Mobile Voice Access in Social Networking Systems

Stan Kurkovsky, David Strimple, Eric Nuzzi, Kerry Verdecchia
Central Connecticut State University

Abstract

This paper presents a multimodal social networking system aimed at sharing geographic information among proximate users. The system provides users with a traditional web-based interface along with a voice-based interface that can be accessed over a mobile phone. We describe the main design features of the system, different aspects of its functionality, as well as a number of design challenges imposed by multimodality on a social networking system. This paper also presents the results of the end-user acceptance study conducted to assess the usability of the system.

Keywords: mobile systems, social networking, voice interface, VoiceXML.

1. Introduction

Proliferate use of mobile telephones has brought a widespread availability of Internet applications to mobile users. The use of these applications, however, may be hindered by some features of the mobile device, such as its screen size which is often too small to effectively handle the graphical content that can be found on the Internet. Mobile phones may also suffer from being too slow, or having an inconvenient keyboard, making it difficult to access lengthy or media-rich information found on the web [1], or a relatively short battery life that may not be sufficient enough for such network traffic-intensive uses as web browsing or viewing mobile video broadcasts. The primary functionality of any phone, no matter how basic, is to enable voice communication, which still remains the most natural and simple method of communication, ideally suited for on the go and hands-free access to information. A voice interface allows the user to speak commands and queries while receiving an audio response. Furthermore, a combination of mobile and voice technologies can lead to new venues for marketing, entertainment, news and information, and business locator services [13]. Applications that so far have enjoyed a wide popularity in the traditional desktop medium, such as social networking systems, may also benefit from voice technologies as a bridge into the mobile market.

A social network may be viewed as a set of people, organizations, or entities that are connected by a set of social relationships, such as, friendships or common interests [5]. In the past, people would ask their neighbors and friends in the community for advice about restaurants, local services and entertainment. Today, they search for the information on the web; in this way, social networking may be viewed as an adaptation of the old concept to the new technology [15]. Social network subscribers can do more than look in the local newspaper or do a search for local reviews or announcements; now they can rely on their network of friends to get information and solicit their opinions.

The vast majority of the most popular social networking systems is web-based and, therefore, may not be easily accessible to mobile users. In this paper, we present See Central Connecticut (SeeCCT), a prototype of a multimodal social networking system designed for sharing of geographical bookmarks; the system is accessible via a traditional web-based interface, as well as via a voice-based interface suitable for mobile phones. The remainder of this paper is organized as follows. Section 2 presents an overview of related work in the areas of voice technologies and their use in mobile applications, as well as background information on social networking systems. Section 3 introduces SeeCCT, its architecture and a number of problems and respective solutions related to the design and implementation of a multimodal social networking system. Section 4 details the end-user acceptance survey conducted to assess the usability of the system. Section 5 concludes the paper with a summary and directions for future research.

2. Related Work

2.1. Voice Technologies

VoiceXML is the de facto standard for implementing voice-based interfaces between computer systems and their human users. A VoiceXML-enabled application may interact with its user by playing fragments of pre-recorded audio and converting text to speech; it can accept user input by recognizing or recording the voice input and accepting touch-tones generated by a telephone. Many VoiceXML portals provide developers with a variety of tools for creating telephone-based applications

