Mobile Field Data Entry for Concrete Quality Control Information
Kondratova, Irina

National Research Council Canada
Institute for Information Technology

Conseil national de recherches Canada
Institut de technologie de l'information

NRC · CNRC

Mobile Field Data Entry for Concrete Quality Control Information *

Kondratova, I.
September 2004

* published in the Proceedings of the European Conference on Products and Processes Modelling (ECPPM 2004). Istanbul, Turkey. September 8-11, 2004. NRC 47158.

Canada

Mobile field data entry for concrete quality control information

I.L. Kondratova
National Research Council of Canada, Institute for Information Technology e-Business, Canada

ABSTRACT: This paper discusses the advantages and challenges of using multimodal and voice technologies for field data collection on the construction site. Current developments in the multimodal, mobile, field data communications are discussed. Multimodal and VoiceXML technology for speech-enabled information retrieval and input using speech over the phone is explained. This paper describes an innovative solution for field entry of quality control data and real-time communication that utilizes concepts of voice and multimodal interaction for field data entry on a handheld device. In this project, a field concrete testing technician can enter test results using variable interaction modes such as speech, stylus and keyboard via a handheld device, or speech-only on a mobile phone. The multimodal prototype application, for data collection in the field, includes a fat wireless client on a Pocket PC that has a multimodal browser, embedded speech recognition technology, and is based on X+V technology. The speech-only prototype, for field data collection, is based on VoiceXML technology that allows data retrieval and input using natural speech on the mobile phone.

1 INRODUCTION

Multimodal interaction can be described as the integration of visual and voice interfaces through the delivery of combined graphics and speech, on handheld devices (Hjelm, 2000). This technology enables more complete information communication and supports effective decision-making. It also helps to overcome the limitations imposed by the small screen of mobile devices. A small screen size, and the need to use a pen to enter data and commands, presents a great inconvenience for field users - especially if their hands are busy using other equipment, or instruments. Speech processing is one of the key technologies to simplifying and expanding the use of handheld devices by mobile workers (Burkhahardt et al, 2002; IDC Viewpoint 2002).

2 MULTIMODAL INTERACTION TECHNOLOGY

There are different models for implementing multimodal interaction on mobile devices. The fat client model employs embedded speech recognition on the mobile device and allows conducting speech processing locally. The thin client model involves speech processing on a portal server and is suitable for mobile phones.

Currently there are two markup languages proposed for creating applications that use voice input (speech recognition) and output (speech synthesis) and support multimodal interaction. Speech Application Language Tags language (SALT) is a lightweight set of extensions to existing markup languages, in particular to HTML and XHTML (XHTML is essentially HTML 4.0 adjusted to comply with the rules of XML), that enables multimodal and telephony access to information, applications and Web services from PCs, telephones, tablet PCs and handheld devices. SALT applications can be implemented using the thin client model with speech processing done on the speech server (Moraes, 2002).

Another markup language that is currently proposed for developing multimodal Web applications is VoiceXML + XHTML (X+V) (W3C Multimodal Activity, 2004). It combines XHTML and a subset of VoiceXML (Voice Extensible Markup Language). Currently VoiceXML is the major W3C standards effort for voice-based services (W3C Voice Browser Activity, 2004). VoiceXML provides an easy, standardized format for building speech-based applications. Together, XHTML and VoiceXML (X+V) enable Web developers to add voice input and output to traditional, graphically based Web pages. This allows the development of multimodal applications for mobile devices based on the fat client model that includes a multimodal browser and embedded speech recognition on a mobile device, and a Web application server (Figure 1).

While both X+V and SALT use W3C standards for grammar and speech synthesis, only X+V is based entirely on standardized languages. X+V's modular architecture makes it very simple to separate an X+V application into different components. As a result, X+V applications can be developed in parts, with experts in voice programming developing voice elements and experts in visual programming developing visual ones. X+V's modularity also makes it adaptable to stand-alone voice application development. Another feature of X+V is that it leverages open industry APIs like the W3C DOM to create interoperable web content that can be deployed across a variety of end-user devices (VoiceXML Forum, 2004).

