
Talking with Conversational Agents in Collaborative Action

Martin Porcheron**Joel E. Fischer**

The Mixed Reality Laboratory
University of Nottingham, UK
porcheron@acm.org
joel.fischer@nottingham.ac.uk

Moira McGregor**Barry Brown**

Mobile Life Centre
Stockholm University, Sweden
moira@mobilelifecentre.org
barry@mobilelifecentre.org

Ewa Luger

Design Informatics
University of Edinburgh, UK
eluger@exseed.ed.ac.uk

Heloisa Candello

Social Data Analytics Group
IBM Research Lab, Brazil
heloisacandello@br.ibm.com

Kenton O'Hara

Microsoft Research, UK
keohar@microsoft.com

Abstract

This one-day workshop intends to bring together both academics and industry practitioners to explore collaborative challenges in speech interaction. Recent improvements in speech recognition and computing power has led to conversational interfaces being introduced to many of the devices we use every day, such as smartphones, watches, and even televisions. These interfaces allow us to get things done, often by just speaking commands, relying on a reasonably well understood single-user model. While research on speech recognition is well established, the social implications of these interfaces remain underexplored, such as how we socialise, work, and play around such technologies, and how these might be better designed to support collaborative collocated talk-in-action. Moreover, the advent of new products such as the Amazon Echo and Google Home, which are positioned as supporting multi-user interaction in collocated environments such as the home, makes exploring the social and collaborative challenges around these products, a timely topic. In the workshop, we will review current practices and reflect upon prior work on studying talk-in-action and collocated interaction. We wish to begin a dialogue that takes on the renewed interest in research on spoken interaction with devices, grounded in the existing practices of the CSCW community.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
CSCW '17 Companion, February 25 - March 01, 2017, Portland, OR, USA
ACM 978-1-4503-4688-7/17/02.
<http://dx.doi.org/10.1145/3022198.3022666>

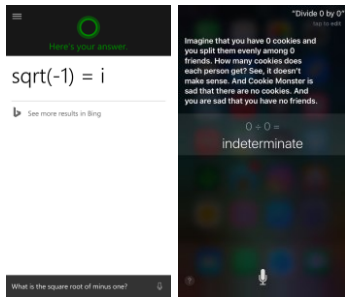


Figure 1. Example screenshots of responses from Cortana and Siri, running on smartphones.



Figure 2. Amazon Echo, a standalone commercially-available device that can only be interacted with by talking to the inbuilt conversational agent Alexa. Configuration for the Echo occurs either through a smartphone app or website.

Author Keywords

conversational agents; intelligent personal assistants; multi-party conversation; collocated interaction; CSCW.

ACM Classification Keywords

H.5.3. Information interfaces and presentation (e.g., HCI): Computer-supported cooperative work; H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

Many of the recent personal mobile devices released to market, such as smartphones, tablets, and watches, have embraced the use of automatic speech recognition as a viable form of device interaction. The devices typically feature a speech interface, often referred to as a conversational agent or intelligent personal assistant, that embodies the idea of a virtual butler [14]. These interfaces listen to spoken commands and queries, and respond accordingly by performing a broad range of tasks on the user's behalf, such as to provide facts and news, and set reminders and alarms. Furthermore, the systems are often anthropomorphised by being given names (e.g. Cortana or Siri, see Figure 1 for screenshots) and endowed with humanlike behaviours [21] such as humour. However, recent research shows that despite grand promises made by manufacturers, existing commercial systems fail to meet users' expectations [10].

Amazon and Google have further embraced this trend by launching standalone hardware in the form of the Amazon Echo (see Figure 2) and Google Home. These devices are specifically designed to be placed in social spaces such as kitchens tables, within everyday settings such as the home. While these devices function

in much the same way as their mobile equivalents, they rely entirely upon speech for interaction to support their broad range of abilities. Given the growing IoT trend, it is hardly surprising that these devices also support the ability to control connected domestic appliances such as lights, thermostats, and kettles.

Given the ever-expanding abilities of the conversational agents, the scope for their use and thus the range of social and collaborative contexts in which their use can become occasioned is also expanding. This provides a renewed impetus for research in HCI and CSCW to explore the practical and social implications of the use of these systems not just in the lab, but to explore the everyday use of conversational agents *in vivo*. Recently, researchers have begun to explore the use of personal assistants in public settings such as cafés [15], or workplaces [11]. Inspired by this recent work, this workshop seeks to explore the use, study, and design of conversational agents in social and collaborative settings. In the following, we highlight some understudied features of conversational agent use in social settings, to topicalise and energise research activity in this space.

