NRC Publications Archive Archives des publications du CNRC

We need to talk : HCI and the delicate topic of spoken language interaction

Munteanu, Cosmin; Jones, Matt; Oviatt, Sharon; Brewster, Stephen; Penn, Gerald; Whittaker, Steve; Rajput, Nitendra; Nanavati, Amit

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1145/2468356.2468803>

ACM digital library in the "extended abstract" companion of the CHI conference proceedings, pp. 2459-2463, 2013-04-01

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=734b5156-3597-4576-95fa-73a2db0e9cd6>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=734b5156-3597-4576-95fa-73a2db0e9cd6>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

We Need to Talk: HCI and the Delicate Topic of Spoken Language Interaction

Moderator:

Cosmin Munteanu

National Research Council and
University of Toronto, Canada
Cosmin.Munteanu@nrc-cnrc.gc.ca

Panelists:

Matt Jones

Swansea University,
United Kingdom
matt.jones@swansea.ac.uk

Steve Whittaker

University of California at Santa
Cruz, USA
swhittak@ucsc.edu

Sharon Oviatt

Incaa Designs, USA
oviatt@incaadesigns.org

Nitendra Rajput

IBM India Research Laboratory
rnitendra@in.ibm.com

Stephen Brewster

University of Glasgow,
United Kingdom
stephen.brewster@glasgow.ac.uk

Contributing authors:

Gerald Penn

University of Toronto, Canada
gpenn@cs.toronto.edu

Amit Nanavati

IBM India Research Laboratory
namit@in.ibm.com

Abstract

Speech and natural language remain our most natural form of interaction; yet the HCI community have been very timid about focusing their attention on designing and developing spoken language interaction techniques. This may be due to a widespread perception that perfect domain-independent speech recognition is an unattainable goal. Progress is continuously being made in the engineering and science of speech and natural language processing, however, and there is also recent research that suggests that many applications of speech require far less than 100% accuracy to be useful in many contexts. Engaging the CHI community now is timely – many recent commercial applications, especially in the mobile space, are already tapping the increased interest in and need for natural user interfaces (NUIs) by enabling speech interaction in their products. As such, the goal of this panel is to bring together interaction designers, usability researchers, and general HCI practitioners to discuss the opportunities and directions to take in designing more natural interactions based on spoken language, and to look at how we can leverage recent advances in speech processing in order to gain widespread acceptance of speech and natural language interaction.

Copyright is held by the author/owner(s).

CHI'13, April 27 – May 2, 2013, Paris, France.

ACM 978-1-XXXX-XXXX-X/XX/XX.

Keywords

Automatic Speech Recognition; Speech Interaction.

ACM Classification Keywords

H.5.2 [User interfaces]: Voice I/O, Natural language, User-centered design, and Evaluation/methodology.

General Terms

Human Factors; Languages.

Introduction and Motivation

Our senses, such as touch, sight, hearing, or speech, allow us to interact with objects, information, or other humans. While such interactions are only slightly altered by technological progress, digital technologies are now reshaping the way we interact with our environment. We are no longer in direct control over such interactions; instead, we project them through a virtual layer. During the past decade we have witnessed dramatic changes in the way people access information and store knowledge, mainly due to the ubiquity of mobile and pervasive computing and affordable broadband Internet. Such recent developments have presented us the opportunities to reclaim naturalness as a central theme for interaction. We have seen this happen with touch for mobile computing; it is now time to see this for speech as well.

Unfortunately, humans' most natural forms of communication, speech and language, are also among the most difficult modalities for machines – despite, and perhaps, because these are the highest-bandwidth communication channel we have. While significant efforts, from engineering, linguistic, and cognitive sciences, have been spent on improving machines' ability to understand speech and natural language,

these have often been neglected as interaction modalities, mainly due to the usability challenges arising from their inherently high error rates and computational complexity.

