



This is a repository copy of *Interactivism in spoken dialogue systems*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/210103/>

Version: Submitted Version

Preprint:

Muñoz, T.R., Ip, E.Y.J., Huang, G. orcid.org/0000-0002-0266-6987 et al. (1 more author)
(Submitted: 2022) *Interactivism in spoken dialogue systems*. [Preprint - arXiv] (Submitted)

<https://doi.org/10.48550/arXiv.2209.13547>

© 2022 The Author(s). This preprint is made available under a Creative Commons Attribution 4.0 International License. (<https://creativecommons.org/licenses/by/4.0/>)

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Interactivism in Spoken Dialogue Systems

Teresa Rodríguez Muñoz*, Emily Ip*, Guanyu Huang*, and Roger K. Moore*

*Department of Computer Science, The University of Sheffield

*{trodriguezmunoz1, eyjipl, ghuang10, r.k.moore}@sheffield.ac.uk

Abstract

The interactivism model introduces a dynamic approach to language, communication and cognition. In this work, we explore this fundamental theory in the context of dialogue modelling for spoken dialogue systems (SDS). To extend such a theoretical framework, we present a set of design principles which adhere to central psycholinguistic and communication theories to achieve interactivism in SDS. From these, key ideas are linked to constitute the basis of our proposed design principles.

Keywords: Spoken Dialogue System, interactivism, incremental dialogue, transactional model.

1 Introduction

In recent years, with the exponential growth of speech technologies such as Siri and Alexa, users have grown accustomed to the rigid dialogue schemes these devices offer. Thus, current human-robot interactions (HRI) are far from being conversational (Moore et al., 2016). To optimise the effectiveness of HRI dialogues, researchers have worked on the accuracy of Automatic Speech Recognition (ASR), the naturalness of Text-to-Speech (TTS) modules and alternative dialogue frameworks. Incremental dialogue systems are one such alternative to attaining natural timing in conversation (Schlangen and Skantze, 2011).

However, the quality of spoken interactions goes beyond increasing ASR accuracy and delivering timely responses. One must also have a better understanding of dialogue as a process, which is not linear, but rather, transactional (Pierce and Corey, 2009). Moreover, dialogue involves continuous, bi-directional interactions between the conversational agent, the contextual environment and the interlocutor via verbal and non-verbal signals (Moore, 2016). Hence, in this work, we suggest that high-performing SDS must embrace interactivism to demonstrate situational and social awareness.

2 Background Theories

2.1 Interactivism: Dialogue as a Process

Interactivism has favoured frameworks of process over models of substance (Bickhard, 2009). This idea may translate to dialogue modelling, as dialogue is the process whereby ideas are exchanged among multiple social actors. It is also suggested that dialogue modelling is only one part of the interconnected modular system of an ‘intelligent’ agent, which coexists in a given environment (Maturana and Varela, 1987). Thus, the system behaves in an enactivist manner as a continuously and autonomously self-producing autopoietic entity (Moore, 2016). Therefore, future SDS design efforts should include *autopoiesis* (Maturana and Varela, 1987) in the form of self-monitoring of the system’s output (i.e., utterances produced by the agent) and its own current status. This has been attempted in incremental dialogue frameworks (Skantze and Schlangen, 2009; Schlangen and Skantze, 2011), where self-monitoring feedback loops between the Contextualiser and Dialogue Modelling modules are employed to self-monitor, self-repair and monitor the user’s speech production and non-verbal feedback signals.

Consequentially, the proposed interactivist framework presents implications for language and its use in human interaction. The framework is inherently social and interactional. Thus, the conversational agent needs to be imbued with the knowledge of these conventions to respond appropriately. Furthermore, future SDS cannot exclusively take in human responses in isolation; they must establish an awareness of the environment as it evolves through such an interaction and makes changes to itself accordingly (Moore, 2016).

2.2 Transactional Model of Communication

As an interactive process, spoken dialogue could be seen as transactional (Pierce and Corey, 2009).

In comparison with linear and interactive dialogue models, the transactional model is the most dynamic. It considers dialogue as a cooperative process in which interlocutors exchange messages simultaneously. The dialogue is built upon shared experiences in culture, language and/or environment, allowing one to use less speech or even a single sound to achieve a successful interaction (Hawkins, 2003). We propose that the transactional model may be the most preferred paradigm to achieve incremental dialogue. The conversational agent needs to be attentive and adaptive, which requires it to be able to adjust its language behaviours according to changes in the user and the environment.

How can we make this adjustment happen? The inner workings of the brain reveal that thought is not linear; it is a process in which we constantly produce and shape ideas (Clark, 2014). While the brain receives information from external resources, it tries to piece things together in a bottom-up way. It also attempts to guess incoming sensory data on the basis of what it knows about how the world is likely to be in a top-down manner. This Predictive Processing allows the agent to anticipate which actions to take, and to adjust its prediction upon its perception of embodied and environmental information (Clark, 2015). Hence, a predictive function should be employed to enrich SDS design.

