Revisiting Habitability in Conversational Systems

A. Baki Kocaballi
Centre for Health Informatics
Macquarie University
baki.kocaballi@mq.edu.au

Enrico Coiera
Centre for Health Informatics
Macquarie University
enrico.coiera@mq.edu.au

Shlomo Berkovsky
Centre for Health Informatics
Macquarie University
shlomo.berkovsky@mq.edu.au

CHI'20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.3383014

Abstract

Conversational systems are inherently disadvantaged when indicating either what capabilities they have or the state they are in. The notion of habitability, the appropriate balancing in design between the language people use and the language a system can accept, emerged out of these early difficulties with conversational systems. This literature review aims to summarize progress in habitability research and explore implications for the design of current AI-enabled conversational systems. We found that i) the definitions of habitability focus mostly on matching between user expectations and system capabilities by employing well-balanced restrictions on language use; ii) there are two comprehensive design perspectives on different domains of habitability; iii) there is one standardized questionnaire with a sub-scale to measure habitability in a limited way. The review has allowed us to propose a working definition of habitability and some design implications that may prove useful for guiding future research and practice in this field.

Author Keywords

Conversational interfaces; conversational agents; chatbots; habitability; evaluation; design.

CSS Concepts

• **Human-centered computing~Human computer interaction (HCI)** → Natural language interfaces

Introduction

Conversational systems are the systems that employ user interfaces supporting conversational interaction using speech or other modalities [21]. Despite having advantages in providing users with ways to interact with interactive systems in their natural language, conversational systems, especially the voice-only ones, are inherently disadvantaged in terms of presenting the capabilities of a system and its status. The visibility of system status, and what is possible or impossible to do at any stage of interaction with the system are essential to build common ground [3, 5]–mutual knowledge required for successful communication–and improve system usability [25]. This paper aims to address the visibility problem by examining the prior work with a focus on a relevant design concept of *habitability*. Watt defined a habitable computer language as "one in which its users can express themselves without straying over the language's boundaries into unallowed sentences" [31]. While a minimally habitable system is limited to understanding only a single way to express an idea, question or meaning, a fully habitable system can accept various unconstrained ways [13, 31]. Since understanding every utterance is a hardly achievable task [11], the work on habitability has focused on applying restrictions on language use and facilitating effective error recovery strategies.

Habitability as Visibility

Moore explains that the capabilities of conversational systems often may not be easily perceived by users, causing a mismatch or habitability gap between user expectations and system capabilities [23]. Narrowing down this habitability gap involves making visible the hidden system capabilities and limitations through effective feedback and visibility methods [26, 27]. Hone and Baber draw analogies between the concept of visibility in graphical user interfaces (GUIs) and that of habitability in conversational interfaces [13]. They point out that conversational interfaces need a design concept equivalent to the visibility of GUIs and suggest that habitability can be a suitable means to express the visibility in conversational user interfaces, ensuring a good match between users' conceptual model of the system and the actual system.

Although the concept of habitability was introduced in the late 1960s, the research interest on the concept has not been sustained. Hone and Baber claim that the rapid advancements in direct manipulation GUIs resulted in a lack of interest in conversational interfaces [13]. But, the recent developments in artificial intelligence technologies, improvements in processing power, and availability of large amounts of data have triggered more interest in conversational interfaces [21]. In parallel to this increasing interest, it may prove useful to revisit the concept of habitability and obtain some insights from previous research studies. To this end, this paper provides an understanding of the current state of research on habitability and draws attention to this neglected and potentially useful concept by performing a literature review. Our aim is to i) develop a working definition of habitability for the current AI-enabled conversational systems, ii) examine habitability evaluation methods; and iii) derive some design implications that may prove useful when

A Habitability Subscale

The following assessment items on habitability are from the SASSI questionnaire [14].

- I sometimes wondered if I was using the right word.
- I always knew what to say to the system.
- I was not always sure what the system was doing.
- It is easy to lose track of where you are in an interaction with the system.

designing and evaluating current conversational systems.

Methods

We performed a search on the ACM Digital Library, SCOPUS, Web of Science, and IEEE Xplore databases with the following keywords either in the paper titles or abstracts: ("habitability" OR "habitable") AND ("voice" OR "speech" OR "conversation" OR "dialogue" OR "chatbot"). Our analysis focused on how habitability was defined and evaluated, and what design principles were employed to support habitability.

