Voice and Multimodal Access to AEC Project Information

Kondratova, Irina

Publisher's version / Version de l'éditeur:

10th ISPE International Conference on Concurrent Engineering: the Vision for Future Generations in Research and Applications [Proceedings], 2003

National Research Council Canada
Institute for Information Technology

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

NRC · CNRC

*Voice and Multimodal Access to AEC Project Information **

Kondratova, I.
July 2003

* published in The 10th ISPE International Conference on Concurrent Engineering: The Vision for Future Generations in Research and Applications. J. Cha et al, (eds), Swets & Zeitlinger. Lisse, Portugal. July 26-30, 2003. pp. 755-760. NRC 45829.

Canada

Voice and multimodal access to AEC project information

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

ABSTRACT: 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, 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 with the project information repository is critical for decision making in the construction field, as information frequently has to be transmitted to and received from the project repository right on site. Whereas construction sites are often established for limited periods of time in locations where wired telecommunication infrastructure is often unavailable or limited. Therefore, it becomes important to establish a framework for augmenting the existing integrated project repository environment with mobile wireless devices. Mobile workers on a construction site will be able to use smart mobile devices to communicate with the project information repository in real time thus enabling timely and informed decision making on the project. This paper discusses the advantages of using VoiceXML technology for mobile industrial applications, presents a pilot industrial application of voice technology, and underlines the direction of future research in the area of mobile multimodal communication of AEC project information.

1 BACKGROUND

Multimodal communications 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 is projected that by 2004 more than 1.5 billion handsets, Personal Digital Assistants (PDA's) and Internet appliances will have wireless capability (Gartner Research, 2002).

We are 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, hybrid devices (phone-enabled Pocket PC), and wearable computers will become powerful enough to replace laptop computers in the field and will be widely used for real time communication of project information to project repositories or between project participants.

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 without the engineer making a costly additional field trip (Rankin, 2002; Meissner, 2002).

2 VOICE AND MOBILE TECHNOLOGIES FOR AEC/FM INDUSTRY

2.1 AEC Industry in Canada

The Architectural, Engineering and Construction (AEC) industry in Canada is a small and medium size enterprise dominant and geographically dispersed industry. It represents a large industrial sector of Canada (12% of GDP) but with a very high degree of fragmentation and with no dominant players to enforce Information and Communication Technology (ICT) solutions on projects (Industry Canada, 1998). In addition, the industry is project oriented, this means that ICT must be deployable within one project to all partners. Most projects are "virtual" organizations set up for the duration of the contract with temporary and short-term business relations.

In this environment, ICT is seen in the future as a key enabler and instrument to support leading edge, innovative and powerful solutions (Filos, 2002). Construction sites are established for limited periods of time in locations where wired telecommunication infrastructure is often unavailable or limited, and wireless communications appear attractive to support communications between the Head office and a construction site. Many challenges in today's construction processes arise from poor access to the right information at the right time for timely decision-making and from a general communication breakdown between the project participants. It is estimated by the industry that potential time and cost saving through improved processes, by applying ICTs, can constitute from 10% to 20% range (Rankin, 2002).

2.2 *Mobility of information to support decision-making*

Research and development efforts in computer-integrated construction that are focused on solving the above mentioned issues have established the foundation and standards for an integrated project repository that allows all construction project participants to exchange information in an open industry standard environment (Russel & Froese, 1997). However, the capability of timely and informed decision-making is not a current reality. Access to the project information on a construction site is often limited by the on-site network and Internet infrastructure that is often not available.

There are some experimental research projects on the use of mobile devices for field data collection (Mrawira et al., 2002), but in these projects the communication with the project data repository is conducted asynchronously by downloading field data from mobile devices onto desktop computers and then transferring this information into the integrated project information repository.

To take full advantage of the recent developments in wireless communications and mobile computing it is important to establish a framework for augmenting the existing integrated project repository environment with mobile wireless devices. Mobile workers on a construction site will be able to use these smart devices to communicate with the project information repository in real time thus enabling more timely and informed decision making on the project.

2.3 *Current research on mobile devices for field applications*

Currently there are several research projects investigating the use of mobile devices for synchronous project information communication. The EU IST 5th Framework funded SABARECO project demonstrates satellite-based communication between remote construction sites using standards such as TCP/IP and IFC/XML (Boehling, H., 2002). An ongoing Stanford University's project on Mobile Computing and Active Mediation Technology in a Ubiquitous Computing Environment focuses mostly on the optimization of Internet use by low end handheld mobile computing devices through remote preprocessing of jobs (Liu et al., 2001).

