Let's Talk about Dialogue: The Case of Alignment and Partner Modelling in Human-Computer Dialogue

Benjamin R. Cowan
School of Information &
Communication Studies
University College Dublin
Belfield
Dublin 4, Ireland
benjamin.cowan@ucd.ie

Abstract
As speech interfaces become more prominent, there is an urgent need for the HCI community to more fully understand the user dimension of this interaction paradigm. The author proposes that viewing speech interface interactions through the prism of dialogue research will aid in establishing speech research within HCI as well as lead to breakthroughs in understanding what governs our language choices in speech interface interaction. The paper highlights work where the use of psycholinguistic theory and the investigation of dialogue phenomena, specifically alignment in dialogue, has led to a more nuanced view of the role of partner models in speech based human-computer dialogue.

Author Keywords
Speech interfaces, psycholinguistics, dialogue research, human-computer dialogue, alignment

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

Introduction
Speech interfaces are now being used in a number of domains and on a number of devices. From traditional

forms of automated telephone services to personal virtual assistants such as Siri and Google Now built in to mobile and wearable platforms, speech is growing in popularity as an interaction modality. Such growth in prominence is only likely to continue as robotic and embodied conversation agents become more mainstream. As interest in speech and natural language interactions grows we as a discipline need to identify the directions to be taken in understanding the human side of such an interaction. It is the author's view that an opportunity exists to compliment works being conducted in the speech technology domain on recognition and synthesis by understanding how design and system characteristics (in terms of recognition and synthesis) affect user perceptions and language choices in interaction. This incorporates a number of research avenues, most notably understanding how these impact how we perceive our partner's abilities (i.e. build partner models) in interaction with speech interfaces, as well as how these perceptions affect dialogue phenomena in a speech interface context. This involves taking a more dialogue-centered view of such interactions, using the knowledge from wider dialogue based disciplines (e.g. pragmatics, psycholinguistics, sociolinguistics). By incorporating work from such domains we can better understand the nuances of speech interaction between user and computer, and use established theoretical insight as a base to understand user behaviours, choices and experiences in speech interface interactions. Of course, such knowledge will also be of fundamental importance in informing the design of natural language speech interfaces as we grow to understand just how the design decisions and system choices affect users in dialogue interaction.

It is worth noting that speech interface research has embraced this view in the past [e.g. 10] yet it has over the years fallen out of favour within the CHI community, gaining more traction within the core disciplines researching dialogue more generally. The author feels that this is a fundamental direction to take in understanding our interactions with speech interfaces and could lead us to make more natural speech interactions. The rest of the paper will focus on the author's findings from cross-disciplinary research across psycholinguistics and HCI domains and how this has led to theoretical insights in speech based interactions with automated partners. It is the authors view that focusing on such a collaborative cross-disciplinary research approach when studying speech interactions will lead the HCI community 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 user.

Insights From Dialogue Research- The Case of Alignment and Partner Modelling

The effect of convergence on syntax [2] and lexical choice [7,13] is well documented in psycholinguistic research. This convergence, or alignment, is important for successful, effective and efficient dialogue [14].

Recent research has highlighted that people also converge at the syntactic and lexical levels when interacting with automated partners [3–5]. Such an effect is proposed to be influenced by speakers' beliefs about their interlocutors as communication partners [4,7,12]. In other words, people form and adapt their speech choices based on what they believe the listener will be able to process in dialogue [8,12]. In dialogue research, these models are thought to be based on

assumptions about the presumed knowledge that communities of which the partner is assumed to be a member are likely to have [8,12], assessment of their likely understanding based on previous language use in the interaction [4] as well as our own levels of knowledge [11]. Although little is known about how we form our partner models in HCD, the impact of partner models in speech choices is echoed within HCD research, where speech choices are made based on people seeing an automated partner (either through identity or behavior) as a limited interlocutor, especially in comparison to a human.

Looking to research highlighting the importance of such models for alignment levels in HCD, the author and psycholinguistic researchers at the University of Edinburgh explored whether the design of the system impacted such partner models, and as such whether those models then affected both lexical and syntactic alignment levels. For instance, based on previous work [4], users may align more heavily with systems that include design components that lead people to perceive the speech interface as basic or limited. This would be evidence that the influence that design has on a person's partner model leads to an impact on the levels of alignment seen in HCD.

Our research manipulated the anthropomorphism of the synthesized speech used by the computer that users interacted with in a referential communication task (a typical task used in psycholinguistics to investigate alignment). We found that the type of synthesized speech used by the automated partner influenced partner models but that this did not influence alignment. In a pre-study people rated a *robotic* synthesized voice used in the communication task as less competent, more basic and less flexible as a dialogue partner compared to a more *anthropomorphic* synthesized voice. This corroborates previous work that suggests anthropomorphic robot partners are seen as more intelligent and capable [16]. People did align both syntactically [10] and lexically [9] in speech interactions with computer partners. However, there was no effect of partner modeling seen in the work. That is, people did not align to different levels depending on the synthesized voices used. Indeed there was no difference between the levels of alignment seen with computer partners than with human partners on both syntax [10] and lexical choice [9]. Such a pattern of results, rather than suggesting a role for partner modeling, can be explained more by priming whereby people use the same syntactic and lexical representations as their partner because they are more activated in the cognitive architectures involved in language processing and production.

