natural tendency of people to change the meaning of real-world objects, and thereby to change relevant context, with speech and writing. We conclude the essay with a discussion of how the interplay of language and context will influence the design of future context-aware systems.

LANGUAGE AND CONTEXT
The work in context-aware computing views context as comprised of a limited set of primitive, unambiguous, and easily sensed properties, location being the most obvious. Consequently, most systems built under the banner of context-awareness, like Olivetti's original Active Badge research [13], are location-focused. Systems that are sensory-restricted in this manner have a very limited perspective on situated use and thus are impoverished communicators. However, human-computer interaction will increasingly involve the processing of language directed towards the system and towards other people.

In human conversation, people use context to derive meaning and to cope with uncertainty when interpreting each other's intentions, utterances, eye movement, prosody, body language, facial expressions, and changes in lexicon. People do not attempt to understand these inputs individu-

ally or without important background information such as common sense knowledge [8], common ground [4], the center of the discourse [7], or environmental cues like shared artifacts, the setting, persons present, etc. Nor do they treat these features of the situation as perfectly understood. Although these perceptual modalities may each individually be inherently ambiguous to a system, if combined, the communication becomes disambiguated and efficient [5, 11]. Thus, context derived from several inputs can be used to reduce and manage uncertainty.

Consider the simple word, "fire." Alone, the word has at least 35 meanings [1]. Suppose that you are wearing a blindfold, smoking a cigarette, and have recently defected from the Foreign Legion. Hearing someone say, "fire," from twenty feet away would be disquieting. On the other hand, if you were a Legionnaire in the same general vicinity, and instead of wearing a blindfold, you were holding a rifle, the utterance would have a markedly different effect. In both cases, the speaker's intent was that of a command to the Legionnaires to begin firing. However, one can readily see that the reaction to the same declaration will vary dramatically depending upon from which side of the firing line you hear it. Location is an important piece of the context, but not *the only* important piece. Other critical aspects in the interpretation of the utterance are each party's point of view, any shared understanding of the situation, such as the firing squad's knowledge regarding who is the prisoner, what activity is taking place, who is giving the orders, etc. As it turns out, the speaker is standing directly next to the prisoner and a major conflagration is headed towards the assembled parties. The speaker has in fact accompanied his utterance of "fire" with a gesture towards the flames. If the firing squad relied on the prisoner's location alone to interpret the utterance, there certainly may be unintended consequences.

One can see that a bare utterance can be interpreted in vastly different ways depending on the context. Expressions that are more complex would furnish even more ambiguous and potentially conflicting interpretations depending upon how much of the context the utterance itself provides. Insofar as the example here is simple and incomplete, the reader may wish to refer to more complete discussions of this subject [2, 3, 6, 12].

USING LANGUAGE TO BRIDGE CONTEXTS

With language, we have the ability to transform an arbitrary physical object into something entirely different: a combination of the original object and an associated one. For example, suppose we want to give directions to the other side of town, and we know that a prominent water tower and its physical relation to streets, buildings, etc. will aid us. To do this, we often use objects at hand and create an association, using spoken or written language, to combine the two. For instance, we can point at the coffee cup and say, "Let's suppose that this is the water tower on the other side of town." Likewise, we can use a pencil to represent Main Street. Alternatively, we can draw a symbol of a water tower on a piece of paper, next to a pair of lines labeled "Main Street." In each case, spoken, written, diagrammatic, and/or iconic language alters the meaning of the physical objects or symbols, such that they become stand-ins for other objects in a different context (i.e., the cup for the tower). Language, specifically the abilities to name and to create analogical relationships, can be used as a *bridge* between contexts: in this case, between the context of the objects in front of me and that of the real world. These features of language are extremely useful, especially when the referents are themselves too large, too small, too far away, too cumbersome, or otherwise inaccessible. Perceptual systems that can interpret these shifts of interpretation may have advantages, such as the ability to understand the denotations occurring and take appropriate action. In the next section, we will present one such system, Rasa—a tangible, mixed reality system that understands the way that officers in a military command post use language to build tools for managing situational awareness [9].

RASA

One way officers maintain situational awareness is by tracking the movement, relation, and disposition of friendly and enemy forces. They do this by building a physical model using Post-it notes and a map (Figure 1). Rather than build this model with available computers, officers choose to use physical objects. They do so because those objects are robust, as well as easy to use and to share during face-to-face collaboration. They choose paper, because it is high in resolution and malleable (i.e., writing on it can quickly change its meaning). They choose Post-it notes, because they are small and can be easily placed, lifted, and moved without marring the surface of the map.

However, by solely employing physical objects to track situational awareness the officers deprive themselves of any benefits of digitization. In an attempt to provide both the benefits of paper and of digital systems, we identified a set of interface constraints that would support existing pa-

Figure 1. Officers using Post-it notes to track unit positions on a map.

per tools and work practices [10]. We observed that if Rasa could leverage the pre-existing language of the task, including both the spoken language and the symbolic language on Post-it notes, officers could employ the same techniques to interact with digital systems that they use to manipulate the paper tools and to collaborate with each other.

As permanent tangible, malleable, and portable representations used for the purposes of collaborative planning, task execution, and dialogue, the paper artifacts in the command post serve as analogs for real world objects (the map for the terrain and the Post-it notes for the units). With Rasa, officers draw familiar military symbols on ordinary Post-it notes, that then stand-in for military units. This artificial, symbolic, compositional language is composed of text combined with analogical diagrammatics that compactly specify the size and type of each unit (see Figure 2). The system attempts to recognize each symbol, as it is drawn on the Post-it note, while listening for speech as well.

