MULTIMODAL FIELD DATA ENTRY: PERFORMANCE AND USABILITY ISSUES

Irina Kondratova[1], Joanna Lumsden[2], and Nathan Langton[3]

ABSTRACT

Mobile technologies have yet to be widely adopted by the Architectural, Engineering, and Construction (AEC) industry despite being one of the major growth areas in computing in recent years. This lack of uptake in the AEC industry is likely due, in large part, to the combination of small screen size and inappropriate interaction demands of current mobile technologies. This paper discusses the scope for multimodal interaction design – with a specific focus on speech-based interaction – to enhance the suitability of mobile technology use within the AEC industry by broadening the field data input capabilities of such technologies.

To investigate the appropriateness of using multimodal technology for field data collection in the AEC industry, we have developed a prototype Multimodal Field Data Entry (MFDE) application. This application, which allows concrete testing technicians to record quality control data in the field, has been designed to support two different modalities of data input – speech-based data entry and stylus-based data entry. To compare the effectiveness or usability of, and user preference for, the different input options, we have designed a comprehensive lab-based evaluation of the application. To appropriately reflect the anticipated context of use within the study design, careful consideration had to be given to the key elements of a construction site that would potentially influence a test technician's ability to use the input techniques. These considerations and the resultant evaluation design are discussed in detail in this paper.

KEY WORDS

mobile field data collection, multimodal interaction, speech recognition, mobile usability.

[1] 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.

[2] Research Officer, National Research Council of Canada Institute for Information Technology e-Business & Adjunct Professor, Faculty of Computer Science, University of New Brunswick, 46 Dineen Drive, Fredericton, Canada, E3B 9W4; PHONE: +1 (506) 444-0382; FAX: +1 (506) 444-6114; email: Jo.Lumsden@nrc-cnrc.gc.ca.

[3] Visiting Worker, National Research Council of Canada Institute for Information Technology e-Business & honors thesis student, Faculty of Computer Science, University of New Brunswick, 46 Dineen Drive, Fredericton, Canada, E3B 9W4; FAX: +1 (506) 444-6114; email: Nathan.Langton@nrc-cnrc.gc.ca.

INTRODUCTION

Although in recent years mobile technology has been one of the major growth areas in computing, it is not yet widely accepted by the Architectural, Engineering, and Construction (AEC) industry (Bowden et al., 2005). The realization of widespread adoption of mobile technology within the AEC industry will require, along with some changes in the industry itself, further research to address the industry-specific usability of the technology, as well as continued investigation into the applicability of mobile technologies within different AEC-specific usage scenarios. Focusing on context-based mobile computing in construction, Menzel et al. (2004) have addressed the issue of different usage scenarios by 'mapping' actors, roles, and processes on the construction site into the functional requirements for mobile technology. Similarly, Bürgy and Garrett (2002) have investigated 'situation-aware' mobile computing for industrial mobile applications; their focus has mostly been on the usability aspects of mobile technology used for data collection and communication in the field.

The potential for using mobile handheld devices in the field is currently limited by cumbersome interfaces and inappropriate interaction demands (Saidi et al., 2002). For example, most handheld devices have a small screen size of about 3.5"x5", and mobile phones have even smaller screens (Pham and Wong, 2004); furthermore, interaction with such devices is typically limited to the use of a stylus and soft keyboard (stemming from the tried-and-tested desktop design paradigm (Lumsden, 2005)). The combination of small screen size and stylus-based interaction presents an inconvenience for field users – especially if their hands are otherwise engaged using field equipment or instruments. To assist users in managing mobile devices, mobile user interface designers are starting to combine the traditional stylus-based and soft keyboard input with 'hands-free' and 'eyes-free' speech input (Wilson, 2004). Consequently, speech processing is becoming one of the key technologies for expanding the use of handheld devices by mobile workers (Picardi, 2002).