The challenges in enabling such natural interactions have often led to these modalities being considered, at best, error-prone alternatives to “traditional” input or output mechanisms. However, this should not be a reason to abandon speech interaction¹ – in fact, people are now exposed to many more situations in which they need to interact hands- and eyes-free with a computing device. Furthermore, achieving perfectly accurate speech processing is a lofty goal that is often nothing short of a fairy tale. There is significant research evidence pointing to the fact that proper interaction design can complement speech processing in ways that compensate for its less-than-perfect accuracy (Oviatt, 2003, and Munteanu, 2006), or that in many tasks where users interact with spoken information, verbatim transcription of speech is not relevant at all (Penn and Zhu, 2008).

While recent commercial applications (e.g. personal digital assistants) have brought renewed attention to speech-based interaction, speech can assist with a wider range of tasks that are not limited to direct interactions. If we are to pick just one example where such research is desperately needed, it would be the area of access to multimedia repositories. 72 hours of video are uploaded to Youtube each minute (Youtube, 2012). At this rate, it is humanly not possible to

¹ Throughout the rest of this document we will use the term speech and speech interaction to denote both verbal and text-based interaction, where the textual representation has been obtained from its original spoken source.

consume the amount of data being generated, and it is becoming increasingly difficult to search for information or navigate through such large and often multilingual collections. Technologies that assist with such tasks include summarization of text or audio/video documents, browsing/searching through and indexing of large multimedia repositories, secure user authentication, natural language generation, speech synthesis, or speech-to-speech machine translation. Unfortunately, there is very little HCI research on how to leverage the progress being made in these areas into developing more natural, effective, or accessible user interfaces.

Goals

In light of such barriers and opportunities, this panel aims to foster an interdisciplinary dialogue on:

- Identifying and discussing the challenges to the widespread adoption of speech interaction,
- Engaging the CHI community in finding user-centric design guidelines for speech-based interactive systems, grounded in good usability practices, and
- Opening up further research opportunities in designing more natural interactions that make use of speech.

Topics

Several topics will be advanced for discussion among panellists and between panellists and audience members:

- What are the important challenges in using speech as a “mainstream” modality?

- What opportunities are presented by the rapidly evolving mobile and pervasive computing areas?
- Given the penetration of mobile computing in emerging markets, are there any specific usability or technology adoption issues surrounding speech interaction?
- What opportunities exist to improve users experiences by using speech-based technologies that are not limited to input or output?
- How can we bridge the divide between the evaluation methods used in HCI and those in speech processing (which are mostly based on Artificial Intelligence practice)?
- What can the CHI community learn from the Automatic Speech Recognition (ASR) and the Natural Language Processing (NLP) research, and in turn, how can it help the ASR and NLP communities improve the user-acceptance of such technologies?
- What are the contexts in which we can expect to see spoken language processing expand the most commercially in the future?

Audience and panel format

Expected audience

The intended audience for this panel is the entire CHI community. The proposed topics address speech interaction, which is currently receiving significant attention in the commercial space and in areas outside HCI, yet it is a marginal topic at CHI. For this, we expect a diverse audience with varying interests. While

we have offered a course at CHI for the past two years (Munteanu and Penn, 2011, 2012) introducing the topic of speech recognition and speech-based interaction, the panel proposed here does not require a knowledge of the inner workings of speech recognition or natural language processing. In fact, we expect that the panel will encourage contributions and discussions from audience members who do not consider themselves “experts” within these areas.

Our panellists have backgrounds that complement each other, ranging from significant work in core speech processing issues, to natural and multimodal interfaces, and to testing the social acceptance of speech applications. Our hope is that this diversity will ensure a vigorous and interesting debate on issues related to the usability, usefulness, and general acceptance of speech as an interaction modality that will surely engage the audience.

Panel proceedings

After a brief introduction and opening statements from each panellist, we will discuss the topics listed in this proposal. The questions raised in the Topics section will be posed to the panellists – each question/topic will be answered by our panellists, followed by questions or brief comments from the audience. Some of the questions may be addressed specifically at some panellists, aiming to elicit diverse (and occasionally, opposite) views on the topic, with the remaining panellists contributing follow-up comments. The panel will conclude with open discussions between panellists and audience members followed by brief concluding statements from the panel members.