Results

The search resulted in 203 papers. The studies without any conversational interfaces or without a consideration of habitability as a design factor were excluded. The final review list included 12 studies across different domains including automotive, banking, book ordering, information retrieval, and travel.

Definitions of Habitability

Our search revealed few studies considering habitability as a factor in conversational system design. Although habitability definitions varied (see Table 1), we identified several important themes including: a right match between user expectations and system capabilities [13, 15, 23, 26, 27], well-balanced language restrictions [1, 7, 16, 26, 27], the degree of support to expressive range of users [7, 26, 27], robust error recovery [24, 32], predictable system behavior [30], and shared dialog control [9]. Based on these common themes, we will offer a working definition of habitability in the discussion section.

Designing for Habitability

Two studies provided comprehensive perspectives on designing habitable interfaces [13, 26]. Ogden and Bernick proposed four domains to characterize habitability: conceptual (objects and actions coverage), functional (expressive coverage), syntactic (paraphrase coverage), and lexical (vocabulary coverage) [26]. A successful user-system interaction requires i) systems to make visible their capabilities in all four domains, and ii) users to learn to stay within the capabilities communicated. Drawing on Ogden and Bernick's earlier work, Hone and Baber offered a design perspective on habitability in terms of different types of constraints operating on user utterances [13]. This perspective included five constraints: semantic (on entire interaction), dialogue (between pairs of utterances), syntactic (within utterance), lexical (single words), and recognition (single phonemes, or words/phrases). The remaining studies offered design principles focusing on feedback and clarification methods [1, 7, 16, 23], system restrictions and constraints [1, 13, 16], a system's awareness of its performance [9], and predictability of system behavior [32].

Evaluating Habitability

In terms of evaluation, only one study focused on the direct evaluation of habitability by developing a standardized questionnaire referred to as the Subjective Assessment of Speech System Interfaces (SASSI) with a subscale to evaluate habitability [14] (see the sidebar). While one study performed a qualitative assessment of user errors according to the four domains of constraints on habitability [13], another study quantitatively assessed effectiveness, helpfulness, and perceived difficulty as factors supporting habitability [7].

Study	Domain	Habitability Description
Moore 2017	Voice User Interfaces	Habitability requires matching capabilities and expectations of users with the features and benefits provided by systems.
Damljanović et al. 2013	Information Retrieval	"How easily, naturally and effectively users can use language to express themselves within the constraints imposed by the system. If users can express everything they need for their tasks, using the constrained system language, then such a language is habitable."
Epstein et al. 2010	Book Ordering	"Habitable dialogue requires mixed initiative, where the user and the system share control of the path the dialogue takes."
Ogden et al. 2008; Ogden & Bernick 1997	Various	"How easily, naturally, and effectively people can use language to express themselves within the constraints of a system language." "Habitable systems enable users to express everything that is needed for a task using language they think the system will understand." "A language's habitability depends on how well it matches user knowledge about the domain of discourse."
Hone & Baber 2001	Banking	"Habitability refers to the match between the language people employ when using a computer system and the language that the system can accept."
Hone & Graham 2000	Voice Interface Evaluation	"A habitable system may be defined as one in which there is a good match between the user's conceptual model of the system and the actual system". "Habitability refers to the extent to which the user knows what to do and knows what the system is doing".
Jönsson 1997; Ahrenberg 1996	Automotive and Travel	"Models that correctly and efficiently handle those phenomena that actually occur in human-computer interaction without having the user feel constrained or restricted when using the interface." "The user should conveniently be able to express the commands and requests that the background system can deal with, without transgressing the linguistic capabilities of the interface."
Morin & Jungua 1993	Generic User Interfaces	Habitability appears to be associated with error robustness and ability to provide assistance while improving the naturalness of user-system dialogues.
Young & Proctor 1989	Information Retrieval	Key ability of a habitable system is to be able to recover from recognition errors effectively.
Trost et al. 1987	Information Retrieval	"Very important though is that the system behaves in a predictable way, i.e. is habitable, so that users can learn the types of acceptable queries very fast."

