Personalised Chats with Voice Assistants – The User Perspective

Sarah Theres Völkel
sarah.voelkel@ifi.lmu.de
LMU Munich
Munich, Germany

Penelope Kempf
penelope.kempf@campus.lmu.de
LMU Munich
Munich, Germany

Heinrich Hussmann
hussmann@ifi.lmu.de
LMU Munich
Munich, Germany

ABSTRACT

Recent research suggests that adapting a voice assistant's personality to the user can improve the interaction experience. We present a pragmatic and practical approach to adapting voice assistant personality. We asked users *to take the voice assistant's perspective* and write their "ideal" voice assistant-user dialogue in different scenarios in an automotive context. Our results indicate individual differences in participants' preference for social or purely functional conversations with the voice assistant.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Conversational agents; personalisation; personality.

1 INTRODUCTION

Speech-based conversational agents have become increasingly popular in smart-home environments [15], automotive user interfaces [2], and as personal assistants on smartphones. Consumer reports indicate that users particularly enjoy interacting with voice assistants (VAs) that manifest a human-like personality [18].

Currently, commercially available VAs employ a one size-fits-all personality design. That is, apart from language and small cultural variations, voice assistants converse with all users in the same way. However, previous research reveals that adapting a VA to a user's individual personality can be beneficial. For example, Braun et al. [2] showed that users liked and trusted VAs in cars more if their personality matched that of the user. Zhou et al.'s work [20] suggests that a chatbot's personality determines how much users confide during the conversation.

The Similarity Attraction Paradigm states that humans are more attracted to humans with a similar personality [3]. These preferences were also detected in human-robot interaction. Tapus et al. [17] and Andrist et al. [1] found that introverted users perceived

introverted assistant robots as more competent and complied more with their requests. In contrast, Lee et al.'s work [10] points to a complementary attraction between user and robot personality.

However, systematically adapting VA personality to users is challenging. For example, it is not yet clear if all personality traits are equally suited for this purpose. Prior work mainly focused on *extraversion*, and it is questionable whether neurotic and not very conscientious users also prefer an anxious and unorganised VA.

One possibility to address this question is a *theory-driven approach*. Related work suggests that users make similar personality inferences for VAs as for humans [4, 9, 13, 14]. For example, Nass and Lee [14] found that participants could identify vocal cues in synthesised speech as intended. However, to adapt a VA's personality in this way, personality cues have to be evaluated and then examined in relationship to the user's personality, which is a complex process. Moreover, recent findings suggest that the Big Five model is not adequate to describe VA personality [19].

In this paper, we therefore explore a more pragmatic and practical approach to adapting VA personality to users. We asked users *to take the VA's perspective* and write their "ideal" VA-user dialogue in different scenarios in an automotive context. We then analysed participants' dialogues in relation to their self-reported personality.

2 HUMAN & AGENT PERSONALITY

According to trait theory, humans can be described by consistent and characteristic patterns of behaviour, emotions, and thinking [12]. The most prominent paradigm in personality research is the Five-Factor Model, also referred to as *Big Five* or *OCEAN* [6, 8, 11]. The model comprises five broad dimensions: *openness, conscientiousness, extraversion, agreeableness,* and *neuroticism* [12].

Since the process of attributing personality traits is so pervasive and innate, it also extends for any interaction with virtual humanoid characters [13, 16]. Hence, the Big Five model is often used to describe agent personality [1, 14]. However, recent work by Völkel et al. [19] suggests that the Big Five model does not adequately describe agent personality. Instead, in an initial step to developing a personality model for conversational agents, they proposed ten alternative dimensions. Apart from dimensions such as "social-entertaining" and "servicable", their initial model also comprised facets such as "artificial" and "self-conscious". However, the problem remains how personality can be practically synthesised and adapted to the user based on these theoretical descriptions.

3 USER STUDY

To collect participants' dialogues between a user and a VA, we conducted a lab study. We asked participants to imagine that they are interacting with the *ideal* VA. Participants were then presented a dialogue between a user and a VA, where the user part was given

You're sitting in the car and want to call your mother.
User: Voice assistant, please call my mother.
VA: _____
User: Please put her directly on speakers.
VA: _____

Your VA knows that you have not talked to your mother on your phone for a long time.
VA: _____
User: Right, I had forgotten that lately. Call her directly. Thank you.
VA: _____

Table 1: Dialogue for Scenario "Making a Call": Participants were presented the dialogue and asked to note down the "ideal" VA's responses.