Researchers explore the use of a handheld computer for 3D visualization of design components and assemblies of construction projects (Shiratuddin et al., 2002) and work on the mobile product models to structure the project information so as it could be effectively displayed on mobile devices (Rebolj et al., 2002). Others investigate the possibility of context sensitive data management on mobile devices, adapting the amount of information delivered to the device based on the computing power and the size of the screen (Menzel et al., 2002). A current feasibility study for using wearable computers in construction (Fuller & Sattineni, 2002) showed that the relatively high cost of rugged industrial wearable computers prevents wide spread use on construction sites.

Because of the importance of location in fieldwork applications, in several mobile projects handheld computers were augmented with a GPS receiver to reference the location. For example, a handheld computer with a GPS receiver was used for construction damage assessment after the September 11 terrorist attack (Bacheldor, 2002). It is also possible to connect other sensors, such as temperature and moisture sensors, or accelerometers to a handheld device. Some examples of the location-referenced field applications include field data collection forms, control of environmental sampling during site inspection, and on-site training (Giroux et al., 2002), etc.

However, more research needs to be conducted on technologies that improve the usability of handheld computing devices in the field (Hjelm, 2000). The small screen size and the need to use a pen to enter data and commands present a great inconvenience for field users - especially if their hands are busy using other equipment, or instruments. The investigation of current technologies that could potentially help overcome these limitations is one of the goals of the ongoing research program at the National Research Council of Canada Institute for Information Technology e-Business. The author believes that speech recognition, along with VoiceXML technology and multimodal communications on handheld smart devices will play a major part in overcoming user interface limitations for mobile devices.

2.4 *Voice enabled Web paradigm*

In spite of the recent progress in the introduction of information technology in the AEC/FM industry, the telephone is still the most widely used information communication tool in the industry (Flood et al., 2002). Thus, a technology that can combine the convenience of telephone use and real-time access to the wealth of information stored in the construction project information repository or a corporate database deserves special attention.

The technology under investigation is VoiceXML technology for voice-enabled information access. VoiceXML stands for Voice Extensible Markup Language. Currently VoiceXML is the major W3C standards effort for voice-based services (W3C, 2002).

Voice XML technology follows the same model as the HTML and Web browser technologies. Similar to HTML, 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 Interpreter that incorporates Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) engines). The VoiceXML architecture model is given in Fig.1. It uses the familiar client – server paradigm.

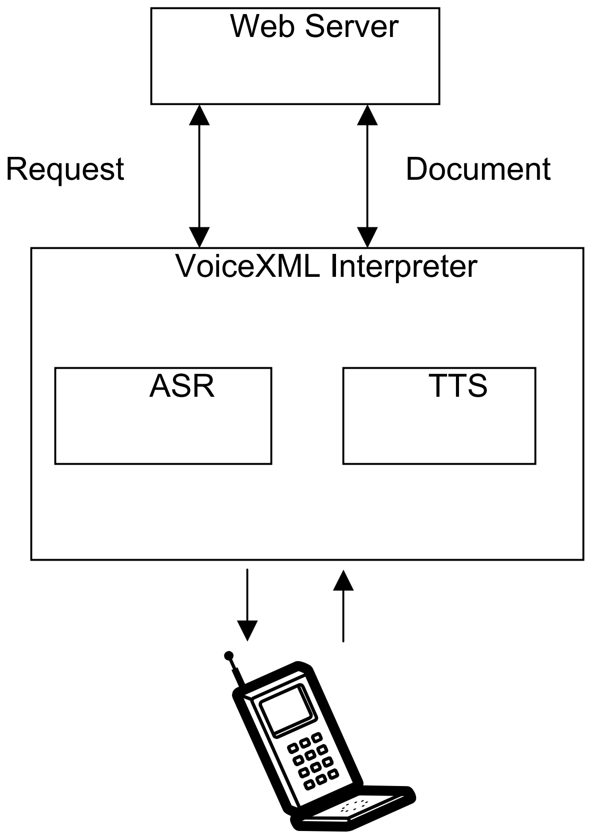

Figure 1. VoiceXML architecture model

VoiceXML allows providers to open 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 (Srinivasan & Brown, 2002). The words or phrases that 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 (Beasley et al., 2001).

The advantage of using 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 voice portal provides telephone users, including mobile phone users, with speech interface to access and retrieve Web content. The potential of voice portals is as big as the reach of the telephone, and, according to Kelsey Group predictions, by 2005 45 million users of wireless phones in North America will regularly use voice portals (Wireless week, 2000).

