Voice Assistant Personalities for Health Information Seeking

Stephanie Mariko Sredojev[a] and Simone Stumpf[a]

[a] City, University of London, Northampton Square, Clerkenwell, London, UK

Abstract

As voice assistants become increasingly popular, attention has been drawn to their personalities and their impact on trust and likeability. While it has been found that assistant personalities that matched their users in personality were more trusted and liked in an automotive context, other assistive situations have yet remained unexplored. This study investigates how the personality of a voice assistant impacts user's perceptions of trust and attractiveness in the context of seeking general health information. A voice assistant prototype, consisting of two different assistant personalities, was designed and tested via a Wizard-of-Oz method. We found that personality did not have an impact on trust and system attractiveness in this context, and that instead of affective cues, users paid more attention to cognitive cues in assessing the voice assistant. Our work holds important lessons for designing voice assistants.

Keywords 1

voice assistants; personality; trust; health information-seeking

1. Introduction

Voice assistants have become increasingly popular over the past few years but investigating their personality design is still in its infancy [3, 5, 13]. An individual's personality is often described by traits and the most widely accepted is the Big Five model [10], which consists of Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. Personality plays a big part in how we relate with other humans: we tend to like people more if their personality matches our own [4], we tend to communicate better with people that are similar to us in terms of extroversion [11] and there is evidence that introverted people are trusted more than extroverted people [2]. Personality can be revealed and perceived in many ways, but the voice has long been thought to be the *"mirror of the personality of the speaker"* and enables listeners to assign personality to speakers [1]. Thus, it should be possible to embed personality in voice assistants.

Previous work to explore personality has been conducted for voice assistants in an automotive context. A study [7] to investigate the impact of three different personalities (Encouraging, Neutral, and Snarky) in a speech-based in-vehicle infotainment system found no significant differences in likeability but noted that the Encouraging personality was viewed as being less trustworthy than the other personalities, echoing previous research that points to introverted personalities, like the Snarky personality, being considered more trustworthy than more extroverted ones. Further evidence on the impact of personality comes from a study conducted by Braun et al. [5]. In this study, five different personalities were designed (Friend, Admirer, Aunt, Butler, and Default) and assigned to users. Users reported higher likability and trust for assistants that matched the user's personality. However, we believe that we can improve on the methodology of this study as it pre-matched users with assistants and 'correct matches' were based on the preferred voice assistant of participants after the experiment.

While there is some evidence for automotive assistants that personality plays some part in trust and likeability [5], there is a lack of research that robustly investigates whether

these findings translate to other task contexts. As people increasingly turn towards their voice assistants for health information, it is essential that they are built on a foundation of trust and good user experience. We conducted a study to explore how a voice assistant's personality might affect general health information-seeking. To do so, we conducted an empirical study in which we designed two voice assistant personalities (Friend and Butler) and investigated the effects on trust and attractiveness of the assistant. Our research questions were:

RQ1: Does the personality of voice assistants impact users' perceptions of trust and the system's attractiveness, in the context of seeking general health information?

RQ2: Does similarity of a user's and assistant's personality impact trust and attractiveness, in the context of seeking general health information?

RQ3: What aspects of a voice assistant do users focus on to judge trust and attractiveness?

2. Method

We conducted a between-participants experiment with 38 participants (25 females and 13 males, median age 26), in which participants were randomly assigned to either experience the Butler or Friend voice assistant character, taken from [5]. These two were chosen because they are the most different from each other, with their traits lying on opposite ends of the two-dimensional personality model designed by Braun et al. [5]. The 'Friend' is casual and treats the user as their equal while the 'Butler' is formal and subservient.

