

Alexa, Are You Human? Investigating the Anthropomorphism of Digital Voice Assistants – A Qualitative Approach

Completed Research Paper

Katja Wagner

University of Siegen
Unteres Schloss 3, 57072 Siegen,
Germany
katja.wagner@uni-siegen.de

Hanna Schramm-Klein

University of Siegen
Unteres Schloss 3, 57072 Siegen,
Germany
hsk@marketing.uni-siegen.de

Abstract

Digital voice assistants, often associated with artificial intelligence in integrated applications or as a stationary stand-alone speaker, are on the rise. By integrating humanlike characteristics, they differ from conventional technologies. This qualitative study investigates the role of anthropomorphism concerning digital voice assistants and examines which humanlike characteristics are the most powerful drivers. By relying on the three-factor theory of anthropomorphism, the media equation theory and the uncanny valley paradox, we present a category system based on N=20 interviews. The results show that anthropomorphism is positively perceived if the voice assistant shows socially adapted behavior and embodies a certain role, as well as through the implementation of personality, independence and interaction, with the voice playing a decisive role. This study not only supports previous research but also contributes to the field of technology acceptance. Developers should, therefore, integrate humanlike characteristics but not a too humanlike appearance due to possible rejection.

Keywords: Digital voice assistants, anthropomorphism, artificial intelligence, humanlike characteristics, technology acceptance, human-computer interaction

Introduction

Currently, almost one billion people worldwide already use digital voice assistants, also called conversational agents. The number of users is forecasted to increase in the upcoming years (Tractica 2018). Voice and speech recognition software sales will increase up to \$6.9 billion by 2025 in comparison to \$1.1 billion in 2017. Starting with installed software on smartphones and computers, digital voice assistants have now also moved into living rooms in the form of stationary smart speakers. A digital voice assistant is a software system that performs actions based on pronounced commands. It supports users in their everyday tasks and through questions to make it easier to cope with their lives.

In general, a digital voice assistant is a software system that is started by pronouncing a specific activation word to receive voice commands from the user. These commands are then recorded by voice and data processing and interpreted using the Internet and existing databases. The information obtained can then be made available to the user as output either in the form of linguistic responses, recommendations or by carrying out actions (Hauswald et al. 2015). On the one hand, there is reactive support, such as answering user questions about the weather; on the other hand, there is proactive support, which also includes reminding users of upcoming events or recommending certain products and services. Some of the best-

known assistants include Siri, released by Apple in 2011, followed by Microsoft's Cortana, Google's Google Assistant and Amazon's Alexa (Shum et al. 2018).

Through the use of artificial intelligence and the possibility of machine learning, digital voice assistants are also in a position to expand their abilities and to anticipate the wishes and needs of their users after a certain time, so that a personalized online environment adapted to the respective user is created (Stucke 2017). By processing artificial intelligence, the digital assistant is able to learn, understand and react to linguistic commands. This new way of controlling devices via human speech therefore enables a very natural, intuitive and revolutionary way of communication between man and machine (Tadeusiewicz 2010), in which the interface between them is moving away from manual operation via keyboard or touchpad to a linguistic operation that brings the technology closer to the private user through increased user-friendliness (Bickmore et al. 2018).

Due to the high current relevance, some previous studies have focused on the research of digital voice assistants and have examined it more closely (Lee and Choi 2017). The investigation of the factors influencing the acceptance of the assistants seems to be particularly relevant to being able to reduce the still existing reservations and inhibitions during use. More specifically, it is interesting to identify possible drivers and barriers of use to draw conclusions for both practice and further research. One approach to improve user acceptance is to simulate human traits in the digital voice assistant design to increase familiarity and satisfaction with the assistant by suggesting a humanlike counterpart (Purinton et al. 2017). This phenomenon of humanizing technical devices is called anthropomorphism and is a main characteristic and part of the new technology of digital voice assistants (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).

With the increasing presence of anthropomorphic elements, the relevance of elaborating on the potential benefits of an anthropomorphic appearance and behavior of a digital voice assistant increases. Appearance is based on visual and haptic appearance, tone, smell and taste, while behavior focuses on social behavior, interactive behavior, movement, non-verbal and verbal communication. The developers of digital voice assistants especially focus on and promote social behavior (Zitzewitz et al. 2013). The most important indication for the acoustic appearance of a person is the characteristic voice. This voice embodies the emotion, intention and mood of a person (Zitzewitz et al. 2013). In the literature, anthropomorphism is considered to have a major influence on robot design and has been empirically proven by several studies (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 a complex and multidimensional construct that 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 show vocal communication can have an influence on the user and his perception (Kiesler and Goetz 2002).

In contrast to embodied conversational agents, where movement and appearance play an important role (Castillo et al. 2018), the main characteristics of non-embodied conversational agents are primarily the voice, especially the use of natural language. Previous research has often concentrated on the design, functional components or acceptance of technical devices in general (Venkatesh et al. 2012). Other studies stress that interaction with an assistant or a robot not only fulfills the technical need but can also trigger emotions (Horstmann et al. 2018). This outcome means that digital voice assistants differ from other technologies in terms of spoken language. 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 to gain a deeper understanding of the influence of anthropomorphism on digital voice assistants. Therefore, our research questions are as follows:

RQ1: How is anthropomorphism manifested in digital voice assistants in general?

RQ2: Which characteristics will lead to a better perception of digital voice assistants and therefore to a higher behavioral intention?

RQ3: Which characteristics are considered to be the most important in referring to digital voice assistants?

In the following section, we provide a short overview of the theoretical framework, explaining the three-factor theory of anthropomorphism, the media equation theory and the uncanny valley paradox. Then we further extend the method and results and end with the discussion and conclusion.

Theoretical Background of Anthropomorphism

As mentioned above, anthropomorphism generally describes the tendency to attribute human characteristics, emotions and behavior to non-human objects with the aim of making the actions of the other person comprehensible and explainable (Duffy 2003). Here, an additional distinction is made between implicit and explicit anthropomorphism, whereby the implicit anthropomorphism takes place spontaneously and unconsciously, while the explicit anthropomorphism takes place consciously and is reflected (Zlotowski et al. 2018). With digital voice assistants, implicit anthropomorphism can occur through observation and interaction with the assistant, while explicit anthropomorphism can be verified through questioning.

The phenomenon of anthropomorphism is a multidimensional construct consisting of different factors (Zitzewitz et al. 2013). In his research, Zitzewitz (2012) introduces a network of parameters, which he first divides into the two main categories: appearance (static) and behavior (dynamic). To appearance, he assigns parameters such as visual appearance and sound, while to the category behavior, he assigns parameters such as movement, interaction, social behavior, verbal and non-verbal communication. Digital voice assistants have many of the dynamic parameters described by Zitzewitz due to their voice-based design, which leads to the fact that the voice assistants are attributed a liveliness perception. According to Ortony (2003), the perception of a constant personality through the inclusion of character traits is also an important influencing factor for anthropomorphism. In their research, Goetz et al. (2003) introduce personality as a subcomponent of anthropomorphism alongside social behavior and intelligence. In addition, this is indicated 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 a 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 of speech quality, speech timing and voice height as manipulation variables through which emotions, competence and intelligence can be transported via the voice. Further findings show that users prefer computer-generated voices that are most similar to their own personality (Nass and 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 and Scassellati 2010). It is also assumed that perceived animation plays a role in humanization. Animation can be understood as the ability to move independently as well as the ability to react to environmental influences (Bartneck et al. 2009). In the case of a suction robot, for example, the former is true, since it has already been established that anthropomorphic perception is triggered by the independent movement and independent visit of the charging station. The latter applies to the digital voice assistant. The voice assistant could be given a kind of liveliness by reacting appropriately to his environment and its stimuli. This appropriate response may also have an effect on perceived intelligence, which, according to Zlotowski (2014), has a lesser effect on perceived anthropomorphism than perceived emotionality.

