WHO ARE YOU TALKING TO?
CONSIDERATIONS ON DESIGNING GENDER AMBIGUOUS VOICE USER INTERFACES

Matheus TYMBURIBA ELIAN\textsuperscript{1,}, Soh MASUKO\textsuperscript{b}, Toshimasa YAMANAKA\textsuperscript{c}

\textsuperscript{a} Graduate School of Comprehensive Human Sciences, University of Tsukuba, Japan, s2030496@s.tsukuba.ac.jp
\textsuperscript{b} Rakuten Institute of Technology, Rakuten Inc. Rakuten Crimson House, Japan
\textsuperscript{c} Faculty of Arts and Design, University of Tsukuba, Japan

ABSTRACT

With the widespread availability of Voice Assistants in smart devices, the usage of Voice User Interfaces has highly increased in the recent years. Although the anthropomorphized Voice Assistants present in these systems can aid users in many tasks, it also activates harmful gender bias and stereotypes. Since the usage of Gender Ambiguous voice agents in these interfaces is considered as solution for mitigating these gender effects, this paper analyzed different studies in the field of Voice User Interface design, while proposing a theoretical framework to design gender ambiguous voice agents, considering the type of recording, method of sound manipulation, method of evaluation for gender identification and contextual characteristics.

Keywords: Voice User Interface Design, Gender Ambiguous Voice, UX Design

1 INTRODUCTION

Recently, the Voice User Interface (VUI) has grown as one of main ways to interact with devices and systems. One of the main reasons for the fast growth in the development and usage of this type of interface was the widespread availability of Voice Assistants (VA) in smartphones and other smart devices (Bentley et al. 2018). Throughout the years these voice assistants are embedded to more and more systems related to different activities, from navigation to shopping, and even health care (Kinsela, 2021, 2022). In the automotive scenario, the addition of voice
assistant to a vehicle interface has even become an important purchase factor by users (Kinsela, 2020).

The interaction through VUIs has advantages when compared to other types of interfaces (e.g., visual or vibrotactile), such as its omnidirectionality, which allows the user to receive the information disregard its body position, and the reduction of visual workload (Shahab et al., 2010; Beattie et al., 2015). In addition, when anthropomorphic characteristics such as name, voice and gender are added to the interface it may increase the level of perceived usability and acceptance towards automated systems, when compared to pure abstract sounds cues (Forster et al., 2017). This can be observed in the fact that the most popular Voice Assistants available (Alexa, Siri, Google Assistant) are marketed as virtual personas, in which the voice agents have their own name, voice, and traits that lead users to attribute certain personalities to them (Kuzminykh et al., 2020).

Even though, if a user asks one of these voice assistants what is its gender, the answers will deny the presence of gender attribution, for example, Alexa’s answer is “As an AI, I don’t have a gender” (Abercrombie et al., 2021). These statements by the VAs are contradictory, since all of them have a female voice by default and, except for Google Assistant, they also have female names (Sutton, 2020). Furthermore, when observing Cortana, Microsoft’s Voice Assistant, it not only has a female voice and name, but it was inspired by a character from the game series Halo, a virtual intelligence that presented itself with the appearance of a woman (Blair, 2014).

Since the Human-Computer Interactions (HCI) are fundamentally social, it is expected that the same rules, expectations and actions might be replicated with real people, as stated by the Computer as Social Actors (CASA) paradigm (Nass et al, 1994). Observing that the interactions with VAs are basically expressed through commands with the expectation of an assistance in a certain matter, the presence of female gender bias in the core of the VAs characters may lead users to have a harmful behavior towards women, not only reinforcing a servant stereotype usually ascribed to women, but also leading to a replication of a commanding tone when interacting with women (Gustavson, 2005). Even with the popularization of VAs in the last few years, the effects of this harmful behavior can already be observed, as women named “Alexa” reported being treated as bots, and even leading women to change their names to avoid a series of distressing situations. (Ard, 2022).

