Mapping Theoretical and Methodological Perspectives for Understanding Speech Interface Interactions

Leigh Clark
Benjamin R Cowan
Justin Edwards
University College Dublin
{leigh.clark;benjamin.cowan}@ucd.ie
justin.edwards@ucdconnect.ie

Cosmin Munteanu
Christine Murad
University of Toronto Mississauga;
University of Toronto
{cosmin;cmurad}@taglab.ca

Matthew Aylett
CereProc Ltd.
matthewa@cereproc.com

Roger K Moore
Sheffield University
r.k.moore@sheffield.ac.uk

Jens Edlund
Eva Szekely
KTH Stockholm
{edlund@speech;szekely@}.kth.se

Patrick Healey
Queen Mary
University of London
p.healey@qmul.ac.uk

Naomi Harte
Ilaria Torre
Trinity College Dublin
{nharte;torrei}@tcd.ie

Philip Doyle
Voysis Ltd.
pdoyle@voysis.com

ABSTRACT

The use of speech as an interaction modality has grown considerably through the integration of Intelligent Personal Assistants (IPAs- e.g. Siri, Google Assistant) into smartphones and voice based devices (e.g. Amazon Echo). However, there remain significant gaps in using theoretical frameworks to understand user behaviours and choices and how they may applied to specific speech interface interactions. This part-day multidisciplinary workshop aims to critically map out and evaluate theoretical frameworks and methodological approaches across a number of disciplines and establish directions for new paradigms in understanding speech interface user behaviour. In doing so, we will bring together participants from HCI and other speech related domains to establish a cohesive, diverse and collaborative community of researchers from academia and industry with interest in exploring theoretical and methodological issues in the field.

CCS CONCEPTS

• **Human-centered computing** → *Natural language interfaces*; *HCI theory, concepts and models*; User centered design;

KEYWORDS

speech interface, voice user interface, theory, method, design, intelligent personal assistants

BACKGROUND

Speech interfaces (aka Voice User Interface [VUIs]) are now commonplace, driven by the wide proliferation of IPAs like Amazon Alexa and Google Assistant in mobile and standalone home-based devices. Despite the technical knowledge underpinning speech interface technology advancing rapidly in recent years [3], research on the user side is limited by comparison [2]. Furthermore, work in the field is highly fragmented, both in the topics studied and the methods used to explore the user interaction [5]. Recent work has identified a significant gap in theoretically grounded, human-centered speech interface research [5]. Although work has highlighted the types of language choices made by users when interacting with speech interfaces [1, 7, 15], there is currently no consistent mapping or development of theory to explain these language choices in the literature. This understanding is fundamental to rigorous investigation of speech technology interactions; much like understanding motor movements and touch icon choice behaviour is to mouse, keyboard or screen-based interactions.

To resolve this we may need to look to other disciplines and existing theories in related fields. Across speech research in HCI, some theoretical concepts and methods have started to be borrowed from other disciplines to explain user behaviour and inform speech interface design [5]. For instance, concepts from the psycholinguistics literature such as perspective taking [8] and alignment (e.g. [9, 14]

have been applied to these interactions [7] as a way to explain and explore the cognitive mechanisms that underpin user behaviour. The concept of common ground and grounding (e.g. [4]) has also been explored to understand user behaviour in human-machine interactions (e.g. [10]). Elsewhere, politeness theory [6, 16] has also been used to motivate the design of new methods for speech interface output.

Studies that adopt these theoretically grounded approaches tend to borrow methods from the disciplines where these core concepts originate. Although this facilitates the comparison of findings to other research disciplines, the contexts and behaviours of speech interface interactions may not be directly relevant to the way these concepts are traditionally researched. It remains unclear what existing theories can be applied from human-human communication, how they can be tested and explored in an HCI domain and how they may need to be re-conceptualised for this type of interaction.

WORKSHOP AIMS & OUTCOMES

Our workshop builds on challenges identified in recent speech and voice workshops in the SIGCHI community (e.g. [11–13]) and aims to:

(1) Map and critically evaluate existing theoretical frameworks and concepts from across disciplines (e.g. linguistics, cognitive sciences, speech technology) and assess their ability to explain and explore speech interface interaction behaviour.
(2) Identify opportunities and establish directions for new paradigms and theories so as to explore speech interface interaction.
(3) Outline strengths and weakness of current methods used to explore these theories and paradigms.
(4) Identify key methodological challenges when applying these to understanding speech interface behaviour, and the need for new methods and data gathering paradigms.
(5) Scope the need for specific datasets and data types for exploring issues identified.
(6) Establish a cohesive, diverse and collaborative research group from academia and industry to build a community to target theoretical and methodological issues in this field.