Table 1. The summary of the definitions of habitability in the reviewed studies

Discussion

Although there have been notable advancements in synthetic speech generation, word error rates, and intent recognition, the current generation of conversational systems is still very limited in understanding the meaning at sentence-level, managing conversations over multiple turns, recovering from errors, and explicating system capabilities and limitations clearly [29]. Therefore, most of the concerns of the earlier conversational systems about habitability are also applicable to the current systems. However, the problem of visibility of the system capabilities has become more critical as the gap between the people's expectations and the capabilities of the AI technologies widens due to the ways in which media reports on the developments in AI technologies [12] and the use of more human-like synthetic voices suggesting higher conversational competence [6]. Since the current conversational systems' capabilities are still limited, the habitability gap is persistent [20]. Therefore, habitability as a design goal is relevant to the current generation of conversational systems.

Drawing on the prior definitions and principles, we have developed a working definition of habitability: *a habitable conversational system is the one in which system and interface capabilities and user expectations are aligned, and adequate accounts of system and interface behavior are provided* (Figure 1). To provide a deeper understanding of the key themes in this definition, the next section will present some high-level design implications. They have been derived from the reviewed studies and also some other recent studies (not explicitly engaging with the notion of habitability but still highly relevant, e.g., [28, 29]).

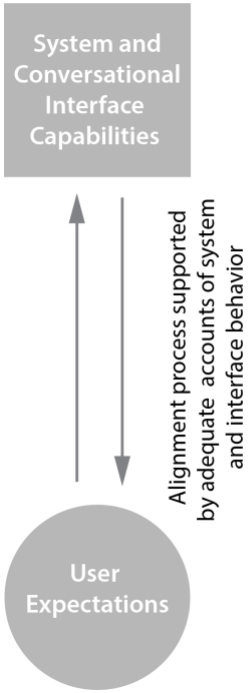

Figure 1. The alignment process towards a habitable conversational system in which system and interface capabilities and user expectations are aligned, and adequate accounts of system and interface behavior are provided.

Design Implications

Habitability is not an instantaneous quality of user-system interaction. The expectation that a user can find a conversational system instantly habitable is very unlikely to be achieved. Habitability requires an ongoing effort of both the user and the system [11]. Therefore, in the first interaction with the system, it is important to communicate with users that although they can use their speech to interact with the system, a training process is needed to explain the system capabilities and limitations and the constraints on the language use. An alignment process between the user and the system is required. Parallel to this understanding, developing a process-oriented approach to evaluating habitability may prove useful (e.g., [17]).

Habitability requires establishing common ground. In a habitable conversational interface, the system capabilities and user expectations need to be aligned. Although training is useful in the first interaction, it may not be possible to align the capabilities and expectations at once. Thus, a dedicated expectation management component that can gradually tune the expectations of users according to system capabilities might be needed. On the user side, this process requires users to play a more active role in tuning by adjusting their use of language in the early phases of their interaction with the system. On the system side, the system should analyze system-user interactions to increasingly accommodate the user's ways of using language over time through personalized conversational styles [18]. Therefore, this process of alignment is a mutual process requiring a clear communication of the alignment process with the users. This is key to establishing common ground [3, 5].

Habitability requires giving accounts of system and interface behavior. Since visibility is a major limitation of conversational systems, mechanisms are needed to make system behavior visible. However, a distinction should be made between system visibility and interface visibility. While the system visibility refers to the visibility of the system's capabilities (i.e., what the system can or cannot do), the interface visibility refers to the visibility of the conversational interface capabilities (i.e., what are the restrictions on the language use). Both types of visibility are important for habitability.

The constraints-based habitability model offered by Hone and Baber is useful to deal with both system and interface visibility [13]. The model has five different domains of constraints operating on user utterances: semantic, dialogue, syntactic, lexical, and recognition. While the semantic domain is concerned with the system capabilities, the other four domains focus on the restrictions on the language use. If a speech system can explain to users which of the five language domains is responsible for an error, this allows users to recover from errors more effectively and develop more accurate expectations of the system capabilities. For example, if it is an error at the lexical level, the system may suggest a different term for the out-of-vocabulary word; or if the error is at the syntactic level, the system may suggest rephrasing the question; or if it is a recognition level error, the system may recommend speaking louder, slower or with fewer pauses.

