What Are You? Investigating the Importance of Human-likeness of Digital Voice Assistants through a Qualitative Approach

Katja Wagner
University of Siegen
Hanna Schramm-Klein
University of Siegen

Paper from the 51st Annual EMAC Conference, Budapest, May 24-27, 2022

What Are You? Investigating the Importance of Human-likeness of Digital Voice Assistants through a Qualitative Approach

Abstract

Conversational agents or digital assistants such as Siri, the Google Assistant or Alexa are moving to the private sector of consumers. They are located e.g., in smartphones, living rooms or general somewhere in the household. This study aims to investigate with a qualitative approach if and what kind of humanlike triggers can weaken or strengthen the perception of a digital voice assistant. 20 participants were interviewed and the results were categorized in four upper categories with different subcategories. Findings show that although the functionality in general is an important factor the integration of humanlike characteristics can influence the perception positively. Especially social behavior and adaptability are strong triggers for anthropomorphizing. So, companies should not only emphasize technical parameters but also the social and humanlike possibilities of digital voice assistants.

Keywords: *Digital Voice Assistant, Anthropomorphism, Artificial Intelligence*

Track: Digital Marketing and Social Media

1. Introduction

Nowadays almost one billion people worldwide use digital voice assistants such as Alexa and Siri, also called conversational agents. Voice and speech recognition software will increase up to $6.9 billion by 2025 in comparison to $1.1 billion in 2017 (Tractica 2018). Starting with installed software on smartphones and computers, digital voice assistants now also move into living rooms in form of stationary smart speakers. Due to the high current relevance, some studies have analyzed the effects of digital voice assistants more closely (Lee and Choi 2017). One approach to improve user acceptance is to simulate human traits in the digital voice assistants design in order to increase familiarity and satisfaction with the assistant by suggesting a human-like counterpart (Purington et al. 2017). This phenomenon of humanizing technical devices is called anthropomorphism (Horstmann et al. 2018). More clearly, anthropomorphism is the tendency to transfer human characteristic traits, emotions or intentions to non-human actors (Epley et al. 2007). In the literature, anthropomorphism is considered to have a major influence on robot design and has been empirically supported (e.g. Fink 2012). However, the phenomenon is largely treated superficially as a single parameter. Zitzewitz et al. (2013), for example, have recognized that the phenomenon as this complex and multidimensional construct requires a closer and more detailed consideration. Particularly in the field of digital voice assistants there is currently a lack of knowledge about anthropomorphism and its triggers, although results are already available that vocal communication can have an influence on the user (Kiesler & Goetz 2002). Surprisingly, as far as we know, no study has proven empirical insights into the personification of digital voice assistants through dimensions of the phenomenon of anthropomorphism. To investigate this phenomenon in a more holistic and fundamental approach we conducted a qualitative study based on the three-factor theory of anthropomorphism, the Media Equation Theory and the Uncanny Valley paradox to gain a deeper understanding about the influence of anthropomorphism on digital voice assistants.

2. Theoretical Framework of Anthropomorphism

In their research, Goetz et al. (2003) introduce personality as a subcomponent of anthropomorphism alongside with social behavior and intelligence. In addition, this is proven by Becker et al. (2007), who state that the integration of emotions leads to an improvement in the credibility, liveliness and personality of an assistant. The digital voice assistant can only express such personality or emotion through its voice, whose mere existence is sufficient to promote anthropomorphic perception (Persson et al. 2000). Murray and Arnott (1993) name the parameters

speech quality, speech timing and voice height as manipulation variables through which emotions, competence and intelligence can be transported. Further findings show that computer-generated voices which are most similar to the one's own personality are preferred (Nass & Lee 2001). This coincides with the similarity-attraction effect, which states that people are attracted to others when they perceive them as similar to themselves. The similarity of a digital voice assistant with the user in terms of dimensions such as behavior or personality, is often sufficient to attract attention, increase sympathy and motivate further interaction (Bernier & Scassellati 2010). The parameter field of social behavior is one of the most important components for embodying anthropomorphization. It is regarded as the peak of human-robot interaction. Social behavior is also made possible by interactions resulting from the perception and interpretation of the world in relation to one's own experiences. As soon as social behavior is reached, the robot appears alive and offers humans a space for anthropomorphizing artificial intelligence (Zitzewitz et al. 2013). The interactive behavior in robotics is formed by providing a limited library of basic behaviors for the robot. This can positively influence the perception of the robot (Cooney et al. 2012). Moreover, various visual and acoustic dimensions are proposed in the literature in this context, such as voice, verbal communication (Sims et al. 2009), gesture, emotions and optics (DiSalvo et al. 2002). This study therefore investigates, if the voice is sufficient to influence the behavioral intention positively and what further humanlike characteristics are important. Epley et al. (2007) formulated three psychological triggers for anthropomorphic thinking: elicited agent knowledge, effectance motivation and sociality motivation. The factor of elicited agent knowledge means that people judge objects through certain signals, by drawing on their experiences with other people and accordingly trying to create a common basis with their counterparts by attributing human qualities. Such a signal can be the voice that evokes knowledge about people and leads to identification with them (Eyssel et al. 2012). Effectance motivation describes the personification due to the need of people to predict the behavior in their environment for helping them to understand and control it. Thus, a perceived personality of the digital voice assistant can contribute to making his behavior comprehensible and predictable. The third determinant sociality motivation describes the anthropomorphization of objects as a result of the human need for social contact and the constant search for social interaction, which is why the non-human counterpart is perceived and treated as a social counterpart (Epley et al. 2007). Social rules are also applied in the use of digital voice assistant, whereby these relate in particular to the linguistic interaction with each other, for example by addressing the digital voice assistant with its name. With