978-0-7695-3099-4/08 $25.00 © 2008 IEEE
DOI 10.1109/ITNG.2008.72

with voice interface. In addition to enabling applications to provide users with a natural language interface to access and retrieve content from the web [3, 8], VoiceXML portals have been successfully used to implement a wide variety of desktop and mobile applications, from transcribing medical records [2] to assisting police officers communicate with computerized information systems hands-free during critical situations [4]. Mobile applications are positioned to gain the most from voice interfaces [9]; for example, a museum in Utah successfully used a VoiceXML solution to let their patrons take a self-guided audio tour [10]. In the past, the patrons would check out a tape player to take on the tour. But the equipment could become damaged, and the information was difficult to update since the tour must be recorded sequentially on a tape. The new system made use of their customers' own mobile phones; the customers called a service hosted by BeVocal (www.bevocal.com), a commercial VoiceXML hosting service. This way, the patrons could tour the museum in any order that they wished. The system was also much easier to maintain, since it was no longer sequential. The museum was not the only one to implement a system such as this. Murmur (murmurtoronto.ca) provides virtual audio tour services. But, in this case, the tours were recorded by the users themselves, and the venues were locations in a number of cities including Toronto, Montreal, Edinburgh, and Dublin.

2.2. Social Networking

The Internet has become a growing and important part of the civic, social and economic lives of many of the world's citizens. Online social communities have evolved from being mere meeting places to serving a diverse range of functions. The desire for companies to capture the economic benefits of the Internet and, in particular, social networks, has driven the search engine giants, Google, MSN, and Yahoo to race to incorporate social searches, because if they do not, they will be surpassed by such social networking services as Friendster and MySpace. The big social networks, such as Friendster (www.friendster.com), which as of January 2008 had more than 54 million members up from 21 million just over two years ago [12], act as a search engine themselves by organizing their users by preferences and similar interests listed in their profiles. Social networks are said to be in their third generation [11, 12]. The first generation of social networks such as sixdegrees.com came into existence in the late 1990's and ended before the decline of the dot-com boom due to a lack of a revenue-producing business model. This generation also lacked an objective; users had no clear idea how they could actually use the concept of social networking. Between the 2001 and 2005, second generation social networking services, such as Friendster, Tribe, and LinkedIn, introduced a viable economic model, but still lacked any direction. The users could do little more than make a profile, invite new users, and search other profiles. As a result, there was a burnout for such sites [14]. The third generation, from the end of the second generation to the present moment, has incorporated tools and mashups that let users upload media such as pictures, audio and video clips, and seamlessly link them with third-party online services (such as Google Maps or Flickr APIs). However, overuse of such heterogeneous multimedia and related services often results in networks becoming cluttered with unstructured information [6].

Today, each social networking system usually focuses on a particular aspect of interests that may bring people together: maintaining professional contacts (LinkedIn), extending online contacts among classmates (FaceBook), or connecting people based on their musical preferences (Last.fm). Murmur is an example of a social network that connects people based on their presence at or an interest in a particular geographical location. Murmur offers its users a chance to make and listen to virtual tours of various cities. Murmur (the company) installs specially designed signs in places of interest; each sign includes a telephone number, which anyone could call and listen for tidbits of information that may be of interest to the visitors of that location. Yellow Arrow (yellowarrow.net) is somewhat similar to Murmur; the company distributes small yellow stickers in the shape of an arrow; each sticker has a unique code, which can be sent by SMS to Yellow Arrow; the system will also reply with a text message describing something that is significant to the original author who created the message and left the yellow arrow at that location. Both Murmur and Yellow Arrow offer its geographically proximate users a chance to tell the community that they have found something potentially interesting or important about a particular place.

Dodgeball, which initially served clients in New York City and currently is available in 22 cities across the US, is another example of a social network focused on serving users from the same locality [16]. Unlike Murmur and Yellow Arrow, Dodgeball is intended to link user with common social interests. Dodgeball started out as a web-based service, but currently it is a network geared towards mobile users. Dodgeball provides a range of social networking tools, which also include text messaging and mapping services. One of its many social network enabling features includes text messaging all of the user's set of preferred friends within a ten-mile radius and notifications about the friends' activities.

