NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Speech-Enabled Handheld Computing for Fieldwork

Kondratova, Irina

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

Publisher's version / Version de l'éditeur:

Proceedings of the International Conference on Computing in CivilEngineering 2005 [Proceedings], 2005

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=54c44509-bcbe-4673-8079-d1c6a832442
https://publications-cnrc.canada.ca/fra/voir/objet/?id=54c44509-bcbe-4673-8079-d1c6a8324426

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

National Research Council Canada
Institute for Information Technology

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

NRC · CNRC

*Speech-Enable Handheld Computing for Fieldwork **

Kondratova, I.
July 2005

* published in the Proceedings of the International Conference on Computing in Civil Engineering 2005. Cancun, Mexico. July 11-15, 2005 . NRC 48238.

Copyright 2005 by
National Research Council of Canada

Permission is granted to quote short excerpts and to reproduce figures and tables from this report, provided that the source of such material is fully acknowledged.

Canada

SPEECH-ENABLED HANDHELD COMPUTING FOR FIELDWORK

I.L. Kondratova[1]

[1]PhD, P. Eng., Group Leader, National Research Council of Canada Institute for Information Technology e-Business & Adjunct Professor, Department of Civil Engineering, University of New Brunswick, 46 Dineen Drive, Fredericton, Canada, E3B 9W4; PHONE: +1 (506) 444-0489; FAX: +1 (506) 444-6114; email: Irina.Kondratova@nrc-cnrc.gc.ca.

ABSTRACT

This paper discusses the advantages and challenges of using speech recognition on mobile devices, for field data collection and real-time communication. Multimodal and voice technology, for speech-enabled information retrieval and input, using mobile phones or handheld computing devices is explained. Multimodal technology enables more complete information communication and supports timely and effective decision-making. It also helps to overcome the limitations imposed by the small screen of mobile devices.

The paper describes several prototype, industrial, mobile solutions developed for the field entry of data and real-time communication that utilize voice and multimodal interaction. In one of these projects, a field concrete inspector can enter inspection results using variable interaction modalities such as speech, stylus, or keyboard on a handheld device, or speech-only on a mobile phone. The author describes prototype applications developed, and presents several usage scenarios for field and maintenance applications, as well as for emergency response.

KEY WORDS

Field data collection, mobile worker, speech recognition, multimodal interaction.

INTRODUCTION

With current developments in information and communication technology, we are rapidly moving away from the Desktop and Laptop Web paradigms towards the Mobile Web paradigm, where mobile smart devices such as Smart phone, Pocket PC, PDA (personal Digital Assistant), hybrid devices (phone-enabled Pocket PC), and wearable computers will become powerful enough to replace laptop computers in the field. The availability of real time, complete information exchange is critical for effective and timely decision making in the construction field, as information frequently has to be transmitted to and received from the corporate database or project repository, right on site. In some cases, when security and safety of the infrastructure are at stake, the importance of real-time communication of field data becomes paramount (Bacheldor, 2002).

However, the potential of using mobile handheld devices in the field is limited by antiquated and cumbersome interfaces. For example, most of the PDAs and Pocket PCs today have a small screen size of about 3.5”x 5”, and mobile phones have even smaller screens (Pham and Wong, 2004). The small screen size, and the need to use a pen to enter data and commands, presents an inconvenience for field users - especially if their hands are busy using other field equipment, or instruments. In order to assist users in managing mobile devices, user interface designers are starting to combine the traditional keyboard or pen input with “hands free” speech input (Wilson, 2004). As a result, speech processing is becoming one of the key technologies for expanding the use of handheld devices by mobile workers (Picardi, 2002).

However, speech technology is limited to only one form of input and output - human voice. In contrast to this, voice input combined with the traditional keyboard-based or pen-based input model permits multimodal interaction in which the user has more that one means of accessing data in his or her device (Wilson, 2004). This new type of interface is called a *multimodal interface*. Multimodal interfaces allow speeder and more efficient communication with mobile devices, and accommodate different input modalities based on user preferences or on the usage context. For example, in a bridge inspection scenario a field engineer, during the bridge inspection, could request information using “hands free” voice input on a hybrid phone-enabled PDA. Then, the requested information would be delivered as a text, picture, CAD drawing, or video, if needed, directly to the PDA screen (Rankin, 2002). Following the inspection, the engineer, back in his vehicle, will be able to enter information using a portable keyboard or a pen.