The parameter field of social behavior is one of the most important components for embodying anthropomorphization. It is seen 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. A large number of basic behaviors can positively influence the perception of the robot (Cooney et al. 2012). Social behavior is also associated with interactive behavior, for which movement is an essential prerequisite and is regarded as the peak of behavior since it is based on the other parameter fields. All these factors lead to increased anthropomorphization and enable an increasingly positive human response in human-robot interaction (Zlotowski, 2015).

Moreover, various visual and acoustic dimensions are proposed in the literature in this context, such as movement (Wang et al. 2006), voice, verbal communication (Sims et al. 2009), gesture, emotions and optics (DiSalvo et al. 2002).

As Zitzewitz et al. (2003) mentions, there are different influencing parameters in the field of anthropomorphism. It is not only about appearance but also about the behavior such as (non) verbal communication or social behavior whereby the latter is seen as the most complex parameter (Goetz et al. 2003). Riek et al. (2009) finds that users feel more sympathy towards human-looking robots than mechanical-looking robots when they were mistreated. Moreover, the evaluation of the appearance seems to be dependent on the personality of the user in the context of domestic robots, but a human-looking robot is only perceived as positive up to a certain point (Walters et al. 2008). This goes along with the findings of Ferrari et al. (2016), where users see androids as a threat to their human uniqueness. Following the results of Airenti (2018), the variability of anthropomorphization in the context of interactions derives from the affective states of the users. Emotions that are integrated into a conversational agent could enhance acceptance as the results of a virtual museum guide robot have shown (Becker et al. 2001). Moreover, Fussell et al. (2008) have shown that users consider a politely and less mechanistic answering robot to be more positive. In a cultural study where three English accents were compared during the use of a robot, the results also demonstrated that a less mechanical voice with a local accent affects the users' perception (Tamagawa et al. 2011). Furthermore, if the robot has a female or a male voice, this influences the perception of the robot (Crowley et al. 2009). Referring to Niculescu et al. (2013), the interaction quality and ease is rated through the voice pitch, and the empathy and humor of the robot provides for a higher enjoyment of special tasks. Digital voice assistants currently lack visual stimuli such as movement, gestures or eye contact. However, these factors could be important for anthropomorphism when integrating voice assistants into household appliances. However, voice is seen as an important influencing factor concerning anthropomorphizing. Among other things, this study investigates whether voice is sufficient to influence the behavioral intention positively and what further humanlike characteristics are important.

In summary, most previous studies concentrate on one parameter of anthropomorphism in a special context or with embodied conversational agents where appearance is also influencing. In the shopping context, it is stated that visual anthropomorphization provides a higher purchase intention (Yuan 2019). Thus, to our knowledge, there are hardly any insights that are taking the sum of the parameters of anthropomorphism into a holistic consideration for non-embodied conversational agents such as digital voice assistants.

In addition to these studies, there are theoretical concepts that describe the reasons for anthropomorphizing (three-factor theory of anthropomorphism), what cues are necessary for users to behave socially with conversational agents (media equation theory) and to which degree humanlike characteristics lead to a negative effect (uncanny valley). It is fundamentally necessary to explain the reasons for users to anthropomorphize to ask our three research questions. Additionally, social behavior seems to be a main aspect influencing the perception of digital voice assistants and as it is assumed that users do not prefer a conversational agent that is totally humanlike, we refer to these theoretical concepts in more detail in the following section.

Three-Factor Theory of Anthropomorphism

Epley et al. (2007) formulates three psychological triggers for anthropomorphic thinking that he calls *elicited agent knowledge*, *effectance motivation* and *sociality motivation*. The factor of elicited agent knowledge means that people judge objects that radiate something similar to human beings through certain signals by drawing on their own experiences with other people and accordingly trying to create a common basis with their counterparts by attributing human qualities to the objects. In the case of digital voice assistants, for example, 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 motivational determinant of personification due to the need of people to understand and predict behavior in their environment. This phenomenon was defined by Robert White (1959) as the tendency of people to explore and influence their own environment. By projecting human traits onto inanimate objects, the human being creates a situation that he can seemingly explain, giving him perceived control subjectively over the situation. Thus, the perceived personality of the digital voice assistant can contribute to making its behavior comprehensible and predictable. The third determinant "sociality motivation" describes the anthropomorphization of objects as a result of the human need for social contact, attachment 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 assistants, whereby these relate in particular to the linguistic interaction with each other, for example, by addressing the digital voice assistant by its name to set the interaction in motion.

Media Equation Theory

With their media equation theory, Nass and Reeves (1996) also determine 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 signals used do not need to be particularly pronounced. As long as the medium shows behaviors that indicate a social presence, people react accordingly by using social, interpersonal interaction rules and transferring them to interaction with the medium (Nass and Reeves 1996). This reaction, however, happens unconsciously and automatically, but it is shown nevertheless, although one is rationally aware that there is no social counterpart and some people even deny their reaction afterwards. The behavior of the medium towards the human being also influences its perception. For example, it was found that a medium that compliments was consistently judged to be more sympathetic and competent (Nass and Reeves 1996). Overall, the theory shows that human social behavior is not evoked by complex processes but can be triggered by minimal stimuli, so that this mechanism leads to an easier handling of new technologies, such as digital voice assistants. The technology of digital voice assistants also possesses such social behavior-triggering elements through the use of natural language and an interactive driver.

The Uncanny Valley

Previous literature could indeed determine increasing familiarity towards an object by adding humanlike elements (Seeger et al. 2017). Mori (1970), investigating the relationship between the degree of anthropomorphism and the familiarity with the robot, finds that at a certain degree, the effect is reversed. This seemingly paradoxical phenomenon is described in the literature as the uncanny valley effect. Accordingly, familiarity with the subject increases steadily with increasing human resemblance, but at a certain level of human resemblance, it undergoes a considerable slump, so that acceptance drops rapidly. This acceptance gap is only closed again when a certain, very high degree of human similarity has been reached. A possible explanation for this phenomenon is that robots with low human similarity are evaluated by standards other than robots with high human similarity. While objects with little human resemblance are rated as having human characteristics, humanlike robots are rated as having weaknesses and deficiencies compared to humans (Merkle 2019). The expectations of the more human robot are higher and can therefore only be disappointing, which leads to a negative reaction. Digital voice assistants can also fall into this uncanny valley, for example, by choosing a voice that is too close to a human being, which could lead to a dissonance compared to the still non-human behavior and thus cause a rejection of the technology. Based on previous literature and these theories, we developed an interview guideline.

Method

To investigate the phenomenon of anthropomorphism within digital voice assistants in a holistic approach, a qualitative study was conducted.

