Performance and Usability of VoiceXML Application
Kondratova, Irina

Proceedings of the Eighth World Multi-Conference on Systemics, Cybernetics and Informatics, 2004

National Research Council Canada
Institute for Information Technology

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

NRC · CNRC

*Performance and Usability of VoiceXML Application **

Kondratova, I.
June 2004

* published in the Proceedings of the Eighth World Multi-Conference on Systemics, Cybernetics and Informatics. Orlando, Florida, USA. July 18-21, 2004. NRC 47154.

Canada

Performance and Usability of VoiceXML Application

Irina Kondratova
NRC IIT e-Business
Fredericton, NB, E3B 9W4 Canada

ABSTRACT

This paper discusses the advantages and challenges of using speech-enabled interactive voice response (IVR) technology for mobile industrial applications. These IVR systems can be built using VoiceXML technology for voice-enabled information communication. VoiceXML technology enables voice-based information retrieval and input and uses the familiar client – server paradigm. Two prototype applications of speech-enabled IVR services developed using VoiceXML technology are discussed in this paper, including voice-based inventory management and time management services for field users. Author presents the results of the performance studies for the prototype Voice Inventory Management System, and discusses the pilot usability study for this prototype application conducted on a group of potential users. Future research directions in the area of voice and multimodal mobile communication of field information are also discussed.

Keywords: Mobile Users, Speech recognition, Voice User Interface, Voice Services.

1. 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, 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 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 [1]. However, the potential of using mobile handheld devices in the field is limited by antiquated and cumbersome interfaces. Speech processing is one of the key technologies to simplifying and expanding the use of handheld devices by mobile workers [8].

This paper discusses the advantages of using VoiceXML technology for mobile industrial applications, presents pilot performance and usability studies of a prototype application of voice technology for warehouse inventory management, and underlines the direction of future research in the area of mobile multimodal communication of field information.

2. VOICE ENABLED WEB PARADIGM

In spite of the recent progress in the introduction of information technology in the industrial fieldwork, the telephone is still the most widely used information communication tool in the industry [4, 5]. Thus, a technology that can combine the convenience of telephone use and real-time access to the wealth of information stored in the project information repository or a corporate database deserves special attention.

2.1 VoiceXML Technology

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 [16].

VoiceXML 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 Gateway that incorporates Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) engines). The VoiceXML Gateway architecture model is given in Figure1. It uses the familiar client – server paradigm.

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 [15]. 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 [2]. However, small grammars could be used successfully in designing applications for industrial users that are trained in using the application [9].

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 [18].

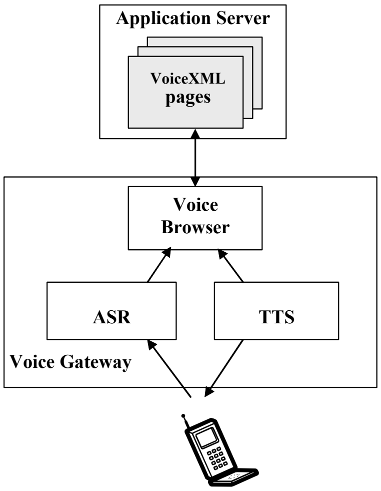

Figure 1. VoiceXML architecture model

2.2 Voice User Interface Design

In spite of the current improvement of speech recognition technology, it is still best suited for producing niche market solutions for people that have their hands busy in some activity or people that cannot use their hands [17].

Speech technology brings its own unique set of human factor issues. These issues include handling unpredictable errors produced by the speech recognition system; managing high human expectations of computers and speech by maintaining realistic outlook on what computer can understand and respond to and teaching users to use constrained speech to get correct results; minimizing the memory load while using the speech-only interface by limiting the number of menu choices that the user has to memorize; as well as providing feedback to the user during processing delays in speech applications [11].