Speech technology is, however, limited to only one form of input and output – that is, human voice. In contrast to this, combining voice input with traditional stylus-based and soft keyboard input techniques permits multimodal interaction in which the user has more than one means of accessing and entering data on a mobile device (Wilson, 2004). This type of interface – known as a multimodal interface – not only allows for faster and more efficient communication with mobile devices, but it also supports the use of different input modalities based on user preferences and/or on the usage context. This is especially important in field work, where environmental conditions (e.g., lighting, noise, and physical distractions) vary throughout the day based on field conditions and/or task context, and therefore require that a user be able to choose the most appropriate interaction modalities at any given point in time. To investigate the effectiveness of multimodal technology for field data collection in the AEC industry, we developed a prototype mobile Multimodal Field Data Entry (MFDE) application. The design functionality and usage scenarios for this field data entry application are described in the following section of this paper.

MULTIMODAL FIELD DATA COLLECTION

To facilitate efficient 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, audio, graphics, and even video. For example, 'hands free' voice input could be used on-site by a concrete technician to enter inspection information using a phone-enabled Personal Digital Assistant (PDA) and a wireless headset. This information could be entered directly into inspection forms on a handheld device and stored locally in an embedded database or wirelessly transmitted to a backend database server. Thus, field-based inspection information could be communicated in real-time to facilitate timely decision making on a construction site and at a ready-mix plant. This information would 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 recorded 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 to simplify the data collection process (Giroux et al, 2002).

Our current research in the area of wireless field quality control data collection focuses on multimodal (including voice) field data collection for concrete test results. We are investigating the use of technologies that will allow a field-based concrete testing technician to enter quality control information into a concrete quality control database using various interaction modes such as speech, stylus, and keyboard on a handheld device, or speech on a mobile phone.

As part of this research, we have developed a prototype mobile multimodal field data entry (MFDE) application to run on a Pocket PC that is equipped with a multimodal browser and embedded speech recognition capabilities. This proof-of-concept application was developed for the wireless Pocket PC utilizing the multimodal NetFront 3.1 Web browser and a fat wireless client with an embedded IBM ViaVoice Speech recognition engine. An embedded relational database (IBM DB2 everyplace) was used for local data storage on the mobile device. Detailed information on the multimodal technology that was used to develop the prototype is presented elsewhere (Kondratova, 2005).

FIELD USAGE SCENARIO

Figure 1 shows the infrastructure necessary to support the following two usage scenarios which were elaborated to guide the development of the MFDE application. Imagine that, on a construction site, a quality control inspector is using a wireless handheld device to record concrete inspection data. Given the multimodality of the MFDE application, the inspector can fill in the report form using speech or stylus-based input. The information could then be communicated as follows:

- On a site with wireless network coverage, the inspector has the option to update the information in the concrete quality control database directly, and immediately, through the synchronization server. In this case, inspection information would be communicated in real-time and, if necessary, adjustments could be made to the concrete at the ready-mix plant before being shipped to the site.

- If there is no wireless network coverage available, an inspector would use the MFDE application as a stand-alone application on the handheld device. This application utilizes an embedded database to store data and access past records stored on the handheld device. Once back at the office, the inspector would synchronize information stored in the embedded database with the backend concrete quality control database via the synchronization cradle, desktop computer, and synchronization server.

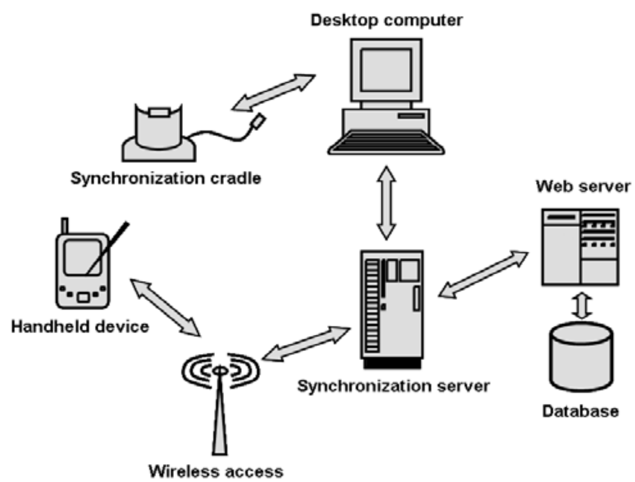

Figure 1: Multimodal Field Data Entry Infrastructure.

