Transcending the display size: the case for speech interaction in educational applications

Munteanu, Cosmin

CHI 2012: ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 1-4, 2012-05-10

Transcending the display size: the case for speech interaction in educational applications

Cosmin Munteanu

Institute for Information Technology
National Research Council Canada

and

Department of Computer Science
University of Toronto

46 Dineen Dr.
Fredericton, NB E3B 9W4
CANADA

cosmin.munteanu@nrc-cnrc.gc.ca

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract

Speech is the most natural form of communication that humans employ, and is one of the main modalities through which we acquire and share knowledge. Moreover, speech is used not only to deliver knowledge, but as a modality that supports learning, such as student-teacher interactions around printed materials. During the past decade, we have witnessed significant advances mainly in preserving spoken educational materials, from informal how-to videos to full academic lectures being stored and available through a variety of online channels. Unfortunately, there is proportionately less research on enabling access to such multimedia knowledge repositories (e.g. searching, indexing) or on facilitating spoken, natural interaction between learners and digital interactive media (such as automated tutors or interactive learning resources). By enabling speech as a modality, learners become less constrained by the physical properties of the educational materials and can interact more naturally with the educational software, be it in the form of a mobile language assistant, a desktop-based online lecture browsing system, or a mixed-reality serious gaming system. In this paper I present examples of such recent research on improving the way we interact with educational resources through speech and natural language, and make the case for the need to conduct further research in this area.

Author Keywords
Educational interfaces, automatic speech recognition, interface design, multimodal interaction.

ACM Classification Keywords
H5.2 User interfaces: Voice I/O, Natural language, User-centered design, Evaluation/methodology. K3.1 Computer Uses in Education: Computer-assisted instruction.

General Terms
Human Factors, Languages.

Introduction

Humankind has for long relied on written texts to preserve knowledge, and for speech to share it directly. The advent of affordable broadband Internet and personal (and portable) computing devices has contributed to dramatic changes in the way people exchange information and store knowledge, such as our ability to easily preserve educational materials in spoken form. As the availability and diversity of storage and interactive media increases, we are witnessing an explosion of spoken information being archived, from informal how-to videos to full academic lectures being stored and available through a variety of online channels. Unfortunately, large-scale preservation of materials in spoken form is one of the few areas of educational applications that has undergone significant transformations with respect to speech-based interaction. Other research areas, such as multimedia indexing or searching, are comparatively less prominent. For example, a user must listen to or watch a long recording in order to locate a specific passage, instead of quickly skimming through a document looking for visual landmarks and textual cues. This represents an important hurdle in making multimedia recordings the digital equivalent of textbooks.

Speech is used in education not only to transfer knowledge, but also to support essential interactions between students and teachers, such as discussions and explanations. Intelligent automated tutors has been for long a focus of research within both Human-Computer Interaction (HCI) and Automatic Speech Recognition (ASR). However, speech has often been neglected as an interaction modality, mainly due to the usability challenges arising from its inherently high error rates. In this position paper I present three examples of research on speech-enabled educational interfaces. These illustrate the point that through careful user-centric design and consideration of ASR's strengths and limitations, speech can become a successful modality in educational interfaces.

Speech Interaction in Educational Applications

A Mobile Language Learning Assistant

In countries like Canada, low-literacy adults represent a sizeable ratio of the adult population. Unfortunately, due to a variety of economic and socio-demographic reasons, the current programs designed to provide learning support and resources to low-literacy adults have difficulty reaching and retaining those that would benefit most from them. For this, we have developed ALEX – a mobile language assistant for use both in the classroom and in daily life, in order to help low-literacy adults become increasingly literate and independent. It is an application running on ultra-mobile devices, designed to help develop language skills and knowledge acquisition pertaining to real life by providing intuitive access to various language-based tools (dictionaries, thesauri, etc.).

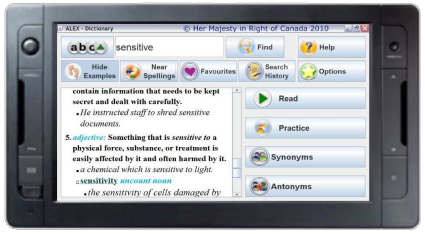

Figure 1: The ALEX mobile language learning assistant. A complete description of the study and its key findings (as well as of the ALEX system) can be found in [2].

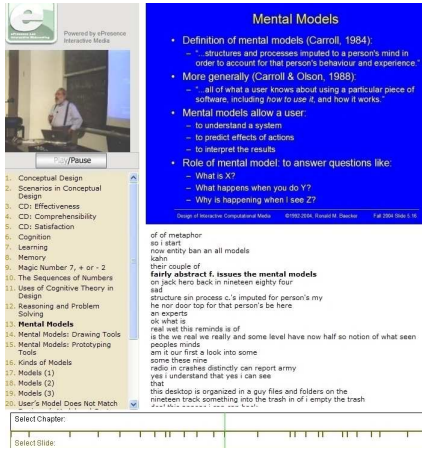

Figure 2: Our transcript-enhanced lecture webcast system.

