Just Like a Human? The Consequences of Human-like Voice User Interfaces

Suzanne Tolmeijer
University of Zurich
Zurich, Switzerland
tolmeijer@ifi.uzh.ch

Jan Marco Leimeister
University of St. Gallen
St. Gallen, Switzerland
janmarco.leimeister@unisg.ch

Naim Zierau
University of St. Gallen
St. Gallen, Switzerland
naim.zierau@unisg.ch

Abraham Bernstein
University of Zurich
Zurich, Switzerland
bernstein@ifi.uzh.ch

Andreas Janson
University of St. Gallen
St. Gallen, Switzerland
andreas.janson@unisg.ch

CHI'22,, April 29 – May 5, 2022, New Orleans, LA, USA
ACM XXXXX.
https://doi.org/XXXXXXX

Abstract
The design of Voice User Interfaces (VUI) is often inspired by human-human communication. While the use of human-like cues can make VUI interactions more enjoyable, it also comes with possible risks: it can trigger incorrect assumptions about the capabilities and traits of VUIs, as well as reinforce existing stereotypes about humans. In this paper, we explore the tension between conscious preferences users have and subconscious assumptions that are triggered by different VUI designs, and how they relate to the humanness of VUIs. After exploring anthropomorphism in VUIs, we analyze the trade-off between more human-like and machine-like VUI design, and the consequences for the mental models of users. We conclude by suggesting future work directions related to the humanness of VUIs that could inform future VUI design.

Author Keywords
voice assistants; voice user interface; computers as social actors; user preference; anthropomorphism

CCS Concepts
•**Human-centered computing** → **Natural language interfaces;** Empirical studies in HCI; *User studies;*

Introduction

Voice User Interfaces (VUIs) are gaining popularity — they can be found in most smartphones and are increasingly used as smart home applications [3]. Using voice as a technological interaction medium has many benefits: it allows for interaction when limited hardware is available, when the users hands or eyes are busy, and when natural language interaction is preferred [4]. However, the use and design of VUIs has its own challenges.

Before the pervasiveness of VUIs, speech had only been used as a reciprocative communication medium between humans. As a result, often, VUIs are being modeled using human communication as a starting point, such as by matching existing social norms [1]. While this human style of interaction, including the use of human-like voice (personalities) and language style, can be intuitive and thus require less cognitive processing by the user, building on the evolutionary nature of interpersonal communications, it tends to trigger subconscious assumptions. In part, this is caused by how our brains have evolved to function: the famous model by Kahneman [11] proposes we have a conscious, rational decision process which requires attention (i.e., *'system 2'*), and an automatic, quick and involuntary response that saves us cognitive effort to process everything in depth (i.e., *'system 1'*). While humans can consciously attempt to correct the immediate output of 'system 1', it still has influence on perception, interpretation, and action. Arguably related to this process, in interaction with technology, this effect is described in the Computers as Social Actors (CASA) paradigm [16]: while people consciously know machine are just machines, they tend to subconsciously respond to them as social actors, especially when they have some human-like features such as speech capabilities. This tension between conscious preferences and subconscious responses of VUI users has consequences for the design of VUIs.

Despite the existence of this tension, current VUI design has placed much emphasis on making the user experience more pleasant through the extensive use of design elements that trigger the tendency to anthropomorphise VUIs. For example, design guidelines explicitly include *'Give the agent a persona through language, sounds, and other styles'* [15, p 39]. We argue that in this race towards human-like VUIs, designers are losing track of the possibly harmful effects and ethical consequences of optimizing for human-likeness.

In this paper, we discuss the ethical dilemma related to the design goal of making VUIs more human-like: should we optimize conversational fluency and pleasantness by mimicking human traits, thereby potentially invoking or even exploiting subconscious responses, or explicitly stay away from human-like CUI design, possibly at the expense of user experience? Additionally, assuming we cannot completely avoid subconscious assumptions of VUI users, how can we design VUIs in a manner that balances intuitiveness of interactions while avoiding inappropriate mental models? We reflect on the design challenges related to these questions and propose future work directions that can inform ethical VUI design.