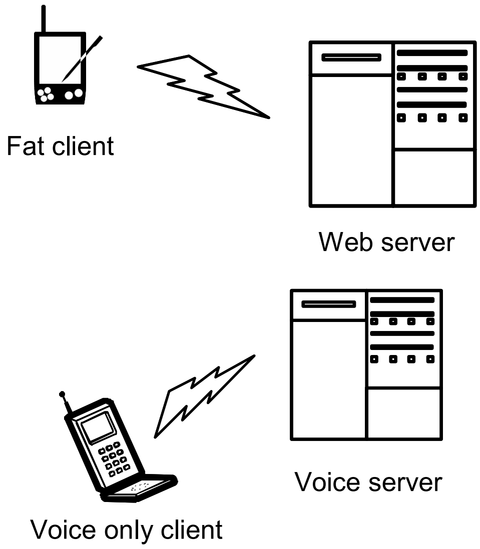

Figure 1. X+V architecture for multiple devices

At the same time, SALT's reliance on the containing environment makes it very difficult to disconnect its coding functions, and makes the language insufficient for the task of stand-alone application development. As a result, the application developer must generate different versions of the application for each execution environment (for example, mobile phones or PDAs from different manufacturers).

X+V technology for multimodal interaction with mobile devices is based on VoiceXML technology for voice access to Web services. VoiceXML technology and examples of VoiceXML field services prototypes are described in the following section of the paper.

3 VOICEXML TECHNOLOGY

VoiceXML technology follows the same model as the HTML and Web browser technologies. Similar to HTML, a VoiceXML application does not contain any platform specific knowledge for processing the content; it also does not have platform specific processing capability. This ability is provided through the Voice XML Gateway that incorporates Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) engines (Kondratova, 2003).

VoiceXML allows providers to deliver Web services using voice user interfaces (VUIs). Developers can use VoiceXML to create audio dialogues that feature synthesized speech, digitized audio, recognition of spoken and touchtone key input (DMTF), recording of spoken input, telephony, and mixed-initiative conversations (Beasly et al., 2002). The words or phrases that a VoiceXML application must recognize are included in a grammar. Large grammars can cause application problems because they can result in recognition errors. Small grammars can cause VUI problems because they require prescriptive prompts that limit the use of natural language dialog. However, small grammars could be used successfully in designing applications for industrial users that are trained in using the application (Kondratova, 2003).

The advantage of using the VoiceXML language to build voice-enabled services is that companies can build automated voice services using the same technology they use to create visual Web sites, significantly reducing the cost of construction of corporate voice portals. A voice portal provides telephone users, including mobile phone users, with a speech interface to access and retrieve Web content. In the next section of the paper, to demonstrate the capabilities and mobile applications of voice technology, the author describes several prototype systems developed for the mobile workforce.

3.1 *VoiceXML for Field Applications*

The full potential of speech-based information retrieval for industrial purposes is not yet harnessed and there are only a handful of existing field applications of VoiceXML technology. For example, Florida USA Power and Light Co. is using a VoiceXML based system for field restoration crews. Using mobile phones, restoration crews can find out about storm-damaged equipment, and report back to the system on the status of the job.

Considering the widespread use of the mobile phone in industrial field applications, there is an opportunity to apply VoiceXML technology for field applications in construction, manufacturing, power and resource industries. These industries can benefit from voice-enabling their operations. The ongoing NRC research program on Voice and Multimodal communications specifically targets industrial field applications of Voice Web technologies.

3.2 *Voice Inventory and Time Management prototypes*

The Voice Inventory Management System (VIMS) prototype, developed at NRC IIT e-Business, allows a mobile worker to easily retrieve product and warehouse information out of the Web-based warehouse database, in real-time, using a regular, mobile phone, or phone-enabled handheld device and natural speech dialog.

The VIMS application keeps track of a series of products and warehouses in a database. All products in the database are entered into the VIMS speech recognition grammar, so that the grammar is updated dynamically with the information on current products in the database. Each product and warehouse has a number of attributes. Each product has a price, product number and description and is associated with the warehouses that product is located in. Each warehouse has an address, and information on the contents of that warehouse. The system also keeps track of product types, represented by a tree that links particular types of products together (Kondratova, 2003).

The Voice Time Management System (VTMS) prototype was developed to allow field crewmembers to enter their time in the time management application, on the corporate server, using a mobile phone. This technology, if implemented, could potentially allow timely billing of the client for the field services provided and bring substantial savings to service providers. The system was designed to fill information as required by the standard timesheet for a construction project.