The qualitative research approaches are characterized not only by a much greater openness but also by greater flexibility (Lamnek 2010). In this way, situations can be better taken into account, and a more in-depth examination of the research object can be carried out against the background of the research question. A higher content validity, but a deeper generalizability in comparison to the quantitative approach, is achieved (Myers 1997, Kaplan and Maxwell 2005). This method is particularly suitable because hardly any knowledge is available so far about the triggers of anthropomorphism with regard to digital voice assistants, so that initial approaches must first be generated, and qualitative research is particularly suitable for this. In the first step, a problem-oriented semistructured interview guideline was developed taking into account previous literature. To offer enough space for the presentation of one's own subjective views, open questions were formulated. In our study, a total of 20 interview partners (IP) were selected through a purposive sampling strategy and independently interviewed, namely, 10 females and 10 males (average age $M=33.25$, $SD =13.27$) (see Table 1). We monitored for heterogeneity in age, balanced gender distribution and in different kinds of knowledge and experiences. Heterogeneity of the sample helps to support evidence that results cannot only be found within one group and thus ensures a higher generalization (Robinson 2014). Thus, our intention was to explore the attitudes of a general sample and not the attitudes based on use. After 15 interviews, data saturation was reached (Saunders et al. 2017). Both non-users (nine) as well as users (11)

were included to obtain broad results whereby all participants had already heard about different voice assistants and tested different assistants. The more experienced participants were using conversational agents daily or at least several times a week and for more than one year, whereas the non-users had only tried it or used it less than five times a year.

These participants were interviewed in the following nine categories: general experience, technical aspects, intelligence/competence, anthropomorphism, animation, sympathy, perceived security, communication skills and trust. Afterwards, an audio file of all interviews was created using a smartphone and then written with the assistance of the transcription software "Express Scribe" following the transcription system according to Kuckartz et al. (2008). The interviews were conducted between November 2018 and February 2019 and had an average length of approximately 30 minutes. The reasons for choosing these nine categories were based on different literature and the named theories. Therefore, we could, for example, assume that experience and technical aspects are important drivers in the field of technology acceptance concerning behavioral intention when we especially consider models such as technology acceptance model (TAM) or the unified theory of acceptance of use of technology (UTAUT2) (e.g., Venkatesh et al. 2012). Since digital voice assistants are technical devices as well, functionality is a basic prerequisite for a positive behavior intention. However, they differ from classical technologies through their use of spoken language, for example; this is why we, furthermore, rely on Bartneck et al. (2009), who developed measurement tools such as intelligence, perceived safety or anthropomorphism for humanlike characteristics in human computer interactions. Further, previous studies have concentrated especially on the difference between artificial and human communication and the simulation of empathetic or fact-oriented communication, which can also influence behavioral intention and trustworthiness (e.g., Bittner and Shoury 2019). In addition, especially in the virtual and digital environment, trust is a decisive factor with regard to purchase or use intentions in general (e.g., de Visser et al. 2016). Moreover, there are further characteristics that could influence the behavioral intention and the functionality of digital voice assistants; however, based on these exemplary named studies, the prior categories are the main influencing constructs for a holistic approach. Further characteristics could be interpreted as a dimension of these main categories.

To ensure that the participants' answers were reliable and valid, we added control questions at the beginning of the interview. For this purpose, we asked about their knowledge of digital voice assistants (e.g., "Which digital voice assistants do you know?"), their experience (e.g., "Have you ever used a digital voice assistant?", "How long have you used digital voice assistants?"), their attitude towards them in general (e.g., "What is your general opinion towards digital voice assistants?") and their future prognosis ("Where do you see digital voice assistants in 20 years?"). Therefore, questions about actual use were asked as well as questions about desirable and undesirable characteristics of digital voice assistants. Next, we were able to identify preferences about anthropomorphic characteristics for digital voice assistants. Because all non-users have at least tried digital voice assistants, there were no striking differences in forming the questions for them.

The resulting data material in the form of transcripts was then evaluated after qualitative content analysis according to Mayring (2010). This very structured evaluation method uses categories to evaluate the transcripts in several fixed steps. However, the categories need to be created first. This creation was done after the first viewing of the transcripts by looking for similarities and differences of the data and formulating the relevant aspects for the research question in new categories. Furthermore, a summarizing content analysis was conducted for each individual interview to find the core elements for the research question. Relevant text elements were limited to the main aspects, and verbal ornaments were dropped to concentrate on the content-bearing components (Guest et al. 2012). Thus, a clarity through reduction was produced. Statements that referred to each other or that had the same content were coded and summarized as a theme since they were more precise as a new category. Afterwards, either the statements were assigned to already existing categories or new categories were created if they did not fit into previous categories. Thus, upper categories (UC) and subcategories (SC) were subdivided and formed inductively from the data material. The subsequent coding of the interviews involved assigning all relevant text passages to the appropriate, previously formulated categories so that at the end, a category-based overview was created in which all relevant text passages were assigned to their appropriate categories. In our study, both text passages that confirm the category and text passages that negate it were assigned so that conclusions could be drawn about the importance of a category. After the development of the category system, we carried out a re-examination of the structure through a second material run to double check that each interview was categorized correctly.

To ensure validity, we used two coders who not only doubled checked the results but also discussed them to reach an intercoder agreement that provided accuracy and credibility (Guest et al. 2012, Creswell and Miller 2000). Moreover, we performed a peer debriefing, which means that questions and results were discussed with another researcher to ensure that no overinterpretation or misleading evaluation took place (Creswell 2008). Although triangulation is controversial (Bollen et al. 1998), we also included contrary statements in the categories to get different views if and moreover, the group of participants were heterogenic in age, working position and educational level (Creswell 2008).

| 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 |

Table 1. Overview of Participants

Results

The results of the study are presented in Table 2 for a better overview. Afterwards, the analysis will be shown in more detail.

| Table 2. Category System | | | |
|---------------------------------|----------------------------|---------------|----------------------------------|
| UC | | SC | |
| UC1 | Experience | | |
| UC2 | Functionality | SC 2.1 | Preference over other technology |
| | | SC 2.2 | Function/Role |
| UC3 | Behavior | SC 3.1 | Social Behavior |
| | | SC 3.2 | Adaptability |
| UC4 | Anthropomorphic Attributes | SC 4.1 | Personality |
| | | SC 4.2 | Independence |
| | | SC 4.3 | Interaction |
| | | SC 4.4 | Voice |
| | | SC 4.5 | Appearance |
| | | SC 4.6 | Competence and Intelligence |
| UC5 | Relationship | SC 5.1 | Proximity and Distance |
| | | SC 5.2 | Emotions during use |

Table 2. Category System

UC1 “Experience”

The first UC, "Experience", describes the previous knowledge and attitudes of the interviewed persons towards digital voice assistants in general. Among the best-known voice assistants are Alexa, Siri, Google Assistant and Cortana (IP8), which are offered by the largest technology companies currently on the market. It is striking that digital voice assistants that have their own names are more likely to be mentioned and

therefore more memorable (IP7). The use of digital voice assistants seems to be different depending on the gadget used. While people who have a preinstalled voice assistant on their smartphones stated that they rarely (IP6) use it, people who have an Alexa or Google Home Assistant as a stationary device stated that they use it several times a day (IP3). However, in terms of the tasks assigned, the stationary devices do not seem to differ from the preinstalled voice assistant in smartphones. In general, simple tasks such as obtaining information, navigation, information about weather and dates, translations, setting alarms and playing music are left to the assistant (IP8). Very few people currently use Alexa and Co. to control their smart homes (IP15). The statements about the recommendation are different; although some describe the digital voice assistant as being helpful, they would not yet recommend it to their acquaintances either because they are not yet behind it (IP1) or because there are still concerns about data protection (IP3). The majority are positive about recommending a digital voice assistant and justify this with its usefulness and ease in everyday situations (IP6). The ease of use and convenience resulting from the use lead to the recommendation (IP20: *“Comfortable, easy. It's the easiest to use. You can even do that if you have broken hands.”*). Despite concerns expressed, everyone is convinced that digital voice assistants will be more mature, more widespread and integrated into many areas and devices in the future, so that they will take on increasingly complex tasks and partially replace humans (IP9).

