

Understanding speech and language interactions in HCI: The importance of theory-based human-human dialogue research

Benjamin R. Cowan
HCI Centre
School of Computer Science
University of Birmingham
Edgbaston Campus
B15 2TT
b.r.cowan@cs.bham.ac.uk

ABSTRACT

With the growth of speech and language interactions in many commercially available devices and applications, it is timely that the HCI community should engage more fully with such a modality. This position paper proposes that HCI should utilize theory-based insight from dialogue research to develop in depth understanding about our behaviors in speech and language interactions with computers. The paper highlights an example of the use of dialogue research to understand effects in speech based HCI interactions and highlights that incorporating such research and its practices will lead to fundamental scientific understanding, that is beneficial for design and cross disciplinary collaboration with relevant human-human interaction disciplines.

Author Keywords

Speech HCI; Psycholinguistics; Alignment; Dialogue

ACM Classification Keywords

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

INTRODUCTION

More natural language orientated speech interfaces such as Siri and Google Now are now becoming commonplace in a range of devices. In addition speech and language as a human-computer interaction modality is of extreme importance in the growth of robotics and embodied conversation agent developments. These interfaces, compared to system initiative interactions, aim to act more

as dialogue partners than command based speech interfaces. As research interest in speech and language interactions in HCI grows it is the position of the author that we should see the interaction as a dialogue between user and computer, and develop theoretical insight as to why people behave the way they do in speech based human-computer dialogue (HCD). Armed with such knowledge, we can then more firmly understand how to design natural dialogue interactions in speech and language scenarios. Language based HCI research has embraced this view in the past [e.g. 10] and further co-operation with disciplines such as psycholinguistics and dialogue based researchers should be sought to grow the insights we have about HCD. An example of the outcomes of such a collaboration is used to show how such an approach can lead to important HCI based theoretical insights. The author believes that engraining such a collaborative approach into the speech and language HCI community will not only lead to solid scientific foundation and paradigms from which to grow research efforts but will also lead us to understand the challenges and impact of design decisions on the human side of speech interactions, as well as hold potential for cross disciplinary impact.

THEORY MOTIVATED RESEARCH

The position of the author is that we should aim to focus on the scientific study of user behaviour in HCD interactions, with design insight flowing from such work. The role of theoretical investigation in HCI has acted as an important foundation for the field in general, yet has become less prominent in the community of late. Recent efforts at CHI 2014 to promote such approaches through the Interaction Science spotlight [18] highlight a growing need for the type of research focus proposed. Interaction Science is mainly motivated by looking to understand, first and foremost, users' interactions with technology [12] rather than focusing on usability or design. However, that is not to say that design and usability cannot be informed by such a research approach [18]. On the contrary, a deep understanding of why users interact in certain ways with

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in TimesNewRoman 8 point font. Please do not change or modify the size of this text box.

these interfaces will likely lead to more effective spoken dialogue systems as well as innovative design interactions that manipulate the effects seen in scientific investigation.

HCD research has a tradition of conducting such studies, whereby theory-led research leads to benefits in system development. For instance work on user behavior in error resolution in spoken dialogue system interactions [21] highlighted that users adapt their speech after experiencing an error by using behaviors such as hyperarticulation, lexical modification of previous statements as well as stabilising prosody, using these as strategies to recover in the interaction (even though many things like hyperarticulation actually make the issue worse [22]). From such work the Computer-elicited Hyperarticulate Adaptation model (CHAM) was developed to explain such behaviors with this insight being used to improve the abilities of spoken language systems to deal with such errors.

