Voice-based Conversational UX Studies and Design

Stuart Reeves
Martin Porcheron
Joel E. Fischer
Mixed Reality Laboratory
University of Nottingham, UK
{str,map,jef}@cs.nott.ac.uk

Heloisa Candello
Social Data Analytics Group
IBM Research Lab - Brazil
heloisacandello@br.ibm.com

Donald McMillan
Moira McGregor
Stockholm University, Sweden
{donald.mcmillan,
mcgregor}@dsv.su.se

Robert J. Moore
IBM Research-Almaden
San Jose, CA 95120, USA
rjmoore@us.ibm.com

Rein Sikveland
Loughborough University,
Loughborough, UK
r.o.sikveland@lboro.ac.uk

Alex Taylor
Centre for Human Computer
Interaction Design, City University,
London, UK
alex.taylor@city.ac.uk

Julia Velkovska
Moustafa Zouinar
Orange Labs, Sociology and
Economics Department (SENSE)
Paris, France
{julia.velkovska,
moustapha.zouinar}@orange.com

Abstract
Voice user interfaces are becoming ubiquitous, providing unprecedented opportunities to advance our understanding of voice interaction in a burgeoning array of practices and settings. We invite participants to contribute work-in-progress in voice interaction to come together to reflect on related methodological matters, social uses and design issues. This one-day workshop will be geared specifically to present and discuss methodologies for, and data emerging from ongoing empirical studies of voice interfaces in use and connected emerging design insights. We seek to draw on participants' (alongside organisers') contributions to explore ways of operationalising findings from such studies for the purposes of design. As part of this we will seek to identify what can be done to improve user experience, and consider creative approaches to how we might ameliorate the many challenges that are faced in the design of voice UIs.

Author Keywords
Voice interfaces; voice interaction design; conversational agents; speech; personal assistants.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

Voice UI, also called conversational UIs or conversational agents, have become mainstream technology: they are integrated into many modern smartphones, and are sold as standalone appliances for the home. Recent advances in hardware processing capabilities, coupled with improvements in machine learning techniques and the availability of significant data sets for fuelling those techniques, has led to increasing accessibility and commoditisation, meaning that broader design interest in 'conversational' interfaces has also flourished.

Although there is a body of prior work related to conversational interfaces (e.g., [3]), HCI has yet to respond substantially to this recent proliferation of voice and—aspirationally—conversational interactions. Not only is there a paucity of studies of conversational UI in actual use (particularly from fieldwork), but also a growing need to understand how findings from such practice-oriented empirical studies might feed into design in terms of guidelines, frameworks, and so on. Thus, this workshop seeks to directly address this by bringing together HCI researchers investigating one or both aspects. We aim to use the workshop as a way to share, and draw connections between the following:

- **qualitative studies** of different devices in use (existing consumer products, research prototypes, etc.) to help consolidate our understandings of user practices and experience;
- **emerging design knowledge** (e.g., principles, heuristics, frameworks, etc.) that can address and apply to future iterations of such conversational interfaces.

Background

The design and study of voice-based and conversational interaction is diverse and interdisciplinary, drawing upon work from many backgrounds including sociolinguistics [8], design and user experience [2], sociology [16] and computational paralinguistics [15]. Furthermore, the work lends itself to a myriad research approaches, including interviews [4], ethnomethodological and conversation analytic work [12, 17, 11, 10, 7], and Wizard of Oz and technical probe approaches [5]. In this workshop we particularly seek to encourage fostering an understanding of voice interaction *as it happens*, in order to generate design insights from observations of voice interaction *practices*. This practice-oriented approach, coupled with an interest in how such accounts map to design outcomes, is underexplored. Many challenges exist: how users develop new forms of talk that fit in with voice UI characteristics, how difficulties with transcription accuracy can be better handled through design, or how users deal with misunderstandings and unintended actions of voice-based interfaces. In addressing this and empirical studies of use, we are interested in a sense of design pragmatism: for instance, rather than relying on future technical improvements that seek to eliminate 'error' we are interested in errors and repairs as a fundamental feature of everyday talk, and how design can provide more resources for users of voice UI to repair and progress interactions.