Our workshop brings together world-leading researchers and industry representatives from a broad range of communities related to speech, dialogue, human-machine interaction, speech interface design and voice UX to bring a multidisciplinary approach in solving these issues. Through engaging across these communities, we aim to highlight the relevance and broaden the reach of speech interface work at CHI to other research communities (e.g. speech technology, linguistics, dialogue research, cognitive sciences), whilst also building a collaborative, diverse and cross-disciplinary speech community in CHI.

ORGANISERS

Leigh Clark is a Postdoctoral Research Fellow in the School of Information & Communication Studies at University College Dublin. His research examines the communicative aspects of user interactions with speech interfaces, how context impacts perceptions of computer speech and how linguistic theories can be implemented and redefined in speech-based HCI.

Benjamin R Cowan is an Assistant Professor at University College Dublin's School of Information & Communication Studies. His research lies at the juncture between psychology, HCI and computer science in investigating how theoretical perspectives in human communication can be applied to understand phenomena in speech based human-machine communication.

Justin Edwards is a graduate student in the School of Information & Communication Studies at University College Dublin. His research focuses on multitasking and interruptions as they relate to conversational interactions and subsequent design challenges for agent-initiated interactions.

Cosmin Munteanu is an Assistant Professor at the Institute for Communication, Culture, Information, and Technology at University of Toronto Mississauga. His research includes speech and natural language interaction for mobile devices, mixed reality systems, learning technologies for marginalized users, and assistive technologies for older adults, and ethics in HCI research.

Christine Murad is a graduate student at the Technologies for Aging Gracefully lab in the Department of Computer Science at the University of Toronto. Her research looks at the usability and design of conversational voice interfaces, and exploring design heuristics to aid intuitive and user-friendly conversational voice interaction.

Matthew Aylett is Chief Science Officer at a founder of CereProc Ltd which offers unique emotional and characterful synthesis solutions and has been awarded a Royal Society Industrial Fellowship to explore the role of speech synthesis in the perception of character in artificial agents.

Roger K Moore is Professor of Spoken Language Processing at the University of Sheffield. Much of his research has been based on insights from human speech perception and production. He was recently awarded the 2016 LREC Antonio Zampoli Prize for "Outstanding Contributions to the Advancement of Language Resources & Language Technology Evaluation within Human Language Technologies".

Jens Edlund is an Assistant Professor at KTH Royal Institute of Technology in Stockholm. His research combines linguistics and phonetics with engineering and speech technology. His published papers focus on many aspects of speech technology, with a particular focus on analysis of human behaviour in conversations and on the generation of humanlike behaviours in computers that speak.

Eva Szekely is a Postdoctoral Research Fellow at KTH Royal Institute of Technology in Stockholm. Her research focuses on modelling expressive and conversational phenomena in speech synthesis, and their application in situational contexts.

Patrick Healey is Professor of Human Interaction and leader of the Cognitive Science Research Group in the School of Electronic Engineering and Computer Science, Queen Mary University of London. His research analyses miscommunication; the processes by which people detect and recover from misunderstandings and its implications in designing technologies to support human interaction.

Naomi Harte is an Associate Professor in Digital Media Systems in the School of Engineering at Trinity College Dublin. She is a Co-Founder of the ADAPT Centre and holds a Google Faculty Award. Her specialist area is Human Speech Communication. Her work involves the design and application of algorithms to enhance or augment speech communication between humans and technology.

Ilaria Torre is a Marie Skłodowska-Curie postdoctoral research fellow in the Department of Electronic and Electrical Engineering at Trinity College Dublin. Her research analyses how speech and other cues influence trusting behaviour in human-machine interaction, drawing inspiration from games and behavioural economics.

Philip Doyle is a speech UX researcher at Voysis Ltd with experience constructing development sets for training AI and examining user language behaviour. His work focuses on understanding partner models in human-computer dialogue contexts.

WEBSITE

Details of the workshop will be hosted on the following website: http://speech-interaction.org/chi2019.