3. Social Networking with Voice Technology

This paper presents See Central Connecticut (SeeCCT), a system that combines many features of social networking services geared towards mobile users

[15]. SeeCCT focuses on implementing features similar to those of Yellow Arrow, but instead of using text messages, it returns an audio response when accessed by phone. To limit the current scope of the system, its user-generated content focuses mainly on attractions and notable places in Connecticut. As a multimodal system, SeeCCT has two front-end interfaces: a web-based interface, which is typical of systems, such as Yellow Arrow and Murmur, and a voice-based interface operating over the telephone using the voice portal provided by BeVocal. This way, users can access or post information using a land-line telephone, mobile phone, or the Internet. The SeeCCT system specifically allows the users to comment and respond to comments left by other users regarding the cities in Connecticut and their amenities, such as parks, schools, etc.

SeeCCT users can register and access the system over the Internet at www.seecentralct.com, as shown in Figure 1. They can also use this website to get entries, called bookmarks, by searching by region (the town name) or by type, such as park, restaurant, school, etc. They can search by user identification number in case they want to find their own entries to view or edit, or to see the bookmarks left by a user that they have discovered to be a particularly useful source of information. Users can leave new bookmarks or leave feedback to an existing bookmark on both the website and phone portal. On the phone, users can record and retrieve audio information. SeeCCT users can access the system over the phone by responding to a number of audio menus and prompts.

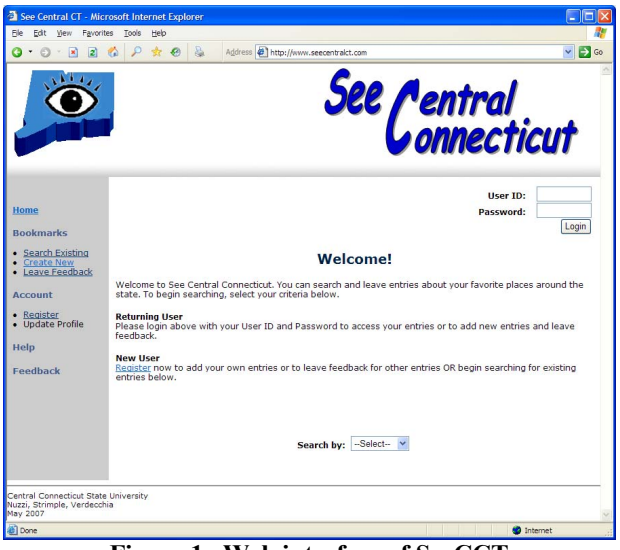

Figure 1. Web interface of SeeCCT

3.1. Architecture of SeeCCT

SeeCCT is a social networking application somewhat similar to Yellow Arrow where people can post just about any information that they would like, as long as it is not profane or libel, about their communities and respond to one another. Unlike Yellow Arrow, the voice interface accepts and returns voice responses, not text. To achieve this, SeeCCT has two front-end interfaces, a phone interface using a VoiceXML portal and a web interface. The web interface employs the search processes, and would most likely be the medium for a SeeCCT user's introduction to the system, whereas the phone interface is more ideal for a mobile, hands free access to the system.