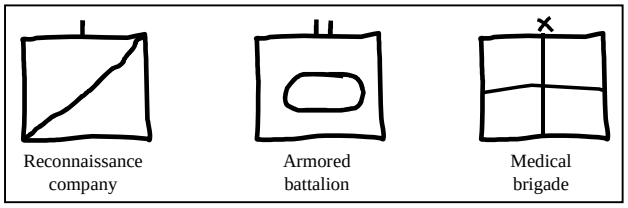

Figure 2. Sample military symbology

When users place a Post-it note bearing a symbol on the map, Rasa automatically adds the unit that it denotes to a shared computer database. The system supports many types of multimodal interactions with the Post-it and map. For example, with accompanying speech, users can draw lines directly on a plastic overlay and their fused meaning is interpreted by the system. Moreover, all of these digitally-augmented physical objects can be named or further specified with spoken language. Collaborators using more conventional computing equipment or another Rasa system can view both the objects (e.g., lines and units) and any officer's multimodal interactions, on the augmented map surface.

Using digitizer technologies and a set of multimodal understanding software components, Rasa works invisibly beneath the paper to capture, recognize, and understand the creation of new contexts, such as registering a map to a location in the world, creating new "units" to manipulate, etc. Rasa works with maps, aerial photographs, and sketches on ordinary pieces of paper. In fact, before map registration, Rasa assumes nothing about the paper on its surface. To register one of these, an officer points at two locations on the map (providing input to the digitizing tablet) and with speech provides accompanying real-world coordinates. The system extrapolates the remainder of the world based on these two utterances. It is this act of registering the paper (or in the future, seeing and understanding it) in relation to the world that puts it in perspective; this act provides context for future use of the paper. It signifies to the system and anyone present that the paper represents a particular part of the Earth. Similarly, adding a symbol to a Post-it note extends not only a particular meaning, but also the context of its use in coordination with others and in relation to the paper map, and thus the world. These actions that produce new context are possible in virtue of the shared multimodal language that approximates the natural spoken and written one that was observed in the field.

In sum, Rasa observes language used as a natural byproduct of teams of people attempting to establish situational awareness. It translates their intentional use of language to create placeholders for real-world objects—extending the meaning/context of the Post-its in the process—into digital representations of the same objects. Rasa communicates its understanding using both conventional display technologies and spoken language output. Officers can then use the digital information it creates to expand the collaboration beyond the physical confines of the command post.

CONCLUSIONS

We claim that the future of human-computer interaction rests in part with perceptual devices that interpret the language and activity of humans in context. This essay has argued that (1) language cannot be understood devoid of context, (2) context is not a set of static properties, but an evolving set of relationships between people and the world in which they interact, (3) language can shift the context in many ways, (4) one way that language shifts context is by extending the meaning associated with physical objects with naming and referring, and (5) these shifts are themselves ambiguous and systems that rely on observation of natural human action must deal with the inherent uncertainty and ambiguity of those inputs. Finally, we have shown how language can be used to establish bridges from one context to another, and how a new system, Rasa, which employs mixed reality tools that combine paper and computational elements, relies on this natural use of language to extend physical objects into new contexts. This aspect of language—the ability to name and refer, thereby creating linguistic placeholders for later use— can be made the basis for the design of interfaces that are aware of broader aspects of the context of use.

ACKNOWLEDGMENTS

We thank the reviewers and our colleagues in the Center for Human-Computer Communication at the Oregon Graduate Institute and everyone who contributed to this work.

REFERENCES

1. *Webster's*(Eds.): Random House, 1991.
2. Austin, J.L., *How to do things with words*. Second ed., J.O. Urmson and M. Sbisa (Eds.). Cambridge, MA: Harvard University Press, 1975.
3. Barwise, J. and Perry, J., *Situations and Attitudes*. Second ed.(Eds.). Menlo Park, CA: CSLI Publications, 1999.
4. Clark, H.H. and Brennan, S.E. *Grounding in communication*, in *Perspectives on socially shared*

cognition, L.B. Resnick, J.M. Levine, and S.D. Teasley, Editors. APA Books: Washington, D.C., 1991, 127-149.

5. Cohen, P.R., McGee, D.R., and Clow, J. The efficiency of multimodal interaction, in the *Proceedings of the Applied Natural Language Programming Conference*, (Seattle, WA, April 2000), ACL.
6. Grice, P., *Studies in the way of words*(Eds.). Cambridge, MA: Harvard University Press, 1989.
7. Grosz, B., Joshi, A., and Weinstein, S., Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 2(21 1995): 203-225.
8. Lenat, D. (1998) *The Dimensions of Context Space*: Austin, TX.
9. McGee, D.R. and Cohen, P.R. Creating tangible interfaces by transforming physical objects with multimodal language, in the *Proceedings of the International Conference on Intelligent User Interfaces*, (Santa Fe, NM, Jan. 14-17 2001), ACM Press, 113-119.
10. McGee, D.R., Cohen, P.R., and Wu, L. Something from nothing: Augmenting a paper-based work practice with multimodal interaction, in the *Proceedings of the Conference on Designing Augmented Reality Environments*, (Helsingor, Denmark, April 12-14 2000), ACM Press, 71-80.
11. Oviatt, S.L. Mutual disambiguation of recognition errors in a multimodal architecture, in the *Proceedings of the Conference on Human Factors in Computing Systems*, (Pittsburgh, PA, 1999), ACM Press, 576-583.
12. Searle, J.R., *Speech Acts: An Essay in the Philosophy of language*(Eds.). Cambridge, MA: Cambridge University Press, 1969.
13. Want, R., Hopper, A., Falcao, V., and Gibbons, J., The active badge location system. *ACM Transactions on Computer-Human Interaction*, 10(1 1992): 91-102.