

# Mobile Information Re-Finding Using a Voice Interface

Robert Capra

Department of Computer Science  
Virginia Tech  
Blacksburg, VA 24061 USA  
rcapra@vt.edu

## ABSTRACT

Mobile users have many needs to re-find information. This research explores how shared context between a user and computer system can be established and used in a cross-modal information re-finding application. Both implicit and explicit context are gathered while a user browses the web with a graphical web browser. This context is then used in a telephone-based voice interface to allow mobile users to re-find the information found while web browsing. This application domain is being used to explore how people approach information re-finding, what information they recall and recognize when trying to re-find information, and how they converse about re-finding in an effort to extend the capabilities of information re-finding systems for mobile users.

## Keywords

Information re-finding, voice user interfaces, mobile computing, shared context, speech recognition

## MOTIVATION

People today have many needs to *re-find* electronic information. A common problem for many computer users is not being able to effectively store, organize, and re-access information that was first found on the web [5, 3]. This problem is especially difficult for mobile users. A recent study found that mobile workers rely heavily on the use of cell phones to enlist the help of co-workers “back at the office” to retrieve information [6]. Mobile users have needs to work seamlessly with information across a variety of computing devices, environments, and situations [6].

A possible solution to the information re-finding needs of mobile users is to provide telephone-based access to personal repositories of information. Telephone-based voice user interfaces are being successfully used in applications with limited domains and can be used to provide mobile users access to personal information. Recent advances in speech recognition and voice user interface design have enabled a wide range of successful

telephone-based voice interfaces such as voice portals, voice dialing services, and services that provide information about stock quotes, train timetables, and airline schedules.

Because of the inherently serial nature of voice communication, voice interfaces (especially for large sets of information) need to streamline and personalize communication in order to provide users access to the information they need quickly and efficiently [1]. The use of *context* can assist with this personalization. *Shared context* that is known to both the system and user can be especially valuable in improving future interactions with a user interface.

A goal of this research is to examine how shared context developed with a user in a graphical interface can be used to help improve interactions with a counterpart voice interface for re-finding information.

## PROBLEM

One of the problems with designing a voice interface for information re-finding is that of *habitability*. Habitability refers to how well the language that users employ to interact with a system matches the language that the system is configured to recognize [4]. Habitability has been shown to be a potential problem in voice interfaces with limited, fixed vocabularies and is often addressed through the careful design of system prompts to help users know what they can say [8]. In a system such as the re-finding application investigated here, users need to converse with the system about a large, dynamic set of items (e.g. the things they are trying to re-find from a repository of personal information). Good prompt design can help solve some habitability problems in such systems, but additional techniques may be required to address habitability for dynamically and implicitly-generated vocabularies. The research described here focuses on exploring how implicit and explicit context can be established in a graphical interface and then transferred for use in a counterpart voice interface. The particular application being used to explore these problems is a telephone-accessible voice interface to help users re-find information that was first found using a graphical web browser.

## APPROACH AND RESULTS

Two main phases of this research are underway: an exploration of voice interfaces for information re-finding,

LEAVE BLANK THE LAST 2.5 cm (1") OF THE LEFT  
COLUMN ON THE FIRST PAGE FOR THE  
COPYRIGHT NOTICE.

and an investigation to examine how people converse about re-finding requests. These phases are described below.

### Exploration of the Concept

I built a prototype system called WebContext [2] as part of an initial exploration into the use of context in voice interfaces and re-finding. WebContext captures information (such as phone numbers and addresses) from web pages browsed by a user, stores the information into an information repository, and then allows users to make re-finding requests using a VoiceXML-based voice interface. Development and testing of WebContext led to the identification of habitability problems in the interface and to ideas for ways to use shared context to improve habitability in voice interfaces.

### Modeling Information Re-Finding and Language Use

In order to address habitability in a voice interface, it is important to have a good understanding of how people speak about the domain of the system. This phase of research focuses on empirically examining how people speak when engaging in collaborative conversations to re-find information.

I have completed an experiment that engaged pairs of human participants in a set of finding and re-finding tasks. During an initial session, the first participant (called the *user*) browsed the web to find information requested by a set of tasks. Then, in a subsequent session, the user was given a new set of tasks that involved re-finding the information found during the first session by talking to a second participant over a telephone. The second participant (called the *retriever*) was given access to a complete history of all the web pages viewed by the user during the first session. The results of this experiment showed that, in general, participants approached collaborative re-finding as a two-stage process – first *searching* for the location where the information is stored, and then *browsing* to retrieve the specific information. This collaborative process involved extensive conversational grounding between the participants through the use of URLs, page titles, and descriptions of page contents. Participants made use of these grounding elements as *waypoints* [7] to help guide their searching and browsing.

The two-stage process of *searching* followed by *browsing* appeared to be linked to what users *recalled* and *recognized*. Users utilized information that they recalled to help form spoken search requests and goal descriptions. Retrievers then assisted the users in browsing pages by providing verbal summaries of information that the user might recognize. Typically, this process was iterated until either the information was found or the participants decided that the information could not be found. From examining the conversations, it is clear that context plays an important role in this process of searching and browsing.

### FUTURE WORK

I am using these findings and observations to develop models of how people speak about and approach re-finding information found on the web. I also plan to examine how users' engagement at the time context is developed can affect the role of that context in future interactions. Gaining a better understanding of what people recall and recognize and how they use it to re-find information will provide guidelines for the design of both voice and non-voice interfaces for information re-finding.

### ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Manuel Pérez-Quiñones, Dr. Naren Ramakrishnan, and Miranda Capra for their continued support and feedback. This work was supported in part by the National Science Foundation under Grant No. IIS-0049075, and by a grant from IBM to explore the use of VoiceXML within their WebSphere product.

### REFERENCES

1. Billsus, D., Brunk, C.A., Evans, C., Gladish, B., and Pazzani, M. (2002). Adaptive Interfaces for Ubiquitous Web Access. *Communications of the ACM*, 45(5):34-38, May 2002.
2. Capra, R., Pérez-Quiñones, M., and Ramakrishnan, N. (2001). WebContext: Remote Access to Shared Context. In *Proceedings of the Workshop on Perceptual User Interfaces* (PUI 01, November 15-16, 2001, Orlando, FL).
3. GVU Center at Georgia Tech. (1998). GVU's 10th WWW User Survey (October 1998). Accessed: March 19, 2002, from [http://www.gvu.gatech.edu/user\\_surveys/survey-1998-10/](http://www.gvu.gatech.edu/user_surveys/survey-1998-10/)
4. Hone, K.S., and Baber, C. (2001). Designing habitable dialogues for speech-based interaction with computers. *International Journal of Human-Computer Studies*, 54: 637-662.
5. Jones, W., Bruce, H., and Dumais, S. (2001). Keeping Found Things Found on the Web. In *Proceedings of the Tenth International Conference on Conference on Information and Knowledge Management*. ACM Press, New York, NY, 119-126.
6. Perry, M., O'Hara, K., Sellen, A., Brown, B., and Harper, R. (2001). Dealing with Mobility: Understanding Access Anytime, Anywhere. *ACM Transactions on Computer-Human Interaction*, 8(4): 323-347.
7. Maglio, P. and Barrett, R. (1997). How to Build Modeling Agents to Support Web Searchers. In *Proceedings of the Sixth International Conference on User Modeling* (UM97, June 2-5, 1997). Springer Wien New York, 5-16.
8. Yankelovich, N. (1996). How Do Users Know What to Say? *Interactions*, 3: 32-43.