Yet as well as identifying our own frameworks, the use of frameworks and knowledge from human-centered dialogue disciplines (such as psycholinguistics) should also be used to help us explain patterns of behavior in language interactions with technology, as well as giving us design insight. One such example is given in the following section with research by [7] also being a great example of HCI interactions being used to explore psychological theories of human-human interaction whilst also giving clues as to how design could impact the user in HCD. Work by [11] highlights there is scope to include such work in the study of human-computer dialogue, yet notes that there will be nuances to the use of human-human dialogue research in human-computer interactions based on fundamental differences in the assumptions of the user as to the nature of interaction. When conversing with humans we are conversing with a partner whereby negotiation about what terms to use can be easily achieved. The perceptions of human-computer dialogue lead the user to assume a uni-directional interaction with a partner whom they assume has little ability to negotiate [11], in effect an inflexible interlocutor. Researchers have also identified that studies of native language speaker-second language speaker dyads could also hold fruitful insights into the nature of behaviors apparent in human-computer dialogue [9]. Although considerations need to be taken on board about the nuances and idiosyncrasies of human-human and human-computer dialogue contexts, human-human dialogue research can be used as a strong theoretical guide in behavior interpretation and explanation in HCD scenarios.

INSIGHT FROM DIALOGUE RESEARCH

An example of how such research can be valuable to understanding human-computer dialogue is given based on the author's collaboration with psycholinguistic researchers at the University of Edinburgh. The project aims to identify the impact design and perceptions have on user language

behavior, specifically the tendency for users to converge lexically and syntactically with their partners.

It is well known in human-human dialogue research that we tend to converge on the syntax we use [6] as well as lexical choice [10,15] in dialogue. This alignment or convergence behavior comes from forming shared mental representations [9], with these shared mental representations being crucial to successful, effective and efficient dialogue [16]. Indeed this convergence is beneficial for both listener and speaker in that there is less chance for misunderstanding as well as reducing cognitive effort for the speaker [6].

Recent experimental research has highlighted this converge at syntactic and lexical levels in HCD interactions [7,8,9]. The prevalent view within HCD research is that the perceptions of the user in terms of their thoughts of the computers ability as a communication partner is important to many language behaviors in dialogue. Indeed the *mediated account* of alignment in human-human dialogue, where linguistic choices are influenced by speakers' beliefs about their interlocutors [7,10,14], also termed *audience design* [1] relies on the same premise that people choose their utterances based on their beliefs about what the listener will be able to understand [13,14]. This mental model can be based on assumptions about the presumed knowledge that communities of which the partner is assumed to be a member are likely to have [13,14], as well as assessment of their likely understanding based on previous language use in the interaction [7]. Aspects that are likely to affect users' perceptions of the partner's abilities as an interlocutor in HCD (such as design and speech system behavior or even greeting prompts [5]) could therefore impact alignment behavior in this interaction. For instance users may align more heavily with systems that they perceive as basic or limited in their abilities as dialogue partners (affected by system design or behavioral cues) so as to guarantee communication success.

There is no doubt that an element of human-computer dialogue behaviors are based on people's perceived limitations of the system, demonstrated by many studies on the encounter of errors [3,21] as well as more generally with user language behavior in HCD interactions where limited lexical choice, simplistic syntactic structures and short command like sentences are used [19,23]. In addition, recent research looking at the effect of user perceptions on lexical alignment levels, supports this view in that, not only did people align more with computer partners but that people aligned more with partners they were informed were basic compared to advanced in their abilities [7].

Yet this may not be the full story. Recent research by the author and colleagues at the University of Edinburgh using the referential communication task and confederate scripting paradigms commonly used in psycholinguistics suggests that this mediated mechanism may not be as influential in all cases. The research investigated the effect design (such as voice anthropomorphism) has on user