Conversational Agents in Social Settings

The growing pervasiveness of speech-enabled technologies revitalises existing research on the practices of face-to-face conversation, on interacting with speech-enabled technologies, and on the use of ubiquitous technologies in multi-device ecologies. To explore this space, the workshop invites contributions across a range of topical interests to examine and collect exemplars of current and future challenges the CSCW community face in researching interactions with and around conversational agents. The broad and

Themes and Goals

We invite contributions related but not limited to any of the following:

- Studies of social settings where speech-enabled technologies are used, or may be used in future
- Design, deployments, and studies of speech-enabled technologies for single – and multi-person use
- Experiments that show implications of design choices (tone, gender, etc.) that might impact on socialising with speech-enabled technologies
- Examples or considerations of how to support the discovery of speech functionality *in situ*
- Considerations of how sensitive social contexts and privacy might be managed when voice is the interface
- Examples of where speech-enabled technologies are used for co-operative or collaborative purposes

diverse sociotechnical topics that exist with new speech-enabled technologies are illustrated by two exemplar challenges relating to the transition of conversational agent technologies from touchscreen to screenless devices, and from single-user to multi-user.

From Touch to Speech

Interacting with personal assistants on mobile devices relies on both touching the device and speaking to it, allowing users to hold and interact with devices in tune with the prevalent interaction paradigm for touchscreen devices. However, with the transition to screenless devices (see Figure 2) fundamentally different qualities emerge that common visual-touch does not occasion [15], related for instance to the potentially more 'natural' accountability of the ongoing interaction to everyone within earshot. By forming a more in depth understanding of the conversational use of people talking to agents, HCI designers can inform and tailor design [13], and potentially support collaborative collocated interactions [5] in different environments. In turn, how such mundane practice occurs with devices that feature no screen remains as-yet unexplored.

To answer how researchers can uncover the social and technological challenges of interacting with conversational agents in multi-party settings, we can examine existing CSCW work on collocated interactions with technology. For example, numerous pieces explore the everyday social practice and the challenges around group interactions with personal devices in settings as diverse as the pub [16], around the television [19], and during family mealtimes [2,12], drawing upon analytic perspectives such as EMCA (Ethnomethodology and Conversation Analysis). This work serves as an important methodological grounding to understanding

how people use speech-enabled technologies, how others can co-manage the use of the technology in a face-to-face setting, and of the broader social implications of the use of these systems and devices.

From Single-User to Multi-User

For devices that feature visual displays, the display may be used to communicate state with the user. However, in multi-party settings this state is likely not observable-reportable to those nearby, and therefore relies on 'the user' to account for the on-screen information themselves [15]. Collocated interactions research has long-explored how to support multi-device and multi-user interactions for collaborative activity (e.g. [1,7,17]). Wooffitt et al. remind us, engaging in dialogue is itself a co-operative effort in negotiation of meaning [4:14], but how this cooperation can be embraced between multiple users while engaging with speech-enabled technologies remains an open question. Although speaking is naturally observable and reportable to all present to its production, it is also transient [20] and requires attention from members to listen and interpret as the action unfolds.

Devices that adopt a multi-user perspective in their design, such as Amazon Echo and Google Home, provide even greater deviation from accustomed mobile devices by eschewing displays altogether. Additionally, through the combination of multiple far-field microphones and speakers, and a tubular design, the devices support interactions by multiple users from different angles, even those who aren't close to the device in all directions. This new-wave of speech-only technologies encourages us to consider how their use becomes embedded within everyday talk. By revealing how interactions with speech-enabled technologies are

Themes and Goals (continued)

- Approaches and examples of how studies of face-to-face interaction inform design of devices
- Studies that explore how such systems could bridge the gulf between expectation and experience
- Techniques of sensing 'social context', e.g. collocation, conversation, and bodily orientation
- Concepts and design examples of systems that support collocated group-awareness and coordination
- Discussions of methods and tools to study and evaluate socio-technical systems with a focus on collocated settings
- Conceptual frames aiding the understanding of collocated interaction
- Case studies and lessons learned from evaluating the impact of technology on collocated interactions

cooperatively attended to by users and spectators, and how talking to a speech-enabled technology is collaboratively managed though talk-in-interaction in face-to-face settings, we can shape the design of future technologies. Given the significant potential that speech-only technologies possess to enrich collaborative work, play, and living—it seems particularly pertinent and timely topic for CSCW to explore the challenges involved.