Mobile computing is a relatively new technology and is not yet widely accepted by the Architectural, Engineering and Construction Industry. The widespread adaptation of this technology will require, along with some changes in the industry, a significant amount of work on usability and applicability of mobile technology to the different usage scenarios in the industry. Menzel et al. (2004) are currently conducting research work on context-based mobile computing in construction, in which they are “mapping” actors, roles and processes on the construction site into the functional requirements for mobile technology. Similarly to this research work, a significant amount of research on “situation-aware” mobile computing for industrial mobile applications was done by Bürgy (Bürgy and Garrett, 2002). This paper presents research work that focuses mostly on technological and usability aspects of voice and multimodal mobile technologies, for data collection and communication in the field.

VOICE TECHNOLOGY

Modern voice technology offers the convenience of telephone usage and real-time access to the wealth of information, stored in the Web-based construction project information repository or a corporate database. Considering the fact that in spite of the recent progress in the introduction of information technology in to industrial fieldwork, the telephone is still the most widely used information communication tool in the construction industry (Egbu and Boterill, 2000; Flood et al, 2002), voice technology has a great potential as a field technology in construction. Modern voice technology is based on *VoiceXML*.

Background on VoiceXML

VoiceXML stands for Voice Extensible Markup Language. Currently VoiceXML is the major W3C standards effort for voice-based services (W3C, 2004a). 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 capability is provided through the Voice Gateway (sometimes called speech server) that incorporates Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) engines. The Voice Gateway architecture is represented in Figure 1.

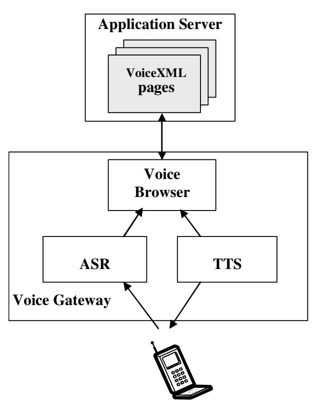

Figure 1: VoiceXML architecture model

VoiceXML technology provides an easy, standardized format for building speech-based applications and allows delivering 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. The words or phrases that VoiceXML application must recognize are included in a grammar. 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 cost of construction of corporate voice portals. A corporate voice portal provides telephone users, including mobile phone users, with a speech interface to access and retrieve Web content.

Voice Technologies: Current Applications

There are quite a few existing speech recognition applications that use VoiceXML technology to provide voice-enabled services to customers. Most frequently VoiceXML is used for building information services that allow wireless phone customers to access news, email, weather, tourist and entertainment information, and business directories (Srinivasan and Brown, 2002). However, at present there are only a handful of existing applications that utilize the potential of voice-based information retrieval for industrial purposes. For example, Florida USA Power and Light Co. used a VoiceXML based system for field restoration crews that were helping to restore power to customers affected by Hurricanes Charley and Frances. In this particular application, using mobile phones, restoration crews could find out about storm-damaged equipment, and report back to the system on the status of the job.

Considering the widespread and preferred use of the mobile phone in industrial field conditions, there is an opportunity to apply VoiceXML technology for other industrial applications, including applications in construction, manufacturing, power and resource industries that can benefit from voice-enabling their operations. The ongoing NRC research program on Voice and Mobile Multimodal communications specifically targets industrial applications of speech recognition technologies. In the framework of this program, several prototype field applications employing VoiceXML technology, including a Voice Inventory Management System and a Voice Time Management System, were developed and tested in a laboratory setting (Kondratova, 2003; Kondratova, 2004b).

Prototype Voice Field Applications

The Voice Inventory Management System (VIMS) allows a mobile worker to easily retrieve product and warehouse information out of the Web-based warehouse database, in real-time, using a regular or mobile phone and natural speech dialog. The user of the system can ask directly about a particular product or warehouse and request information regarding that product or warehouse. For example, a user looking for information on aluminum window frames might ask, "What is the price of an aluminum window frame?" or "What warehouses are the aluminum window frames located in?" In addition to this, an extensive listing system incorporated into the program allows the user to browse an alphabetical listing of all products or warehouses stored in the system (Kondratova, 2003).