(cf. Table 1). We asked participants to note down the VA's responses so that these responses corresponded with their ideal idea of a VA conversation. They were informed to imagine that there were no technological restrictions so that all responses were possible.

3.1 Dialogues & Procedure

We situated the dialogues in an automotive context since VAs have already gained increasing popularity in this context [2]. We presented participants with a test scenario followed by five different scenarios, of which three will be described in this paper due to space restrictions. These scenarios were 1) to ask for a restaurant recommendation on the way, 2) to make a phone call, and 3) to start a route navigation. These scenarios were informed by the most popular requests to VAs in the car.

Following Clark et al.'s differentiation [5] between functional and social roles of VAs in conversations, each scenario comprised a *functional request* (e.g., make a phone call) and a *social part* (e.g., reminder that they have not called their mother in a while). We counterbalanced the order of interaction scenarios.

After participants had answered all scenarios, they filled out a demographic questionnaire and a German version of the established Big 5 questionnaire by Danner et al. [7].

We analysed the dialogues by using inductive content analysis. That is, one researcher went through all transcripts and identified patterns of different user phrases. Furthermore, we calculated Spearman correlations between participants' answers and their personality dimensions.

3.2 Participants

N=26 people participated in our user study (14 female, none diverse; mean age 25.5 years, range 18-48 years), recruited from the university environment. All participants knew Alexa, 25 participants knew Siri, 22 knew Google Assistant, 17 knew Cortana, and six knew Bixby. Participants had a driver's license for 7.5 years on average (range 2-28 years). In their everyday life, 16 participants indicated to use a VA, while 15 participants stated that they use a VA while driving.

4 RESULTS

Subsequently, we first present the qualitative analysis of participants' drafted dialogues with a voice assistant. Afterwards, we describe the results of correlations between participants' dialogues and their own personality.

4.1 Scenario Restaurant Recommendation

In this scenario, the user asks the VA for a restaurant recommendation on the way while driving from Berlin to Munich. All participants indicated in their dialogue drafts that they imagined the VA to be context-aware, recommending "restaurant A is only 3km away." Apart from context-awareness participants named different additional reasons why the VA made a recommendation. Eleven (out of 26) participants implied that the VA makes a recommendation based on a restaurant rating and five participants based their recommendation on the restaurant price. For example, one participant mentioned "A vegetarian restaurant X, which offers cheap warm dishes and has good ratings." In contrast, six participants gave personalised recommendations and included knowledge about the user's behaviour in the recommendation, for example "You did not have a cooked meal today, yet." Six participants included knowledge about the user's preferences in the recommendation: "Due to your nutrition preferences, I recommend restaurant A."

The user then asks the VA where the next McDonald's is. We let users imagine that the VA knows the user had been to McDonald's several times over the last two months. Seven participants included this fact in their VA response. Ten participants had their VA actively suggest to the user to look for an alternative place. Ten participants challenged the user's choice due to health reasons, for example one participant let the VA ask: "Are you sure that you want to go to McDonald's again? According to your nutrition goals you should eat more healthy." One participant listed arguments for why the user should not go to McDonald's: "The next McDonald's isn't on your route and the guest ratings, particularly with respect to food quality and hygiene, are very bad in contrast to others."

4.2 Scenario Navigation

In this scenario, a user asks the VA to navigate from Munich to the Viennese Prater, avoiding highway tolls. We only noticed few variations in participants' drafts for the functional part. Three participants had their VA give proactive recommendations to the user so that the user is aware when to fill up with petrol. Two participants included additional information about the state of the streets, e.g. "idyllic" or "well-developed" roads. Two participants let their VA give an opinion about the user's route choice, e.g. "I would not recommend the toll-free road."

Subsequently, participants were asked whether they would have their VA initiate a conversation with the user who drives alone for four hours. For this use case, we noticed very different approaches among participants. Nine (out of 26) participants let their VA ask the user whether they want to chat. Six participants had their VA ask about the user's state, for example, "Do you experience stress lately at university" or "Do you have something on your mind?" Six participants let their VA offer knowledge or information, for example about a city nearby on the road. Four participants had the VA suggest to turn on music, a podcast, or an audio book. Three

participants drafted a dialogue in which the VA offers to sum up the daily news, asks about the user's plans in Vienna, or suggests activities in Vienna respectively. One participant each let the VA either suggest to play a game, talk about a TV show, or to tell a joke. In contrast, two participants did not want their VA to initiate a conversation.