This adds a layer of theoretical complexity to the view that we currently have of language choices in HCD. The author proposes that our language choices in HCD are not purely guided by user perceptions and adapted by strategies to achieve communication success based on these models. Rather, at points, we may be governed by the cognitive architectures that we use to process and create language. To be clear, that is not to say that these perceptions (i.e. our partner models) may never play a role in human-computer dialogue, or alignment in such an interaction. It may be that only when our partner's abilities are made obvious to us (such as after an error) or when it is critical for the user be successful in communicating efficiently with their partner (e.g. an emergency scenario) that alignment becomes more influenced by such effects. There is no doubt that at

certain times people's perceptions of an automated partner as a dialogue actor lead to adaptation. This is demonstrated by many studies on the encounter of errors [1,17] as well as more generally with user language choices in HCD interactions where limited lexical choice, simplistic syntactic structures and short command like sentences are used [15,18]. The HCI community has a large contribution to make to both theory and design insight in speech interface interaction by understanding under what circumstances such adaptation is driven by the model we have of an automated partner as a dialogue actor and what can be explained by other mechanisms shown to be used in human-human dialogue interaction, such as priming in the case of alignment. What's more, research into what causes us to form our partner models in speech interface interactions in the first place and how these develop over interaction(s) is limited, needing much further and more rigorous study. This is something that the author is currently working towards. Such research may be key to discovering and understanding the fundamental challenges of widespread and prolonged adoption of speech and natural language interactions.

Conclusions

With the popularity and prominence of speech as an interface modality growing it is important that we as a community engage fully with the variety of disciplines that allow us to understand this interaction. This includes, but is not restricted to, the speech technology community. It is the author's position that viewing speech interface interactions, be it personal virtual agents on wearable or mobile devices, robot interactions or automated telephone services, through the prism of dialogue research will lead to breakthroughs in our theoretical and design based understanding of speech interface interactions. The research highlighted above serves as an example of how existing dialogue research can be used to add nuance to our current view of user speech choices in such interactions. Working closely with other dialogue based researchers and using this type of view will allow us to build on a foundation of existing theories and accounts as well as influence those foundational theories. As in the research showcased in this paper, the role of design in impacting our speech choices, as well as the dialogue phenomena that appear both in human-human and human-computer dialogue can be identified which in turn can feed into the development of more effective speech technology.

References

1. Linda Bell and Joakim Gustafson. 1999. Interaction with an animated agent in a spoken dialogue system. *Proceedings of the Sixth European Conference on Speech Communication and Technology*, ISCA, 1143–1146.
2. H. P. Branigan, M. J. Pickering, and A. Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition* 75: B 13–25.
3. H. P. Branigan, M. J. Pickering, J. M. Pearson, and J. F. McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42, 9: 2355–2368.
4. H. P. Branigan, M. J. Pickering, J. M. Pearson, J. F. McLean, and A. Brown. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121, 1: 41–57.
5. H. P. Branigan, M. J. Pickering, J. M. Pearson, J. F. McLean, and C. Nass. 2003. 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, 186–191.

6. S.E. Brennan. 1998. The grounding problem in conversations with and through computers. *Social and cognitive psychological approaches to interpersonal communication*: 201–225.
7. S. E. Brennan and H. H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 6: 1482–1493. http://doi.org/10.1037/0278-7393.22.6.1482
8. H. H. Clark. 1996. *Using Language*. Cambridge University Press.
9. B. R. Cowan and H. P. Branigan. In Press. Does voice anthropomorphism affect lexical alignment in speech-based human- computer dialogue? *Proceedings of Interspeech 2015*, ISCA.
10. B. R. Cowan, H. P. Branigan, E. Bugis, M. Obregon, and R. Beale. In Press. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic alignment in human-computer dialogue. *International Journal of Human-Computer Studies*.
11. Susan R. Fussell and Robert M. Krauss. 1991. Accuracy and bias in estimates of others' knowledge. *European Journal of Social Psychology* 21: 445–454.
12. 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: 378–391. http://doi.org/10.1037/0022-3514.62.3.378
13. S. Garrod and A. Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27, 2: 181–218. http://doi.org/10.1016/0010-0277(87)90018-7
14. S. Garrod and M. J. Pickering. 2009. Joint action, interactive alignment, and dialogue. *Topics in Cognitive Science* 1: 292–304.
15. Alan Kennedy, Alan Wilkes, Leona Elder, and Wayne S. Murray. 1988. Dialogue with machines. *Cognition* 30, 1: 37–72. http://doi.org/10.1016/0010-0277(88)90003-0
16. William Joseph King and Jun Ohya. 1996. The representation of agents: anthropomorphism, agency, and intelligence. *Conference Companion on Human Factors in Computing Systems*, ACM, 289–290. http://doi.org/10.1145/257089.257326
17. S Oviatt, J Bernard, and G A Levow. 1998. Linguistic adaptations during spoken and multimodal error resolution. *Language and speech* 41 (Pt 3-4): 419–442.
18. Elizabeth Zoltan-Ford. 1984. Reducing Variability in Natural-Language Interactions with Computers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 28, 9: 768–772. http://doi.org/10.1177/154193128402800905