A Voice Time Management System (VTMS) was designed for data entry using speech (Kondratova, 2004b). The VTMS prototype allows construction crewmembers to input their time, in to the time management application on the corporate server, using a mobile phone. VTMS keeps track of the user's hours for each job, work number and day, and stores this information in the corporate database. Using voice commands with VTMS, a field worker can retrieve and input information from a timesheet that is unique to the caller. We believe that this system, if implemented, would allow timely billing of the client for the field services provided and could bring substantial savings to the construction, facility management, and field service companies.

Both prototype systems have undergone in-house performance and usability testing in an industrial noise environment (about 60-70 dB) and performed quite well in terms of accuracy of speech recognition (Kondratova, 2004a). The testing for the VIMS application was conducted using a computer product database containing more than 1000 items. Testing results show no decrease in the accuracy of speech recognition for this large database, as opposed to the pilot database with less than 50 items. We found that the speech processing time increases with the increase in the size of the database, due to increases in the size of the grammar. It is possible to shorten grammar-matching time by creating sub-grammars in the VoiceXML application; however this will make the navigation through the voice user interface more complex. Pilot usability testing, for a VIMS system that provides voice access to a database containing entertainment products such as videos and books, was carried out on a group of Web proficient, university student users (Kondratova, 2004b). Participants feel it was easy to get information on the desired products and found it easy to navigate the Main menu and Warehouse information. However, it was harder for the participants to navigate the Product Information section that had a more complicated navigational structure. Thus, reducing speech-processing time by creating a more complex navigational structure is not the best choice in terms of usability of the voice application. This, in a way, limits the applications of voice technology to simple navigational structures that users can easily visualize.

MULTIMODAL INTERACTION TECHNOLOGY

VoiceXML technology is limited to only one form of input and output - human voice. To overcome this limitation, *multimodal technology* could be used. Multimodal technology allows the use of multiple modes of interaction with the computing device such as speech, gesture, pen, mouse and keyboard. Multimodal interaction is sometimes described as the integration of visual and voice interfaces through the delivery of combined graphics and speech, on handheld devices (Hjelm, 2000). 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 speech 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. The 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 speech access to information, applications and

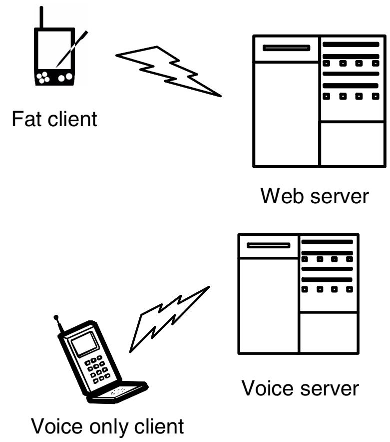

Figure 2: X+V system architecture for multiple devices

Web services from PCs, telephones, Tablet PCs and handheld devices. SALT applications can be implemented using the thin client model with the speech processing done on the speech server (Moraes, 2002), similar to the Voice Gateway represented in Figure 1.

Another markup language that is currently proposed for developing multimodal Web applications is VoiceXML + XHTML (X+V) (W3C, 2004b). It combines XHTML and a subset of VoiceXML. 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 multimodal browser, embedded speech recognition on a mobile device, and a Web application server - as it is presented in Figure 2.

While both X+V and SALT technologies 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.

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 phone-enabled PDA and a wireless headset. 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, to simplify the data collection process (Giroux et al, 2002).

Multimodal Field Data Entry Application

Our current research in the area of wireless, field quality control data collection is based on concepts of both multimodal and voice field data collection. As a result of 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 prototype application (MFDE), developed as a part of the project, 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 will be 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. The high-level system architecture for the prototype MFDE application is similar to the one presented in Figure 2. The proof of concept application was developed for the wireless Pocket PC utilizing multimodal NetFront 3.1 Web browser and a fat client with embedded IBM Via Voice Speech recognition engine. An embedded relational database (IBM DB2 everyplace) was used for local data storage on the mobile device.

MFDE functionality

Two usage scenarios, developed for the field concrete quality control multimodal application implemented on a Pocket PC are presented in Figure 3. In addition to this, 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 in this paper. 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.

1. 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, at the ready-mix plant.
2. 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.