Anthropomorphism and VUIs

Long before the arrival of computers, humans have anthropomorphized their surroundings. Anthropomorphism, which is the *"attribution of human mental states (thoughts, feelings, motivations and beliefs) to nonhuman[s]"* [19, p 437], has for example been seen to help people better predict and explain animal behavior [9]. This natural tendency to anthropomorphize has also been investigated and utilized in human-computer interactions. For example, in avatar de-

sign, higher levels of anthropomorphism have been found to increase the user's perceptions of competency, trustworthiness, and conformity in a dilemma task [8]. For chatbots, human-like cues increase perception of social presence and emotional connection [2].

Regarding the design of VUIs, mixed results regarding anthropomorphism have been found. On the one hand, it has led to positive user experiences: users feel more engaged with VUIs and are more likely to show re-use intention when the VUI has more human-like features [14]. Anthropomorphic features can also lead users to feel more social connection [21] and trust the VUI more [7], which can lead to more disclosure of information by the user [18]. On the other hand, some anthropomorphic features can have an adverse effect. For example, the usage of fillerwords like 'uh' and 'um' lead to lower perceived intelligence and likeability in a task-oriented settings [10]. When VUIs give overly emotional or extreme expressions, it can cause reactance in users [21]. Specifically, too high levels of empathy can lead to rejection of VUIs, possibly because of an 'uncanny valley' effect [23, 13]. A demonstration of Google Duplex, a VUI made to sound as human as possible, got responses from both researchers and social commentators that 'impersonating humans to interact with other humans is likely not ethical or appropriate' [17, p 50]. Culley and Madhaven even specifically warn about potential undue trust as a result of anthropomorphic design [5].

In summary, these results suggest that the CASA paradigm [16] also applies to VUIs: even though users relate to VUIs as being 'just a machine' [12, p 3333], their responses show a blurring of the clear categorical differences between human and machine [12]. In the same way, while users consciously do not assign any stereotypical traits to VUIs, VUI gender can activate subconscious gender stereotyping [22]. In other words, while some level of anthropomorphism can benefit the general user experience, it can subconsciously mislead users (e.g., by increasing trust) or lead them to assign incorrect capabilities, e.g., human-like intelligence. These results raise the question how to consciously apply anthropomorphic VUI designs that both elicit an enjoyable user experience as well as adhere to fundamental beliefs about machine ethics.

What should VUI Design Optimize for?

Striking the right balance between different VUI design options can be challenging. Firstly, there is the question whether designers should optimize for conscious or subconscious preferences that users indicate. In our ongoing work on VUI designs (see [22] for our initial results on trait attribution towards VUIs), we find that while some people state they prefer a functional or task-oriented VUI that does not have 'fake' social cues, they end up responding more positively and display more trust towards social VUI designs. Should VUI design adhere to rational mental models of potential users, indicating that the VUI is just a machine and nothing more? Or should designers anticipate this mismatch between user intention and behavior [20] and design for social pleasantness?

While these questions are often posted in a dichotomous fashion, there could be a middle ground where social cues are used to support specific user experiences, while the VUI is still transparent on its design. This could lead to appropriate expectations and less subconscious assumptions. However, being transparent about VUI design can also backfire; Ferland and Koutstaal found that when a VUI was more transparent on its user modeling, this reduced user engagement [6]. Additionally, it can be challenging to decide which type of expectations to optimize for. It can be considered paternalistic to decide which mental model is

most appropriate for users. Nevertheless, designers should try to stay away from design choices that invoke or reinforce harmful biases (such as that women should serve as assistants to men) [22].

While we do not attempt to answer the difficult and context-dependent question of which design goal to optimize for, we want to increase attention to the potential consequences of anthropomorphism in VUI design. Specifically, we argue that blind pursuit of human-like interactions can (un)intentionally come at the expense of users and exploit subconscious processes. We call upon CUI researchers to further investigate the various consequences of designing for human-likeness. Specifically, we ask designers to take into account unintended consequence of harmful mental models being induced and to make explicit potential stereotypes embedded in its design.