their Media Equation Theory, Nass and Reeves (1996) also address this effect of sociality. Their theory is based on the assumption that people implicitly treat computers and other technologies socially as soon as the technology exhibits one or more elementary social signals such as the use of natural language, an interactive character, or the fulfillment of a particular social role. The CASA (Computer are social actors) paradigm has also empirically supported that the mere existence of a voice is sufficient to convey a social presence (Nass et al. 1994). Overall, the theory shows that human social behavior is not evoked by complex processes, but can be triggered by minimal stimuli.

3. Method

To investigate the phenomenon of anthropomorphism with a holistic approach within digital voice assistant, we conducted a qualitative study. The qualitative research approaches are characterized not only by a much greater openness, but also by greater flexibility (Lamnek 2010). Through a more in-depth examination, a higher content validity, but a deeper generalizability in comparison to the quantitative approach is achieved (Myers 1997, Kaplan & Maxwell 2005). So far, there is hardly any knowledge available about the triggers of anthropomorphism with regard to digital voice assistants, so that initial approaches must first be generated. In the first step, a problem-oriented semi-structured interview guideline was developed taking into account previous literature. In order to offer enough space for the presentation of one's own subjective views, open questions were formulated. Subsequently, persons of different age, gender, education and profession were recruited. In our survey, a total of 20 persons (IP) were interviewed, 10 of them female and 10 males (average age M=33.25, SD =13.27) (see Table 1). After that the threshold of new answers was reached. These participants were interviewed in nine categories, e.g., general experience, technical aspects, intelligence, animation and sympathy. Audio files of all interviews were created and transcribed via the transcription software "Express Scribe" and applying the transcription system according to Kuckartz et al. (2008). The interviews were conducted between November 2018 and February 2019 and lasted an average time of 30 minutes. Afterwards, the transcripts were evaluated according to the qualitative content analysis of Mayring (2010). This method uses categories to evaluate the transcripts in several fixed steps. The category creation is done by looking for similarities and differences of the data and formulating the relevant aspects for the research question in categories. Both upper (UC) and subcategories (SC) can be subdivided and are formed inductively from the data material. The subsequent coding of the interviews involves assigning all relevant text passages to the appropriate, previously formulated categories, so that at the end a category-based overview is created.

Table 1. Overview of Participants									
Female	Female	Male	Female	Male	Female	Female	Male	Male	Female
51	20	45	38	58	33	21	28	27	22
Language Teacher	Student	Photographer	Office Employee	Self-employed Craftsman	Account Manager	Student	Security Engineer	Penetration Tester	Student
Male	Female	Male	Female	Male	Male	Female	Female	Male	Male
28	19	56	28	26	21	22	23	50	49
Product Manager	Education as a Nurse	Network Administrator	Product Manager	Systems Analyst	Student	Student	Student	Authorized Signatory	Mechanical Installer

4. Results

An overview of the results of the study is presented in table 2, followed by a detailed analysis of the four categories with regard to the role of anthropomorphism.

