Multimodal Interaction for Mobile Learning

Irina Kondratova

National Research Council Canada Institute for Information Technology
46 Dineen Drive, Fredericton, NB, Canada E3B 9W4
{Irina.Kondratova}@nrc-cnrc.gc.ca

Abstract. This paper discusses issues associated with improving usability of user interactions with mobile devices in mobile learning applications. The focus is on using speech recognition and multimodal interaction in order to improve usability of data entry and information management for mobile learners. 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, adding other modes of interaction such as speech-based interfaces that are capable of interpreting voice commands. Several research studies on multimodal mobile technology design and evaluations were carried out within our state-of the art laboratories. Results demonstrate feasibility of incorporating speech and multimodal interaction in designing applications for mobile devices. However, there are some important contextual constrains that limit applications with speech-only interfaces in mobile learning, including social and environmental factors, as well as technology limitations. These factors are discussed in detail.

Keywords: Mobile usability, multimodal interaction, speech recognition, mobile evaluation.

1 Introduction

Many researchers see great value in mobile learning because of portability, low cost and communication capabilities of mobile devices [21]. Mobile devices are becoming an increasingly popular choice in university and school classrooms, and are increasingly being adopted by the "lifelong learners". Several features of mobile technologies make it attractive in learning environments, among them: relatively low cost of mobile devices [25] and good fit within informatics and social layers of classroom communications [20].

Evaluations of mobile technologies within the classroom environment are largely positive [1, 24]. However, widespread use of mobile technology in learning applications is impeded by numerous usability issues with mobile devices.

The gravity of mobile usability problems is highlighted by recent surveys of mobile Internet users [22]. They show that usability is by far the biggest source of frustration among the users of mobile technologies. In particular, for learning applications, research shows that the most important constraining factors for widespread mobile learning adoption, along with battery life, are the screen size and user interface of most portable devices [17].

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2009, LNCS 5615, pp. 327–334, 2009.

This paper explores possible improvements in the usability of mobile devices that is facilitated by utilization of natural user interfaces to enhance interaction with mobile devices. In section two of the paper the author provides background information on speech-based interaction with mobile devices and on technologies involved. This section of the paper also addresses the concept of multimodality and multimodal applications for interaction with mobile devices. The follow-up section discusses several laboratory studies conducted to evaluate efficacy and feasibility of multimodal interactions with mobile devices and their potential applications to mobile learning. The author concludes with observations on the potential for incorporating speech and multimodal technologies in mobile learning domain and some limitations of these technologies.

2 Alternative Interaction Modalities

2.1 Speech as an Interaction Modality

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 [28], adding other modes of interaction such as speech-based interfaces that are capable of interpreting voice commands [23]. As a result, speech processing is becoming one of the key technologies for expanding the use of handheld devices by mobile users [18]. In the eLearning technology foresight, technology-based education guru Tony Bates predicted that: "A new computer interface based on speech recognition will have a major impact on the design of e-learning courses" [15]. Currently, automated speech recognition (ASR) technology is being used in desktop e-learning applications for automated content-based video indexing for interactive e-learning [29], audio–clip retrieval based on student questions [30], and, together with speech synthesis, to improve accessibility of e-learning materials for visually impaired learners [3, 4]. Another novel application of mobile technology for experiential learning is being developed for functionally illiterate adults [14]. This application employs speech recognition and text-to-speech to assist adult literacy learners in improving pronunciation of words they learn.

2.2 Multimodal Interaction

Speech technology seems to be ideally suited for enhancing usability of mobile learning applications designed for the mobile phone. In this domain speech is a natural way of interaction, especially where a small screen size of a mobile device limits the potential for a meaningful visual display of information [2]. 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 permits multimodal interaction where the user has more than one means of accessing data in his or her device [16]. This type of user interface is called a multimodal interface [5].

Multimodal interfaces allow speedier and more efficient communication with mobile devices, and accommodate different input modalities based on user preferences and the usage context. A field trip learning environment, offers the most comprehensive scenario for using of speech and multimodal interaction with mobile device. For example, in a field trip scenario for a group of engineering students, a student can request information about the field structure (bridge, building, road, etc.) from the course repository using “hands free” voice input on a “smart phone” (hybrid phone-enabled PDA). The requested information would then be delivered as a text, picture, CAD drawing, or video, if needed, directly to the PDA screen. The student will be able to enter field notes in the forms using a portable keyboard or a pen, if appropriate or via voice input during field data gathering. In addition to this, free-form verbal field notes could be attached to the data collected as an audio file and later analyzed in class [6].