The software application developed using VoiceXML accepts user input in the form of DTMF (touch tones produced by a phone) and speech and generates output in the form of synthesized speech and pre-recorded audio, thus allowing “hands free” information retrieval and helps to overcome the limitations of the user interface of a field handheld computer (Moraes, 2002).

3 VOICEXML FOR FIELD APPLICATIONS

3.1 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. These services, offered by wireless service providers, allow wireless customers to access news, email, weather, tourist and entertainment information, and business directories (Nuance, 2003). They also include customer service applications where customers can, using natural speech, access their account balances, register their travel reservations, check travel information or find vehicle information and register their vehicles. However, there is a lack of industrial applications of Voice XML technology.

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. is using a VoiceXML based system for field restoration crews (Datria, 2001). Using mobile phones, restoration crews can find out about storm-damaged equipment, and report back to the system on the status of the job.

A crewmember uses voice commands to interact with the application, identifying the work ticket and the status of the job. The data is then automatically entered into the outage management system, while audio commands and online help are provided to assist the crewmember throughout the transaction. After completion of the call, the crewmember can request the status of the next job to be assessed, thus eliminating possible duplication of work assignments and unnecessary travel time.

Considering the widespread use of the mobile phone in industrial field applications, 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.

3.2 Voice Inventory Management System

The Voice Inventory Management System (VIMS) prototype, developed by the NRC IIT e-Business Human Web group, 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 control flow diagram for the VIMS system is presented in Fig. 2.

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.

The VIMS application keeps track of a series of products and warehouses in a 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 the contents of that warehouse. The system also keeps track of product types represented by a tree that links particular types of products together. The user has the option of browsing through a hierarchical menu of product types, which is useful if the name of a product is not known.

The user of the system can also ask directly about a particular product or warehouse and request information regarding that product or warehouse. A natural dialog is used. 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, 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. As searching through an alphabetical list of dozens of items can be time consuming, the user also has the ability to skip to items beginning with a certain letter.

The system also has an authentication menu to only allow authorized access to the system. To enter the system, users must say their name and a personal identification number. This number could be spoken or entered using a touchtone keypad.

The results of the speech recognition accuracy testing for the VIMS prototype are presented in Fig. 3.