To address this matter, UNESCO (2019) released a technical report highlighting the gender bias in Artificial Intelligences (AI), sexism and gender inequality in technology, and suggested possible strategies and solutions for these issues, also calling for researchers and companies for further studies and development in this field. Regarding the matter of the VAs, the report indicates that using a gender-ambiguous voice might decrease the perception and replication of its gender bias and stereotyping.

Since the literature regarding gender-ambiguous voice agents in VUIs is very little, this paper is an attempt to expand the discussion regarding this matter. This paper is presented as an essay, contemplating the dialogue between humanities and HCI, as highlighted by Bardzell & Bardzell (2016), where we analyze what is currently being done in this field of study, observing methods
for the research and design of gender ambiguous voice agents, while proposing possible solutions for the further development and the application of this type of voice agents.

2 THE NEGATIVE GENDER EFFECT FOR FEMALE USERS

To first identify how were the gender effects in Voice User Interfaces, Elian et al. (2021) conducted a study that consisted in a survey where participants, considering an automated driving scenario, should evaluate voice messages, recorded by a male and a female artificial voice agent, according to three different usability factors: Pleasure, Commandingness and Urgency. Each message represented a call to action for four situations, each one with a different level of designed urgency, the lowest level corresponded to waiting for the startup of the vehicle’s system, and the highest one to a Take-Over Request - situation where the automated driving system has reached its boundaries and a transfer of control to driver is necessary (Naujoks et al., 2017).

The analysis of this study not only highlighted that the voice agent gender had an effect on the participants, where the female agent was perceived significantly more pleasant and urgent, according to the given situation, but it was also found that the participant gender had an effect on the perception of each factor. It was observed that throughout all the situations, female participants had lower scores for Pleasure than the male ones. Also, female participants had higher scores for Commandingness and Urgency when compared to the male ones for almost all the observed situations. Even in the situations with a lower designed urgency, the female participants perceived it as considerably more urgent than the male ones.

This study highlights one more issue when discussing the gender in VUIs: not only there is a stereotyping and behavioral problem towards women with scenarios involving the interaction with this type of interface, as aforementioned, but also the overall experience for female users tends to be considerably worse when compared to male users. As an extrapolation of these findings, if we consider that in a real-life situation, like the Take-Over Request, which has a high sensorial and cognitive workload, a negative perception of the alerts might lead the user, in the worst-case scenario, to a traffic accident, what contradicts one of the main purposes of the Design itself. In addition, it also indicates that depending on the situation that the voice agents within the binary gender frame might not be enough to lead to proper experience.

3 THE Q PROJECT

As one of the first initiatives to change the paradigm of the binary gender frame voice agents in VAs was the project named “Q, the genderless voice” (Q, 2019). Q was presented as a voice agent that, by not sounding as male or as female, would reduce the impacts of gender stereotyping, while including people that do not feel represented in the binary gender frame. To achieve such a voice agent, its creators stated that there is a fundamental frequency range between 145Hz and 175Hz, which they called “the neutral zone”, where the male and female
voice frequencies would overlap each other, allowing to find a voice that would sound as none of the binary genders (Nørgaard, 2019).

The method used by Q to achieve their voice agent consisted in recording twenty people that did not ascribe themselves as male or female. These voices went through sound processing, using change of pitch, formant filter and harmonics modulation. From all the voices recorded, they picked one that fit in the neutral zone and that sounded unique when compared to the others. Then they altered this voice into four different versions and screened them through a survey, where participants were asked if the voice sounded as male or female. The voice with the closest ratio between the male/female answers was the one selected to be Q’s voice agent (Nørgaard, 2019).

The objective of Q’s project was to be a “proof-of-concept” to the developers of the main VAs in the market that a “genderless” voice was possible as an alternative to break voice gender stereotypes, while highlighting the necessity to take in consideration the non-binary community when designing Voice Assistants (Q, 2019).