3 Evaluations of Mobile Speech and Multimodal Technologies

This section compares several applications of mobile multimodal technologies (speech-based and keyboard/stylus). In particular, the focus is on user evaluations of these technologies conducted to study the feasibility and efficacy of speech-based and multimodal interactions in different contexts. This comparison will form the basis for author’s estimate for potential of using speech as an interaction technique in various learning contexts.

3.1 Speech vs Stylus Interaction

Comparison of efficacy of speech-based and stylus-based interaction with a mobile device was conducted as a part of our research in the area of mobile field data collection that focus on multimodal (including voice) field data collection for industrial applications. We investigated the use of technologies that 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 and stylus, on a handheld device. A prototype mobile multimodal field data entry (MFDE) application have been developed to run on a Pocket PC that is equipped with a multimodal browser and embedded speech recognition capabilities. The prototype 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. A built-in microphone on a Pocket PC was utilized for speech data input [7].

User evaluation was conducted as a lab-based mobile evaluation of the prototype technology we developed. The detailed description of the study design is given in [8]. Our mobile application was designed to allow concrete technicians to record, while in the field (or more specifically, on a construction site), quality control data. The application supported two different modalities of data input – speech-based data entry and stylus-based data entry. The purpose of the evaluation was 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 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. A total of eighteen participants participated in the study.

The results of the evaluation confirmed, as it was anticipated, that stylus-based input was significantly more accurate than speech under the conditions of use that included construction noise in the range of 60-90 dB (A) [11]. We observed, however, that the stylus-based interaction was, on average, slower than speech-based input and that speech-based input significantly enhanced the participants' ability to be aware of their physical surroundings. In addition, majority of participants expressed preference for using speech as interaction technique with mobile device. As a result, this research study demonstrated significant preference for using speech as an interaction modality, with some limitations imposed by the lower speech recognition accuracy levels due to environmental noise. These findings led us to investigation of several technology factors that can potentially influence the accuracy of speech recognition, such as the type of the microphone and the type of speech recognition engine used.

3.2 Speech-Based Interactions – Technology Evaluations

The choice of microphone technology and speech recognition engine plays an important role in improving quality of speech recognition [19]. Our study described in detail in [12] was designed to evaluate and compare three commercially available microphones – the bone conduction microphone, and the two types of condenser microphones for their effect on accuracy of speech recognition within mobile speech input application. We developed a data input application based on a tablet PC running Windows XP and utilized IBM's ViaVoice embedded speaker-independent speech recognition engine, the same speech recognition engine that was utilized in our previous study [8]. Twenty four people participated in the laboratory-based study. The participants were mobile while entering information requested. The results of the study helped us to prove that the choice of microphone had significant effect on accuracy of mobile speech recognition; in particular, we found that both condenser microphones (QSHI3 and DSP-500 microphones) performed significantly better than bone conduction microphone (Invisio). In addition, we found that there was no significant effect of a background noise (within our evaluation scenario we incorporated street noise of 70 dB (A) level) on the accuracy of speech recognition, indicating that all microphones under evaluation had sufficient noise-cancelling capabilities.

Considering the importance of choosing the best speech recognition engine on the accuracy of results obtained, a complementary laboratory study was conducted to evaluate a number of state-of-the art speech recognition engines as to their effect on the accuracy of speech recognition [13]. This study was based on pre-recorded user speech entries, collected in our previously mentioned study [12]. All speech recognition engines were evaluated in speaker independent mode (e.g. walk-up-and-use). Based on the results of this study, we also proved the importance of proper pairing of microphone systems and speech recognition engines to achieve the best possible accuracy of speech recognition for mobile data entry.

3.3 Feasibility of Using Speech Interaction in Learning Contexts

Our previous research demonstrates that it is technically possible to implement speech-based and multimodal interaction with a mobile device and to achieve significant level of user acceptance and satisfaction with technology. However, if we were to consider implementation of speech-based interfaces within mobile learning domain, we have to look at other important considerations, such as appropriateness of speech as an interaction modality within certain contexts of use and social acceptance of speech-based interactions.

In a classroom environment, when a number of learners could potentially utilize mobile technology to participate in learning and collaboration process, the appropriateness of speech-based interaction is questionable, since simultaneous use of speech by multiple users will introduce high level of environmental noise that could significantly reduce the accuracy of speech recognition for each individual device. Thus, based on contextual considerations, this application of speech-based interfaces is not appropriate.