UC2 “Functionality”

The second UC, "Functionality", describes the usability and role of the digital voice assistant compared to other technologies. It can be divided into the SCs "Preference over other technology" and "Function/Role". Digital voice assistant technology is mostly preferred to other technologies. The reasons given for this outcome are particularly the simple and convenient operability through the language (IP7) and the associated speed (IP2). Moreover, the usability in general and the voice recognition have improved so it is very easy to use (IP7: *“What I actually noticed is that you don't have to speak as explicitly as you used to. It has become more user friendly and as far as speech recognition is concerned, it has become a lot easier.”*). IP3 sees the advantage in the fact that speaking and listening does not prevent them from doing other things as much, and IP20 also finds it *“easier to dictate something than to type it into the keyboard [because] you can do different things at the same time”*. Furthermore, using a digital voice assistant is seen as efficient and as a comfortable and natural kind of communication and interaction (IP12). For some, it is also situation-dependent whether they prefer the voice assistant (IP2); for others, the technology is not yet mature enough to prefer it over other technologies (IP13).

The SC “Function/Role” describes the role attributed to digital voice assistants. IP14 describes that the assigned role depends on the expectations you have of the assistant and how you see it. Some see it as a tool, instrument or technical aid that does not require human characteristics (IP5: *“This is supposed to be my tool, the voice assistant. Nothing more.”*). Others attribute to it the role of the machine friend who takes orders and thinks along with them and whose friendly relationship helps when something needs to be taken care of (IP2: *“[...] is a kind of machine friend who has the position of helper who is always willing to do commands for me, do things for me.”*). However, participants also admit that dealing with digital voice assistants differs from dealing with fellow human beings because they confront voice assistants more ruthlessly and roughly than their fellow human beings (IP18). Therefore, it can be stated that the assigned role and function of the assistant can favor or weaken anthropomorphization. If the voice assistant has the role of a tool, it conveys less humanity than if it is regarded as a technical friend or interlocutor. Therefore, the perceived role of the digital voice assistant in particular seems to have an influence on the attributed human similarity.

UC3 “Behavior”

The third UC, "Behavior", describes the reaction of the digital voice assistant perceived by the user to actions of the user. This category can be divided into the SC “Social Behavior” and “Adaptability” of the assistants. The SC of “Social Behavior” describes through which perceived characteristics the digital voice assistant can be perceived as a social presence and conducts social behavior. While some participants do not attribute social behavior to the assistant (IP4), others see the courtesy of the digital assistants as a characteristic of social behavior. If the voice assistant shows behavior such as making contact with people, apologizing for not understanding, letting a person speak or remembering what has already been discussed, it will be awarded politeness characteristics and thus a social behavior (IP1: *“[...] the voice assistant behaves socially*

to the extent that it is in contact with people (...), that it apologizes, that is already polite. [...] that he lets you speak [...] remembers things”). The use of names during communication is also seen as social behavior that creates a more personal quality with the assistant (IP2). In addition, the omission of behaviors such as insults or lies is perceived as social (IP14); thus, social behavior can also be mediated by the absence of antisocial behavior. A further component of social behavior can also be the empathy shown and the response to the respective emotional state. However, the integration of emotions in digital voice assistants is only expected or preferred by a few (IP8). Most believe the responsibility for feelings is in humans and regard it as questionable if the voice assistant would be able to have any (IP10: “*Emotions? A person is responsible for that, in my eyes*”). For many, this ability of empathy would be too humanlike, and they would feel strange communicating with the assistant on an emotional level (IP5). In addition, integrity and friendliness are expected when dealing with personal data (IP16). The strengthening of the social area is also preferred by a digital voice assistant. This support is realized by means of small talk functions and attractive gimmicks.

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: “*When he realizes it's a child's voice or it's a voice from an older person because you can really hear it out, then he adapts to it.*”) or adapts itself to the respective user with its language peculiarities (IP20). 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.

UC4 “Anthropomorphic Attributes”

The fourth UC, “Anthropomorphic Attributes”, describes characteristics that make the digital voice assistant appear individual and human. The SCs are “Personality”, “Independence”, “Interaction”, “Voice”, “Appearance”, and “Competence and Intelligence”. There are different opinions concerning the integration of a personality. Some advocate a personality 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). In addition, if the assistant is assigned many human similarities, this could lead to disappointment because the device cannot fulfill the same expectations of the user as a human can (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). For a human being, empathy, reliability and patience are mentioned as further important qualities (IP1). In particular, the possibility of communication, which is characterized by mutual understanding and empathy, is seen as a relevant characteristic of a human being (IP18). The interviewees prefer a digital voice assistant who is characterized by patience, reliability and consistency of performance (IP1). Furthermore, attributes such as friendliness and helpfulness are listed, which are also relevant for the interviewees (IP2). It is also noticeable that a character is preferred who is similar to the user or at least largely agrees with him, so that sympathy can develop (IP2). Others indicate that they also find it interesting if the digital voice assistant has different values and opinions so that they could learn from it (IP17: “*I think I would be a person who would also like to have an artificial intelligence, who can also represent other opinions because I can only learn through that.*”). In general, for most participants, functionality is more important than personality (IP7).

With regard to the attributed independence, opinions differ again. Some of the respondents attribute autonomy to the voice assistance because it independently filters and executes information or commands and generates proposals itself (IP2: “*They are already very independent in the sense that when you give the order, they also do what is expected without any queries or anything like that.*”). Others do not regard it as independent because it does not do anything by itself but has to be activated by voice commands (IP7) and then only 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 digital voice assistant has a more lifelike effect due to its interactive character and the possibility of a joint conversation (IP9), thereby lowering the inhibition threshold for communicating with it (IP11). The

interviewees pay particular attention to the speed of response and the integrity of the commands and demands to be executed (IP12). IP20 takes a critical view of interaction and the associated increase in independence and is more reserved in its demand for more interaction with the assistant (*“It depends on how far he is interactive, i.e., independent, because interactive always means a certain independence. And there one would have to know, how far would that go then I would say ‘perhaps?’”*). Therefore, it becomes clear that especially with the integration of human characteristics in the digital voice assistant, opinions differ widely and have a great influence on the perception and evaluation of the it. This could be due to the uncanny valley effect.