ALEX has been designed following guidelines for inclusive design of mobile tools. Several features are provided to assist users in their learning goals: alphabetic and QWERY soft keyboards, a near-spelling feature to facilitate dictionary look-up for users who have difficult spelling, and text-to-speech navigation of menus and functions. Text-to-speech is also used in the learning process, allowing users to hear the correct pronunciation of words or to have a dictionary definition read to them (with corresponding text being highlighted and synchronized at word level). Through ASR, ALEX allows learners to practice their pronunciation. The pronunciation practice functionality provides feedback in the form of a color-based dial accompanied by positive reinforcement messages. Users can hear their own recording and compare it with the correct pronunciation.

We have evaluated ALEX through a six-month study with 11 participants in two adult literacy classes. Participants used the devices both in the classroom and in daily life. Our study revealed that students perceived the device as helpful when doing homework, as well as with the pronunciation of difficult words, which is an essential component of literacy programs. We also found that such technologies can contribute to students' independence with respect to activities requiring the use of literacy skills and can increase students' confidence in their own capabilities and motivation to learn.

Webcast Lecture Browsing

Webcasts are a common vehicle for delivering online presentations, and universities are embracing them in order to stream lectures over the Internet. Most webcast presentations are archived, and several interactive systems exist that allow users to watch the recorded presentations. However, many such systems lack interactive transcripts, impeding information-seeking tasks such as retrieval, browsing, or skimming.

In our research [1], we have investigated how the accuracy of automatically-generated transcripts influences both users' performance in a question-answering task and their perception of transcript quality. For this, we have used a full-featured lecture browsing system (University of Toronto's ePresence) that was augmented with transcripts obtained through ASR – transcript lines were synchronized with the audio and clickable, allowing users to cue the video playback to the corresponding location. A controlled study with 48 participants in a quiz-answering scenario revealed that transcript quality does not need to be perfect – students' quiz scores, even when transcripts had a 25% word error rate, were better than when having no transcripts at all. A similar result was manifested for users' confidence and perception of difficulty. Students perceived transcripts as being very helpful and indicated they would rather have imperfect transcripts than no transcripts.

To mitigate the effects of ASR errors, we have also developed a collaborative editing tool that allows users to correct and edit the transcripts. It extends the basic functionality of the system without burdening the user a the same time. A longitudinal study of this editing extension [1], for the duration of an entire semester with 12 participants, showed that this is a feasible solution for improving the quality of lecture transcripts.

The results of this research are encouraging for the use of automatically-generated transcripts as a navigational aid in online lecture delivery systems, even when ASR accuracy is not perfect. Empowering users to correct ASR errors in the transcripts yielded even more accurate

Figure 3: Speech interaction and touch controls in a mixed-reality training system (MINT).

transcripts and created a different sense of engagement and participation in lectures.

Interaction with Mixed-reality Serious Games

Serious gaming systems, such as mixed-reality training simulators, are considered to be a cost-effective alternative to field-based courses in several complex learning environments, such as military or law enforcement training. A realistic rendering of scenarios and conditions allows trainees to transition from course-based (classroom) instruction to applying their knowledge, skills, and judgement to solving real-life situations. Many of such situations require trainees to interact with people in different roles, and as such, speech should be an essential modality in mixed-reality learning systems.

In collaboration with the Canadian Forces, we have developed MINT, a customizable research platform that supports multimodal interaction between trainees and an immersive serious gaming environment [3]. The gaming environment is projected on multiple walls, allowing the recreation of real-world environments and interaction with life-size characters. MINT has ASR capabilities to facilitate natural interactions with the immersive gaming system. Since several adverse conditions normally present in such settings affect the quality of the ASR, we have introduced complementary modalities to enhance the recognition of spoken commands, in the form of a tablet game control interface. This touch interface allows instructors to create "on-the-fly" training situations that test the responsiveness and judgment of the trainee by animating game characters in response to trainees' actions.

We have evaluated the multimodal training system with a team of instructors testing the system under scenarios they normally design as part of the basic training curriculum. We have found that MINT is well received by its intended users, and that speech-based interaction is critical in ensuring the realism of the interaction with the virtual environment. The ASR accuracy can vary significantly across training conditions, but can be easily compensated for by using complementary modalities, such as the tablet-based trainer interface through which learning scenarios can be controlled and customized.

Conclusions

Speech-based interaction has the potential to improve the naturalness of educational interfaces. Unfortunately, it is only timidly embraced by interface and interaction designers, mainly due to its inherently-high error rates. In this position paper I have presented three examples of research projects within the educational landscape that employ ASR and speech-based interactions. These examples demonstrate that through a proper cross-disciplinary approach, speech can be successfully used as a modality in a wide range of interactive educational applications. Evidence of such successes will hopefully lead to further research being conducted on speech-based interactions for educational interfaces, within the areas of both ASR and HCI.

References

[1] Munteanu, C. et al. (2006). Automatic Speech Recognition for Webcasts: How Good is Good Enough and What to Do When it Isn't. Proc. of ICMI.

[2] Munteanu, C. et al. (2011). Showing off Your Mobile Device: Adult Literacy Learning in the Classroom and Beyond. Proc. of Mobile HCI.

[3] Fournier, H. et al. (2011). A Multidisciplinary Approach to Enhancing Infantry Training through Immersive Technologies. Proc. of IITSEC.