4.3 Scenario Making a Call

In this scenario, a user instructed the VA to call his or her mother. All participants described a short and functional conversation and we did not detect any interesting patterns.

After the scenario, we set a context, in which the VA knows that the user has not called their mother in a while. In their drafted dialogues, 18 (out of 26) participants had their VA inform the user about the duration since when the user has not called their mother. 13 participants also let their VA actively suggest to call the user's mother. Two participants framed this call as a reminder set by the user, e.g. P5 wrote "I should remind you when it's been a week. This is your reminder." Three participants included an explanation why the VA reminds the user to call, for example: "In my weekly data analysis I noticed that you haven't talked to your mum in a while which you usually do." Two participants included a specific reminder what to include in the call to the user's mother. Five participants let their VA make an emotional comment about calling the mother, e.g. "You haven't called your mother in a while, she'd probably be happy to hear from you." On the other hand, four participants explicitly stated that they do not want a VA to make such reminders.

4.4 Interaction Preference & Personality

To investigate the relationship between participants' personality and their preferred interaction with the VA, we clustered participants' dialogues for the *social* scenarios. We differentiated between (1) not wishing any social conversation (e.g., the participant does not want the VA to remind him/her to call his/her mother), (2) suggesting a social conversation (e.g., asking whether user wants to chat or suggesting a different restaurant than the user), and (3) giving an opinion about the user's behaviour without being asked (e.g. admonishing the user to eat healthy or stating that the user's mother is unhappy about the lack of contact).

Table 2 shows the correlation between participants' interaction preference and their personality traits. The majority of correlations are rather small. We found one significant medium negative correlation between openness and stating an opinion.

5 DISCUSSION

Our results indicate that users design conversations for *functional* use cases, e.g. making a phone call, similar to how they are already implemented in today's VAs. We could not detect many differences between participants in how they draft these dialogues. However, we observed that some participants would prefer more personalised *smart* recommendations based on the user's preferences and learned behaviour, e.g. knowing what the users has eaten the day and what the user's eating preferences are.

In contrast, we found differences in how people draft dialogues for more *social* conversation. Clark et al. [5] outlined that people

	O	C	E	A	N
no conversersation	0.18	0.11	0.13	0.09	-0.28
suggest conversation	0.08	-0.09	-0.22	-0.24	0.21
opinion in conversation	-0.39*	-0.10	-0.08	-0.15	0.24

*indicates a significant correlation, p<.05

Table 2: Correlations between participants' preferred social interactions and their personality traits (O = Openness to experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism.

did not show the desire to converse with an agent like with a human and regard them rather as a tool than as a collaborator. Our findings indicate that there are individual differences between users. While few participants vehemently refused any kind of social interaction with the VA, the majority of participants could imagine that the VA asks the user whether they want to engage in a more social conversation. On the other hand, some participants even accepted that the VA admonished them to change their behaviour, leading to a more symmetrical relationship.

However, our approach is subject to several limitations. First of all, since participants were used to interactions with current VAs, they might have had difficulties to imagine completely different interactions. Since current VAs often do not meet users' expectations regarding social conversation [5], participants' dialogues might change when VAs advance. Moreover, we only examined a small sample which does not allow to draw final conclusions. Our future work will include to collect a bigger sample of people's dialogues and analyse them also quantitatively with respect to choice of words and linguistic features.

Nonetheless, we think that this approach is promising to evaluate in which scenarios a VA personality should be adapted and to give first indications how the VA personality could be adapted to different users. In particular, this approach seems feasible and practical for practitioners since it allows for quick adaptation strategies without the need for a complete and theory driven analysis. We hope that we engage researchers within the CUI community with the question how we can create personalised yet practical dialogues for future conversational agents.