Panellists

Dr. Cosmin Munteanu

is a Research Officer with the National Research Council (NRC) Canada and an Adjunct Professor at the University of Toronto. His area of expertise is at the intersection of ASR and HCI, having extensively studied the human factors of using imperfect speech recognition systems, and having designed and evaluated systems that improve humans' access to and interaction with information-rich media and technologies through natural language. Cosmin currently oversees the NRC's Voice and Multimodal Lab where he leads several industrial and academic research projects exploring spoken language interaction for mobile devices and mixed reality systems.

Prof. Matt Jones

Matt Jones is a Professor of Computer Science at the Future Interaction Technology Lab at Swansea University. He has worked on mobile interaction issues for the past seventeen years and has published a large number of articles in this area, including co-authoring the "Mobile Interaction Design" book. He has had many collaborations and interactions with handset and service developers. He has been a Visiting Fellow at Nokia Research and held an IBM Faculty Award to work with the Spoken Web group in IBM Research India. He is an editor of the International Journal of Personal and Ubiquitous Computing and on the steering committee for the Mobile Human Computer Interaction conference series. His research work has focussed on the fusion of physical and digital spaces in challenging contexts, recently being involved in projects such as exploring the role of haptics, gestures and audio in mobile scenarios and storytelling in rural Indian Village.

Prof. Steve Whittaker

Steve Whittaker is Full Professor in Psychology at University of California at Santa Cruz. His research interests are in the theory and design of collaborative systems, CMC, speech browsing and personal information management. He has designed many novel systems, including: lifelogging systems, one of the first IM clients, shared workspaces, social network email clients, meeting capture systems, and various tools for accessing and browsing speech. He has previously worked at Sheffield University, Hewlett Packard, Bell Labs, AT&T, Lotus and IBM Cambridge and IBM Almaden Labs. He is a member of the Association of Computational Machinery Computer Human Interaction Academy. Right now he is working on digital tools to support human memory.

Dr. Sharon Oviatt

Sharon Oviatt is well known for her research on human-centred interfaces, multimodal and mobile interfaces, and educational interfaces. She has published over 130 scientific articles, and is an Associate Editor of the main journals in the field of HCI. She was the recipient of a National Science Foundation Special Creativity Award for pioneering work on mobile multimodal interfaces that combine natural input modes like speech, pen, touch, and gesture. Recently, she founded Incaa Designs (<http://www.incaadesigns.org/>), a nonprofit that researches and evaluates new educational interfaces designed to stimulate thinking and reasoning.

Dr. Nitendra Rajput

Nitendra Rajput is a Senior Researcher and a Research Manager with IBM India Research Lab. His interests and expertise lie in the area of building novel interaction

systems, especially for resource-constrained mobile devices targeting low-literate users. His research on Hindi Speech Recognition and on Spoken Web has been presented awards by the President of India in 2007 and 2009 respectively. He has been active in patent innovation, for which he was recognized as IBM Master Inventor in 2010. He has also been active in organizing multiple workshops related to technologies for developing regions, the latest being IKM4DR and IUI4DR.

Prof. Stephen Brewster

Stephen Brewster is a Professor of Human-Computer Interaction in the Department of Computing Science at the University of Glasgow, UK, where he leads the Multimodal Interaction Group (part of the Glasgow Interactive Systems Group). His main research interest is in Multimodal Human-Computer Interaction, sound and haptics and gestures. Stephen has conducted significant research into Earcons, a particular form of non-speech sounds. He has authored numerous publications in the fields of audio and haptic (touch-based) interaction and mobile computing devices.

References

- [1] Munteanu, C. and Penn, G. (2012). Speech-based interaction. Course, ACM SIGCHI 2011, 2012.
- [2] 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.
- [3] Oviatt, S. (2003). Advances in Robust Multimodal Interface Design. IEEE Comput. Graph. Appl. 23-5.
- [4] Penn, G. and Zhu, X. (2008). A critical reassessment of evaluation baselines for speech summarization. In Proc. of ACL-HLT.
- [5] Youtube (2012). Upload statistics. http://www.youtube.com/t/press_statistics