VTMS is designed to keep track of the user's hours for each job, work number and day and store this information in the corporate database. Using voice commands with VTMS, a field service worker can retrieve and input information from a timesheet that is unique to the caller. A user can retrieve information such as hours worked on a particular day for a unique work number. A user can also find out the total hours worked for the week and the total hours worked on a particular work number for a job and update time information such as hours worked on a particular day (Kondratova, in press).

Both prototype systems have undergone in-house performance and usability testing in an industrial noise environment (about 60-70 dB) and were found to be performing quite well in terms of accuracy of speech recognition and ease of navigation (Kondratova, 2004). However, both prototypes are limited to voice only input and output for data entry and access. To provide users with multimodal interaction capabilities for data entry and retrieval, a multimodal field data collection application on the Pocket PC was developed using X+V technology.

4 MULTIMODAL FIELD DATA COLLECTION

To facilitate speedy field data collection and timely decision making, especially in the case of field quality control inspection it would be highly beneficial to use multimodal wireless handheld devices capable of delivering, voice, text, graphics and even video. For example, "hands free" voice input can be used by a concrete technician in the field to enter inspection information using a hybrid phone-enabled PDA and a wireless, Bluetooth technology enabled headset piece. This information could be entered directly into the inspection forms on the handheld device and stored locally in the embedded database or wirelessly transmitted to the backend database server. Thus, field inspection information could be communicated in real time to facilitate timely decision-making on the construction site and at the ready-mix plant. This information will be stored in the project database and retrieved easily, if needed, in case of litigation. By combining a multimodal mobile handheld device with a GPS receiver and a Pocket GIS system, the gathered inspection information could be automatically linked to its exact geographical location. In addition, other environmental sensors, such as temperature and moisture sensors could also be connected to a handheld device, if needed (Giroux et al, 2002).

4.1 *Wireless field quality control data entry*

Our current project on wireless, field quality control data collection is based on concepts of both multimodal and voice field data collection. In this project, a field concrete testing technician will be able to enter field quality control information into the Concrete Quality Control Database using various interaction modes such as speech, stylus and keyboard on the handheld device or speech on the mobile phone.

The multimodal field data entry application (MFDE) includes a fat wireless client on a Pocket PC that has a multimodal browser and embedded speech recognition, and is based on X+V technology, described previously in this paper. The voice-only data collection application is based on the VoiceXML technology that allows data retrieval and input using natural speech on the mobile phone, similar to the VIMS and VTMS applications described in the previous section. The high-level system architecture for the prototype MFDE application is similar to the one shown in Figure 1. This proof of concept prototype was developed for the wireless Pocket PC utilizing multimodal NetFront 3.1 browser and a fat client with embedded IBM Via Voice Speech recognition engine (IBM Pervasive computing, 2004). An embedded relational database (IBM DB2 everyplace) was used for local data storage on mobile device.

4.2 *System design and functionality*

The two usage scenarios for the field concrete quality control multimodal application are presented in Figure 2.

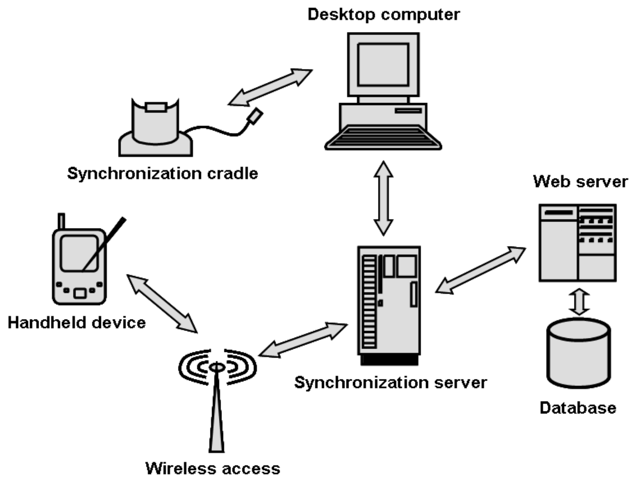

Figure 2. Multimodal field data collection usage scenario

A quality control inspector on a construction site will be using a wireless handheld device to collect field inspection data. Since the application is multimodal, an inspector can fill in the report form by using voice or stylus. On a site with wireless coverage an inspector has the option to update the information in the concrete quality control database directly through the synchronization server. Thus, inspection information is communicated in real time and necessary adjustments to the concrete shipped to the site could be made, if needed.