This workshop brings together organisers from previous workshops at CHI and CSCW which explored the user experience of voice UIs [7] and designing for collaborative action [13]. In this workshop, we welcome both examinations of consumer devices (e.g., Amazon

Echo, Siri, Google Home), as well as the study and design of research prototypes and experimental systems. In terms of domains and situations of use, we are not only interested in established discretionary uses (such as interactions with household devices like the Amazon Echo), but also welcome consideration of design for settings where there are other constraints or challenges which present opportunities or requirements for voice interactions. For example, one might consider situations of use which can potentially benefit from hands-free interaction with technology, such as in-car [5], sterile environments [9], or during more mundane activities in the home, such as cooking [6]. These situations of use also connect with voice UI applications for accessibility [14].

As organisers, our intention is for the workshop to draw upon the strong interdisciplinary foundation both within the CHI community, and on the research of voice and conversational interface design in industry and academia. Through an open exchange of ideas and discussion, we will work to identify and position the role of HCI's more explicitly practice-oriented approaches for contributing to the design of voice UI, and set an agenda for how this interactional modality (both in its design and study) might be addressed in conceptually novel ways.

Organisers
By drawing upon a diverse range of organisers from a broad background of industry and academic practice, this workshop sets out to build upon recent research and workshop experiences at CHI '17 [7] and CSCW '17 [13] on conversational agents.

Stuart Reeves is EPSRC Fellow and Assistant Professor in the School of Computer Science, University of Nottingham, UK. He researches social and collaborative technologies, investigating how people use diverse kinds of interactive devices and systems in real world situations and places. As EPSRC Fellow he is exploring the connections between academic HCI research and the work of practitioners in UX and design professions.

Martin Porcheron is a PhD student in the Mixed Reality Lab at the University of Nottingham. His work examines the use of everyday technologies such as smartphones and VUIs in multi-party settings like pubs and the home from an empirical perspective. He has recently co-organised workshops at MobileHCI '16, CHI '16, CSCW '16 and '17 on topics such as collocated interaction and conversational interfaces.

Joel E. Fischer is Assistant Professor at the School of Computer Science where he teaches Design Ethnography, and member of the Mixed Reality Lab at the University of Nottingham. His practice-focused research in collocated interaction has been published at CHI, CSCW, UbiComp, ECSCW, and he has previously co-organised related workshops at CHI and CSCW, including recently on voice interaction [12].

Heloisa Candello is an interaction designer and a researcher at the IBM Research laboratory in Brazil. She has experience in leading and conducting design research activities to understand people's contexts and motivations to use conversation technologies. She recently co-organized a related workshop at CSCW '17 and previously published her research on UX with conversational systems at various HCI conferences [1, 2].

Moira McGregor is a PhD student in the department of computer and system sciences, at the university of Stockholm. Her research focuses on the use of mobile devices during interaction with co-present others, and has recently published work in CSCW on the use of conversational agents in business meetings.

Donald McMillan is an Assistant Professor (senior lecturer) in the Department of Computer and System Sciences, University of Stockholm, Sweden. His current research focus is on harnessing the interactional opportunities afforded by background conversation and environmental audio.

Robert J. Moore is a research scientist at IBM Research-Almaden, and has previously worked at Yahoo! Labs and Xerox PARC. He examines the intersections of ethnomethodology, Conversation Analysis and technology. He is currently developing a general methodology and framework for conversational UX design that incorporates empirical studies of natural human conversation and has recently co-organized a CHI workshop on the topic.

Rein Sikveland is a Research Associate at Loughborough University, UK. Rein's expertise is in conversation analysis and phonetics, and he applies conversation analytic research to communication skills training, currently as co-developer of the CARM-enterprise (Conversation Analytic Role-play Method). Rein also develops methods for upscaling conversation analytic research, including technological applications to call centre profiling (University of Oxford, UK).

Alex Taylor is a sociologist at the Centre for HCI Design, City, Univ of London. With a broad fasciation for the entanglements between social life and machines, his research ranges from studies of technology in daily life to speculative design interventions—both large and small. Across these realms, he draws on a feminist technoscience to ask questions about the roles human-machine composites play in forms of knowing and being.

Julia Velkovska is a sociologist at the Social Sciences Department of Orange Labs and associate researcher at EHESS in Paris. Drawing on ethnomethodology her research interests are in the study of relationships between social interaction and communication technologies in both personal and organizational settings (e.g.service encounters, collaboration in work settings, web based consumers' communities). Recent work includes analyses of interactions with 'conversational' technologies in customer relationship as well as in home settings.

Moustafa Zouinar is a researcher at Orange labs. His work is focused on understanding how people use and interact with information and communication technologies in various settings (home, work settings) with the practical goal of informing the design of these technologies. He is particularly interested in analyzing people's activities through observational methods (video / audio recordings) in everyday settings. He worked on different technologies (e.g., video-communication systems, social mobile applications, authentication systems) and is currently involved in a research project on voice assistants.