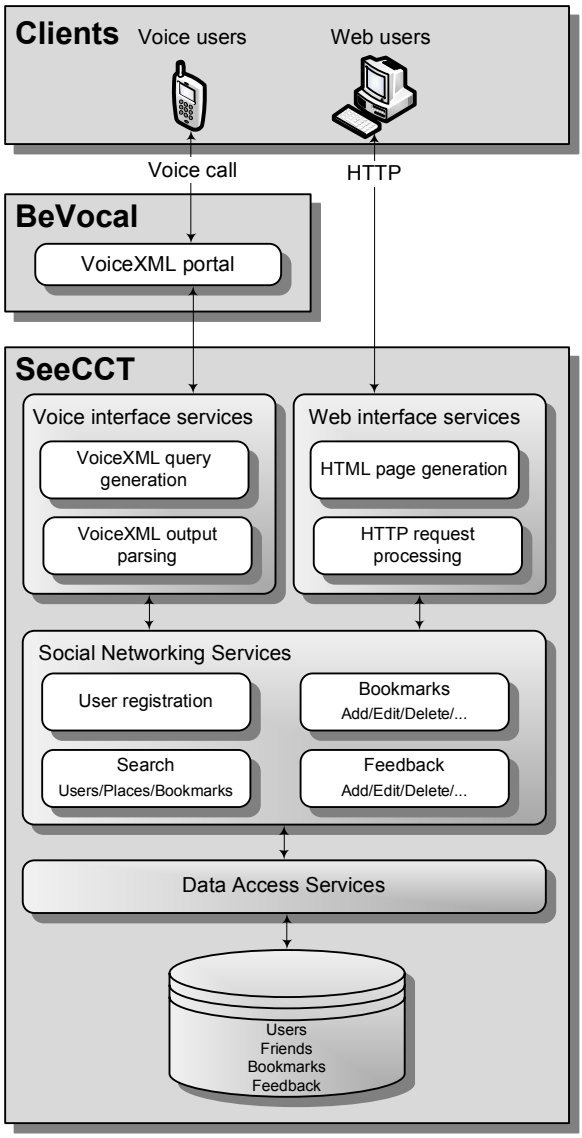

Figure 2. Architecture of SeeCCT

An outline of the SeeCCT architecture is shown in Figure 2. The presentation layer of the system consists of the voice and web interface services implemented using ASP.NET with C#. Voice users interact with the system through the interface provided by a VoiceXML portal. As shown in Figure 2, Social Networking Services

implement all application level logic, including more elementary but necessary underlying functionality, such as user authentication. The purpose of the Data Access Services is to communicate with the data store containing all of the user-generated content of the system.

SeeCCT could be used in many ways via its voice or web interface to access bookmark data stored in the system's database. Suppose that this system was being used to search for all of users who subscribe to a social network that live within a twenty mile radius of the person submitting the search. The user could access the system using its voice interface. After dialing in, the VoiceXML portal's voice gateway would retrieve VoiceXML pages from SeeCCT's Voice Interface Services and begin a dialog, within which the user could make a voice request to get the desired information. The application-level functionality of this request will be served by the Social Networking Services, which, in turn, would be passed on to the Data Access Services to retrieve the corresponding data. Depending on the type of query, the resulting information is likely to contain a mixture of text and audio information. Furthermore, depending on which interface the user is using to access the system, SeeCCT will generate the output suitable for that interface. When the user accesses the system via the voice interface, all text-based output will be played using VoiceXML text-to-speech conversion. If the user accesses the system using the web, all corresponding elements of the audio output will be linked to the resulting HTML pages and available for playback.

3.2. SeeCCT Design Considerations

The SeeCCT database stores information about users, bookmarks, regions, types, and feedback, which are explained in detail below. Users can only register over the web by submitting their first name, last name, email address, and personal identification number (PIN). It would be an arduous process to do this over the phone because names could easily be misinterpreted by the voice application. The alternative to misspelled names would be for the user to spell every entry, but this process would feel artificial and not enjoyable. Users are only required to register if they wish to leave a bookmark or feedback entry. Users do not need to register if they simply want to search for bookmarks or feedback and read them on the web, or listen to them over the phone. Upon registration, the user is assigned a five-digit user identification number (userID). User-selected PIN also must only contain digits. The reasoning behind this is the same as for disallowing telephone registration. Words are more likely than numbers to be misinterpreted and it is not desirable to have different login information for the phone and the web interfaces. Also, this allows the system to take advantage of the DTMF feature of VoiceXML. Research indicates that users might often prefer to use keypad input to speaking their choices when navigating mobile VoiceXML-based applications [10]. It is preferable to give the users the option to use the keypad or the spoken word, and not force them to navigate the phone interface vocally if they do not wish to do so.

Users can search and registered members can post as well as search for information about the towns whose names have been entered into the system. For the purpose of system prototyping, the list of towns was limited to those within a fifteen-mile radius of Hartford, the state capitol of Connecticut. Currently, the system must be bootstrapped with this data; however, the system can easily be extended to allow registered users add new towns to the system.