TESTING THE SPEECH RECOGNITION

As mentioned previously, the MFDE application adopts multimodal interaction techniques to facilitate efficient data input. In particular, it permits the use of voice input in addition to the use of the standard stylus and soft keyboard. As the standard (or default) mobile interaction technique, the latter is well tested. Conversely, as a new technology, speech recognition on mobile devices is less well tested. Before investigating the use of the MFDE application as a whole under context-relevant conditions, we evaluated the efficacy of the speech recognition in its own right.

We had previously tested the performance and accuracy of speech recognition for a *desktop* speech-based warehouse information retrieval system (Kondratova, 2004) and found that, using our prototype system, a native English speaker – even in a noisy environment – was able to retrieve desired product information 95% of the time; a user with a foreign accent was able to achieve a success rate of 80%. We subsequently tested the desktop prototype again after increasing the computer product database population to more than 1000 items. The results of this evaluation showed there to be no decrease in the accuracy of speech recognition with increase in database size; the processing time, however, did increase due to increase in the size of the grammar.

We also tested the accuracy of speech recognition for the *desktop* MFDE prototype at different construction noise levels (Burke, 2005). After testing a total of 2880 utterances, we found that noise levels of 90 dB did not significantly affect the accuracy of speech recognition for the MFDE software. We also found that the accuracy of speech recognition was equal to, or higher than, 97% when we used a restricted numerical vocabulary.

Having established that the speech recognition component of the MFDE application was effective under isolated static test conditions, we turned our attention to designing a comprehensive usability evaluation of the MFDE prototype. The study design is described in detail in the next section of this paper.

USABILITY EVALUATION: EXPERIMENTAL DESIGN

Evaluations of mobile technologies are typically conducted under stationary conditions – that is, with users sitting at a desk to use and evaluate the technology (Kjeldskov and Graham, 2003). Whilst such studies are effective at identifying cosmetic usability issues (Kjeldskov and Stage, 2004) they do not highlight the usability concerns which arise as a result of the environmental and physical context in which the technology will ultimately be used. Although one might, therefore, argue that effective evaluations are only achievable in the field, a recent investigation has shown that there is little or no benefit to undertaking evaluations in the field as opposed to in the lab (Kjeldskov et al., 2004). Not only have lab-based mobile evaluations of mobile technologies been found to identify more usability problems (including context-specific problems) than field-based studies, but also the lab environment allows for greater experimental control and easier data capture than in the field.

While lab-based studies are therefore a viable means by which to assess the usability and suitability of mobile technologies, it is essential that the lab set-up adequately reflect the intended context of use for the technology in order that the results returned are meaningful. The remainder of this section discusses the design of a lab-based mobile evaluation of the MFDE prototype, wherein an abstract representation of the actual context of use is utilized in order to maximize the relevance of the results that will be returned by the experiment. The design described below is made possible as a result of a custom-built mobile HCI lab (8.65m x 17.3m in size) in our research facility.

MODELLING THE CONTEXT OF USE

As previously mentioned, the MFDE application is designed to allow concrete testing technicians to record, while in the field (or more specifically, on a construction site), quality control data. The application has been designed to support two different modalities of data input – speech-based data entry and stylus-based data entry. The purpose of the evaluation described here is to (a) determine and compare the effectiveness and usability of the two different input options and (b) to determine which of the two options is preferred by users in relation to the application's intended context of use. In order, therefore, to appropriately reflect the anticipated context of use within our study design, we had to consider the key elements of a construction site that would potentially influence a test technician's ability to use one or both of the input techniques. We determined these to be: (a) the typical extent of mobility of a technician while using the application; (b) the auditory environmental distractions surrounding a technician – that is, the noise levels inherent on a typical construction site; and (c) the visual or physical environmental distractions surrounding a technician – that is, the need for a technician to be cognizant of his or her physical safety when on-site.