Website
A website will be set up at voiceux.wordpress.com, which, if the workshop proposal is accepted, will be

used to promote the workshop by publishing the call for participation and submission details online. We will also use this as a way of distributing news and updates relating to the workshop, and to maintain a list of (and files for) all accepted position papers.

Pre-Workshop Plans

The call for participation will be distributed in relevant academic mailing lists and through the organisers' existing social networks (e.g., via Twitter). We will also actively solicit submissions (via personal contacts and targeted invitations) from a broad range of researchers whose work relates to the workshop topic or who can make a significant contribution to the workshop.

Potential workshop participants should submit a 3–6 page position paper (including references) describing their interest and/or previous work related to the workshop. We will select papers based on: a) relevance to the workshop, b) quality of the submission, and c) the diversity of the participants. We will limit the size of the workshop to 15-20 people to ensure effective discussion in the second part of the workshop.

All selected papers will be available online on the website prior to the workshop, and participants will be expected to read all accepted submissions ahead of the workshop to help ground the discussion.

Workshop Structure

We are proposing a one-day workshop to present, reflect on, and enrich participants' work-in-progress. To reflect this commitment, we explicitly seek 'unfinished work', such as audio-visual data of voice interfaces in use, and early articulations of design thinking such as

ideas or prototypical frameworks, design cards, and heuristics.

Plan for the Day
The day will consist of three sections, the balance and timing of which will be dependent upon participants' submissions.

Section one will consist of introductions, including summaries of each participants' contribution to the workshop (e.g., audio-visual data, design considerations, etc.).

Section two will focus on collective exploration of a curated selection of data drawn from participants and organisers. We will run this (where appropriate) in the format of an organiser-curated 'data session'. The workshop will divide into various breakout groups (max 6 per group), who will spend time jointly examining a suitable fragment or fragments of data and documenting relevant issues revealed in that examination. For example: failures, 'bugfixing', repair and course-correction; coherence of dialogues / conversations; collaborative use; embodied and visual features of use (video data dependent); the role of simultaneously occurring activities; etc. At the end of section two the findings of each group will be reported back to the workshop. Organisers with more experience in 'data sessions' will take a stronger role in facilitating this section.

('Backup' material will be formulated by the organisers, drawn from their own collections of voice UI data fragments and emergent design principles to seed parts two and three where relevant.)

In **section three**, we will attempt to move findings to design issues by using an affinity diagramming approach to collate and map out possible design responses to issues raised. At this point, we will also bring in existing design guidelines, frameworks or principles from participants and organisers in order to validate, correct, extend, and amalgamate them. Bodystorming could then be used as part of this to 'prototype' expressions of these design considerations.

Resources: the workshop will require minimal resources, as the focus is on the presentation of participants' research, including audio/video data of interaction with voice UIs (*projector and speakers*), to guide discussion (*flipchart paper and pens*).

Post-Workshop Plans

Drawing upon the experience of our authors and our workshop participants, and depending on the outcomes of the workshop, an ideal outcome might focus on plotting out a *Communications of the ACM* magazine article that acts as a call-to-action and demonstrator of the value of synthesising HCI and speech technology / AI approaches.

Call for Participation

This one-day workshop will bring together researchers and practitioners who study and design voice or 'conversational' user interfaces. We invite participants to share and discuss what has been learned from practice-oriented empirical studies of voice interfaces *in use*, to understand design opportunities available for voice and conversation as an interaction modality. We specifically seek presentations of 'unfinished work' that would benefit from discussion and analysis in the workshop. We intend to make use of an interactive data

session to examine a selection of empirical data provided by organisers and participants, and to consider creative approaches to how we might ameliorate or rethink the many challenges that face voice interaction design. The emphasis of the day will be to establish a generative programme of work for turning findings from empirical studies into actionable design implications.

Interested workshop participants should submit a 3–6 page position paper in the CHI 2018 Extended Abstracts format (incl. references) describing their ongoing work related to the workshop including, but not limited to:

- **qualitative research** of voice interfaces in use, with an emphasis on practice-oriented approaches, such as audio-visual data 'vignettes' demonstrating aspects of voice interfaces in use;
- **design methods and approaches** for voice interaction grounded in empirical work.