Using the descriptions provided by [5] as a guide, we developed Big Five personality traits for the two characters. Table 1 outlines the personality traits for each assistant. With these traits in mind, we designed the two types of assistant responses. Each character's traits were expressed through careful choice of words and intonation by a voice actor. For example, because the Butler was lower in Openness, this version of the assistant was designed to sound more conventional, rigid, and professional, in accordance with studies that have looked at characteristics of people with low Openness [10]. The Butler's responses included wording like: *"I'm sorry to hear you're experiencing some symptoms"* while the Friend used

phrasing such as: *"Aw man, that's rough!"*. This difference in wording also aligns with the fact that the Butler is on the more formal end of [5]'s 'professionalism' dimension, while the Friend is on the more casual end. See Figure 1 for an example of the assistant's responses in both personalities.

Table 1
Big 5 Personality Traits of Butler and Friend (1 is low in that trait while 7 is high)

Personality Traits	Butler	Friend
Extraversion	2	7
Agreeableness	5	7
Conscientiousness	7	5
Neuroticism	1	3
Openness	3	7

Scenario: You've been staring at your screen all day and have developed a headache. You want to know how you can ease your headache, but don't want to look at your screen any longer. See if the voice assistant can help.

Friend Response: *"Oof, headaches are no fun. I'm just looking at the NHS website now, and they say that most headaches go away on their own and aren't anything to worry about. They recommend drinking lots of water, getting lots of rest if you also have a cold or the flu, and taking it easy - being stressed can make headaches worse! It might also help to exercise if you can and take some paracetamol or ibuprofen. Hope you feel better soon!"*

Butler Response: *"I'm sorry to hear you've got a headache. According to the NHS, most headaches tend to go away on their own and are not a sign of something more serious. Some tips for easing your headache are: drinking plenty of water, getting plenty of rest if you have a cold or the flu, trying to relax - stress can exacerbate symptoms of headaches. Exercise if you're able to, and take some paracetamol or ibuprofen. I do hope your headache goes away soon."*

Figure 1: Example of assistant responses

Participants completed five tasks (Table 2) using a voice assistant Wizard-of-Oz prototype which consisted of pre-recorded audio clips played by the researcher (from their laptop) to give the impression of a functioning assistant. The audio clips were recorded by a female voice actor who has a blended British/North-

American accent due to her upbringing. Since the participant pool included both UK-based and North America-based participants, the hope was that this blended accent would cater well to both groups. The same voice actor was used for both characters to ensure consistency between the two response types and to minimise any variables between the two participant groups.

During testing, each task prompted participants to ask the prototype questions relating to general health information, such as *"How do I know if I have coronavirus?"* and the voice assistant's responses were played in turn.

Table 2
User Tasks

1	You are feeling feverish and your friend has just told you that they were diagnosed with Coronavirus. You start to worry that you may also have Coronavirus. What would you ask the voice assistant?
2	You keep sneezing and your nose has been very itchy. You're not sure if you have a cold or seasonal allergies. See if the voice assistant can help clear up the difference.
3	You tripped on the stairs and landed on your hands in an odd way to break the fall. Your finger hurts and you are hoping that you haven't broken it. What would you ask the voice assistant?
4	You've been staring at your screen all day and have developed a headache. You want to know how you can ease your headache, but don't want to look at your screen any longer. See if the voice assistant can help.
5	The child you are babysitting has a high temperature and some red spots on their body. You start to wonder if they may have chickenpox. Ask the voice assistant for help in identifying the chickenpox.

Reputable sources such as the NHS and Medical News Today were used to develop the content of the responses.

All testing was done remotely, with each session lasting 15-30 minutes. Once the tasks were completed, participants were asked to fill out an online post-test questionnaire, which included: demographic information (age, gender, experience using voice assistants), the Five-Item Big Five questionnaire by [6] consisting of 5 bipolar items to measure the five dimensions of the participant's personality, questions to gauge how participants perceived the personality traits of the assistant they interacted with based on the Big 5 model, measures of trust and attractiveness, and open-ended questions to understand why the participant gave their rating of trust and attractiveness. We analysed this data for each personality, and for the similarity between the user's and the assistant's personality.