The SC of "Voice" describes the elements that influence the perception of the digital assistant in relation to its liveliness and artificiality, as elements that convey liveliness, voice, appearance and movement are mentioned (IP1). 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 of the interview participants show that an increasingly human voice meets with more acceptance among voice assistants than does an artificial voice (IP2: *“Basically more human because I understand them better. Because if an artificial voice was very clearly developed, I still find through this distortion that it doesn't have this human melody [...]. [...] but it doesn't always have to be human one-to-one. So, there can also be gradations in human-likeness”*). Many of the respondents find a human voice pleasant because it is characterized by a certain softness and a soft flow of speech (IP3). The respondents' statements indicate that the majority have no preference for a female or male voice in certain areas (IP13). However, a minority of respondents stated that the choice of voice should be adapted to the context. The gender-specific voice should be adapted to the respective role model and thus indicates the use of a female voice in the drugstore area and the choice of a male voice in navigation (IP2). In general, however, the preference for a female voice is valid, since this is generally perceived as more pleasant (IP4). In addition, a middle-aged voice is prioritized because the competence of a young voice is perceived as doubtful (IP4). Regarding pitch, respondents agree that comfort is provided by a medium pitch and thus causes an increase in sympathy (IP18). Many of the respondents feel that an unpleasant voice is characterized by a very high pitch. Therefore, the preference of a voice of medium pitch applies (IP13). In addition, the respondents' statements indicate a preference for a successive speed (IP6). The volume is also mentioned by some of the respondents. A volume that is too loud is associated with obtrusiveness and is therefore perceived as unpleasant (IP14). A pleasant voice is also characterized by the attributes of a certain gentleness and friendliness (IP11). However, all these aspects can also constitute artificiality. Artificiality can be evoked, as can liveliness through voice, language and optics. Especially when the voice sounds too robot-like, tinny, indignant and monotonous, the assistant is perceived as artificial (IP2). Language is described as emotionless and repetitions of formulations as fake (IP11).

As far as the appearance is concerned, the voice assistant does not yet convey liveliness through its loudspeaker optic; nevertheless, when the assistant responds, the bright lights are perceived as living elements that imitate the rhythm of the language (IP1, IP3). In addition, it seems important that the digital voice assistant has eyes or an eye-catcher that can be used for orientation during interaction (IP2: *“So technical points for the eyes, as landmarks. Nose and such is not important to me. Mouth is actually also not. But eyes to be able to build up eye contact.”*). Furthermore, an important liveliness mediator is seen in the movement. Although this is not yet present in those digital assistants, it leads to a more human perception, e.g., a robotic vacuum cleaner (IP2). Similar to movement, facial expressions and gestures are considered important for anthropomorphization (IP11). Additionally, the optics can appear artificial from the form and material used and even reduce attributed humanity. (IP13). From this result, it can be concluded that the choice of voice, the appearance and the integration of movements can trigger or weaken anthropomorphism, depending on how the respective aspect is implemented. In general, it can be said that respondents would prefer a digital voice assistant that is more similar to a robot than to a human (IP4). In addition, there is a preference for the size of an assistant. Many of the respondents prefer a small size when choosing a voice assistant (IP17). Moreover, the appearance of the assistant favors an inconspicuous appearance (IP14). In general, respondents also prefer a segregated assistant that stands out as an additional unit and thus provides a fixed point of interaction (IP15). However, a minority feels that a humanlike appearance would promote sympathy for digital language assistants (IP11: *“Rather with a human looking one because that creates direct sympathy”*).

The SC of “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 abilities and reasons for this are manifold, and the idea of which components are considered intelligent is variable and multilayered. Most frequently, however, a foresighted answer to the question and the ability of the voice assistant to link and combine answers are described (IP20: *“For me it is intelligent [...] that it understands what I want and perhaps offers me alternatives that could also be relevant to the solution of my problem or question.”*). The ability to intuitively understand the user or to suggest alternative solutions was also positively mentioned by the respondents (IP16). Some participants also list scenarios such as the individual usage pattern created, processed and stored by the voice assistant, 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 as intelligent aspects (IP13). Only three of the respondents consider an implementation of intelligence to be irrelevant, either because they have concerns about it due to humanity losses or because they feel the devices are not intelligent because there is programming behind the system (IP19: *“I think the word intelligent is wrong in this context because in principle there is a programming behind it”*).

The perception of the participants referring to competence was very different here with regard to the acquisition of information and search functions of the digital voice assistant. The combination and implementation of given information is also often perceived as competent. The assistant should fulfil its purpose to satisfy the user, and the prerequisite for this is the qualitative answering of questions (IP7). The speed of implementation and interaction of the digital assistant and the simple use and operation are also mentioned here concerning competence (IP13). In isolated cases, speech recognition and processing, the presence and a human voice, which is in accordance with the functions of the assistant, is perceived as competent (IP5). In contrast, it was conspicuous that six of the test persons perceived poor quality in answering questions as incompetent (IP3: *“Competent, of course, if you can put it that way, if he can give me answers. Incompetent is when he gives me useless information, things I already know, things that don't help me.”*). In isolated cases, a robotic voice, inadequate speech recognition, the immature state of technology and the lack of security with regard to data protection are regarded as incompetent.

UC5 “Relationship”

This UC was divided into two SCs, namely, “Proximity and Distance” and “Emotions during use”. Some of the respondents mention 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). Some of the respondents also mention an aversion to an emotional attachment to an object and make this clear by the insignificance of parallels between a person and a digital assistant (IP5). Another aspect mentioned by the majority of the respondents describes situations in which interaction with the voice assistant was disrupted. The assistant often provides incorrect or inadequate information and results, which leads to anger (IP3). Speech recognition and processing were also frequently criticized by the respondents (IP14). A few times the lack of ability to recognize foreign languages is mentioned when assistants have been configured for a certain language in advance (IP7). Further points include wrong navigation or the linguistic input of technical terms, which does not function perfectly as yet and is mentioned as annoying. In a few of them, no more precise feeling could be described; it is rather described as neutral and not regarded as relevant (IP4: *“No feeling at all. It is simply accepted”*). In other cases, a negative feeling is explained, which arises when confidential information is passed on or when questions are not answered correctly. Some of these persons describe a feeling of discomfort and skepticism and, in one case, even fear (IP1). The majority of the respondents, however, describe a positive feeling when answering the questions. The descriptions of feelings were very individual and varied. Feelings such as trust, satisfaction, confirmation, understanding, joy, relief, and feeling at home were mentioned (IP15: *“It actually gives me a feeling of relief. It has a bit of the feeling of home when I think of Alexa now.”*).

Discussion and Conclusion

The evaluation of the interview results shows that many of the respondents have already come into contact with the topic of digital voice assistants. In general, all respondents have some form of voice assistant and can categorize their use. Many respondents mention the use of a digital voice assistant through smartphone integration, which supports the statement that the number of smartphone users using digital voice assistants has increased substantially over the years. The use focuses mainly on weather forecast information and assistance with tasks such as writing a text message. This outcome is based on the theoretical principles of reactive support. However, the statements also show a tendency towards proactive assistance. This is characterized, for example, by the fact that the user is reminded of upcoming events.

To answer the research questions from the beginning, (1) in general, anthropomorphism itself is manifested through the spoken language as a human characteristic through visual symbols such as a blinking ring or flashing points during the conversation and through human traits such as politeness, intelligence or social behavior. (2) Furthermore, social behavior, adaptability, similarity to the user, personality, independence, voice, appearance and interaction are all factors that can contribute to the anthropomorphization of a digital voice assistant, which positively influences human-computer interactions. (3) According to our third research question, the most important human characteristics are an interactive character, a humanlike voice, intelligence and social behavior to the extent that it is friendly, helpful, reliable.

Although this study primarily focuses on the triggers for anthropomorphizing rather than the reasons, we can stress that trust in particular plays a decisive role for why users are anthropomorphizing. IP3 stated, e.g., that “human traits create trust, so I know I can tell this bot everything without the data ending up anywhere.” This aspect seems to be a relevant point for future research. Surprisingly, in general, there were no significant differences between the non-users and the more experienced users concerning the evaluation of anthropomorphism probably because either the participants use digital voice assistants regularly or at least they know what they are capable of and tried some. Therefore, non-users can imagine what kind of assistant they would prefer, and which humanlike characteristics are important.