Table 2. Category System
UC1 Functionality: **SC 1.1 -** Preference over other technology, **SC 1.2** - Function/Role
UC2 Behavior: **SC 2.1** - Social Behavior, **SC 2.2** – Adaptability
UC3 Anthropomorphic Attributes: **SC 3.1** - Personality, **SC 3.2** - Independence, **SC 3.3** – Interaction, **SC 3.4** - Voice, **SC 3.5** Appearance, **SC 3.6** - Competence and Intelligence
UC4 Relationship: **SC 5.1** - Proximity and Distance, **SC 5.2** - Emotions during use

UC1 "Functionality"

SC "Preference over other technology" shows that digital voice assistant is mostly preferred to other technologies. The reasons given for this are particularly the simple and convenient operability through the language (IP7) and the associated speed (IP2). IP3, e.g. sees the advantage in the fact that speaking and listening does not prevent him from doing other things as much. Furthermore, using a digital voice assistant is seen as efficient, comfortable and natural kind of communication and interaction (IP12). For others the technology is not yet mature enough to prefer it over other technologies (IP13). Within the SC "Function/Role" IP14 describes that the assigned role depends on the expectations you have of the assistant. Some see it as a tool, instrument or technical aid that does not require human characteristics (IP5). Others attribute the role of the machine friend who takes orders and thinks along with them (IP2). However, it is also admitted, that dealing with digital voice assistants differs from dealing with fellow human beings by confronting voice assistants more ruthlessly and roughly than fellow human beings (IP18). If the voice assistant has the role of a tool, it conveys less humanity than if it is regarded as a technical friend. So, the

perceived role of the digital voice assistant in particular seems to have an influence on the attributed human similarity.

UC2 "Behavior"

The SC of "Social Behavior" describes through which characteristics the digital voice assistant can be perceived as a social presence and conducts social behavior. If the voice assistant shows behavior such as making contact with people, apologizing for not understanding etc., it will be awarded politeness and thus social behavior (IP1). The use of names during communication is also seen as social behavior (IP2). In addition, the omission of behaviors such as insults or lies is perceived as social (IP14), so that social behavior can also be mediated by the absence of antisocial behavior. However, most attribute the competence for feelings to humans and regard it as questionable if the voice assistant would be able to do so (IP10). The SC of "Adaptability" describes the competence of adapting behavior to certain circumstances. An adaptation of the digital voice assistant to the user is advocated in the sense that when a child's voice is recognized, the assistant adapts its response behavior to certain questions in a child-friendly way (IP1). In addition, the voice assistant should adapt itself to the respective user and the situation with regard to the suggestions made and adapt its personality depending on the situation (IP8). It can therefore be stated that the behavior is considered positive by showing social behavior and adaptability and can contribute through the social component to a human perception of the digital voice assistant.

UC3 "Anthropomorphic Attributes"

There are different opinions concerning the integration of a "Personality". Some advocate a personality in order to build a better relationship with the digital voice assistant (IP2). Others see this as a danger because the boundaries between man and machine would become blurred (IP3) and people could be replaced (IP13). If the assistant is assigned with a lot of human similarities this could lead to disappointment because the device cannot fulfill the same expectations of the user like a human (IP15). Nevertheless, the voice assistant is attributed various traits such as reliability, friendliness, helpfulness, humor, intelligence, honesty and impertinence (IP11, 16), so that a certain personality of the assistant is already perceived by the voice and its responses (IP2). Especially the possibility of communication, which is characterized by mutual understanding and empathy, is seen as a relevant characteristic of a human being (IP18). With regard to the attributed "Independence", opinions differ. Some of the respondents' attribute autonomy to the voice assistance because it

independently filters and executes information or commands and also generates proposals itself (IP2). Others do not regard it as independent because it has to be activated by voice commands (IP7) and then just follows its programming (IP19). The repeated mention of programming in this context (IP13) shows that there is still a great awareness of the technical interface, which could inhibit anthropomorphization. "Interaction", on the other hand, is given high importance as a trigger for anthropomorphization, because the assistant has a more lifelike effect due to its interactive character and the possibility of a joint conversation (IP9). IP20 takes a critical view of interaction and the associated increase in independence. This could be due to the Uncanny Valley effect (Mori 1970). "Voice" describes the elements that influence the perception of the digital assistant in relation to its liveliness and artificiality. The voice and language of the assistant is felt to be alive when it is easy to understand, uses slang phrases, pauses to imitate breathing and mimics human imperfections (IP2). The pronunciation of words and the prosody of the voice seem to be important (IP15). In general, the assumptions show that an increasingly human voice meets with more acceptance than an artificial voice (IP2). Many of the respondents find a human voice pleasant because it is characterized by a certain softness (IP3). As far as the "Appearance" is concerned, the voice assistant does not yet convey liveliness through its loudspeaker optic, but when the assistant responds the bright lights are perceived as living elements that imitate the rhythm of the language (IP1). It seems important that the assistant has eyes or an eye-catcher that can be used for orientation during interaction (IP2). From this it can be concluded that the choice of voice and the appearance can trigger or weaken anthropomorphism, depending on how the respective aspect is implemented. In general, respondents would prefer a digital voice assistant that is more similar to a robot than to a human (IP4) but it should be of a small size (IP17) and of an inconspicuous appearance (IP14). "Competence and Intelligence" highlights the perceived relevance of implementation competence and intelligence into digital voice assistants. What is striking about this category is that a large majority of the participants consider an implementation of intelligence to be relevant and meaningful (IP3). The most common scenarios listed as intelligent aspects are: processed and stored individual usage patterns, connectivity with other devices, quality of interaction, speech recognition, drawing logical conclusions and responses, ability to retrieve information, speed of speech processing, feedback capability, and reliability of results (IP20). Referring to competence, the prerequisite for this is the qualitative answering of questions (IP7).