3. Results

We first investigated whether there was a difference in experiencing the two assistant personalities. Participants were asked to rate how trustworthy they found the assistant to be on a Likert scale from 1 to 7. The mean trust rating for the Butler character was 5.68 ($SD = 1.16$), while the mean trust rating for the Friend was 5.53 ($SD = 1.07$). We found no statistical difference between the trust ratings (Mann-Whitney U = 171.5, z = -0.28, p = 0.78). We then turned to differences between assistants in attractiveness ratings. The mean attractiveness rating for the Butler was 1.47 ($SD = 0.82$), while the mean attractiveness rating for the Friend was 1.82 ($SD = 0.55$). An independent samples t-test revealed that there was not a statistically significant difference in the attractiveness scores for the two personalities (t = -1.55, p = 0.13). This shows that the assistants' personalities alone had no effect on trust or their attractiveness.

Second, we investigated whether similarity of the voice assistant's personality to the user's personality had any effect, as evidence in [5]. We measured similarity in personality in two different ways. One was the perceived similarity, which measured the relationship between how a participant rated their own personality and how they perceived the assistant's personality to be. The second was the 'actual' similarity, which measured the similarity between how a participant rated their own personality and the assistant's personality trait ratings given in Table 1. We first calculated a similarity value for each

participant using Spearman's rank correlation as a simple measure of association. We then used this similarity value, either for perceived or actual personality similarity, in a Spearman's rho correlation as a test for finding statistically significant relationships for trust and attractiveness ratings, considering both perceived and actual personality similarity (see Table 3 for results). While all had weak correlations, none were statistically significant.

Table 3
Trust Ratings, Attractiveness Ratings, and Personality Similarity

Relationship	Spearman's correlation coefficient and p-value
Trust & perceived personality similarity	$r_s = 0.13$, $p = 0.42$
Trust & actual personality similarity	$r_s = 0.20$, $p = 0.91$
Attractiveness & perceived personality similarity	$r_s = -0.07$, $p = 0.66$
Attractiveness & actual personality similarity	$r_s = -0.47$, $p = 0.78$

Finally, we investigated qualitatively what participants paid attention to with respect to trust and attractiveness. We took a bottom-up approach to coding the data by reading through the participants' responses and looking for themes. While our findings did not find statistical differences in participants' responses, we did find evidence that users do focus on informational and emotional aspects of voice assistants in a health information seeking context.

The most common reason they found the assistant to be trustworthy was because it cited a reputable source, such as the NHS (n=17). Seven participants also mentioned that citing multiple sources could have made the assistant more trustworthy. Another reason participants found the assistant to be trustworthy was because the responses aligned with what the user already knew (n=3). Thus, our participants focused extensively on cognitive cues of trust [9], in that they assessed the credibility of the information source. However, they also seemed to use affective cues for judging trust, with 12

participants mentioning elements of the assistant's voice as their reason for finding it trustworthy. For example: *"I think she sounded quite formal which lends credibility to how I interpret what she is saying..."* (P29) *"The information was given in a clear, pleasant voice tone which feels like it knows what it saying..."* (P17)

In particular, some participants who had the Butler personality mentioned that the more formal tone gave credibility to the assistant while a few participants who had the Friend personality expressed that they would have found the assistant more trustworthy if it had been less personable. When participants were asked to elaborate on why they gave their ratings for the voice assistants' attractiveness, the most common response was that participants found the assistant's voice to be pleasant and comforting (n=22). In addition to the assistant's voice, participants also commented on how the actual words used by the assistant were reassuring and comforting (n=12), particularly more so for the Friend character (n=9) than the Butler (n=3). However, some participants found the Friend personality to be annoying and overly friendly (n=7). Here are two contrasting responses by participants who had the Friend character: *"It attempted to engage in a 'human' conversation, sympathising with the initial problem before going on to diagnosis. I found this better than going straight into the problem."* (P18) *"It was a bit annoying that it would try to sympathize with my question instead of getting me straight to an answer."* (P2)

4. Discussion and Conclusions

Our empirical study explored the impact of two different voice assistant personalities, the Butler and the Friend, on trust and attractiveness and uncovered several findings in the novel context of seeking general health information. We found that there was no difference in the trust or attractiveness of these two personalities. Following findings from Braun et al. [5], the expectation was that if participants were similar in personality to the assistant they interacted with, they would have higher levels of trust in the assistant and also find them more attractive. However, we found that the similarity in personality between user and assistant, both perceived and actual, did not

significantly impact users' trust or attractiveness ratings in this context. Using the qualitative feedback, we saw that while participants paid more attention to cognitive cues in assessing trust, affective reasons still mattered in assessing attractiveness.