Since the parameter field of social behavior is one of the most important components for embodying anthropomorphization as the interviews have shown, it is assumed that the increasingly human characteristics of a robot are accepted. Therefore, the findings of the study suggest that the perceived social behavior and the adaptability of the behavior of a digital voice assistant can be a possible trigger for personification. 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 suggestion of a social behavior is sufficient as a signal to activate the *elicited agent knowledge* and to trigger a personification so that the user recognizes the voice assistant as a social counterpart and treats it as such. For example, Fussel could also show that humans regard a robot that reacts politely as less mechanistic than a rude one, which contributes to the hypothesis that social robots are perceived as more humanlike (Fussel 2008). 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 too high a human similarity, which has an uncanny effect on the user and can lead to a rejection of the assistant. Another explanation mentioned by some users is that emotions and empathy belong to human beings and not to robots. Therefore, especially for developers, caution seems to be required when integrating emotions into conversational agents.

The functionality factor determined in the study and the significance of the assigned role also influence the triggering of anthropomorphism. These findings are in line with those of Goetz's (2003) and Bernier's and Scasselatti's (2003) fit hypothesis to bring the anthropomorphic design of the digital voice assistant in line with its role, function and expectations. Dautenhahn (2007) formulates in his research that a robot is perceived as human if it can exhibit a certain functional ability and if its role as a machine is pushed into the background by its role as a friend.

Social behavior promotes the appearance of the liveliness of a robot and offers humans a space for the anthropomorphization of artificial intelligence. The appearance of a digital voice assistant, however, shows the strong rejection of a personification of the robot. Thus, users prefer a robot-like appearance, even if they perceive the human characteristics in the form of behavior as increasingly positive. The uncanny valley effect describes this process. The existence of the thin line between the acceptance of a humanlike appearance and the harmony with humanlike behavior and the rejection of it is described as an uncanny valley. Many of the interviewees feel that the increase in human characteristics in the form of behavior in a digital voice assistant

is positive, but they prefer the functionality (such as information presenting, switching on and off the lights or scheduling appointments) over these humanlike characteristics. However, if the voice assistant is a human duplicate, expectations are raised that it cannot meet. This result also clarifies the theory of anthropomorphism and reinforces that the depths of the uncanny valley are reached. Furthermore, many of the interviewees mentioned modifications to enable increasingly human behavior but did not consider this to be absolutely necessary. Moreover, the comparison between the expectations of the attributes of a human being and those of a digital voice assistant shows that there are similarities. Respondents expect patience and reliability from both a person and a voice assistant and would welcome humor. There is also a preference for strengthening the social sphere. In general, respondents find the promotion of small talk functions and gimmicks increasingly sympathetic.

In line with Lee et al. (2006), it was confirmed in our study that the digital voice assistant is assigned a personality that can help to understand its behavior through its similarity to humans. Contrary to our study in which conversational agents with a similar personality were desired by the user, Lee was able to show that a robot with a complementary personality is perceived as more intelligent, more attractive and more socially present than a robot with a similar personality and that the interaction with it is also experienced more positively. In addition to the results of the study, in which the personality is conveyed via the voice and the answer, Fong et al. (2003) finds that the personality can also be conveyed via emotions, the appearance and the way in which a task is completed. According to our study, another trigger of human perception can be an interactive character and thus independence. Kanda et al. (2004) find a positive correlation between an independently interactive robot and a human perception, which leads to the assumption that an interactive digital voice assistant has a more human effect. According to our study, anthropomorphization is additionally influenced by voice, appearance and elements of movement. In his research on robots, Zitzewitz (2013) also identifies these parameters as influencing factors of human similarity. Contrary to the many positive answers concerning digital voice assistants, it was noticed that almost half of the participants did not pass on any tasks to the voice assistant for which they had to disclose personal or private data. In addition, a smaller proportion of respondents stated that they did not want to share financial data, health data, security information such as passwords, and more complex information with an assistant to perform specific tasks in this context (IP8). This outcome means that in some cases, the necessary confidence is still lacking to entrust the voice assistant with complex tasks. A more humanlike behavior could strengthen the basis of trust and thus eliminate data transmission that has so far been viewed critically. However, this does not include the anthropomorphic appearance of a digital voice assistant. A strongly anthropomorphic appearance raises expectations that the robot may not be able to meet, which leads to a strong rejection of human characteristics. 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. There is an interesting contradiction in the statements regarding the appearance of a digital voice assistant that is particularly striking. Many respondents find a human voice to be extremely pleasant with digital voice assistants and experience an increase in sympathy when it is increasingly human. However, respondents perceive the opposite in the case of an increasingly human appearance of a voice assistant. Therefore, if the voice of the digital voice assistant is perceived to be humanlike, the acceptance to use it increases, probably because it appears have the feeling to interact with a real human.

Of course, there is also a dark side for the users. Companies can misuse the users' private data and endanger privacy by listening to private conversations without the users' knowledge. By doing so, they are able to take advantage of users, e.g., marketing strategies or special advertisements to increase their sales. Some users are already aware of the risk of misuse through eavesdropping. Nevertheless, it is important to strengthen the digital competence for all users in general so that they can develop a critical and balanced view. Particularly vulnerable people, such as children, older, disabled or sick people, are easy victims. Another point is that when emotions and empathy are integrated into digital voice assistants, these can also be used to persuade users to do or buy things, they do not truly want by appealing to their feelings. Especially in sensitive areas such as medicine or social fields, ethical guidelines for what a robot is allowed to do and what it is not allowed must be determined. Legal questions such as who is responsible if something happens and what kind of decisions are allowed to be made by an agent should be discussed before implementation.

Overall, it can be concluded and highlighted that anthropomorphism is positively perceived in a digital voice assistant if it shows a social, adaptive behavior, embodies a certain role as well as the implementation of personality, independence and interaction, with the voice playing a decisive role. The more integrated the

assistants were into everyday life, the greater the degree of anthropomorphization, even though all test persons were aware that they were communicating with a machine. Nevertheless, almost all participants find it more pleasant when the digital voice assistant has human features and prefer to use it in that case. For companies, this means that the implementation of human traits in voice assistants can lead to higher acceptance. Users experience a more positive emotional reaction to products and brands when they are anthropomorphized, which can ultimately lead to a preference for the respective product (Van Doorn et al. 2017). However, the findings also show that anthropomorphism is evaluated differently and is more complex: appearance should not be humanlike, whereas the voice should be and exhibit social behavior in the form of friendliness and reliability, but to the contrary, empathy and emotions cause discomfort by users. Therefore, this holistic study contributes that although functionality is an important factor concerning the acceptance of digital voice assistance it is not sufficient and anthropomorphism as a complex construct respectively the integration of chosen human characteristics into conversational agents support a positive evaluation of the agent and drives the acceptance and behavioral intention.