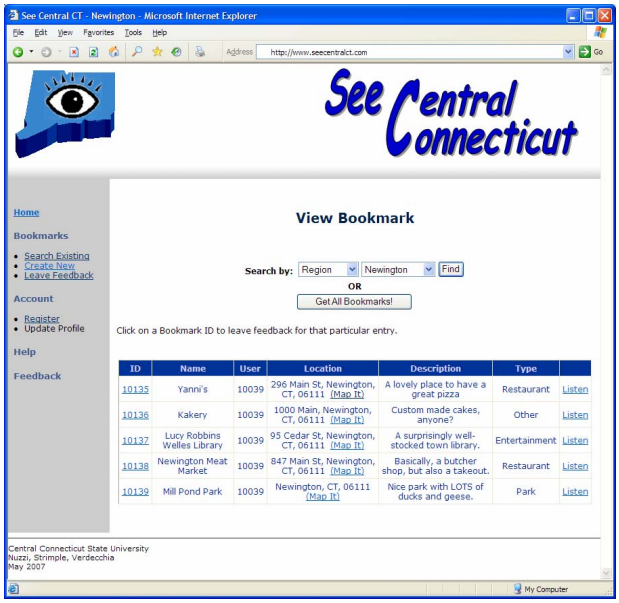

Figure 3. Bookmarks in SeeCCT

All initial entries made by users in SeeCCT are referred to as bookmarks. Currently, the system allows entering bookmarks according to of the following six categories: Government, Restaurant, Park, School, Entertainment, and Other, as shown in Figure 3. In addition to the bookmark itself and its name, each bookmark entry also contains information about the user who entered the bookmark, as well as region/town, bookmark type, and address. On the SeeCCT web site, bookmarks are searchable by each of these categories. While searching by region or type is obvious, searching by user name was made available because of the following reasons. First, the user may want to search all entries they themselves entered in the system. Second, a user may find that another particular user has left a fairly valuable bookmark, and they would like to see what other entries that user had authored. The bookmark itself can contain text, audio, or both.

SeeCCT would not be able to claim having any social networking capabilities if the system didn't facilitate an interaction among its users. The value of the system increases if one can do more than just read what others write. Feedback is the medium through which such an interaction occurs in the system. If, for example, a user leaves a particularly flattering review about a certain restaurant in Hartford, another user may reply with a question or their thanks. Yet another user may share a negative experience with that restaurant. A dialog may form through bookmarks and feedback. And it may be this series of bookmarks and feedback that gives a better overall description of the subject matter, and hence offers the true value of the system.

4. The SeeCCT Usability Survey

To gain insight into the user satisfaction and acceptance of the SeeCCT system, twenty participants volunteered to experiment with the system by completing a set of tasks and then complete the survey presented in Figure 4. The participants were students, educators, and technology professionals ranging in age from 16 to 53. The participants were asked to visit the SeeCCT web site, register, leave a bookmark, leave feedback, and search for bookmark and feedback entries in any manner that they wished. They were then asked to try the phone interface of the system by calling the BeVocal VoiceXML portal. Over the phone they were asked to create bookmark and feedback entries, and listen to either kind of entry.