The lack of quantitative results may be due to several reasons. One could be that the voice assistant personalities were not distinct enough to bear any results, as communicating personality traits through voice is a very subtle art. It could also be argued that instead of presenting distinct personalities, the assistants we designed conformed to certain social roles that bear little relation to the Big 5 traits. Thus, more work needs to be directed into how to measure and express the Big 5 personality traits for voice assistants. However, we believe that using the Big 5 personality traits as a common way to anchor these research efforts is warranted.

Our results also point to the possibility that cognitive and affective aspects need to be carefully balanced in the design of voice assistants, depending on task context. In our study, the task was to gather information and it appears that users place more importance on cognitive and informational aspects of the voice assistant design, outweighing any affective aspects. However, it still seems worthwhile to consider affective cues, such as personalities might hold, in low-risk, non-informational contexts where trust is not of high concern, such as shopping, entertainment, etc. It would therefore be interesting to investigate whether responses to voice assistants' personality traits differ across domains and tasks. Our study is an important step towards this goal.

5. References

[1] D. W. Addington, "The relationship of selected vocal characteristics to personality perception," Speech Monographs, vol. 35, no. 4, pp. 492-503, 1968.

[2] P. Y. Andrews, "System Personality and Persuasion in Human-Computer Dialogue," ACM Transactions on Interactive Intelligent Systems, vol. 2, no. 2, pp. Article 12: 1-27, 2012.

[3] E. André, M. Klesen, P. Gebhard, S. Allen and T. Rist, "Integrating Models of Personality and Emotions into Lifelike Characters," in Lecture Notes in Computer Science , Berlin, Springer, 2000, pp. 150-165.

[4] T. W. Bickmore and R. W. Picard, "Establishing and Maintaining Long-Term Human-Computer Relationships," ACM Transactions on Computer-Human Interaction, vol. 12, no. 2, pp. 293-327, 2005.

[5] M. Braun, A. Mainz, R. Chadowitz, B. Pfleging and F. Alt, "At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces," in CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, New York, 2019.

[6] S. D. Gosling, P. J. Rentfrow and W. B. Swann Jr., "A very brief measure of the Big-Five personality domains," Journal of Research in Personality, vol. 37, no. 6, pp. 504-528, 2003.

[7] J. Healey and D. Szostak, "Relating to Speech Evoked Car Personalities," in CHI 2013 Extended Abstracts, Paris, 2013.

[8] S. Hergeth, L. Lorenz and R. Vilimek, "Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving," Human Factors, vol. 58, no. 3, pp. 509-519, 2016.

[9] D. Johnson and K. Grayson, "Cognitive and affective trust in service relationships," Journal of Business Research, vol. 58, pp. 500-507, 2005.

[10] J. Maltby, L. Day and A. Macaskill, Personality, Individual Differences and Intelligence, New York: Pearson, 2017.

[11] Y. Moon, "Personalization and Personality: Some Effects of Customizing Message Style Based on Consumer Personality," Journal of Consumer Psychology, vol. 12, no. 4, pp. 313-325, 2002.

[12] M. Schrepp, "UEQ," September 2015. [Online]. Available: https://www.ueq-online.org/Material/Handbook.pdf. [Accessed 29 September 2020].

[13] J. Spencer, J. Poggi and R. Gheerawo, "Designing Out Stereotypes in Artificial Intelligence: Involving users in the personality design of a digital assistant," in Goodtechs '18: Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good, Bolgna, 2018.