In summary, we have briefly outlined underexplored concerns of speech-based interaction in social settings relating to recent technological trends, which have in turn raised numerous open-ended questions. By adopting methodical practice within existing CSCW work, many of these questions could be explored. The overarching goal of the workshop is to reflect on the insights and established practice of research across this diverse domain, to understand the impact of interacting with technology through speech, and even more so, how we can design and study such interactions?

Workshop Plan

The one-day workshop is structured into a series of segments, designed to introduce participants to each other's work, and to foster intellectually stimulating interaction. The day will begin with sessions devoted to acquainting attendees with each other and their work. Each participant will be asked to briefly present their position paper or poster to the group – this will be fast-paced and time-limited in relation to the number of submissions. We will follow this activity with activities around mapping the design space, and discussion of the key challenges raised within the submissions and through related work to prepare the development of opportunities and design ideas in the afternoon.

The afternoon will focus on breakout sessions to discuss the challenges emerged in the morning. Each participant will write down the main challenges they face in their work and share their experience. They can share their main methodology choices, experiments and approaches of their work. Main themes, trends, areas will emerge and organisers will take notes (cards, post-its). Participants, in groups, will think and share their experience and knowledge identifying CSCW opportunities and design ideas to be explored in each area previous discussed. This discussion will be materialised in a fictional scenario, and make use of paper prototyping materials. Participants will discuss the main elements present on a collaborative conversational system - task, context, single or multi-user, language, tone, content etc. – and will create a scenario envisioning a design idea or an opportunity.

Towards the end of the day, we will reconvene to discuss and reflect on the outcomes of the workshop activities. This will include a demonstration of the breakout activities, facilitating and stimulating a discussion amongst participants and organisers. This session will allow attendees to consider and reflect upon the challenges and themes raised during the morning, and to explore how the CSCW can respond to these challenges. The outcomes of the day, including this discussion, will be recorded to allow for follow-up activities. We will orient towards forming tangible outcomes of the interactive sessions of workshop, including an agenda for research in this space. Follow up activities from the workshop could include further related workshops, joint publications, and potentially a special-issue of a journal.

Organisers

The organisers have recently co-organised a range of workshops that explored different topics within collocated and social interaction [3,6,8,9,18]. We seek to build on these recent experiences, as well as practice and research on interacting with speech recognition systems from academic and industrial practice.

Acknowledgements

Martin Porcheron is supported by EPSRC grants EP/G037574/1 and EP/G065802/1. Joel E. Fischer is supported by EPSRC grant EP/N014243/1.

Data access statement: No research data was collected or used for this publication.

References

1. James Clawson, Amy Volda, Nirmal Patel, and Kent Lyons. 2008. Mobiphos: A Collocated-Synchronous Mobile Photo Sharing Application. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '08)*, 187. <https://doi.org/10.1145/1409240.1409261>
2. Hasan Shahid Ferdous, Bernd Ploderer, Hilary Davis, Frank Vetere, Kenton O'Hara, Jeremy Farr-Wharton, and Rob Comber. 2016. TableTalk: Integrating Personal Devices and Content for Commensal Experiences at the Family Dinner Table. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*, 132–143. <https://doi.org/10.1145/2971648.2971715>
3. Joel Fischer, Martin Porcheron, Andrés Lucero, Aaron Quigley, Stacey Scott, Luigina Ciolfi, John Rooksby, and Nemanja Memarovic. 2016. Collocated Interaction: New Challenges in “Same Time, Same Place” Research. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*, 465–472. <https://doi.org/10.1145/2818052.2855522>
4. Nigel Gilbert, Robin Wooffitt, and Norman Fraser. 1990. Organising Computer Talk. In *Computers and Conversation (1st Editio)*, Paul Luff, David Frohlich and Nigel Gilber (eds.). Academic Press, 235 – 257.
5. Alistair Jones, Jean-Paul A. Barthes, Claude Moulin, and Dominique Lenne. 2014. A rich multi-agent architecture for collaborative multi-touch multi-user devices. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC '14)*, 1107–1112. <https://doi.org/10.1109/SMC.2014.6974062>
6. Andrés Lucero, James Clawson, Kent Lyons, Joel E Fischer, Daniel Ashbrook, and Simon Robinson. 2015. Mobile Collocated Interactions: From Smartphones to Wearables. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*, 2437–2440. <https://doi.org/10.1145/2702613.2702649>
7. Andrés Lucero, Jaakko Keränen, and Hannu Korhonen. 2010. Collaborative Use of Mobile Phones for Brainstorming. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*, 337. <https://doi.org/10.1145/1851600.1851659>
8. Andrés Lucero, Aaron Quigley, Jun Rekimoto, Anne Roudaut, Martin Porcheron, and Marcos Serrano. 2016. Interaction Techniques for Mobile Collocation. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*. <https://doi.org/10.1145/2957265.2962651>
9. Andrés Lucero, Danielle Wilde, Simon Robinson, Joel E. Fischer, James Clawson, and Oscar Tomico.