If there is no wireless coverage on site, an inspector will be using a stand-alone multimodal application on the handheld device. This application utilizes an embedded database to store data and access past records stored on the handheld device. Back at the office information stored in the embedded database will be synchronized with the backend concrete quality control database through the synchronization cradle, desktop computer and the synchronization server. Field information could also be entered into the backend database through a mobile phone utilizing VoiceXML technology as it is done for the Voice Time Management system described earlier.

4.3 *User interface design*

The work on the development of the proof-of concept MFDE application for concrete quality control data collection is conducted in collaboration with the New Brunswick Department of Transportation (NBDOT) concrete quality control engineers and the University of New Brunswick. A graduate student in Civil Engineering, in consultation with NBDOT staff, developed field data entry forms for concrete quality control information. These forms were placed in the Web-based construction project management system. We adapted the design of the field quality control forms for multimodal data entry on the Pocket PC. A screen shot of the Main Menu for the concrete quality control multimodal field data entry application (MFDE) is given in Figure 3. This menu provides options to search, delete and edit old reports, as well as, an option to create and save new reports using voice and stylus or keyboard.

Figure 3. Main Menu for MFDE (interaction with voice, stylus or keyboard)

A Concrete Placing Report form is shown in Figure 4. It contains drop down menus, as well as entry fields for information such as a sample ID or contract number that could be filled in by using a stylus or by entering information by voice.

A Grammar Menu shown in Figure 5 allows minor editing of the MFDE's VoiceXML grammar, including entering names of new contractors or adding material options. The quality control technician could do this on-site to keep the information in the database current. The grammar for the application is dynamically updated by updating information in the database. Thus, major grammar adjustments will be made using a desktop version of the application and later synchronized with the grammar on the handheld device.

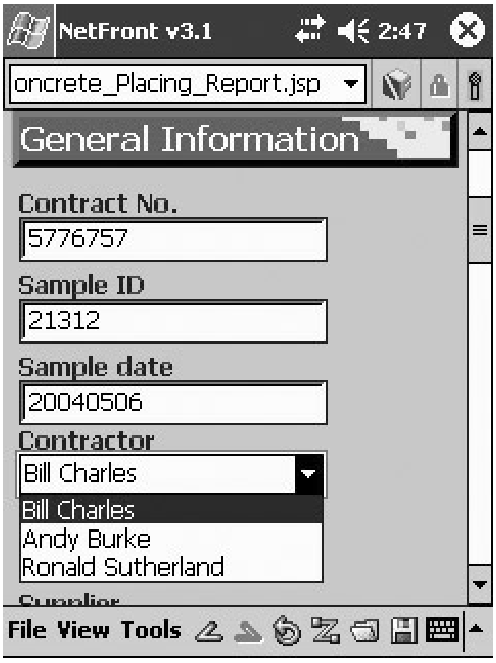

Figure 4. Concrete Placing Report form (interaction with voice and stylus or keyboard)

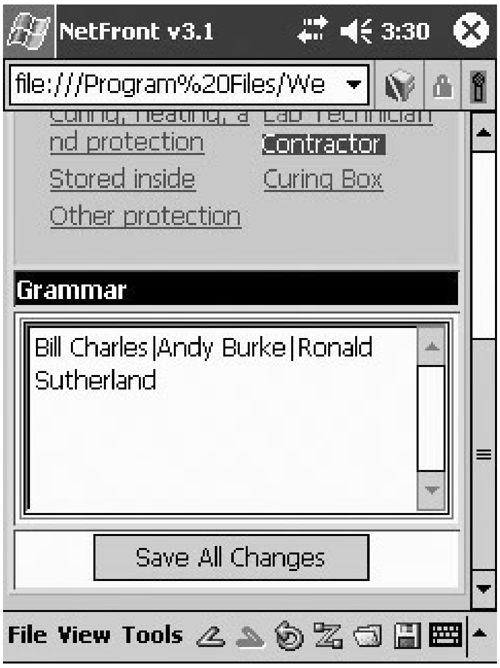

Figure 5. Grammar Menu (interaction with stylus or keyboard)

5 TECHNOLOGY CHALLENGES

In the multimodal prototype development phase we experienced some challenges associated with the novelty of the multimodal technology, including unresolved interoperability issues, mobile OS limitations and also challenges associated with a limited number of multimodal browser vendors. We hope that the mobile industry will resolve these issues in the near future. However, one challenge will remain and will require extensive research and testing: the usability of this technology in the field. The usability evaluation for the multimodal field data entry prototype will be conducted during the next phase of our research project.