Consider, first, the issue of mobility. A concrete test technician typically attends to concrete pour locations on a construction site and, at each location, performs tests on the concrete and then records the results of these tests. Of relevance to our evaluation is the fact that a technician moves between specific locations and, at each, pauses to record data while standing. Our MFDE application is not concerned with the actual concrete testing process, so we decided to reflect the high level work practice simply by requiring our experimental participants to walk to various points in our lab space and, at each point, enter provided data into the MFDE application.

A construction site is inherently noisy; indeed, environmental noise on a typical construction site ranges from 70dB to 100dB (Gilchrist et al., 2003). Speech recognition technologies are very susceptible to interference from environmental noise and as such it is essential that distractions from realistic construction noise be integrated into our experimental design. To achieve this, we will use a 7.1 surround sound system in our lab to play construction noises within the aforementioned range while the experimental participants are performing data entry tasks. Ethically, we cannot expose our participants to such noise levels without provision of hearing protection. As such, participants will be required to wear industrial hearing protectors (specifically, the Intruder EM7202 Head Band Earmuff) during this experimental study. Not only does this mitigate the risks to our participants, but it is also representative of the situation on a construction site where a test technician would be required to wear hearing protection.

Finally, consider the need for a test technician to be cognizant, for safety reasons, of his or her surroundings on a construction site. It is important to ensure that, when using the MFDE application, a test technician's visual resource is not so completely engaged with the application that he cannot attend to the dangers around him. Such dangers might typically include heavy equipment moving around the construction site. When designing our experiment, we therefore had to reflect the requirement that a technician needs to be aware of his surroundings while using the application. To do this, we will use a ceiling mounted projection system in the lab to project photographic images around the walls of the lab space. These images will include a series of 'safe' construction site photographs (that is, with no heavy equipment) and one 'danger' photograph – all of which will be displayed in random sequence, location, and duration around the lab; Figure 2 shows examples of each.

Figure 2: Visual Distractions Used – (a) An Example of a 'Safe' Visual Distraction and (b) the 'Danger' Distraction.

While using the application to enter data, participants will be required to be conscious of the images being projected and maintain a mental tally of the number of 'danger' photographs of which they were aware. Post-experimental analysis of the data will allow us to compare the actual number of 'danger' images projected with the number reported by the participants in order to derive a measure of awareness per participant. Unlike the noise levels, which we consider to be *interfering distractions* because they have the potential to interfere directly with a user's ability to interact with the mobile device, we refer to these visual distractions as *active distractions* (the 'danger' photographs) because they require a participant to react to the distraction in some way – in this case, by tallying the number of instances of a given image – and *passive distractions* (the 'safe' photographs) because they distract the participants but do not require an active response.

Hence, using a combination of simultaneous passive, active, and interfering distractions affecting both the auditory and visual senses, together with a mobile task set-up, it is possible to abstract key or relevant elements of a construction site and meaningfully represent them within our lab environment.

EXPERIMENTAL PROCEDURE

The MFDE application includes data entry fields for numbers/text, decimal numbers, and dates, as well as a series of drop down lists. For the purpose of our experiment, we will use an abridged version of the full MFDE application – that is, we will focus on a subset (seven) of the actual fields in the MFDE forms, to include one or more instances of each of the different field types. This will allow us to adequately compare the two input techniques (speech-based input v. stylus-based input) for each data type. The abridged version of the MFDE application will run on an HP 4700 series iPAQ.

Participants will be asked to walk between tables in our lab space (see Figure 3) to enter data into each of the fields in an abridged MFDE form. Each table in the lab will be labeled according to one of the seven fields in the form. The order in which the participants visit the seven tables will reflect the sequential order of the fields on the form. Each table will display instructions to participants regarding the precise data they are to enter into the corresponding data field in their form.