2015. Mobile Collocated Interactions With Wearables. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (MobileHCI '15), 1138–1141. <https://doi.org/10.1145/2786567.2795401>
10. Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 5286–5297. <https://doi.org/10.1145/2858036.2858288>
 11. Moira McGregor and John Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '17). <https://doi.org/10.1145/2998181.2998335>
 12. Carol Moser, Sarita Y Schoenebeck, and Katharina Reinecke. 2016. Technology at the Table: Attitudes About Mobile Phone Use at Mealtimes. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 1881–1892. <https://doi.org/10.1145/2858036.2858357>
 13. Michael Norman and Peter Thomas. 1990. The Very Idea: Informing HCI Design from Conversation Analysis. In *Computers and Conversation* (1st Editio), Paul Luff, David Frohlich and Nigel Gilber (eds.). Academic Press, 51–65.
 14. Sabine Payr. 2013. Virtual butlers and real people: styles and practices in long-term use of a companion. In *Your Virtual Butler*, Robert Trapp (ed.). Springer-Verlag Berlin, Heidelberg, 134–178. https://doi.org/10.1007/978-3-642-37346-6_11
 15. Martin Porcheron, Joel E Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '17). <https://doi.org/10.1145/2998181.2998298>
 16. Martin Porcheron, Joel E Fischer, and Sarah C. Sharples. 2016. Using Mobile Phones in Pub Talk. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 1647–1659. <https://doi.org/10.1145/2818048.2820014>
 17. Martin Porcheron, Andrés Lucero, and Joel E Fischer. 2016. Co-Curator: Designing for Mobile Ideation in Groups. In *Proceedings of the 20th International Academic Mindtrek Conference* (AcademicMindtrek '16), 226–234. <https://doi.org/10.1145/2994310.2994350>
 18. Martin Porcheron, Andrés Lucero, Aaron Quigley, Nicolai Marquardt, James Clawson, and Kenton O'Hara. 2016. Proxemic Mobile Collocated Interactions. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16), 3309–3316. <https://doi.org/10.1145/2851581.2856471>
 19. John Rooksby, Timothy E Smith, Alistair Morrison, Mattias Rost, and Matthew Chalmers. 2015. Configuring Attention in the Multiscreen Living Room. In *Proceedings of the 14th European Conference on Computer Supported Cooperative Work* (ECSCW '15), 243–261. https://doi.org/10.1007/978-3-319-20499-4_13
 20. Ben Shneiderman. 2000. The Limits of Speech Interaction. *Communications of the ACM* 43, 63–65. <https://doi.org/10.1145/348941.348990>
 21. Giorgio Vassallo, Giovanni Pilato, Agnese Augello, and Salvatore Gaglio. 2010. Phase Coherence in Conceptual Spaces for Conversational Agents. In *Semantic Computing*, Phillip C.-Y. Sheu, Heather Yu, C. V. Ramamoorthy, Arvind K. Joshi and Lotfi A. Zadeh (eds.). John Wiley & Sons, Inc., Hoboken, NJ, USA, 357–371. <https://doi.org/10.1002/9780470588222.ch18>