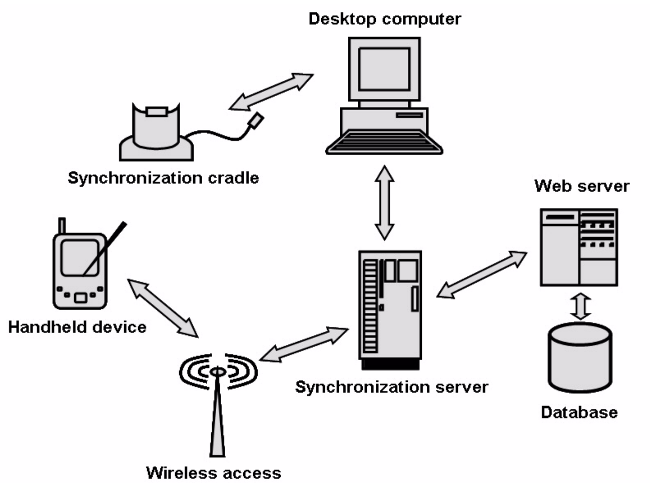

Figure 3: Multimodal field data collection usage scenarios for Pocket PC

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 (Kondratova, 2004c). The field usability study of this application will be the next phase of the project.

VOICE AND MULTIMODAL TECHNOLOGY FOR EMERGENCY RESPONSE

To evaluate the feasibility of using voice technology for government emergency notification services, a prototype Automated Forest Fire Service (AFFS) application was developed in-house using VoiceXML technology. The Automated Forest Fire Service is of interest to people from many different user groups, especially tourists, campers, and people working in the forest.

Automated Forest Fire Notification Services

The first part of the AFFS system is the *Main Menu*, where callers can request particular services. This menu is designed as a mixed-initiative dialog (meaning that the users can ask for a particular service directly or respond to system prompts). Callers could also be led by the system through a directed dialog, if they are having trouble. The menu prompt informs callers about commands available such as “Public Alert”, “Conditions” and “Fire Report”. The voice user interface is designed so that there is a certain amount of flexibility on how callers can enter commands. For instance, the caller can say, “Public Alert” or just “Alert”. While the system does explain what commands are available, some callers may instinctively add words, saying for example, “I would like to hear the public alert please”. The grammar for the Main Menu can handle some of these utterances. Once a caller gives a valid utterance, and has finished his or her work in a particular section of the system, a caller will be automatically returned to the Main Menu. There are several "global" commands that are available anywhere in the AFFS system. If the caller is saying, "Help", the system will provide information on valid commands for the current dialog. Other global commands include "Operator", "Main Menu", and "Exit". To reduce the possibility of callers giving commands that cannot be recognized by the system, several synonyms of "Exit" commands are accepted, such as "Hang up" and "Good bye".

One of the services provided by the AFFS system is the current Fire Weather Index (FWI) information for a particular county in New Brunswick. To eliminate specialized jargon that callers may not understand, the decision was made to call this service Conditions, instead of Fire Weather Index. To choose this service from the Main Menu, callers can give two pieces of information in one utterance, one that indicates that they want to hear conditions, and another specifying a region or a city name. Valid utterances include "I would like to hear the conditions in Moncton", or a shorter command like "Conditions". If a region is not given in the first utterance, callers are sent to a directed dialog that requests this information. Once a valid utterance is spoken, the system will give a message such as "The forest fire weather index is low for Westmoreland County". Another service is the Public Alert, which contains a message indicating FWI values across the province, and other relevant information. The Fire Report service allows callers to report forest fires. If utterances such as "Fire Report" or "I would like to report a forest fire" are given, callers are transferred to a government official who will take the report. Before this occurs, the system will explicitly verify if callers really want to make a report, since they may have arrived at this part of the system by accident.

Quality of Speech Recognition

Two possible issues concerning speech recognition in this public emergency notification system are the audio quality of cellular phones, which most callers will probably use and the accuracy of speech recognition for callers with foreign accents or callers in distress. These issues will have to be considered during future usability testing. To increase the chances of accurate speech recognition, information on how to use the system, including valid utterances, can be published in brochures and travel guides.

In our previous testing on the quality of speech recognition for VIMS prototype (Kondratova, 2004b) we found that a native English speaker using a VIMS prototype system, even in a noisy environment, retrieves required product information 95 times out of 100, and a speaker with a foreign accent has an 80 percent success rate in retrieving correct information.