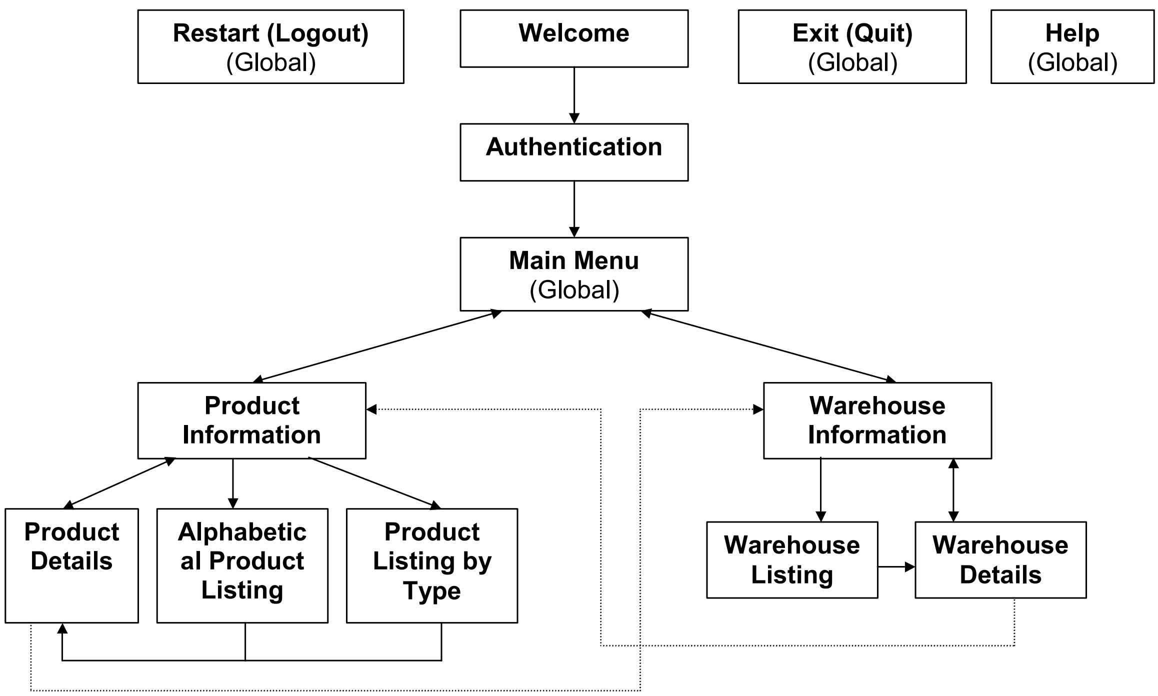

Figure 2. A prototype Voice Inventory Management System control flow diagram

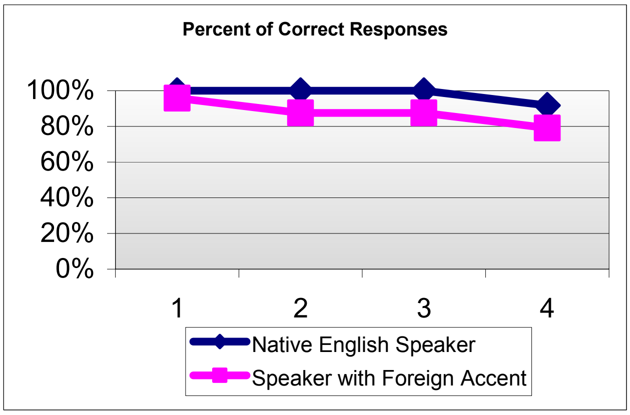

Figure 3. Speech recognition accuracy testing for VIMS:
1 - standard telephone; 2 - speakerphone; 3 - cellular phone;
4 - cellular phone in "noisy" environment

As it can be seen from the percentage of correct responses, a native English speaker using a VIMS prototype system, even in a noisy industrial environment, retrieves required product information 95 times out of 100.

Currently the testing continues using the industrial product database containing more than 1000 items. The initial results show no decrease in the accuracy of speech recognition for a larger database; however, the 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 create sub-grammars in the VoiceXML application that shorten grammar-matching time, but make the navigation through the voice interface more complex. The grammar and the VUI design are very important components of VoiceXML application and will require additional research work and testing.

A pilot usability testing study 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. This testing showed that, with minimal training, the users could easily navigate through the voice interface, draw a sketch of the VIMS navigational structure and retrieve required information on the products and on warehouses where these products are located.

Further, wider scale usability testing program will be carried out on a group of field users to evaluate the feasibility of using VoiceXML technology in the field.

4 MULTIMODAL AND LOCATION BASED SERVICES FOR AEC

The advantages afforded by the field use of the VoiceXML technology to retrieve corporate and project information could be substantial. However, VoiceXML technology is limited to only one form of input and output - human voice (Beasly et al., 2002). To facilitate speedy field data collection and timely decision making, especially in the case of field inspection, where information is largely based on visual observations, it would be 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 an engineer conducting construction site inspection to request information from the project repository using a hybrid phone-PDA and a wireless, Bluetooth technology enabled headset piece. Then the requested information can be delivered as a text, picture, CAD drawing, or video, if needed, directly to the PDA screen.

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. The linkage of information to a specific location could serve multiple purposes: getting precise location of the data collected, helping timely retrieval of information related to the specific location, and providing an option of "pushing" location relevant information to the field crew. Currently commercially available Pocket PC GPS technology allows location accuracy of three to five meters. By utilizing differential GPS technology higher location accuracy (decimeters range) becomes possible.

The migration of speech recognition to smaller devices, such as PDAs, smart phones etc. is enabled by the introduction of efficient speech recognition engines that can better handle noise and variations in speech, as well as, by the development of larger computing power for small devices, and faster, bigger and cheaper processors for speech engines. It was projected by the Kelsey Group in July of 2002 that software licenses from embedded speech applications will grow from US $ 8 million in 2002 to $227 million in 2006. This would make the embedded speech recognition industry one of the fastest-growing segments of the speech market (Kumagai, 2002).

Because of the foreseen benefits of the use of wireless handheld devices to communicate data collected during construction site inspection into the integrated project information repository, the focus of the future research will be on investigation of:

a) the processes and procedures, used during the construction inspection process, that could be augmented using speech recognition technology;
b) the feasibility of VoiceXML technology and multimodal communications on handheld devices to communicate the location-referenced inspection information from the construction site into the project repository;

c) the existing and developing data communication standards, protocols and tools that will support mobile communication of location-referenced project information within the integrated construction project repository environment.