To uncover the problems with the design of speech applications, it is important to conduct usability studies in the environments in which real users will use the application. To design the effective usability study it is customary to conduct a pilot usability study that helps to refine the procedures and the design of user questionnaires. The pilot usability study conducted for the prototype VoiceXML application will be discussed later in this paper.

2.3 VoiceXML for Field Applications

2.3.1 Current applications for VXML

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 [14].

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 VoiceXML 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 [3]. 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.

2.3.2 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 Figure 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.

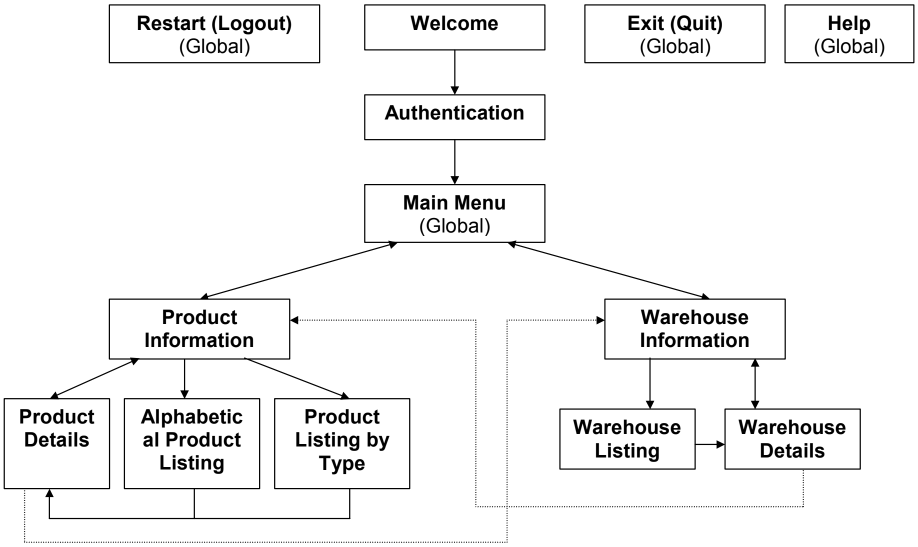

Figure 2. Control flow diagram for the VIMS system

2.4 VIMS Performance Testing

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

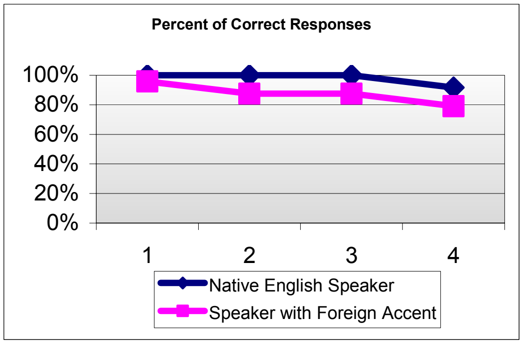

Figure 3. Speech recognition accuracy testing for VIMS:

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.

Subsequent testing was conducted using the computer 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.

2.5 VIMS Usability 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.

2.5.1 Study participants

A study was conducted on a group of university students, with two female and ten male students of 18-21 years old representing a wide range of Faculties (Computer Science, Engineering, Arts etc.). Two out of twelve participants never used an automated telephone system before.

2.5.2 Study design

At the beginning of the test participants were given a short demonstration of the VIMS application and learned about the navigation choices available, without researcher showing the diagram of the navigational structure. After the introduction they were given a simple task of ordering a product and finding out the address of the warehouse where the product is located. They were also asked to sketch the navigation structure of the VIMS application and answer a questionnaire.

2.5.3 Results

This testing showed that, with minimal training, the users could easily navigate through the voice interface, draw a sketch of the VIMS navigational structure (see Figure 4) and retrieve required information on the products and on warehouses where these products are located.