CONCLUSIONS

The advantages afforded by the field use of voice and multimodal technology to retrieve corporate and project information and enter field data could be substantial. However, proper application of these technologies to field work requires a considerable amount of further research work. In the prototype development phase for the MFDE project we experienced some challenges associated with the novelty of the multimodal technology. These included unresolved interoperability issues, mobile OS limitations as well as challenges associated with a limited number of multimodal browser vendors. We hope that the mobile industry will resolve these issues in the near future so that the commercial development of these types of applications will become possible.

In our project, in parallel with MFDE prototype development on Pocket PC running Pocket Windows system software, we successfully developed a prototype MFDE application on a Tablet PC platform thus escaping many of the pitfalls of the Pocket PC development. With recent miniaturization of the Tablet PC (a good example would be the new Sony U750 palmtop model); we believe it would be feasible to successfully utilize a rugged version of the palmtop Tablet PC as a field multimodal computing device. Another plausible technical solution would be to employ a wearable computer that clips on the worker's belt. However, one challenge will remain and will require extensive research and testing: the usability of multimodal 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.

ACKNOWLEDGEMENTS

The author would like to acknowledge the support provided for the project by the National Research Council Canada, valuable input from my colleagues at NRC and the hard work and dedication of students that participated in the development of prototypes and conducted the pilot usability study.

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. and Squire, L. (2002). *Voice Application Development with VoiceXML*, SAM Publishing.

Bürgy, C., J. H. Garrett, Jr. (2002). "Situation-aware Interface Design: An Interaction Constraints Model for Finding the Right Interaction for Mobile and Wearable Computer Systems," in Proceedings of ISARC2002, Washington, D.C., September 23-25, 2002, 563-568.

Egbu C.O. and Boterill K. Information technologies for knowledge management: their usage and effectiveness, ITcon, Vol.7, 2000, 125-136, (available at http://www.itcon.org/).

Flood, I., Issa, R.R.A. & Caglasin, G. 2002 Assessment of e-business implementation in the construction industry, eWork and eBusiness in AEC, Turk & Scherer (eds), Swets & Zietilinger, Lisse, 27-32.

Hjelm, J. (2000). *Designing Wireless Information Services*, Wiley Computer Publishing, John Wiley & Sons, Inc.

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

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

Kondratova, I. (2004a). “Voice And Multimodal Technology For The Mobile Worker”, *Electronic Journal of Information Technology in Construction*, Vol. 9, Special Issue Mobile Computing in Construction. September 2004. p. 345-353.

Kondratova, I. (2004b). Performance and Usability of VoiceXML Application, Proceedings of the Eighth World Multi-Conference on Systemics, Cybernetics and Informatics. Orlando, Florida, USA. July 18-21, 2004.

Kondratova, I. (2004c). “Mobile Field Data Entry for Concrete Quality Control Information.” Proceedings of the European Conference on Products and Processes Modelling (ECPPM 2004), Dikbas and Scherer (eds.), Taylor and Francis Group, London, p.499-504.

Menzel, K., Eisenblatter, K. and Keller, M. (2004). “Issues of context sensitivity in mobile computing: restrictions and challenges in the construction sector”. Proceedings of the European Conference on Products and Processes Modelling (ECPPM 2004). eWork and eBusiness in Architecture, Engineering and Construction, Dikbas and Scherer (eds.), Taylor and Francis Group, London, 505-513.

Moraes, I. (2002). “VoiceXML, CCXML, SALT. Architectural tools for enabling speech applications.” *XML Journal*, 3 (9) 30-34.

Pham, B. and Wong, O. (2004). “Handheld Devices for Applications Using Dynamic Multimedia Data, Computer graphics and interactive techniques in Australasia and South East Asia.” Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia. ACM Press New York, USA.

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

Rankin, J. (2002). “Information mobility for the construction industry”. In Integrated Technologies for Construction, *Canadian Civil Engineer Spring issue,* 2002.

Wilson, L. (2004). “Look Ma Bell, No Hands! – VoiceXML, X+V, and the Mobile Device.” *XML Journal*, August 3, 2004.

Srinivasan, S. and Brown, E. (2002). “Is speech recognition becoming mainstream?” *Computer Magazine*, April 2002, 38-41.

W3C (2004a), Voice Browser Activity, (available at http: www.w3org/Voice).

W3C (2004b). Multimodal Activity, (available at http://www.w3.org/TR/xhtml+voice/).