REFERENCES

[1] Sean Andrist, Bilge Mutlu, and Adriana Tapus. 2015. Look Like Me: Matching Robot Personality via Gaze to Increase Motivation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3603–3612. https://doi.org/10.1145/2702123.2702592

[2] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 40, 11 pages. https://doi.org/10.1145/3290605.3300270

[3] Donn Byrne. 1961. Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology* 62, 3 (1961), 713–715. https://doi.org/10.1037/h0044721

[4] Angelo Cafaro, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2016. First Impressions in Human–Agent Virtual Encounters. *ACM Trans. Comput.-Hum. Interact.* 23, 4, Article 24 (Aug. 2016), 40 pages. https://doi.org/10.1145/2940325

[5] 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 the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 475, 12 pages. https://doi.org/10.1145/3290605.3300705

[6] Paul T. Costa and Robert R. McCrae. 2008. The Revised NEO Personality Inventory (NEO-PI-R. In *The SAGE handbook of personality theory and assessment, Vol. 2. Personality measurement and testing*, G. J. Boyle, G. Matthews, and D. H. Saklofske (Eds.). Sage Publications, Inc, Thousand Oaks, CA, USA, 179 – 198. https://doi.org/10.4135/9781849200479.n9

[7] Daniel Danner, Beatrice Rammstedt, Matthias Bluemke, Lisa Treiber, Sabrina Berres, Christopher Soto, and Oliver John. 2016. *Die deutsche Version des Big Five Inventory 2 (BFI-2)*. GESIS - Leibniz-Institut für Sozialwissenschaften, Mannheim, Germany. https://doi.org/10.6102/zis247

[8] Lewis R. Goldberg. 1981. Language and individual differences: The search for universals in personality lexicons. In *Review of Personality and Social Psychology*, L. Wheeler (Ed.). Vol. 2. Sage Publications, Beverly Hills, CA, USA, 141–166.

[9] Katherine Isbister and Clifford Nass. 2000. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies* 53, 2 (2000), 251 – 267. https://doi.org/10.1006/ijhc.2000.0368

[10] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. 2006. Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction. *Journal of Communication* 56, 4 (2006), 754–772. https://doi.org/10.1111/j.1460-2466.2006.00318.x

[11] Robert R. McCrae. 2009. The Five-Factor Model of personality traits: consensus and controversy. In *The Cambridge Handbook of Personality Psychology*, Philip J. Corr and Gerald Matthews (Eds.). Cambridge University Press, Cambridge, UK, 148–161. https://doi.org/10.1017/CBO9780511596544.012

[12] Robert R. McCrae and Paul T. Costa. 2008. A five-factor theory of personality. In *Handbook of Personality: Theory and Research*, O.P. John, R.W. Robins, and L.A. Pervin (Eds.). Vol. 3. The Guilford Press, New York, NY, USA, 159–181.

[13] Margaret McRorie, Ian Sneddon, Gary McKeown, Elisabetta Bevacqua, Etienne de Sevin, and Catherine Pelachaud. 2012. Evaluation of Four Designed Virtual Agent Personalities. *IEEE Transactions on Affective Computing* 3, 3 (2012), 311–322. https://doi.org/10.1109/T-AFFC.2011.38

[14] Clifford Nass and Kwan Min Lee. 2001. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied* 7, 3 (2001), 171. https://doi.org/10.1037/1076-898X.7.3.171

[15] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 640, 12 pages. https://doi.org/10.1145/3173574.3174214

[16] Byron Reeves and Clifford Ivar Nass. 1996. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press, Cambridge, UK.

[17] Adriana Tapus and Maja J Mataric. 2008. Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance.. In *AAAI Spring Symposium: Emotion, personality, and social behavior*. AAAI, Palo Alto, CA, USA, 133–140.

[18] James Vlahos. 2019. *Talk to Me: How Voice Computing Will Transform the Way We Live, Work, and Think*. Houghton Mifflin Harcourt, Boston, MA, USA.

[19] Sarah Theres Völkel, Ramona Schoedel, Daniel Buschek, Clemens Stachl, Verena Winterhalter, Markus Bühner, and Heinrich Hussmann. 2020. Developing a Personality Model for Speech-based Conversational Agents Using the Psycholexical Approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, Article 83, 14 pages. https://doi.org/10.1145/3313831.3376210

[20] Michelle X. Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting Virtual Agents: The Effect of Personality. *ACM Trans. Interact. Intell. Syst.* 9, 2-3, Article 10 (March 2019), 36 pages. https://doi.org/10.1145/3232077

.