5 CONCLUSIONS

Rapidly increasing computing power of mobile devices allows their utilization as efficient field communication tools for construction project information. However, the limitations of a small screen size and a small (or absent) keyboard present usability challenges for field users. Speech recognition, along with VoiceXML technology and multimodal communications should play a major role in overcoming the user interface limitations for mobile devices. The research program on mobility of information in AEC industry targets development of applications that allow communication of site information, including site inspection results, quality control results, etc. into the integrated construction project repository using different modalities that are appropriate for the mobile device used, the user's abilities and the environmental context. The outcome of this research program will bring solutions that provide better access to the right information at the right time for numerous AEC project participants.

6 ACKNOWLEDGEMENTS

The author would like to acknowledge the support provided for the project by the National Research Council Canada.

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, SAM Publishing, 2002.

Boehling, H. 2002 Satellite-based remote multi-project reporting and controlling in construction industry applying IFCs and XML standardization, *eWork and eBusiness in Architecture, Engineering and Construction*, Turk & Scherer (eds), Swets & Zietilinger, Lisse, p. 635-636.

Datria.2001.http://www.datria.com/company/press/floridapowerandlight.htm

Filos, E. 2002. European collaborative R&D projects related to the "Smart organization". A first evaluation of activities and implications for construction, *eWork and eBusiness in AEC*, Turk & Scherer (eds), Swets & Zietilinger, Lisse, p. 27-32.

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, p. 27-32.

Fuller, S. & Sattineni, A. 2002 Investigating the use of wearable computers in construction, *Proceedings of the First International Conference on Construction in the 21st Century, Challenges and Opportunities in Management and Technology*, Miami, Florida, p. 583-588.

Gartner Research. 2002. Gartner's Glossary of Wireless and Mobile Terms: 2002 update, R-17-0193, Strategic Analysis Report, 12 June 2002, 32 pp.

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

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

Industry Canada. 1998. Sector Competitiveness Framework Series; Construction Overview and Prospects.

Kumagai, J. 2002. Talk to the machine, *IEEE Spectrum Magazine Special R&D Report on Leading Edge Technologies*, September 2002, p. 60-64.

Meissner, A., Mathes, I., Baxavanaki, L., Dore, G. & Branki, C. 2002. The COSMOS integrated IT solution at railway and motorway construction sites - a case study, *eWork and eBusiness in AEC*, Turk & Scherer (eds), Swets & Zietilinger, Lisse, p. 623-626.

Menzel, K., Eisenblatter, K., Keller, M. & Scherer, R.J. 2002. Context-sensitive process and data management on mobile devices, *eWork and eBusiness in AEC*, Turk & Scherer (eds), Swets & Zietilinger, Lisse, p. 627-634.

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

Mrawira, D., Rankin, J. & Nunoo, C. 2002. Web-based tool in project quality management, Proceedings of the First International Conference on Construction in the 21st Century, Miami, Florida, p.959-965.

Liu, D. W., Law, K. H., Liston, K., & Wiederhold, G. 2001. Mobile computing and active mediation technology in a ubiquitous computing environment, CIFE Seed Proposal 2001, http://mediator.stanford.edu/archive.htm

Rankin, J. 2002. Information mobility for the construction industry, in *Integrated Technologies for Construction, Canadian Civil Engineer Spring issue*, 2002.

Nuance. 2003. Case study, Qwest Wireless, http://www.nuance.com/corp/customers/casestudies/qwest.html

Rivard, H. 2000. A survey on the impact of information technology on the Canadian architecture, engineering and construction industry. *Electronic Journal of Information Technology in Construction 5*, http://www.itcon.org/2000/3

Rebolj, D., Cus-Babic, N., Tibault, A., Magdic, A., & Rodosavljevic, M. 2002. Mobile product models, *eWork and eBusiness in AEC*, Turk & Scherer (eds), Swets & Zietilinger, Lisse, p. 637-638.

Russel, A. & Froese, T. 1997. Challenges and a vision for computer-integrated management systems for medium-sized contractors, *Canadian Journal of Civil Engineering*, 24, p. 180-190.

Srinivasan, S. & Brown, E. 2002. Is speech recognition becoming mainstream? *Computer Magazine*, April 2002, p.38-41.

W3C. 2003. Voice Browser Activity - Voice enabling the Web!, http: www.w3org/Voice

Shiratuddin, M.F., Perdomo, J. L., & Thabet, W. 2002. 3D visualization using the pocket PC, *eWork and eBusiness in AEC*, Turk & Scherer (eds), Swets & Zietilinger, Lisse, p. 639-644.

Wireless Week. 2000. http://www.inetnow.com/home/press/2000-09-18_WirelessWeek/