At the same time, utilization of speech interaction by a single mobile learner is very much appropriate and could significantly improve experience of his/her "learning on the go". Research has proven that mobile speech-based interaction could be successfully designed for users on the go, such as city tourist guides or in-car speech-based interfaces [10]. Most frequently these types of applications utilize a constrained vocabulary of user commands and a constrained grammar of possible user entries. This functionality enables menu navigation, information retrieval and some basic data entry capabilities. The same principles apply to utilization of speech-based and multimodal interfaces for student field trips, where students are mobile and take notes "on the go" [9].

Another interesting and rapidly developing research area is an application of speech-based and multimodal interfaces within various training scenarios, including industrial and military training. Within these scenarios, when training is conducted in the field or in the simulated field environment, voice command could enable efficient "hands free, eyes free" information retrieval, menu navigation and basic data entry. Another application of speech-based interfaces is within the domain of gaming, including "serious gaming" in education and training domains [27]. A major challenge for speech-based interfaces within "serious gaming" domain is to improve the accuracy of speech recognition within environmentally challenging conditions (high level of noise, people possibly being under stress thus affecting the way they speak and reducing the accuracy of command recognition, etc). Within this usage domain, we see an opportunity to successfully deploy multimodal interaction so that multiple channels of input would assist in improving accuracy and usability of the system [26].

4 Conclusions

Our research on speech-based and multimodal user interaction with mobile devices has proven that it is technically feasible to implement speech-based (or multimodal) interaction with a mobile device and to achieve significant level of user acceptance and satisfaction with this technology. We also identified some challenges associated with use of speech-based and multimodal interaction within the learning and training domains. Our future research efforts will be focused on exploring ways to better incorporate multimodal (including speech-based) interfaces within "serous gaming" scenarios, where this technology has potential to significantly improve usability of user interactions with technology, especially in cases where "hands-free and eyes free" interaction is a must, such as military and industrial training applications.

Acknowledgements. The author would like to acknowledge the support for this research program provided by the National Research Council Canada.

References

1. Crawford, V., Vahey, P.: Palm Education Pioneers Program: Evaluation Report. SRI International, Menlo Park (2002)
2. de Freitas, S., Levene, M.: Evaluating the Development of Wearable Devices, Personal Data Assistants and the Use of Other Mobile Devices in Further and Higher Education Institutions. JISC Technology and Standards Watch Report: Wearable Technology (2002)
3. Guenaga, M.L., Burger, D., Oliver, J.: Accessibility for e-Learning Environments. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 157–163. Springer, Heidelberg (2004)
4. Jahankhani, H., Lynch, J.A., Stephenson, J.: The Current Legislation Covering E-learning Provisions for the Visually Impaired in the EU. In: Shafazand, H., Tjoa, A.M. (eds.) EurAsia-ICT 2002. LNCS, vol. 2510, pp. 552–559. Springer, Heidelberg (2002)
5. Jokinen, K., Raike, A.: Multimodality – Technology, Visions and Demands for the Future. In: Proceedings of the 1st Nordic Symposium on Multimodal Interfaces, Copenhagen (2000)
6. Kondratova, I., Goldfarb, I.: M-learning: Overcoming the Usability Challenges of Mobile Devices. In: Proceedings International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICNICONSMCL 2006), p. 223. IEEE Computer Society Press, Los Alamitos (2006)
7. Kondratova, I.: Speech-Enabled Handheld Computing for Fieldwork. In: Proceedings of the International Conference on Computing in Civil Engineering 2005, Cancun, Mexico (2005)
8. Kondratova, I., Lumsden, J., Langton, N.: Multimodal Field Data Entry: Performance and Usability Issues. In: Proceedings of the Joint International Conference on Computing and Decision Making in Civil and Building Engineering, Montréal, Québec, Canada, June 14-16 (2006)
9. Kravcik, M., Kaibel, A., Specht, M., Terrenghi, L.: Mobile Collector for Field Trips. Educational Technology & Society 7(2), 25–33 (2004)
10. Larsen, L.B., Jensen, K.L., Larsen, S., Rasmussen, M.H.: Affordance in Mobile Speech-based User Interaction. In: Proceedings of the 9th international Conference on Human Computer interaction with Mobile Devices and Services, MobileHCI 2007, pp. 285–288. ACM, New York (2007)