6 CONCLUSIONS

The advantages afforded by the field use of multimodal and VoiceXML technology to retrieve corporate and project information and enter field data could be substantial. The availability of real-time, complete information exchange with the project information repository is critical for decision-making in the field of construction site inspection, as information frequently has to be transmitted to and received from the project repository right on-site.

In some cases, when the security and safety of people and infrastructure are at stake, the importance of real-time communication of field data becomes paramount, as, for example, in assessing the damage to buildings in emergency situation (Bacheldor, 2002). In such cases multimodal applications for data collection could be used to collect other types of information such as digital pictures or video of the site.

In this paper the author described applications of VoiceXML and multimodal technology for field data collection on the construction site. The multimodal field data entry prototype allows concrete technician in the field to enter testing data using speech and stylus via the handheld device, as appropriate. On a site with wireless connectivity the testing results are transmitted in real time and entered in to the concrete quality control database, enhancing decision-making on the construction project.

In spite of the current multimodal technology limitations, mentioned previously in this paper, this technology has great potential to overcome user interface weaknesses for mobile devices, in the field, and speed up the data collection and communication process. However, the usability issues for this novel interaction technology require special attention as user acceptance of this technology in the field will be, to a large extent, determined by its ease of use.

7 ACKNOWLEDGEMENTS

The author would like to acknowledge the support provided for the project by the National Research Council Canada and by New Brunswick Innovation Foundation, and the hard work and dedication of the University of New Brunswick Computer Science students that participated in the development of the multimodal prototype.

REFERENCES

Bacheldor, B. 2002. Handheld system assesses damage to see how buildings survived, *Information Week*, March 18, 2002.

Beasly, R., Farley, M., O'Reilly, J. & Squire, L. 2002. *Voice Application Development with VoiceXML* . New York: SAM Publishing.

Burkhahardt, J., Henn, H., Hepper, S. and Rintdorff, K. 2002. *Pervasive Computing*. Boston, NJ: Addison –Wesley.

Hjelm, J. 2000. Research Applications in the M obile Environment, in Wireless Information Services, John Wiley & Sons, 2000.

Giroux, S., Moulin, C., Sanna, R. & Pintus, A. 2002. Mobile Lessons: Lessons based on geo-referenced information, *Proceedings of E-Learn 2002*: 331-338.

IBM Pervasive Computing. 2004. Online. http://www-306.ibm.com/software/pervasive/

IDC Viewpoint. 2002. Five Segments Will Lead Software Out of the Complexity Crisis, by A.C. Picardi, December 2002, Doc #VWP000148.

Kondratova, I. 2003. Voice and Multimodal Access to AEC Project Information, Mobile Computing in Architectural, Engineering and Construction, *10th ISPE International Conference On Concurrent Engineering*, J. Cha et al. (eds), Swets & Zeitlinger, Lisse, Portugal: 755-760.

Kondratova, I. 2004. Speech-enabled mobile field applications. *Proc. of the International IASTED conference on Internet and Multimedia Systems and Applications, IMSA 2004*, Kauai, Hawaii, August 16-18, 2004 (in press).

Meissner A., Mathes I., Baxavanaki L., Dore G. and Branki C. The COSMOS integrated IT solution at railway and motorway construction sites - a case study, *Proc. of the Conference on eWork and eBusiness in AEC* (Turk and Scherer, editors), Swets & Zietilinger, Lisse: 623-626.

Moraes. 2002. VoiceXML, CCXML, SALT. Architectural tools for enabling speech applications, *XML Journal*, Sept. 2002: 30-25.

W3C. Multimodal Activity, X+V, http://www.w3.org/TR/xhtml+voice/

W3C. Voice Browser Activity - Voice enabling the Web! http: www.w3org/Voice

VoiceXMLForum. XHTML+Voice Profile 1.2. http://www.voicexml.org/specs/multimodal/x+v/12/