Two design approaches may prove useful for implementing 'visibility' features in conversational systems over multiple habitability domains: seamful design [2] and the notion of accounts [8]. While seamful design advocates the deliberate use of seams–

Acknowledgements

This work is supported by the National Health and Medical Research Council (NHMRC) grant APP1134919 (Centre for Research Excellence in Digital Health) and Programme Grant APP1054146.

gaps and breaks in functionality or system components–as a resource for action, the notion of accounts advocates the provision of a reflexive and situated information on the system behavior. Rather than trying to hide the complexities involved in different domains of habitability through various abstractions, systems can embrace the idea of seamful integration of those domains, which can allow users to understand a problem or error without needing additional feedback. For example, when there is a problem with the engine of a car, the engine makes various noises, suggesting a potential problem with the engine without necessarily requiring a warning signal on the dashboard. A promising research direction might be to explore how such situated accounts can be provided in conversational interfaces [28]. Such a direction can also prove useful for supporting explainable AI research agenda [10].

Habitability requirements may be different for task-oriented and non-task-oriented cases. Users expect natural language interfaces to handle a variety of different expressions that may sound unnatural or unkind. Ogden et al. [13] found that users preferred to use only keywords instead of full question statements when performing a voice-based search with a task-oriented conversational system. This suggests that naturalness of natural language interfaces for task-oriented cases may not need to fully resemble the rules, conventions, and structures observed in naturally occurring conversations between humans. While users expect efficient and effective completion of tasks in their interactions with conversational systems for task-oriented cases, their expectations may require higher dialogue management capabilities for the non-task-oriented social interactions with the agents, where the enjoyment of the interaction may play a more important role than achieving an end goal [4]. Therefore, the human-human conversational model may not be the right model of interaction for all situations. Similarly, the current conversational systems with their limitations in sentence-level language understanding and dialogue context management may benefit from a model of "sequentially organized moves around request and response" [29].

Conclusion

When AI systems with their black-box characteristics are coupled with an often-invisible conversational interface, the gap between user expectations and system capabilities can get wider. Therefore, habitability as a design goal is highly relevant to the current generation of AI-enabled conversational systems. The definitions, perspectives, and principles presented in this study provide a useful basis to support the habitability of conversational systems. Future work can focus on developing i) low-level design principles (e.g., conversational user experience patterns [22]) for the five different domains of habitability; ii) more comprehensive and process-oriented methods to evaluate habitability within its different domains [13]; iii) an expectation management module responsible for monitoring and adjusting the expectations of users [19]; and iv) new interaction models not fully relying on human-human conversational interaction rules and conventions [4]. Furthermore, future work can extend this preliminary review by including additional keywords relevant to habitability, such as discoverability and common ground, to provide a more comprehensive understanding and establish new connections with other well-established design concepts.

References

[1] Lars Ahrenberg. 1996. Customizing Interaction for Natural Language Interfaces. Computer Information Science 1, 1.

[2] Matthew Chalmers and Areti Galani. 2004. Seamful Interweaving: Heterogeneity in the Theory and Design of Interactive Systems. In Proceedings of Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques. ACM, Cambridge, MA, USA, 243-252.

[3] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in Communication. In Perspectives on Socially Shared Cognition., American Psychological Association, Washington, DC, US, 127-149.

[4] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade and Benjamin R. Cowan. 2019. What Makes a Good Conversation?: Challenges in Designing Truly Conversational Agents. In Proceedings of Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, Glasgow, Scotland UK, 1-12.

[5] E. Coiera. 2000. When Conversation Is Better Than Computation. Journal of the American Medical Informatics Association : JAMIA 7, 3, 277-286.

[6] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley and Natasha Bandeira. 2017. "What Can I Help You With?": Infrequent Users' Experiences of Intelligent Personal Assistants. In Proceedings of Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, Vienna, Austria, 1-12.

[7] Danica Damljanović, Milan Agatonović, Hamish Cunningham and Kalina Bontcheva. 2013. Improving Habitability of Natural Language Interfaces for Querying Ontologies with Feedback and Clarification Dialogues. Journal of Web Semantics 19, 1-21.

[8] Paul Dourish and Graham Button. 1998. On" Technomethodology": Foundational Relationships between Ethnomethodology and System Design. Human-computer interaction 13, 4, 395-432.

[9] Susan L. Epstein, Joshua Gordon, Rebecca Passonneau and Tiziana Ligorio. 2010. Toward Spoken Dialogue as Mutual Agreement. In Proceedings of Proceedings of the 4th AAAI Conference on Metacognition for Robust Social Systems. AAAI Press, 14-21.