Although Q represented a well-designed initiative to attempt a change of paradigm, certain aspects of its development must be taken in consideration. As highlighted by Sutton (2020), using terms like “genderless” or “gender neutral” are unhelpful not only when considering the sexism in technology, but also because it implies that gender should not be taken in consideration in the design. Sutton suggests that the term “gender ambiguous” is the most appropriate when designing or researching this type of voice agent.

4 FEMALE BY DEFAULT?

This question that entitles the study conducted by Tolmeijer et al. (2021) is relevant provocation for the discussion about the voice gender in VAs. This study explored the effects of gender and pitch of VAs on trait and trust attribution by users. For the contextualization, it is important to highlight that, currently, there is no Text-To-Speech (TTS) generator tool with gender ambiguous voice agents, which creates a challenge for researchers in this field: how to effectively design these voice agents? Differently from Q (2019), that had as its outcome only a voice agent, Tolmeijer et al. presents a study that further develops this concept, not only finding a simple but effective method to design gender ambiguous voice agents, while empirically testing users’ perceptions and its feasibility of its implementation as part of a VA.

Firstly, the study presents how the method to design its gender ambiguous voice agents for further application in their experiments. Instead of using natural voice recordings like Q, the authors used a series of recordings Wavenet TTS voice agents. Parallel Wavenet is currently the state-of-the-art production model for fast and high-fidelity speech synthesis, it is used by the Google Assistant in few languages, and it is also available as part of the TTS generation tool developed by the company (Oord et al., 2018). Tolmeijer et al. used recordings from different male and female Google Wavenet voice agents, manipulating its pitch afterwards. The pitch was shifted in a certain number of semitones towards the opposite gender from the original voice
agent. Also, they pitch shifted the original voice agents to have a high-pitch and a low-pitch male and female voices, for further comparison.

To screen which of these voice agents should be used in the study continuation, the authors conducted a survey where the participants were asked to hear each one of the agents, and identify it between three options: male, female and unsure. The agent with the highest division between the answers was the one selected as its gender ambiguous voice. In that case, the selected voice had 58% of male answers and 42% for female.

Secondly, the study presents two experimental tasks, an assistant, and a compliance task, that simulated realistic ones in the usage of VAs. Both were inspired by classical gender stereotyping, where women would be better in assistant roles, and males in leadership roles. The assistant task was a simulation of a flight ticket booking, and the compliance one consisted of personal questions in the context of a customer survey. In these tasks, participants were asked to answer questions regarding trust and the level of agreement to a series of trait ascriptions related to the gender stereotyping for men and women (e.g., “Authoritative”, “Cold”, “Dominant”, “Delicate”, “Friendly”, “Sensitive”, etc.).

It is relevant to clarify that these “classical” gender stereotypes are recurrent in the literature. The work by Gustavsson (2005) describes how this assistant role ascribed to women has been reinforced by the service industry and how it is constantly replicated in applications for virtual environments. Considering the stereotype for the male as leader, Mayew et al. (2013) found an association between a lower pitch voice and economical success for men in leadership corporate positions.

The results of Tolmeijer et al. (2021) study regarding the gender ambiguous showed that even though there were direct effects of the voice pitch on the trust attribution, a trend where female participants had a higher trust was observed. For the trait ascription, they found the voice pitch and voice gender to subtly influence the participants’ perceptions, also revealing that the context of the task had an effect on the perception and activation of stereotypes. For the gender ambiguous voice, it presented no significant difference in trust when compared to the gendered voices. In addition, it had both effects found for the female and male voices in the compliance task, which indicates that it was not assigned to one specific gender, but the authors also state that since it does not fit to one stereotypical response, it might be sensitive to multiple ones.