PRE-WORKSHOP PLAN

Attendees & Recruitment: We aim to bring together 20-25 participants from a diverse pool of researchers in academia and industry. The workshop will be advertised via mailing lists in the HCI (CHI List), speech technology (ISCA & SIGDIAL list) linguistics and dialogue (CUNY, CogSci, AMLaP & LINGUISTLIST lists) communities, internal institutional email, authors' social media and personal websites, and in-person within the networks of organising committee members. Participants will be selected based on the relevance of their position papers to the workshop's goals.

Submissions: Participants are asked to submit a 2 page position paper (CHI EA format) focusing on a proposed theory/concept/paradigm, the methods used to explore the paradigm and how this could be applied to understanding speech interface interaction. Participants are asked to submit these to the organisers via email at speechhci@gmail.com.

Important dates: [12th December 2018]: Call for participants released. [12th February 2019]: Position paper deadline. [1st March 2019]: Notification of acceptance.

Collaboration & Cohesion: For accepted authors and confirmed attendees we will set up collaborative resources for participants to use before, during and after the workshop: include a Zotero references group, a Collaborative Reading Repository (including all accepted papers) through Google Drive and a Workshop Slack group to help frame issues to explore before attending the workshop.

Table 1: Workshop structure

Order	Activity
14:30-14:40	Introductions
14:40-15:30	Presentations
15:30-16:45	Keynote speakers
16:45-17:15	Coffee break
17:15-18:15	User behaviours & theory mapping
18:15-19:15	Food & refreshments
19:15-20:15	Methods
20:15-20:30	Summary & future planning
20:30-late	Social event

WORKSHOP STRUCTURE

(1) **Introductions.** Brief introductions outlining workshop structure, setting collaborative tone, and reiterating workshop goals.

(2) **Participant presentations.** Participants will conduct 3 minute presentations of their papers, focusing on relevant theories, methods and applications around speech interfaces.

(3) **Keynote speakers: Prof. Holly Branigan and Prof. Michael McTear** Prof. Branigan will give a keynote presentation on dialogue theory use in human-machine dialogue. She is Professor of Psychology of Language & Cognition at the University of Edinburgh and a leading researcher in the field of psycholinguistics with experience of researching speech interfaces.
Prof. McTear will give a keynote about methods used through the speech domain to explore user behaviour in interaction. He is Emeritus Professor at Ulster University, Belfast and a leading researcher in the field of spoken dialogue systems, having written a leading book on the subject.

(4) **User behaviours & theory mapping.** Using recorded interactions with speech interfaces as data probes, groups will be asked to categorise and discuss causes of user behaviours identified. Groups will then be asked to reflect on theoretical perspectives to understand user behaviour. Taking the user behaviours identified above, the same groups will take the theories identified in presentations and discuss: 1) how they can be used to explain or explore user behaviours; 2) the challenges faced; and 3) the opportunities for new theory. From this we will produce a graphical theory challenge map depicting existing theory issues and opportunities for speech interface interaction.

(5) **Methods for testing hypotheses in speech interface design.** Participants will explore how to research the challenges identified earlier using their chosen methods. This will use the "Disney Brainstorming Method" - thinking styles of outsiders, dreamers, realisers and critics.

(6) **Workshop summary & future planning.** We will discuss the need for unique theoretical frameworks, methods and potential speech interaction datasets and how these need to be validated and applied in other application areas of speech interfaces. We will also plan how to communicate our ideas to others outside of our workshops, including to other workshops accepted at CHI 2019.

POST-WORKSHOP PLANS

After the CHI workshop, we aim to:

(1) Organise a special issue in the journal Human-Computer Interaction where participants will be encouraged to submit their work.

(2) Plan to organise further workshops on this topic targeting the annual Interspeech, SIGDIAL and CogSci conferences.

(3) Scope potential for a) a cross disciplinary conference series focused on speech and conversational interfaces, in collaboration with ACM and ISCA; b) CHI Special Interest Group on Speech and Language Interfaces.

CALL FOR PARTICIPATION

Speech as an interaction modality has grown through the integration of Intelligent Personal Assistants (IPAs) like Google Assistant and Siri into mobile and home-based devices. This collaborative workshop focuses on theories and methods we can use to understand user behaviours with speech interfaces.