11. Lumsden, J., Kondratova, I., Langton, N.: Bringing A Construction Site Into The Lab: A Context-Relevant Lab-Based Evaluation Of A Multimodal Mobile Application. In: Proceedings of the 1st International Workshop on Multimodal and Pervasive Services (MAPS 2006), Lyon, France (2006)
12. Lumsden, J., Kondratova, I., Durling, S.: Investigating Microphone Efficacy for Facilitation of Mobile Speech-Based Data Entry. In: Proceedings of the British HCI Conference, Lancaster, UK, September 3-7 (2007)
13. Lumsden, J., Durling, S., Kondratova, I.: A Comparison of Microphone and Speech Recognition Engine Efficacy for Mobile Data Entry. In: The International Workshop on MObile and NEtworking Technologies for social applications (MONET 2008), part of the LNCS OnTheMove (OTM) Federated Conferences and Workshops, Monterrey, Mexico, November 9-14 (2008)
14. Lumsden, J., Leung, R., Fritz, J.: Designing a Mobile Transcriber Application for Adult Literacy Education: A Case Study. In: Proceedings of the International Association for Development of the Information Society (IADIS) International Conference Mobile Learning 2005, Qawra, Malta, June 28 – 30 (2005)
15. Neal, L.: Predictions for 2002: e-learning Visionaries Share Their Thoughts. eLearn Magazine 2002(1), 2 (2002)
16. Oviatt, S., Cohen, P.: Multimodal Interfaces that Process What Comes Naturally. Communications of the ACM 43(3) (March 2000)
17. Pham, B., Wong, O.: 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 (2004)
18. Picardi, A.C.: IDC Viewpoint. Five Segments Will Lead Software Out of the Complexity Crisis, Doc #VWP000148 (December 2002)
19. Quek, F., MCNeill, D., Bryll, R., Dunkan, S., Ma, X.-F., Kirbas, C., MCCullough, K.E., Ansari, R.: Multimodal Human Discourse: Gesture and Speech. ACM Transactions on Computer-Human Interaction 9(3), 171–193 (2002)
20. Roschelle, J., Pea, R.: A Walk on the WILD Side: How Wireless Handhelds May Change Computer-supported Collaborative Learning. International Journal of Cognition and Technology 1(1), 145–168 (2002)
21. Roschelle, J.: Keynote paper: Unlocking the learning value of wireless mobile devices. J. of Computer Assisted Learning 19, 260–272 (2003)
22. Sadeh, N.: M-Commerce: Technology, Services, and Business Model. John Wiley & Sons, Inc., Chichester (2002)
23. Sawhney, N., Schmandt, C.: Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. ACM Transactions on Computer-Human Interaction 7(3), 353–383 (2000)
24. Smordal, O., Gregory, J.: Personal Digital Assistants in Medical Education and Practice. Journal of Computer Assisted Learning 19(3), 320–329 (2003)
25. Soloway, E., Norris, C., Blumenfeld, P., Fishman, B.J.K., Marx, R.: Devices are Ready-at-Hand. Communications of the ACM 44(6), 15–20 (2001)
26. Tse, E., Greenberg, S., Shen, C.: Exploring Interaction with Multi User Speech and Whole Handed Gestures on a Digital Table. In: Proceedings of ACM UIST 2006, Montreux, Switzerland, October 15–18 (2006)
27. Wang, X., Yun, R.: Design and Implement of Game Speech Interaction Based on Speech Synthesis Technique. In: Pan, Z., Zhang, X., El Rhalibi, A., Woo, W., Li, Y. (eds.) Edutainment 2008. LNCS, vol. 5093, pp. 371–380. Springer, Heidelberg (2008)

28. Wilson, L.: Look Ma Bell, No Hands! – VoiceXML, X+V, and the Mobile Device. XML Journal, August 3 (2004)
29. Zhang, D., Nunamaker, J.F.: A Natural Language Approach to Content-Based Video Indexing and Retrieval for Interactive E-Learning. IEEE Transactions on Multimedia 6(3), 450–458 (2004)
30. Zhuang, Y., Liu, X.: Multimedia Knowledge Exploitation for E-Learning: Some Enabling Techniques. In: Fong, J., Cheung, C.T., Leong, H.V., Li, Q. (eds.) ICWL 2002. LNCS, vol. 2436, pp. 411–422. Springer, Heidelberg (2002)