Conclusion

The research on VUI design has spiked in the last years, but not enough attention has been given to the ethical consequences of attempting to approach near-human-like communication by VUIs. We argue that both ends of the spectrum, completely human-like vs. very machine-like, have their pros and cons. A middle ground can be found, but designers need to reflect on the consequences of their design, and attempt to steer away from any design choices that could invoke or reinforce mental models with negative consequences. Furthermore, more insights and future work is needed on the trade-off between higher and lower levels of anthropomorphism.

REFERENCES

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, and others. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[2] Theo Araujo. 2018. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior* 85 (2018), 183–189.

[3] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–24.

[4] Philip R Cohen and Sharon L Oviatt. 1995. The role of voice input for human-machine communication. *proceedings of the National Academy of Sciences* 92, 22 (1995), 9921–9927.

[5] Kimberly E Culley and Poornima Madhavan. 2013. A note of caution regarding anthropomorphism in HCI agents. *Computers in Human Behavior* 29, 3 (2013), 577–579.

[6] Libby Ferland and Wilma Koutstaal. 2020. How's your day look? The (un) expected sociolinguistic effects of user modeling in a conversational agent. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[7] Jonas Foehr and Claas Christian Germelmann. 2020. Alexa, can I trust you? Exploring consumer paths to trust in smart voice-interaction technologies. *Journal of the Association for Consumer Research* 5, 2 (2020), 181–205.

[8] Li Gong. 2008. How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior* 24, 4 (2008), 1494–1509.

[9] Alexandra C Horowitz and Marc Bekoff. 2007. Naturalizing anthropomorphism: Behavioral prompts to our humanizing of animals. *Anthrozoös* 20, 1 (2007), 23–35.

[10] Yuin Jeong, Juho Lee, and Younah Kang. 2019. Exploring effects of conversational fillers on user perception of conversational agents. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.

[11] Daniel Kahneman. 2011. *Thinking, fast and slow*. Macmillan.

[12] Lucian Leahu, Marisa Cohn, and Wendy March. 2013. How categories come to matter. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3331–3334.

[13] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

[14] Emi Moriuchi. 2021. An empirical study on anthropomorphism and engagement with disembodied AIs and consumers' re-use behavior. *Psychology & Marketing* 38, 1 (2021), 21–42.

[15] Christine Murad, Cosmin Munteanu, Benjamin R Cowan, and Leigh Clark. 2019. Revolution or evolution? Speech interaction and HCI design guidelines. *IEEE Pervasive Computing* 18, 2 (2019), 33–45.

[16] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[17] Daniel E O'Leary. 2019. GOOGLE'S Duplex: Pretending to be human. *Intelligent Systems in Accounting, Finance and Management* 26, 1 (2019), 46–53.

[18] Kambiz Saffarizadeh, Maheshwar Boodraj, Tawfiq M Alashoor, and others. 2017. Conversational Assistants: Investigating Privacy Concerns, Trust, and Self-Disclosure.. In *ICIS*.

[19] James Serpell. 2003. Anthropomorphism and anthropomorphic selection—beyond the" cute response". *Society & Animals* 11, 1 (2003), 83–100.

[20] Paschal Sheeran and Thomas L Webb. 2016. The intention–behavior gap. *Social and personality psychology compass* 10, 9 (2016), 503–518.

[21] Hai Son Nguyen, Andreas Mladenow, Christine Strauss, and Katharina Auer-Srnka. 2021. Voice Commerce: Anthropomorphism Using Voice Assistants. In *The 23rd International Conference on Information Integration and Web Intelligence*. 434–442.

[22] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Sebastian Wahdatehagh, Jan Marco Marco Leimeister, and Abraham Bernstein. 2021. Female by Default?–Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

[23] Katja Wagner and Hanna Schramm-Klein. 2019. Alexa, are you human? Investigating

anthropomorphism of digital voice assistants–a qualitative approach. (2019).