Through the workshop we will establish what theories, concepts, and paradigms are important in understanding our interaction choices with speech interfaces, and potential new theoretical frameworks required to explore them. We will critically evaluate existing methodologies used to explore these interactions and determine priority areas for improving methods to explore the theories and paradigms identified.

This workshop brings together researchers and practitioners from any domain interested in understanding user interaction behaviours with speech interfaces. Interested researchers are invited to submit position papers outlining a theory, concept or paradigm and related research methods and highlight their importance in understanding user behaviour in speech interface interactions.

Position papers should be 2 pages maximum (excluding references) in the CHI EA format and submitted by February 12th. Papers will be reviewed by the organisers and chosen based on 1) relevance to the workshop; 2) quality of the contribution; 3) ability to contribute to discussion. Once accepted, at least one author must register for the workshop and at least one day of the CHI 2019 conference. Queries should be sent to speechhci@gmail.com. See the workshop website for more information.

Important Dates: [12th December 2018]: Call for participants released. [12th February 2019]: Position paper deadline. [1st March 2019]: Notification of acceptance.

ACKNOWLEDGMENTS

REFERENCES

[1] René Amalberti, Noëlle Carbonell, and Pierre Falzon. 1993. User representations of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies* 38, 4 (1993), 547–566.

[2] Matthew P. Aylett, Per Ola Kristensson, Steve Whittaker, and Yolanda Vazquez-Alvarez. None of a CHInd: relationship counselling for HCI and speech technology. ACM Press, 749–760. DOI:http://dx.doi.org/10.1145/2559206.2578868

[3] W. Chan, N. Jaitly, Q. Le, and O. Vinyals. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *ICASSP 2016* (2016-03). 4960–4964.

[4] H.H. Clark. *Using Language.* Cambridge University Press.

[5] Leigh Clark, Phillip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, and Benjamin Cowan. 2018. The State of Speech in HCI: Trends, Themes and Challenges. *Accepted with Interacting with Computers* (2018). http://arxiv.org/abs/1810.06828

[6] Leigh Clark, Abdulmalik Ofemile, Svenja Adolphs, and Tom Rodden. 2016. A multimodal approach to assessing user experiences with agent helpers. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 4 (2016), 29.

[7] Benjamin R Cowan, Holly P Branigan, Mateo Obregón, Enas Bugis, and Russell Beale. 2015. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human- computer dialogue. *International Journal of Human-Computer Studies* 83 (2015), 27–42.

[8] Susan R Fussell and Robert M Krauss. 1992. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of personality and Social Psychology* 62, 3 (1992), 378.

[9] Patrick GT Healey, Matthew Purver, and Christine Howes. 2014. Divergence in dialogue. *PloS one* 9, 6 (2014), e98598.

[10] S. Kiesler. Fostering common ground in human-robot interaction. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* (2005-08). 729–734. DOI : http://dx.doi.org/10.1109/ROMAN.2005.1513866

[11] Cosmin Munteanu, Pourang Irani, Sharon Oviatt, Matthew Aylett, Gerald Penn, Shimei Pan, Nikhil Sharma, Frank Rudzicz, Randy Gomez, Ben Cowan, and Keisuke Nakamura. Designing Speech, Acoustic and Multimodal Interactions. In *Proceedings of the 2017 CHI Conference Extended Abstracts* (2017) *(CHI EA '17)*. ACM, 601–608.

[12] Cosmin Munteanu, Matt Jones, Steve Whittaker, Sharon Oviatt, Matthew Aylett, Gerald Penn, Stephen Brewster, and Nicolas d'Alessandro. Designing speech and language interactions. In *CHI'14 Extended Abstracts* (2014). ACM, 75–78.

[13] Cosmin Munteanu and Gerald Penn. Speech-based Interaction: Myths, Challenges, and Opportunities. ACM Press, 992–995. DOI : http://dx.doi.org/10.1145/2851581.2856689

[14] Martin J Pickering and Holly P Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and language* 39, 4 (1998), 633–651.

[15] Martin Porcheron, Joel E. Fischer, and Sarah Sharples. "Do Animals Have A ccents?": Talking with Agents in Multi-Party Conversation. ACM Press, 207–219. DOI : http://dx.doi.org/10.1145/2998181.2998298

[16] Megan Strait, Cody Canning, and Matthias Scheutz. Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. ACM Press, 479–486.