[10] Randy Goebel, Ajay Chander, Katharina Holzinger, Freddy Lecue, Zeynep Akata, Simone Stumpf, Peter Kieseberg and Andreas Holzinger. 2018. Explainable Ai: The New 42? In Proceedings of Machine Learning and Knowledge Extraction. Springer International Publishing, 295-303.

[11] Randy Allen Harris. 2004. Voice Interaction Design: Crafting the New Conversational Speech Systems. Elsevier,

[12] J. Hecht. 2018. Managing Expectations of Artificial Intelligence. Nature 563, 7733, S141-s143.

[13] Kate S Hone and Chris Baber. 2001. Designing Habitable Dialogues for Speech-Based Interaction with Computers. International Journal of Human-Computer Studies 54, 4, 637-662.

[14] Kate S Hone and Robert Graham. 2000. Towards a Tool for the Subjective Assessment of Speech System Interfaces (SASSI). Natural Language Engineering 6, 3-4, 287-303.

[15] Kate S Hone and Robert Graham. 2001. Subjective Assessment of Speech-System Interface Usability. In Proceedings of Seventh European Conference on Speech Communication and Technology.

[16] Arne Jönsson. 1997. A Model for Habitable and Efficient Dialogue Management for Natural Language Interaction. Nat. Lang. Eng. 3, 2, 103-122.
[17] Evangelos Karapanos, Jean-Bernard Martens and Marc Hassenzahl. 2012. Reconstructing Experiences with Iscale. International Journal of Human-Computer Studies 70, 11, 849-865.
[18] Ahmet Baki Kocaballi, S Berkovsky, JC Quiroz, L Laranjo, HL Tong, D Rezazadegan, A Briatore and E Coiera. 2019. Personalization of Conversational Agents in Healthcare: A Systematic Review. Journal of Medical Internet Research, 21(11), e15360.
[19] Rafal Kocielnik, Saleema Amershi and Paul N. Bennett. 2019. Will You Accept an Imperfect Ai?: Exploring Designs for Adjusting End-User Expectations of Ai Systems. In Proceedings of Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, Glasgow, Scotland Uk, 1-14.
[20] 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 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 5286-5297.
[21] Michael McTear, Zoraida Callejas and David Griol. 2016. The Conversational Interface: Talking to Smart Devices. Springer,
[22] Robert J. Moore and Raphael Arar. 2019. Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework. ACM,
[23] Roger K Moore. 2017. Is Spoken Language All-or-Nothing? Implications for Future Speech-Based Human-Machine Interaction. In Dialogues with Social Robots, Springer, 281-291.
[24] Philippe Morin and Jean-Claude Junqua. 1993. Habitable Interaction in Goal-Oriented Multimodal Dialogue Systems. In Proceedings of Third European Conference on Speech Communication and Technology.
[25] Jakob Nielsen. 1993. Usability Engineering. Morgan Kaufmann Publishers Inc.,
[26] William C Ogden and Philip Bernick. 1997. Using Natural Language Interfaces. In Handbook of Human-Computer Interaction, Elsevier, 137-161.
[27] William Ogden, James Mcdonald, Philip Bernick and Roger Chadwick. 2008. Habitability in Question-Answering Systems. In Advances in Open Domain Question Answering, Springer, 457-473.
[28] Martin Porcheron, Joel E Fischer, Stuart Reeves and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In Proceedings of Proceedings of CHI Conference on Human Factors in Computing Systems. ACM, 640.
[29] Stuart Reeves, Martin Porcheron and Joel Fischer. 2018. 'This Is Not What We Wanted': Designing for Conversation with Voice Interfaces. Interactions 26, 1, 46-51.
[30] Harald Trost, Wolfgang Heinz, Johannes Matiasek and Ernst Buchberger. 1992. Datenbank-Dialog and the Relevance of Habitability. In Proceedings of Proceedings of the third conference on Applied natural language processing. Association for Computational Linguistics, Trento, Italy, 241-242.
[31] William Watt. 1968. Habitability. American Documentation 19, 3, 338-351.
[32] S. J. Young and C. E. Proctor. 1989. The Design and Implementation of Dialogue Control in Voice Operated Database Inquiry Systems. Computer Speech & Language 3, 4, 329-353.