We will select papers based on: a) relevance to the workshop, b) quality of the submission, and c) the diversity of the participants. We will limit the size of the workshop to 15-20 people to ensure effective and focussed discussion in the second and third parts of the workshop. At least one author of each accepted position paper must attend the workshop and that all participants must register for both the workshop and for at least one day of the conference.

More information about the workshop, organisers and submission is at voiceux.wordpress.com.

References

1. Heloisa Candello, Claudio Pinhanez, and Flavio Figueiredo. 2017. Typefaces and the Perception of Humanness in Natural Language Chatbots. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17). ACM, New York, NY, USA, 3476-3487. DOI: https://doi.org/10.1145/3025453.3025919

2. Heloisa Candello, Claudio Pinhanez, David Millen, and Bruna Daniele Andrade. 2017. Shaping the Experience of a Cognitive Investment Adviser. In *Design, User Experience, and Usability: Understanding Users and Contexts: 6th International Conference, DUXU 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part III*, Aaron Marcus and Wentao Wang (eds.). Springer International Publishing, Cham, 594–613. https://doi.org/10.1007/978-3-319-58640-3_43

3. Paul Luff, David Frohlich, and Nigel Gilber (eds.). 1990. *Computers and Conversation*. Academic Press.

4. 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

5. Nikolas Martelaro and Wendy Ju. 2017. WoZ Way: Enabling Real-time Remote Interaction Prototyping & Observation in On-road Vehicles. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '17). ACM, New York, NY, USA, 169-182. https://doi.org/10.1145/2998181.2998293

6. Sarah Mennicken, Oliver Zihler, Frida Juldaschewa, Veronika Molnar, David Aggeler, and Elaine May Huang. 2016. "It's Like Living with a Friendly Stranger": Perceptions of Personality Traits in a Smart Home. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '16), 120–131. https://doi.org/10.1145/2971648.2971757

7. Robert J. Moore, Raphael Arar, Guang-Jie Ren, and Margaret H. Szymanski. 2017. Conversational UX Design. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17). ACM, New York, NY, USA, 492-497. https://doi.org/10.1145/3027063.3027077

8. Dong Nguyen, A. Seza Doğruöz, Carolyn P. Rosé, and Franciska de Jong. 2016. Computational Sociolinguistics: A Survey. *Computational Linguistics* 42, 3: 537–593. https://doi.org/10.1162/COLI_a_00258

9. Kenton O'Hara, Gerardo Gonzalez, Graeme Penney, Abigail Sellen, Robert Corish, Helena Mentis, Andreas Varnavas, Antonio Criminisi, Mark Rouncefield, Neville Dastur, and Tom Carrell. 2014. Interactional order and constructed ways of seeing with touchless imaging systems in surgery. *Computer Supported Cooperative Work: CSCW: An International Journal* 23, 3: 299–337. https://doi.org/10.1007/s10606-014-9203-4

10. Hannah R.M. Pelikan and Mathias Broth. 2016. Why That Nao?: How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 4921–4932. https://doi.org/10.1145/2858036.2858478

11. Martin Porcheron, Joel Fischer, Stuart Reeves, Sarah Sharples. (unpublished) Voice User Interfaces in Everyday Life.

12. 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

13. Martin Porcheron, Joel E. Fischer, Moira McGregor, Barry Brown, Ewa Luger, Heloisa Candello, and Kenton O'Hara. 2017. Talking with Conversational Agents in Collaborative Action. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '17 Companion), 431–436.
https://doi.org/10.1145/3022198.3022666

14. François Portet, Michel Vacher, Caroline Golanski, Camille Roux, and Brigitte Meillon. 2013. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Pers Ubiquit Comput* 17, 127: 127–144.
https://doi.org/10.1007/s00779-011-0470-5

15. Björn Schuller, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christian Müller, and Shrikanth Narayanan. 2013. Paralinguistics in speech and language—State-of-the-art and the challenge. *Computer Speech & Language* 27, 1: 4–39. https://doi.org/10.1016/j.csl.2012.02.005

16. Julia Velkovska and Valérie Beauduoin. 2014. « Parler aux machines, coproduire un service. Intelligence artificielle et nouvelles formes de contribution du client dans les services téléphoniques de l'après-vente » in Kessous E., Mallard A., La Fabrique de la vente. Le travail commercial dans les télécommunications, Paris, Presses des Mines, p. 97-128.

17. Julia Velkovska and Moustafa Zouinar. 2017. "Making sense of interaction with smart assistants in home settings". IIEMCA 2017, July 10-13 – Westerville, Ohio, USA;
https://iiemca2017.wordpress.com/