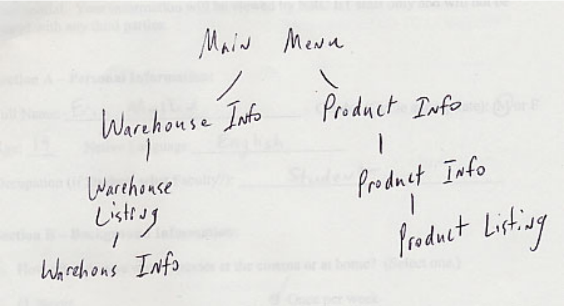

Figure 4. Participant's sketch of navigational structure

Most participants responded that VIMS application was overall equal or better than their favorite automated voice response system. They thought that is was easy to get information on the desired products, but found the Product Information section hard to navigate. However, respondents found it easy to navigate the Main menu and Warehouse information.

All study participants were satisfied with the quality of speech recognition, half of participants found the quality of speech recognition to be “fair” and the rest found it to be “good” or “very good”. Most negative comments were about the time of the VIMS response that sometimes was too long because of problems with the host BeVocal Café server being overloaded.

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. To conduct this usability studies in a carefully controlled environment, a Voice and Multimodal laboratory at NRC IIT e-Business is currently being equipped with VoiceGenie VoiceXML Gateway software and server and with the soundproof test cabins that allow complete control of the testing environment including simulation of different levels of industrial noise.

3. MULTIMODAL FIELD SERVICES

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 [2]. 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 to request information using a hybrid phone-enabled 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 [13].

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. Because of the importance of location in fieldwork applications, handheld computers can be connected to 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 [1]. In addition, other environmental sensors, such as temperature and moisture sensors, or accelerometers could be also connected to a handheld device, if needed. Some examples of the location-referenced field applications include field data collection forms, control of environmental sampling during site inspection, on-site training [6], etc.

However, little research is done on technologies that improve the usability of handheld computing devices in the field [8]. A 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 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 can help to overcome the limitations of the user interface of a field handheld computer [12].

The investigation of current technologies that will help overcome these limitations is one of the goals of the ongoing research program. The author believes that speech recognition, along with VoiceXML technology and multimodal wireless communications using handheld smart devices will play a major part in overcoming user interface limitations for mobile handheld devices.

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 would 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 [10].

Because of the foreseen benefits of the use of wireless handheld devices to communicate field data the focus of the future research should focus on investigation of field processes and procedures that could be augmented using speech recognition technology, including investigating the feasibility of VoiceXML technology and multimodal communications on wireless handheld devices to communicate field location-referenced information.

4. ACKNOWLEDGMENTS

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 on the VIMS application and the study design, and hard work and dedication of the Acadia University and the University of New Brunswick students that participated in the application development and conducted the usability survey.

5. REFERENCES

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

[2] Beasly, R., Farley, M., O'Reilly, J. & Squire, L. 2002. **Voice Application Development with VoiceXML**, SAM Publishing, 2002.

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

[4] Egbu C.O. and Boterill K. Information technologies for knowledge management: their usage and effectiveness, **ITcon**, Vol.7, 2000, 125-136, http://www.itcon.org/ [Last accessed 09/10/03]

[5] 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.

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

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

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

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

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

[11] Lai, J. & Yankelovich, N. Conversational Speech Interfaces, in **The Human Computer Interaction Handbook**, Lawrence Erlbaum Associates Publishers, 2000, 698-713.

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

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

[14] Nuance. 2003. Case study, Qwest Wireless, http://www.nuance.com/corp/customers/casestudies/qwest.html [Last accessed 09/10/03]

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

[16] W3C. 2003. Voice Browser Activity - Voice enabling the Web!, http: www.w3org/Voice [Last accessed 09/10/03]

[17] Weinschenk, S. & Barker, D.T., **Designing Effective Speech Interfaces**, 2000.

[18] **Wireless Week**. 2000. http://www.inetnow.com/home/press/2000-09-18_WirelessWeek/Science Computing Review, 2 (Winter 1992), 453-469.