perceptions of the system as a dialogue partner and as such how this affects alignment behavior in interaction. The research shows that anthropomorphism does affect user assessment of automated partner's intelligence and capability as dialogue partners, in that people rated a more robotic voice as less competent, more basic and less flexible as a dialogue partner compared to a more anthropomorphic computer voice. This supports work that highlights that anthropomorphic artificial partners are seen as more intelligent and capable [20]. We also found a tendency for people to align with their partner's syntactic choice as well as their lexical choice in spoken human-computer dialogues, supporting previous research identifying the presence of such an effect in HCD interactions (for a review see [9]). However, contrary to effects predicted by a mediated account of HCD, there was no significant impact on alignment of the voices used. By using insight from psycholinguistics we can find a theory-based explanation for such a finding. The pattern found fits more towards an alternative *unmediated or priming account* of alignment by which people use the same grammatical and syntactic representations as their partner because they are more activated in the cognitive architectures involved in language processing and production. It seems therefore that the mechanisms that govern our human-computer dialogue behavior may not all be driven by our perceptions of our partner's abilities, with some (such as lexical and syntactic alignment behavior) governed more by priming.

That is not to say that these perceptions do not play a role. It may be that only when our partner's abilities become particularly salient (such as after an error, a previous poor experience with the same service, or where interaction success is critical) that alignment becomes mediated rather than unmediated in nature. The authors propose that alignment can show us that in fact our language behaviors in HCD are not purely guided by user perceptions and strategies to achieve communication success, but some are governed by the cognitive architectures that we use to process and create language. It may be that only when needed (such as in interactions where the computers limitations are highly salient as well as when success of interaction is critical) do we utilize our perceptions.

Such research is an example of how human-human dialogue work can be used to expand our theoretical view of HCD interactions and add nuance to our current view of user behaviors in speech and language interactions. This type of work not only has important theoretical implications but also has important messages for spoken dialogue system design. For instance in the case of alignment, if we understand the mechanisms and causal indicators of such behavior we can take this into account in the design of the spoken dialogue system, making recognition and comprehension more efficient. Indeed if we know by which conditions participants do take system perceptions into account then these can also be modeled and taken into account in the comprehension system. For instance, spoken

dialogue systems could include indicators of their design state and the behaviors encountered by the user so that the probability of user behaviors such as alignment of input with previous output can be expected and modeled. The scientific understanding that such work gives about how we interact with computer dialogue partners will lead to a more considered, informed and natural design within this modality. In addition such work can be used to inform theoretical frameworks in other related disciplines, as demonstrated in existing research by [7]. By embracing lessons, theoretical frameworks and research approaches from other related linguistic disciplines we can not only build on existing research but can also inform and break new boundaries in these other disciplines.

METHODOLOGICAL CONSIDERATIONS

It is important that we remain open to using methodological approaches common in other dialogue research. Of course consideration of the trade offs in methodological approaches between corpus-based, naturalistic data gathering and controlled dialogue experiments are needed (such trade offs are also apparent in human-human dialogue research). The freeform nature of corpora and naturalistic dialogue allows for a rich understanding of user behavior in specific scenarios encountered within that dialogue. Within the speech and language HCI community we need to embrace the value of laboratory experimentation and methods common in dialogue research in understanding the role design and system behavior plays in human-computer dialogue interactions. The use of controlled experimentation, as shown in work by [7] allows us to truly investigate the causal role specific aspects have in influencing our language choices in human-computer dialogue, by controlling potential confounds to behaviors being observed. The author feels that we should consider controlled experiment based language studies as a principal tool to produce causal, generalizable evidence of what aspects of human-computer dialogue affect user language behavior, which can then be observed and verified in more naturalistic settings. Indeed the cycle could be used the other way around, whereby aspects seen in natural corpora can be used as hypothesis generation for laboratory based explorations. To do this, experimental and corpus based methods currently used to develop insight into human-human dialogue such as in [4,6,17] should be embraced, experimented with and applied in human-computer dialogue contexts. This would allow us to build on an existing knowledge base and develop methodological common ground with highly related disciplines.