2.3 Adjustment of Dialogue Behaviours

In tandem with these *intrapersonal* adjustments within the system, *interpersonal* dynamics evolve as well. Entrainment describes how interlocutors become more similar to each other in their speech throughout a conversation (Levitan, 2013), as speakers' conversational behaviour tends to be influenced by that of the other interlocutor. For this reason, effort-based models (Lindblom, 1990; Moore and Nicolao, 2017) have been developed to account for humans' regulatory behaviour in everyday speech. These models consider what the speaker and the listener(s) share in common to adjust that effort. This closely relates to the Theory of Mind, which involves the ability to discern the mental states, including emotions, knowledge and beliefs of oneself and others (Woodruff and Premack, 1978). A conversational agent with such capabilities would be able to adjust to the user and build a rapport. This type of closed-loop dialogue system facilitates adaptive behaviours which are emergent, and not choreographed.

Furthermore, Gricean pragmatics can be considered to achieve more interactive behaviours in SDS. To more closely emulate human dialogue, Grice's maxims dictate that one should be as informative, concise, and relevant to the discussion as possible (Grice, 1975). Moreover, ostensive communication (i.e., communication involving the expression and recognition of intentions via verbal or non-verbal cues) should be further explored in SDS design (Scott-Phillips, 2017). This would allow the conversational agent to be aware of the multimodal protocols to open and close channels of communication and engagement.

3 Design Principles for Transactional SDS

Based on the theoretical background discussed, several design principles have been suggested below:

- (i) **Conversational agents must have an incremental dialogue framework:** incremental SDS employ self-monitoring feedback loops to perform revisions on the system's output (either covertly or overtly) and determine whether an utterance was spoken, interrupted by the user or revoked as a failed hypothesis.
- (ii) **Agents should adjust their communicative effort in dialogue:** if the conversational agent can identify the user's abilities and adjust itself, then it could achieve autonomous, progressive learning of a user incrementally.
- (iii) **Conversational agents must be aware of the context of the conversation and have memory of past interactions with a user.** Short- and long-term information needs are essential in natural conversation.
- (iv) Further design of **multi-party SDS should consider ostensive behaviour** when engaging with different social actors.

4 Conclusion and Future Work

This work has briefly discussed interactivism and other central psycholinguistic and communication theories to improve the performance of current SDS. We have identified key design principles that the community may employ to design novel conversational agents, founded on the interactivist theoretical framework. Future work will involve the study of incremental dialogue systems and adapting them to align with the principles identified in this work.

Acknowledgements

This work was supported by the Centre for Doctoral Training in Speech and Language Technologies (SLT) and their Applications funded by UK Research and Innovation [EP/S023062/1]. Moreover, this work was partially supported by The Fulbright University of Sheffield Postgraduate Award.

References

- Mark H Bickhard. 2009. Interactivism: A manifesto. *New Ideas in Psychology*, 27(1):85–95.
- Andy Clark. 2014. Perceiving as predicting. *Perception and its modalities*, pages 23–43.
- Andy Clark. 2015. *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.
- Herbert P Grice. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.
- Sarah Hawkins. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of phonetics*, 31(3-4):373–405.
- Rivka Levitan. 2013. Entrainment in spoken dialogue systems: Adopting, predicting and influencing user behavior. In *Proceedings of the 2013 NAACL HLT Student Research Workshop*, pages 84–90.
- Björn Lindblom. 1990. Explaining phonetic variation: A sketch of the h&h theory. In *Speech production and speech modelling*, pages 403–439. Springer.
- Humberto R Maturana and Francisco J Varela. 1987. *The tree of knowledge: The biological roots of human understanding*. New Science Library/Shambhala Publications.
- Roger K Moore. 2016. Introducing a pictographic language for envisioning a rich variety of enactive systems with different degrees of complexity. *International Journal of Advanced Robotic Systems*, 13(2):74.
- Roger K Moore, Hui Li, and Shih-Hao Liao. 2016. Progress and prospects for spoken language technology: What ordinary people think. In *INTER-SPEECH*, pages 3007–3011. San Francisco, CA.
- Roger K Moore and Mauro Nicolao. 2017. Toward a needs-based architecture for ‘intelligent’ communicative agents: Speaking with intention. *Frontiers in Robotics and AI*, 4:66.
- Teresa Pierce and Amy M Corey. 2009. *The evolution of human communication: From theory to practice*. EtrePress.
- David Schlangen and Gabriel Skantze. 2011. A general, abstract model of incremental dialogue processing. *Dialogue & Discourse*, 2(1):83–111.
- Thomas C Scott-Phillips. 2017. Pragmatics and the aims of language evolution. *Psychonomic Bulletin & Review*, 24(1):186–189.
- Gabriel Skantze and David Schlangen. 2009. Incremental dialogue processing in a micro-domain. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 745–753.
- Guy Woodruff and David Premack. 1978. Does the chimpanzee have a theory of mind. *Behavioral and Brain Sciences*, 4(1):515–526.