5 DESIGNING GENDER AMBIGUOUS VOICE AGENTS FOR RESEARCH PURPOSES

It was possible to observe in the work of Q (2019) and Tolmeijer et al. (2021) that defining gender ambiguity is more related to the perception of those who listen to the voice than to a theoretical definition. Sutton (2020) clarifies that perceiving which gender a voice should sound as, is not only dependent on the anatomy, but this perception is also built through experiencing the speech community the listener is inserted in, indicating that the cultural context must be taken in consideration when designing voice agents.
The challenge imposed by the unavailability of TTS generator tool that included a gender ambiguous voice agent has not a sole answer. In this sense, it is possible to contrast both Q and Tolmeijer et al. methods. The method developed by Q included the recording of several people’s voices outside the binary gender frame, where the recordings went through a complex sound manipulation process to achieve the result that the authors found to be satisfactory for the continuation of their study. The method developed by Tolmeijer et al. used artificial voices from a TTS tool, that were originally gendered, and used a pitch shift method to make the voices sound towards the opposite gender. In terms of complexity, replicating the Q’s method for research purposes might be quite difficult since it demands a higher level of knowledge on acoustics and sound engineering than the method used by Tolmeijer et al. once the pitch shift technique was empirically demonstrated as effective while being simpler to be replicated. When comparing the usage of natural and artificial voice recordings it is possible to observe that there are only originally gendered artificial voices that might not lead to the same perceptions as using recordings from real people that do not fit in the binary gender frame. Currently there is no study comparing both types of voices, but since this matter of the originally gendered voices was indicated by Tolmeijer et al. as a constraint of their study, this should be considered when designing the gender ambiguous voice agents.

The next step after creating these voices is the evaluation by users regarding their gender perception. Both mentioned studies used surveys asking the participants to identify the gender of each voice message, which seems to be effective, but there are points that must be observed. The survey structured by Q for the gender identification of their voice agents, as mentioned before, had two options “male” and “female”, and the number of answers were even for each option (Yates, 2020). The contradiction that the evaluation of a voice that was supposed to be “genderless” (or gender ambiguous, using the most appropriated term) had only binary gender frame options for its gender identification was highlighted by Sutton (2020), suggesting that a third gender option should be added in this type of evaluation, to avoid the participants to be locked in the binary gender frame. To address this matter Tolmeijer et al. (2021) added a third option to their gender identification survey which was the “unsure” one. The fact that their gender ambiguous voice agent with highest division of answers had no “unsure” answers highlights an issue on the wording of these options. The word “unsure” in a context where the other two options are “male” and “female” suggests to the participant that this option is the adequate if it is not sure between male or female, inducing the answer to fit in the previous options. It is possible to consider that no “unsure” answer assigned by the participants was a reflection of this issue. Thus, to avoid undesirable answering bias, we consider that using the word “other” as a third option might be a better solution for this matter, since it implies something else than the other “male” and “female” options available, and it is also a common term for gender questions in surveys.

Further, it is relevant to underline that context or situation where the gender ambiguous voice agent is presented. Not only due to the impact of the cultural variable on its perception, but how the context might activate certain stereotypes, or even the own nature of the situation, like the Take-Over Request presented in the study by Elian et al. (2021), that had a high level of urgency. Presenting different versions of the voice agent (with different levels of pitch or volume, for
example) according to the context or situation, might be a solution for reduce the impacts of gender bias on the Voice User Interface, while it adjusts the level of situational awareness of the users for the given task.

6 CONCLUSION

This paper focused on highlighting the current gender bias and stereotyping issues to be tackled in the field of Voice User Interfaces, while supporting the application of gender ambiguous voice agents as a possible solution. Through the analysis of different studies in the field of Voice User Interface design, we propose a theoretical framework to design gender ambiguous voice agents, considering the type of recording, method of sound manipulation, method of evaluation for gender identification and contextual characteristics. These observations can aid researchers and designers to further develop studies and projects regarding this type of voice agent, mainly for real life tasks, which are still absent in the literature.

ACKNOWLEDGMENTS

We thank all members of the Kansei Information Design Laboratory at the University of Tsukuba for supporting this project. This study was supported by the program for “Research and Design for Future Merchant” funded by Rakuten Institute of Technology.

REFERENCES


