

Karen, the Interrupting Customer

Mirre VAN DEN BOS, Gergana DZHONDZHOROVA, Ioana FRINCU,
Ella VELNER, Thomas BEELEN and Mariët THEUNE

University of Twente, Enschede, The Netherlands

Abstract. Karen is a conversational agent taking the role of an angry customer in a retail context. While the user (retail employee) tries to convince Karen to follow the rules, the agent interrupts the user, and verbally and nonverbally reacts to the user's sentiments.

Keywords. embodied conversational agent, robot, interruptions, customer interaction

1. Introduction

We present Karen, a prototype of an embodied conversational agent aimed at training retail employees by simulating conflict situations with customers. The agent takes the role of an angry customer that refuses to put on a face mask, in spite of Covid regulations. The employee's goal is to convince the agent to follow the rules and manage the unexpected situation. The agent's task is to immerse the employee in a scenario that requires fast reactions and reasonable behaviour to calm the customer down. To create an uncomfortable conversation, the agent frequently interrupts the user. Uncontrollable and unpredictable interruptions can negatively affect performance, memory and problem solving [1], increase error rate and induce stress and anxiety [2].

The inspiration for the agent stems from the stereotype "Karen", a derogatory slang term which represents "an obnoxious, angry, entitled, and often racist middle-aged white woman who uses her privilege to get her way or police other people's behaviours" [3]. We based the agent's nonverbal behaviour and dialogue responses on real videos made during the pandemic which show anti-mask behaviour in supermarkets.

Karen was developed using the Furhat [4] software development kit, and can be run on a physical Furhat robot as well as a on-screen agent (see Figure 1). We use the "Ursula" character for its face. Its voice is female and US-English speaking with a high pitch and an increased speaking rate, to mimic the Karen stereotype.

2. Dialogue Design

The user (employee) initiates the conversation to confront the agent (customer) about not wearing a mask. The agent first replies in a neutral way by saying, "I have the personal preference to not wear a mask." Further responses from Karen will depend on the sentiment detected in the user's utterances. If the user's sentiment is negative, the agent will also give a (randomized) negative reply and if it is positive, the agent's reply will

also be congruent. For sentiment analysis, we use the VADER lexicon [5]. The user's sentiment also influences Karen's interruption speed, through varying maximum lengths of the user's utterance (in ms) before Karen will interrupt the user:

- Neutral: 3000 ms. Given the average time it takes to produce a turn in human conversations [6], this length simulates a rapid response rather than an interruption, as long as the user knows what he/she wants to say.
- Positive: 4000ms. Longer than neutral to make Karen seem positively affected.
- Negative: 2000ms. The agent does not show any restraint and keeps talking more and more, hardly giving a chance to the user to speak up. If the user talks over the agent with a positive or neutral response, the negative loop might be broken.

Finally, Karen shows nonverbal behaviour (head roll, surprised or angry facial expression) that also depends on the sentiment expressed by the user. Microexpressions such as blinking and shifting gaze are used to assure smoothness in gesture transition as well as making the agent mimic interest towards the user utterances.



User enters
 Karen: Huh?!What stupid shit are you gonna spew at me?
 User: endearment you need to wear a must please
 Karen: So, what if I promise to stay away from the elderly?
 User: now that's not our policy you just need to wear masks story
 Karen: Yeah, I know the rules, but I am here for only 2 minutes, no one will notice that I am not wearing a mask.
 User: you are really wasting my time right now yes or if it if you don't want to wear a mask I need to
 Karen: If only one of us is wearing the mask, it will be fine.
 User: yeah that's true but it's not our policy I didn't make the rules

Figure 1. Karen with a dialogue fragment from one of the user tests (user utterances are ASR output)

3. User testing and future work

Karen is fully functional¹ and has been evaluated in a small-scale (N=5) user test. Figure 1 shows a fragment from one of the user interactions. Note that for the user utterances, the output of automatic speech recognition (ASR) is shown, including ASR errors that make it more challenging for the agent to react appropriately. The results of the user tests showed that users found Karen unlikable and were annoyed and frustrated by her, as they felt the agent did not listen to them. Some also thought the agent felt like a real customer, since it was similar in the way it argued. This is in line with our goal, which was to create a conversational agent that imitates a human customer being difficult.

Future work includes improving the agent's robustness to ASR errors, restructuring the dialogue to avoid repetitiveness in the agent's responses, and the expansion to non-Covid related use cases for retail employee training. Another important future step is to further investigate the implications of propagating potentially harmful stereotypes. It may be beneficial to diversify the appearance of the agent to avoid potential negative outcomes as well as provide a more extensive training for employees.

¹Video showing an exemplary interaction with Karen: <https://youtu.be/Eh1acu6PS5Y>.

References

- [1] Iqbal ST, Bailey BP. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. CHI '05 Extended Abstracts on Human Factors in Computing Systems; 2005. p. 1489–92.
- [2] McFarlane D, Latorella K. The scope and importance of human interruption in human-computer interaction design. *Human Computer Interaction*. 2002 Mar;17(1):1–61.
- [3] Dictionary.com, Karen; Jul 2020. <https://www.dictionary.com/e/slang/karen/> (accessed Apr 19, 2022).
- [4] Al Moubayed S, Beskow J, Skantze G, Granström B. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In: Esposito A et al., editors. *Cognitive Behavioural Systems*. Berlin, Heidelberg: Springer; 2012. p. 114–30.
- [5] Hutto CJ, Gilbert EE. VADER: a parsimonious rule-based Model for sentiment analysis of social media text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI; 2014. p. 216–25.
- [6] Heldner M, Edlund J. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*. 2010 Oct;38(4):555–68.