Implications for Management, Practice and Design

Our study provides many interesting findings for management, practice and design. The identified factors that trigger anthropomorphism and lead to higher acceptance could be useful for voice assistant developers and other researchers in further development. It seems to be important to first develop a uniform role of the assistant that is adapted to the functionality and adapt the design accordingly with regard to its anthropomorphic elements. For the developers of digital voice assistants, this outcome means that conversational agents should have an integrated imitation of social behaviors, demonstrating adaptability and an appearance appropriate to its role or task. Especially for companies, it is important to ensure that a combination of first the functionality and second the integration of human behavior is given to achieve the highest acceptance rate. Because functionality is also an important point concerning the perception of digital voice assistants, our study has shown that anthropomorphic elements foster successful acceptance. As the findings have shown, the relationship between the user and conversational agent is stronger when the voice assistant is used regularly. Thus, companies can increase their market share through habitualized use, e.g., through downstream services such as offering smart-home applications in connection with a digital voice assistant. However, the human likeness should not be adopted in regard to the appearance, so the design of the digital voice assistant should either be robot-like or abstract and inconspicuous. Otherwise, the users will feel uncomfortable, resulting in a rejection of the device. Just as no human appearance is preferred, caution seems to be required in regard to integrating emotions and empathy into digital voice assistants because users attribute these characteristics primarily to human beings and not to robots. However, it would be interesting for companies to integrate different degrees of emotions, so users can customize their individual preferences. This customization can also be senseful in terms of the agent's personality. It can be equipped with different personalities to be perceived as similar or contrary to the user himself. Additionally, companies have to build up the trust of the users concerning their data and protect their privacy through a high-level and understandable transparency policy, but since human characteristics can provide trust, these can make it much easier for companies to reach users.

Implications for Future Research and Limitations

Since emotions and empathy are seen as a critical integrated part of digital voice assistants, this result could be investigated in more detail when further studies examine how emotional a digital voice assistant should be, and in which situations and what kind of users prefer emotions, and which do not. It could also be dependent on the user character and the context of the preference they have. Additionally, it would be interesting to check whether humanlike characteristics such as apologies would overcome the effect of a partial dysfunction. As functionality is seen as an important factor, an experimental study could investigate whether this assumption can be leveraged through more humanlike traits. Moreover, further studies could examine whether users are more forgiving of humanlike agents than of real humans, e.g., in the context of service failures. Furthermore, one essential area of investigation could be that if companies succeed in increasing trust are users more willing to share private information, such as doing e-banking transactions.

Unfortunately, the study also has 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 in different age groups, and using a balanced gender distribution, and different types of occupations. Nevertheless, it would

be interesting to investigate these factors of anthropomorphism with a larger sample and to investigate the degree of humanlike characteristics in different contexts. As mentioned by some participants, there are privacy concerns that can be overcome through using a real human. Therefore, a further study can examine if the more humanlike a digital voice assistant produces less privacy concerns. In addition, despite that we included both users and nonusers in our study, every participant in our study had at least a general knowledge on digital voice assistants and their functions. We therefore believe that it would be interesting to replicate this study with completely inexperienced users who have never tried a digital voice assistant before and to compare the results with this study. Furthermore, this study was conducted in Germany. Germany is an IT savvy country, but according to the culture dimensions of Hofstede (2010) also, a country with a high “uncertainty avoidance” score resulting in skepticism towards innovations such as digital voice assistants. That is why we argue that these results can also be interesting for other countries that have a similar score. However, it would be interesting to conduct a cross-cultural study and examine whether other cultures are seeing anthropomorphism differently and to determine different roles anthropomorphizing plays depending on the country and to investigate this phenomenon with countries that differ clearly from the German “uncertainty avoidance” score.

References

- Airenti G. 2018. “The Development of Anthropomorphism in Interaction: Intersubjectivity, Imagination, and Theory of Mind”, *Frontiers in psychology* (9), pp. 2136.
- Bartneck, C., Kulic, D., Croft, E., Zoghbi, S. 2009. “Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots”, *International Journal of Social Robotics*, (1:1), pp. 71–81.
- Becker, C., Kopp, S., Wachsmuth, I. 2007. „Why emotions should be integrated into conversational agents”, *Conversational Informatics: An engineering approach*, pp.1–25.
- Bernier, E.P. and B. Scassellati 2010. "The Similarity-Attraction Effect in Human-Robot Interaction", in: *IEEE 9TH International Conference on Development and Learning*, pp. 286–290.
- Bickmore, TW., Utami, D., Matsuyama, R., Paasche-Orlow MK. 2016. “Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment”, *Journal of Medical Internet Research* (18:1).
- Bittner, E.A.C., Shoury, O. 2019: “Designing Automated Facilitation for Design Thinking: A Chatbot for Supporting Teams in the Empathy Map Method”, in: *Proceedings of the 52nd Hawaii International Conference on System Science*, pp. 227-236.
- Bollen, K. A., Paxton, P. 1998. “Detection and Determinants of Bias in Subjective Measures”, *American Sociological Review* (63:3), pp. 465–478.
- Castillo, S., Hahn, P., Ledge, K., Cunningham D.W. 2018. “Personality Analysis of Embodied Conversational Agents”, *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pp.227-232.
- Cooney, M., Zanzungo, F., Nishio, S. 2012. “Designing a Flying Humanoid Robot (FHR): Effects of Flight on Interactive Communication”, in: *IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pp. 364-371.
- Creswell, J. W. 2008. “Research Design: Qualitative, Quantitative, and Mixed Methods Approaches”, Thousand Oaks, CA: Sage.
- Creswell, J. W., Miller, D. 2000. “Determining validity in qualitative inquiry”, *Theory Into Practice* (39:3), pp. 124–130.
- Crowelly, C. R., Villanoy, M., Scheutzz, M., & Schermerhornz, P. 2009. “Gendered voice and robot entities: perceptions and reactions of male and female subjects”, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3735-3741.
- Dautenhahn, K. 2007. “Socially intelligent robots: Dimensions of human-robot interaction”, *Philosophical Transactions of the Royal Society B: Biological Sciences* (362:1480), pp. 679–704.
- De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., Parasuraman, R. 2016. “Almost human: Anthropomorphism increases trust resilience in cognitive agents”, *Journal of Experimental Psychology: Applied* (22:3), pp. 331-349.
- DiSalvo, C.F., Gemperle, F., Forlizzi, J., Kiesler, S. 2002. „All robots are not created equal: the design and perception of humanoid robot heads.” in: *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, pp. 321– 326. ACM, New York.
- Duffy, B.R. 2003. “Anthropomorphism and Robotics”, presented at the Symposium on Animating Expressive Characters of Social Interactions”, London.