UC4 "Relationship"

The two SCs consist of: "Proximity and Distance" and "Emotions during use". Some of the respondents mentioned an increasing strengthening of a friendly relationship to the digital voice assistant by means of an increase in human characteristics (IP2). Thus, many predict an increase in the emotional attachment to a voice assistant (IP8). The second SC deals with the softening of the boundaries between a human being and a voice assistant. The feelings of the interviewees illustrate a clear differentiation between a digital voice assistant and a human being and illustrate this by noting an assistant as a machine (IP5). The assistant often provides incorrect or inadequate information and results which leads to anger (IP3). Some of these persons describe a feeling of discomfort and skepticism, and in one case even fear (IP1).

5. General Discussion and Conclusion

The results describe the influence of anthropomorphism of a digital voice assistant, which can be done via voice and character attributes and show the possibility of influencing the perception of a voice assistant and the increase of sympathy towards it. Almost all participants find it more pleasant when the digital voice assistant has human features and prefer to use them. For companies, this means that the implementation of human traits in voice assistants can lead to higher acceptance. Since the parameter field of social behavior is one of the most important components for embodying anthropomorphization, it is assumed that the increasing human characteristics of a robot are accepted and can be possible triggers. This finding is based on the CASA paradigm (Nass et al. 1994) as well as on the triggers of anthropomorphism formulated by Epley et al. (2007). The skepticism towards the voice assistant's empathy in the study can possibly be explained by the Uncanny Valley mentioned by Mori (1970), so that the empathy produces a too high human similarity, which has an uncanny effect on the user and can lead to a rejection of the assistant. However, many respondents find a human voice to be extremely pleasant and experience an increase in sympathy. Many of the interviewees felt that the increase in human characteristics is positive, but they prefer the functionality over these humanlike characteristics. But it would be interesting to check in further research if maybe characteristics like apologies would overcome the effect of a partly dysfunction. In line with Lee et al. (2006), our study confirms that digital voice assistants are assigned a personality that can help to understand the behavior of it through its similarity to humans. According to our study, another trigger of human perception is an interactive character and thus independence. Kanda et al. (2004) have found a positive correlation between an independently interactive robot and a human perception, so an interactive digital voice assistant has a more human effect. According to our study,

anthropomorphization is additionally influenced by voice and appearance. In his research on robots, Zitzewitz (2013) has also identified these parameters as influencing factors of human similarity. In summary, this holistic study has shown that anthropomorphism plays a significant role and can lead to a better perception and higher acceptance of the digital voice assistant. Furthermore, social behavior, adaptability, similarity to the user, personality, independence, voice, appearance and interaction are all factors that can contribute to the anthropomophization of a digital voice assistant, which positively influence the human-computer-interaction. These identified factors could be useful for voice assistant developers and other researchers. This would mean for the design that conversational agents should have integrated imitation of social behaviors, demonstrating adaptability and an appearance appropriate to its role or task. Unfortunately, the study has also some limitations as it is a qualitative investigation which means that the sample was small. We tried to eliminate some important factors by depicting participants with different age group, a balanced gender distribution, and different types of occupations. But nevertheless, is would be interesting to investigate the found factors of anthropomorphism with a larger sample and to investigate the degree of humanlike characteristics in different contexts and cultures.