	Website		1	2	3	4	5	
1	When using the website, searching for bookmarks is…	**Simple**	5	5	5	5	0	**Complicated**
2	When using the website, the ways in which I can search for bookmarks are	**Not relevant**	1	0	5	9	5	**Relevant**
3	When using the website, registering is…	**Easy**	13	5	2	0	0	**Difficult**
4	When using the website, leaving a bookmark or feedback is…	**Easy**	6	6	6	1	1	**Difficult**
	Phone							
5	When using the phone, I experienced inconvenient delays…	**Frequently**	0	2	3	6	8	**Never**
6	When using the phone, I find that navigating through the system menus is…	**Complex**	1	1	4	1	12	**Straightforward**
7	When using the phone, leaving a bookmark or feedback is…	**Easy**	10	4	1	1	4	**Difficult**
8	When using the phone, I find the prompts are…	**Difficult to understand**	0	1	3	3	12	**Clear and understandable**
9	I find that having a mobile interface is…	**Useful**	6	8	3	2	0	**Not useful**
	General							
10	How effective of a forum do you feel this system is for voicing your opinions about your community?	**Not relevant**	0	3	6	9	2	**Very relevant**
11	I would recommend the SeeCCT to my friends...	**Definitely**	1	8	4	6	1	**Not at all**
12	If this system were available for my community, I would use it …	**Frequently**	0	7	6	5	1	**Never**
13	I feel that the potential information that could be found using this system, such as searching for restaurants and favorite locations would be …	**An asset to the community**	10	5	2	2	0	**Of no importance**
14	To access SeeCCT, I would prefer using the web interface as opposed to the voice interface.	**Agree**	11	2	3	0	3	**Disagree**
15	To access SeeCCT, I would prefer using the voice interface as opposed to the web interface.	**Agree**	4	0	3	5	10	**Disagree**

Figure 4. SeeCCT user survey results

The survey responses in general were positive. Answers to questions 1, 3, and 4 regarding web site usability suggest that the site was sufficient, if not stellar for initial testing. Answers to question 1 and 2 had somewhat unexpected results. Searching for bookmarks is the most robust feature of the SeeCCT web site. User responses were very positive regarding the relevance of the bookmarks search, but not quite as positive about the ease-of-use, in searching for bookmarks.

Acceptance of the phone interface was very apparent from the survey results. Four users, however, reported that leaving a bookmark or feedback entry was difficult. Comments left by the users reported error messages or disconnects. The BeVocal logs were not clear as to why these problems occurred. It is likely that more testing is needed to find out why some errors occur.

Questions 10-13 determine how relevant the users find the SeeCCT system. The results to these questions suggest that the participants felt that a system like this would be a useful asset to the community and a relevant forum for voicing their opinions. But, the results also reflect only a moderately positive opinion regarding how often the participants would use the system, or suggest it to their friends.

The last two questions on the survey were provided to gauge which of the two interfaces, voice or web, was preferable by the users. If a user responded with a 1 or a 2 for both question, the response was discarded. This was seen as a contradiction, showing that the user preferred

the web system over the phone while at the same time they preferred the phone system over the web. However, if a participant were to answer both questions with a 4 or a 5, this was not seen as a contradiction. This could have meant that they did not really prefer either application at all. Of the twenty responses only one was discarded. The overall trend suggested that over half of the participants preferred the web interface over the phone, while twenty percent preferred the phone. This result was not surprising as the majority if not all of the participants were very Internet literate, while for some the phone interface was not as familiar. It would be interesting to see if the younger participants, who likely grew up with a mobile phone, were the ones who preferred the phone interface, but that data was not available in this survey.

5. Conclusions

This paper presented SeeCCT, a social networking system with multimodal interface designed for web-based and voice access. The user generated content of SeeCCT revolves around common interests of the users located in the same geographical area. SeeCCT implements a basis for a social networking system creating a forum for users to comment about restaurants, schools, parks, governments, and more in their neighborhoods, and to respond to the comments made by other users. Using the web-based interface, users can register, login, view and search bookmark or feedback entries, and leave new bookmark or feedback entries if they are logged in. Using the voice interface, the users can navigate voice menus in order to listen to bookmark or feedback entries, or leave new entries once logged in. The current target area of the SeeCCT project is the area of the greater Hartford, the capital of Connecticut.