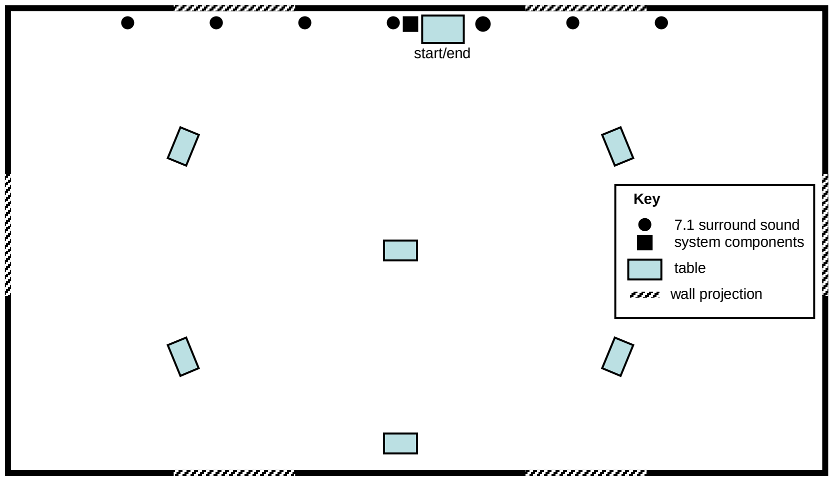

Figure 3: Physical Lab Set-Up.

The length of time it takes each participant to complete the form will be recorded. Participants will be instructed not to correct erroneous or omitted fields since we want to determine which particular types of data entry are most problematic. As participants are completing the sequence of data entry tasks, they will be surrounded by construction level noise as well as projections of ‘safe’ and ‘danger’ images; as previously indicated, they will be required to be cognizant of the ‘danger’ images and report the number they believe they saw to the evaluator once they have completed the form.

As mentioned previously, our aim is to determine and compare the effectiveness and usability of the two data entry techniques. We therefore plan to use a fully counterbalanced, between-groups design for this experiment. Basically, each participant will be required to complete an MFDE form twice, once with each input technique. By counterbalancing the order in which they use the two input techniques, we can mitigate against learning effects. Although the form used in each session will contain the same fields, the order of these fields – and hence the path participants will walk between tables in the lab – and the precise data elements to be entered, will be different between the two sessions to further mitigate against learning effects. Additionally, we would like to determine noise thresholds at which the speech-based input does and does not work. To this end, we will divide the participants into 3 groups, with each group being exposed to a different noise range (see Table 1). In total, we will use 18 participants divided into 3 groups of 6 participants. Before each session, participants will be trained in the use of the input technique they are about to use, and will be given an opportunity to practice with it under conditions similar to the actual session. After each session, participants will be asked to complete a NASA Task-Load Index (TLX) questionnaire which is designed to determine their perception of workload as they performed the data entry tasks (Hart and Wickens, 1990). After both sessions, participants will be asked to complete a short questionnaire to determine their preference for one or other of the input techniques. Table 1 outlines the experimental organization.

Table 1: Experimental Organization, Showing Between Groups Design and Counterbalancing of Conditions.

<table>
<tr><th></th><th># Participants</th><th rowspan="5">Background Data Collection</th><th>Input Style 1</th><th rowspan="4">TLX Workload Test</th><th>Input Style 2</th><th rowspan="4">TLX Workload Test</th><th rowspan="5">Preference Ratings</th></tr>
<tr><td>Group A
70 dB – 80 dB</td><td>6</td><td>Train & Test</td><td>Train & Test</td></tr>
<tr><td>Group B
80 dB – 90 dB</td><td>6</td><td>Train & Test</td><td>Train & Test</td></tr>
<tr><td>Group C
90 dB – 100 dB</td><td>6</td><td>Train & Test</td><td>Train & Test</td></tr>
<tr><td></td><td></td><td colspan="4">Input styles delivered in counterbalanced order per group</td></tr>
</table>