- Eyssel, F., Kuchenbrandt, D., Hegel, F., De Ruyter, L. 2012. „Activating elicited agent knowledge: How robot and user features shape the perception of social robots”, in: *Proceedings of the 21th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2012)*, Paris. pp.851-857.
- Epley, N., A. Waytz, and J.T. Cacioppo, 2007. "On seeing human a three-factor theory of Anthropomorphism", *Psychological Review* (114:4), pp. 864–886.
- Ferrari, F., Paladino, M.P., Jetten, J. 2016., “Blurring human–machine distinctions: anthropomorphic appearance in social robots as a threat to human distinctiveness”, *International Journal of Social Robotics* (8:2), pp. 287-302.
- Fink, J. 2012. “Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction,” in: *Social Robotics. ICSR 2012. Lecture Notes in Computer Science*, Ge S.S., Khatib O., Cabibihan JJ., Simmons R., Williams MA. (eds), Vol. 7621. Springer, Berlin, Heidelberg, pp. 199-208.
- Fussell, S.R., Kiesler, S., Setlock, L.D., Yew, V. 2008. “How people anthropomorphize robots”, *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, pp.145–152.
- Goetz, J., Kiesler, S., Powers, A. 2003. “Matching robot appearance and behavior to tasks to improve human-robot cooperation”, in: *Proc. IEEE Roman*, pp. 55-60.
- Guest, G., MacQueen, K. M., Namey, E. E. 2012. “Applied thematic analysis“, Thousand Oaks, CA: Sage.
- Hauswald, J., Laurenzano, M.A., Zhan, Y., Li, C., Rovinski, A., Khurana, A., Dreslinski, R.G., Mudge, T., Petrucci, V., Tang, L., Mars, J. 2015. “An Open End End Voice and Vision Personal Assistant and Ist Implications for Future Warehouse Scale Computers”, in: *ACM SIGPLAN Notices* (50:4), pp.223–238.
- 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), pp. 1-25.
- Kaplan, B., Maxwell, J.A. 2005. “Qualitative Research Methods for Evaluating Computer Information Systems“, in: *Evaluating the organizational Impact of Healthcare Information Systems*, Hannah, K.J., Ball, M.J. (eds.), 2nd edition, Springer, New York, pp. 30-55.
- Kiesler, S., Goetz, J. 2002. “Mental models and cooperation with robotic assistants”, in: *Proceedings of CHI’02 on Human Factors in Computing Systems*, Minneapolis, MN: ACM Press, pp. 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), pp .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), pp. 95-105.
- Lester, J.C., Converse, S.A, Kahler, S.E., Barlow, S.T., Stones, B.A, Bhogal, R.S. 1997. “The persona effect Affective impact of animated pedagogical agents”, in: *Proceedings of CHI’97*, pp.359-366.
- Mayring, P. 2010, “Qualitative Inhaltsanalyse”, in: *Handbuch Qualitative Forschung in der Psychologie*, Mey, G. Mruck K. (eds.), Springer, Wiesbaden, pp. 601-613.
- Merkle, M. 2019. “Customer Responses to Service Robots Comparing Human-Robot Interaction with Human-Human Interaction”, in: *Proceedings of the 52nd Hawaii International Conference on System Sciences*, pp. 1396-1405.
- Mori, M. 1970. “The uncanny valley”, in: *Energy* (7:4), pp.33–35.
- Murray, I.R and 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, pp. 1097–1108.
- Myers, M.D. 1997. “Qualitative research in information Systems”, *MIS Quarterly* (21:2), pp.241-242.
- Nass, C., Steuer, J.S., Tauber, E. 1994. „Computers are social actors”, in: *Proceeding of the CHI ’94 Proceedings*, Boston, MA, pp. 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*, New York.
- Nass, C. and 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), pp. 171–181.
- Niculescu, A., van Dijk, B., Nijholt, A., Li, H., Lan See, S. 2013. ”Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy”, *International Journal of Social Robotics* (5:2), pp. 171 -191.
- Ortony, A. 2002. “On making believable emotional agents believable”, in: *Emotions in humans and artifacts*, R. Trapp, P. Petta, S. Payr (eds.), Cambridge, pp. 189-211.

- Persson, P., Laaksolahti, J., Lönnqvist, P. 2000. "Anthropomorphism—A multi-layered phenomenon", in: *Proc. Socially Intelligent Agents-the Human in the Loop*, AAAI Fall Symposium, Technical Report FS-00-04, pp. 131–135.
- Purinton, 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", in: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2853-2859.
- Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., Robinson, P. 2009. "How anthropomorphism affects empathy toward robots", in: *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 245-246.
- Robinson, O. C. 2014. "Sampling in interview-based qualitative research: A theoretical and practical guide", *Qualitative research in psychology*, (11:1), pp. 25-41.
- Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., Jinks, C. 2018. "Saturation in qualitative research: exploring its conceptualization and operationalization", *Quality & Quantity* (52:4), pp. 1893-1907.
- Seeger, A.-M., Pfeiffer, J., Heinzl, A. 2017: "When Do We Need a Human? Anthropomorphic Design and Trustworthiness of Conversational Agents", in: *SIGHCI 2017 Proceedings 15*.
- Shum, H.-Y., He, X.-D., Li, Di. 2018. "From Eliza to XiaoIce: Challenges and opportunities with social chatbots", *Frontiers of Information Technology & Electronic Engineering* (19), pp. 10-26.
- 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", in: *Proceedings of the Human Factors and Ergonomics Society*, pp.1418–1421.
- Stucke, M. E., Ezrachi, A. 2017. "How Digital Assistants Can Harm Our Economy, Privacy, and Democracy", *Berkeley Technology Law Journal*, 32., pp.1 239.
- Tadeusiewicz, R. 2010. "Speech in Human System Interaction", in: *3rd Conference on Human System Interactions*, pp.2–13.
- Tamagawa, R., Watson, C. I., Kuo, I. H., MacDonald, B. A., Broadbent, E. 2011. "The effects of synthesized voice accents on user perceptions of robots", *International Journal of Social Robotics* (3:3), pp. 253-262.
- Tractica 2018. "Voice and Speech Recognition Software Market to Reach \$6.9 Billion by 2025", <https://www.tractica.com/newsroom/press-releases/voice-and-speech-recognition-software-market-to-reach-6-9-billion-by-2025/> (accessed 10/4/2019).
- Van Doorn, J., Mende, M., Noble, S., Hulland, J., Ostrom, A.L. Grewal, D., Petersen, A. 2017. "Domo Arigato Mr. Roboto: Emergence of Automated Social Presence in Organizational Frontlines and Customers' Service Experiences", *Journal of Service Research* (20:1), pp. 43-58.
- Venkatesh, V., Thong, J. Y. L., & Xu, X. 2012. "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology", *Management Information Systems: MIS Quarterly* (36:1), pp. 157–178.
- Walters, M. L., Syrdal, D. S., Dautenhahn, K., Te Boekhorst, R., Koay, K. L. 2008. "Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion", *Autonomous Robots*, (24:2), pp. 159-178.
- Wang, E., Lignos, C., Vatsal, A., Scassellati, B. 2006. "Effects of head movement on perceptions of humanoid robot behavior", in: *HRI 2006: proceedings of the 2006 ACM conference on human-robot interaction*, pp.180–185.
- White, R. 1959. „Motivation reconsidered: the concept of competence“, *Psychological Review* (66:5), pp. 297-333.
- Yuan, L., Dennis, A. R. 2019. "Acting Like Humans? Anthropomorphism and Consumer's Willingness to Pay in Electronic Commerce", *Journal of Management Information Systems* (36:2), pp. 450-477.
- Zitzewitz, J. von, Boesch, P. M., Wolf, P., & Riemer, R. 2013. „Quantifying the Human Likeness of a Humanoid Robot“, *International Journal of Social Robotics* (5:2), 263–276.
- Zlotowski, J., Strasser, E., Bartneck, C. 2014. "Dimensions of anthropomorphism: from humanness to humanlikeness", in: *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction-HRI '14*, pp. 66–73.
- Zlotowski, J., Proudfoot, D., Yogeewaran, K., Bartneck, C. 2015. "Anthropomorphism. Opportunities and Challenges in Human–Robot Interaction", *International Journal of Social Robotics* (7:3), pp. 347–360, 348-350.
- Zlotowski, J., Sumioka, H., Eyssel, F., Nishio, S., Bartneck, C., Ishiguro, H. 2018. "Model of Dual Anthropomorphism: The Relationship between the Media Equation Effect and Implicit Anthropomorphism", *International Journal of Social Robotics* (10:5), pp.701–714.