A survey given to the users of SeeCCT indicates that the system was well received by its users. Some users encountered occasional problems such as errors or disconnects, but in large were able to navigate the system without problems. They found potential value in the system as a forum to voice their opinions about their communities, and in general thought that they would use it often or suggest it to their friends. The users showed a preference to the web site over the voice application, possibly due to the demographics of the users participating in the study

There are a number of possibilities for future development in the SeeCCT project. The web-based interface could have more detailed information about each of the screens and entry forms, so that users have a better understanding of their actions. Currently, SeeCCT incorporates a mapping technology by linking each bookmark with a street address; this can be further improved by using Google Maps API for simultaneously displaying several nearby bookmarks on the same map. As mentioned earlier, the scope of the system was artificially limited in order to assess the viability of the concept. However, there are no design limitations that would prohibit expanding the system's scope; in order to accomplish this, the users will be allowed to create new regions and types of bookmarks. Additionally, there are many avenues in which the voice-based interface can be improved; for example, using mixed-initiative scripts would allow for more natural and dynamic interactions between the user and the voice interface of the system.

6. References

[1] S.E. Chang, M.S.H. Heng, "An Empirical Study on Voice-Enabled Web Applications," *IEEE Pervasive Computing*, IEEE, Jul.-Sept. 2006, pp. 76-81.

[2] J. Chugh, V. Jagannathan. "Voice-Enabling Enterprise Applications", *Proc. 11th IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, IEEE CS, pp. 188-189.

[3] R. A. Frost, "Call for a Public-Domain Speechweb," *Communications of the ACM*, vol. 48, no. 11, pp. 45-9, Nov. 2005.

[4] J. Gilbert et al. "VoiceLETS Backs Up First Responders", *Pervasive Computing*, IEEE CS, Jul-Sep 2005, pp. 92-96.

[5] M. Jamali, H. Abolhassani, "Different Aspects of Social Network Analysis," *Proc. IEEE/WIC/ACM Int'l. Conf. Web Intelligence*, Dec. 2006, pp. 66-72.

[6] S. Foll, et. al. "Classifying Multimedia Resources Using Social Relationships," *Proc. 8th IEEE Int'l Symposium on Multimedia*, IEEE CS, pp. 690-5, Dec. 2006.

[7] N. Leavitt, "Two Technologies Vie for Recognition in Speech Market," Computer, IEEE, June 2003, pp. 13-16.

[8] J. Lin, Y. Yeh, "A Research on Telephone-Supported Multimodal Accessible Website," *Proc. IEEE Int'l. Conf. Sensor Networks, Ubiquitous, and Trustworthy Computing*, IEEE CS, vol. 2, pp. 118-23, June, 2006.

[9] M. Lucente, "Conversational Interfaces for E-commerce Applications", *Communications of the ACM*, ACM, Sep 2000, p. 59.

[10] M. Nickerson, "History Calls: Delivering Automated Audio Tours to Visitors' Cell Phones," Proc. 2005 Information Technology: Coding and Computing, IEEE CS, vol. 2, pp. 30-34, Apr. 2005.

[11] D. Rosenblum. "What Anyone Can Know," IEEE Security & Privacy, May-June 2007, pp. 40-49.

[12] W. Rousch, "Social Networking 3.0," *Technology Review*, Nov. 2005.

[13] J.A.Q. Ruiz, J.R.M. Sanchez. "Design of a VoiceXML Gateway," *Proc. 2003 4th Mexican Int'l Conf. Computer Science,* IEEE CS, Sept. 2003, pp. 49-53.

[14] L. Sullivan, "Mobile Social Networking Has Not Yet Come of Age," TechWeb Technology News, May, 2006, http://www.techweb.com/wire/188103062.

[15] J. Thom-Santelli, "Mobile Social Software: Facilitating Serendipity or Encouraging Homogeneity?," *IEEE Pervasive Computing*, IEEE, Jul.-Sept. 2007, pp. 46-51.

[16] N.D. Ziv, B. Mulloth, "An Exploration on Mobile Social Networking: Dodgeball as a Case in Point," *Proc. Int'l. Conf. Mobile Business*, IEEE CS, pp. 26-32, June, 2006.