CONCLUSIONS AND FURTHER WORK

The potential advantages afforded by the availability of multimodal interaction for construction crews to enter data and retrieve additional project information using mobile technology in the field could be substantial. In particular, by supporting ‘hands free’ and/or ‘eyes free’ interaction, the speech-enabled mobile MFDE application that we developed has the *potential* to enhance user interaction with a handheld computer in the field, thereby facilitating greater user satisfaction and industry wide adoption of mobile technology. To ensure this potential is met, we have designed a comprehensive context-relevant evaluation study which will investigate the suitability and usability of different input modalities under conditions representative of a construction site. After conducting the actual evaluations, we hope to be able to empirically demonstrate that mobile technologies capable of supporting multimodal interaction are suited to use within the AEC industry.

REFERENCES

Bowden, S., Dorr, A., Thorpe, A., Anumba, C. J., and Gooding, P., (2005). “Making the Case for Mobile IT in Construction”, in Proceedings of the 2005 ASCE International Conference on Computing in Civil Engineering, Cancun, Mexico. July 11-15, 2005.

Bürgy, C., and Garrett, J.H., 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.

Burke, M. (2005). “Multimodal Prototype Software Application for Field Data Collection on the Construction Site: Development of a Testing Procedure”, Senior Thesis Report, Department of Civil Engineering, University of New Brunswick, Fredericton, Canada.

Gilchrist, A., Allouche, E.N., and Cowan, C., (2003). “Predication and Mitigation of Construction Noise in the Urban Environment”, in the Canadian Journal of Civil Engineering, 30, 659-672.

Giroux, S., Moulin, C., Sanna, R., and Pintus, A., (2002). "Mobile Lessons: Lessons based on geo-referenced information", in Proceedings of E-Learn 2002, 331-338.
Hart, S.G., and Wickens, C., (1990). "Workload Assessment and Predication" in MANPRINT: An Approach to Systems Integration, Booher, H.R., (Ed), Van Nostrand Reinhold: New York, 257-296.
Kjeldskov, J., and Graham, C., (2003). "A Review of Mobile HCI Research Methods" in Proceedings of the 5th International Symposium on Mobile Human-Computer Interaction (MobileHCI'03), Udine, Italy, Sept. 8 – 11, 317-335.
Kjeldskov, J., Skov, M.B., Als, B.S., and Høegh, R.T., (2004). "Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field", in Proceedings of the 6th International Symposium on Mobile Human-Computer Interaction (MobileHCI'04), Brewster, S., and Dunlop, M., (Eds), Glasgow, Scotland, Sept. 13-16, 61-73.
Kjeldskov, J., and Stage, J., (2004). "New Techniques for Usability Evaluation of Mobile Systems" in International Journal of Human Computer Studies (IJHCS), 60(5-6), 599-520.
Kondratova, I. (2005). "Speech-Enabled Handheld Computing for Fieldwork", in Proceedings of the International Conference on Computing in Civil Engineering 2005. Cancun, Mexico. July 11-15, 2005.
Kondratova, I. (2004). "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.
Lumsden, J., (2005). "The Future of m-Interaction", in the Encyclopedia of Multimedia Technology and Networking, M. Pagani (Ed.), Idea Group Inc., Hershey, USA., 342-347.
Menzel, K., Eisenblatter, K. and Keller, M. (2004). "Issues of context sensitivity in mobile computing: restrictions and challenges in the construction sector," in 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.
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," in 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.
Saidi K. S., Haas C. T., Balli N.A. (2002). "The Value of Handheld Computers in Construction", NIST SP 989, in Proceedings International Symposium on Automation and Robotics in Construction, 19th (ISARC). National Institute of Standards and Technology, Gaithersburg, Maryland. September 23-25, 2002, 557-562.
Wilson, L. (2004). "Look Ma Bell, No Hands! – VoiceXML, X+V, and the Mobile Device," XML Journal, August 3, 2004.