References

- Becker, C., Kopp, S. & Wachsmuth, I. (2007). Why emotions should be integrated into conversational agents, *Conversational Informatics: An engineering approach*, 1–25.
- Bernier, E.P. & B. Scassellati (2010). The Similarity-Attraction Effect in Human-Robot Interaction, *IEEE 9TH International Conference on Development and Learning*, 286–290.
- Cooney, M., Zanlungo, F. & Nishio, S. (2012). Designing a Flying Humanoid Robot (FHR): Effects of Flight on Interactive Communication, *IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 364-371.
- DiSalvo, C.F., Gemperle, F., Forlizzi, J. & Kiesler, S. (2002). All robots are not created equal: the design and perception of humanoid robot heads. *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 321– 326.
- Eyssel, F., Kuchenbrandt, D., Hegel, F. & De Ruiter, L. (2012). Activating elicited agent knowledge: How robot and user features shape the perception of social robots, *Proceedings of the 21th IEEE International Symposium on Robot and Human Interactive Communication*, 851-857.
- Epley, N., A. Waytz, & J.T. Cacioppo (2007). On seeing human a three-factor theory of Anthropomorphism, *Psychological Review*, 114(4), 864–886.
- Fink, J. (2012). Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction, Social Robotics. *ICSR 2012. Lecture Notes in Computer Science* (pp.199-208), Ge S.S., Khatib O., Cabibihan JJ., Simmons R., Williams MA. (eds), Vol. 7621. Springer, Berlin, Heidelberg.
- Goetz, J., Kiesler, S. & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation, *IEEE Roman*, 55-60.

- Horstmann, A.C., Bock, N., Linhuber, E., Szczuka, J.M., Straßmann, C. & Krämer, N.C. (2018). Do a robot's social skills and its objection discourage interactants from switching the robot off? *PLoS ONE*, 13(7), 1-25.
- Kaplan, B. & Maxwell, J.A. (2005). Qualitative Research Methods for Evaluating Computer Information Systems, *Evaluating the organizational Impact of Healthcare Information Systems* (pp. 30-55), Hannah, K.J., Ball, M.J. (eds.), 2nd edition, Springer, New York.
- Kiesler, S. & Goetz, J. (2002). Mental models and cooperation with robotic assistants, *Proceedings of CHI'02 on Human Factors in Computing Systems*, 576 –577.
- Kuckartz, U. (2008). *Qualitative Evaluation: Der Einstieg in die Praxis*, VS-Verlag, Hamburg.
- Lamnek, S. (2010). *Qualitative Sozialforschung*, 5th edition, Beltz, Basel.
- Lee, K. M., Peng, W., Jin, S. A. & Yan, C. (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), 754-772.
- Lee, SY. & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity, *International Journal of Human-computer Studies,* 103, 95-105.
- Mayring, P. (2010). Qualitative Inhaltsanalyse, *Handbuch Qualitative Forschung in der Psychologie (*pp. 601-613), Mey, G. Mruck K. (eds.), Springer, Wiesbaden.
- Mori, M. (1970). The uncanny valley, *Energy,* 7(4), 33–35.
- Murray, I.R & Arnott, J.L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion, *Journal of the Acoustical Society*, 93, 1097–1108.
- Myers, M.D. (1997). Qualitative research in information Systems, *MIS Quarterly,* 21(2), 241-242.
- Nass, C., Steuer, J.S. & Tauber, E. (1994). Computers are social actors, *Proceeding of the CHI '94 Proceedings*, 72–78.
- Nass, C. & Reeves, B. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, *Cambridge University Press*.
- Nass, C. & Lee, K.M. (2001). Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction, *Journal of Experimental Psychology*, 7(3), 171–181.
- Ortony, A. (2002). On making believable emotional agents believable, *Emotions in humans and artifacts (*pp. 189-211), R. Trappl, P. Petta, S. Payr (eds.), Cambridge.
- Persson, P., Laaksolahti, J. & Lönnqvist, P. (2000). Anthropomorphism–A multi-layered phenomenon, *Proceedings of Socially Intelligent Agents-the Human in the Loop*, 131–135.
- Purington, A., Taft, J.G., Sannon, S., Bazarova, N.N. & Taylor, S.H. (2017). Alexa is my new BFF: Social Roles, User Satisfaction, and Personification of the Amazon Echo, *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2853-2859.
- Sims, V.K., Chin, M.G., Lum, H.C., Upham-Ellis, L., Ballion, T. & Lagattuta, N.C. (2009). Robots' auditory cues are subject to anthropomorphism, *Proceedings of the Human Factors and Ergonomics Society*, 1418–1421.
- Tractrica (2018). *Voice and Speech Recognition Software Market to Reach $6.9 Billion by 2025*, Retrieved from https://www.tractica.com/newsroom/press-releases/voice-and-speech-recognition-software-market-to-reach-6-9-billion-by-2025/ (Last accessed March 10, 2019).
- Zitzewitz, J. von, Boesch, P. M., Wolf, P., & Riener, R. (2013). Quantifying the Human Likeness of a Humanoid Robot, *International Journal of Social Robotics,* 5(2), 263–276.