CONCLUSIONS

As speech and language interactions become more common and systems more natural and advanced, it is important that we strive to view speech and language interaction in HCI through the prism of existing dialogue research, exploring

theories and methods from these disciplines. Co-operating with disciplines such as psycholinguistics and dialogue based research will give us an existing set of foundations and paradigms from which to grow research efforts, as well as contributing to our understanding of the challenges of designing for the user in this type of interaction modality. By adopting such a view we also increase the opportunity for HCI based language work to have cross-disciplinary impact, forging new ground in current dialogue disciplines. It is the position of the author that we should first and foremost strive to fundamentally understand scientifically what makes users behave the way they do in HCD, and human-human dialogue research can help in this endeavour. From this, robust design insight will follow.

REFERENCES

1. Bell, A. Language style as audience design. *Language in Society* 13, 02 (1984), 145–204.
2. Bell, L., Gustafson, J., and Heldner, M. Prosodic adaptation in human-computer interaction. *Proceedings ICPHS 2003*, ISCA (2003), 2453–2456.
3. Bell, L. and Gustafson, J. Interaction with an animated agent in a spoken dialogue system. *Proceedings of the Sixth European Conference on Speech Communication and Technology*, ISCA (1999), 1143–1146.
4. Bortfeld, H. and Brennan, S.E. Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes* 23, 2 (1997), 119–147.
5. Boyce, S.J. Spoken Natural Language Dialogue Systems: User Interface Issues for the Future. In D. Gardner-Bonneau, ed., *Human Factors and Voice Interactive Systems*. Springer US, 1999, 37–61.
6. Branigan, H.P., Pickering, M.J., and Cleland, A. Syntactic co-ordination in dialogue. *Cognition* 75, (2000), B 13–25.
7. Branigan, H.P., Pickering, M.J., Pearson, J.M., McLean, J.F., and Brown, A. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121, 1 (2011), 41–57.
8. Branigan, H.P., Pickering, M.J., Pearson, J.M., McLean, J.F., and Nass, C. Syntactic alignment between computers and people: the role of belief about mental states. *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*, Erlbaum, Mahwah (2003), 186–191.
9. Branigan, H.P., Pickering, M.J., Pearson, J.M., and McLean, J.F. Linguistic alignment between people and computers. *Journal of Pragmatics* 42, 9 (2010), 2355–2368.
10. Brennan, S.E. and Clark, H.H. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 6 (1996), 1482–1493.
11. Brennan, S.E. The grounding problem in conversations with and through computers. *Social and cognitive psychological approaches to interpersonal communication*, (1998), 201–225.
12. Card, S. Interaction Science in the age of makers and instructables. *ACM SIGCHI*, (2012).
13. Clark, H.H. *Using Language*. Cambridge University Press, 1996.
14. Fussell, S.R. and Krauss, R.M. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology* 62, 3 (1992), 378–391.
15. Garrod, S. and Anderson, A. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27, 2 (1987), 181–218.
16. Garrod, S. and Pickering, M.J. Joint action, interactive alignment, and dialogue. *Topics in Cognitive Science* 1, (2009), 292–304.
17. Gries, S.T. Syntactic priming: a corpus-based approach. *Journal of psycholinguistic research* 34, 4 (2005), 365–399.
18. Howes, A., Cowan, B.R., Janssen, C., et al. Interaction Science Spotlight. *Proceedings of CHI 2014*, ACM (2014).
19. Kennedy, A., Wilkes, A., Elder, L., and Murray, W.S. Dialogue with machines. *Cognition* 30, 1 (1988), 37–72.
20. King, W.J. and Ohya, J. The representation of agents: anthropomorphism, agency, and intelligence. *Conference Companion on Human Factors in Computing Systems*, ACM (1996), 289–290.
21. Oviatt, S., Bernard, J., and Levow, G.A. Linguistic adaptations during spoken and multimodal error resolution. *Language and speech* 41 (Pt 3-4), (1998), 419–442.
22. Soltau, H. and Waibel, A. Specialized acoustic models for hyperarticulated speech. *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings*, (2000), 1779–1782 vol.3.
23. Zoltan-Ford, E. Reducing Variability in Natural-Language Interactions with Computers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 28, 9 (1984), 768–772.