Match or Mismatch? How Matching Personality and Gender between Voice Assistants and Users Affects Trust in Voice Commerce

Fabian Reinkemeier
University of Goettingen
fabian.reinkemeier@wiwi.uni-goettingen.de

Ulrich Gnewuch
Karlsruhe Institute of Technology (KIT)
ulrich.gnewuch@kit.edu

Abstract

Despite the ubiquity of voice assistants (VAs), they see limited adoption in the form of voice commerce, an online sales channel using natural language. A key barrier to the widespread use of voice commerce is the lack of user trust. To address this problem, we draw on similarity-attraction theory to investigate how trust is affected when VAs match the user's personality and gender. We conducted a scenario-based experiment (N = 380) with four VAs designed to have different personalities and genders by customizing only the auditory cues in their voices. The results indicate that a personality match increases trust, while the effect of a gender match on trust is non-significant. Our findings contribute to research by demonstrating that some types of matches between VAs and users are more effective than others. Moreover, we reveal that it is important for practitioners to consider auditory cues when designing VAs for voice commerce.

1. Introduction

Over the past decade, voice assistants (VAs) such as Amazon's Alexa and Google's Assistant have managed to reach a significant mass of consumers: More than 35% of U.S. adults have adopted smart speaker–based VAs [1]. Their usage shows a consistent pattern, with searching the web, listening to music, or setting a timer consistently among the most common use cases. However, customers can also use them to carry out online shopping activities (so-called voice commerce or voice shopping). These include personalized shopping journeys whereby users can perform tasks such as searching for, comparing, and purchasing products as well as placing reorders or checking delivery statuses. VAs thus enable customers to conveniently enter into a hands-free, online dialogue using natural spoken language—and all this with round-the-clock accessibility. Companies have recognized the potential of this technology and its promising use cases, driving them to prioritize investments in this new customer touchpoint [2].

Despite these many advantages of voice commerce, it is reported that in the United States only 14.1% of those who use smart speaker–based VAs regularly employ them to make purchases [1]. This limited uptake of voice commerce can be attributed to several obstacles: Users experience low interaction quality, limited transparency [3, 4], and an uncomfortable feeling when shopping via voice [5]. In addition, several studies have found that one of the most important barriers to the widespread use of voice commerce is a lack of trust [5, 6], especially when it comes to trust in the competence and benevolence of VAs [3]. Trust has also been recognized as one of the key success factors for adoption in studies on other information systems (IS) artifacts, particularly in the context of online shopping [7, 8]. Addressing this lack of trust is therefore of utmost importance in any endeavor to take advantage of voice commerce and provide customers with a pleasant experience, including greater decision satisfaction [6].

Drawing on similarity-attraction theory [9], research has shown that people tend to place more trust in those who are similar to themselves [10]. This applies not only to human–human interaction but also to IS: Existing approaches have demonstrated that attraction effects can also arise from endowing technologies with "anthropomorphic design elements, so-called social cues," [11] such as providing visual representations that match the user's ethnicity or gender. Designing VAs that resemble their users could therefore be a viable approach to reducing barriers in voice commerce. However, unlike IS such as recommendation agents (e.g., [12]) and chatbots (e.g., [13]), VAs are disembodied assistants. Therefore, the opportunities for designing VAs to resemble users are limited to the VA's voice, including the content of a message (what to say) and the way it is spoken (e.g., characteristics of the voice, how to say it).

When someone cannot be seen, their voice becomes more important; this is the case for podcasts and sales calls, where speakers are physically invisible. The auditory cues of the speaker's voice are then critical for the listener's perception and engagement intention [14,

15]. Auditory cues are those that can be heard but are not expressed in words. As such cues can be particularly powerful in conveying gender [16] and personality traits (especially the individual's degree of extraversion or introversion) [17, 18], research has largely focused on these two traits. However, studies have found varying results, depending on the characteristics and context [14].

In the field of VAs, voices can also be designed to exhibit certain personalities and gender traits: Amazon's Alexa, for example, is a VA with a name, gender, and personality [19]. While previous research has shown that users tend to trust others who are similar to themselves, it is unclear how to create a "match" between users and VAs based on auditory cues alone and whether such a match can increase users' trust [4]. Hence, we pose the following research question:

How do matches in (1) personality and (2) gender between VAs and users influence users' trust in VAs in voice commerce?

To address this research question, we conducted a scenario-based experiment in voice commerce with 380 participants and four differently designed VAs. We created distinct personalities (extraverted vs. introverted) and genders (male vs. female) by adjusting only the VAs' auditory cues (pitch, tempo, and volume of the voice).

Our study offers three major contributions. First, we advance our understanding of how (mis)matches between a VA and a user affect trust. More specifically, we reveal that certain matches (personality) are more effective than others (gender). Second, we suggest that modifying only the auditory cues of a VA is sufficient to signal a different personality without changing the content of its responses (e.g., language style). Third, we take a first step toward overcoming barriers in voice commerce adoption by analyzing user trust in VAs, which is particularly useful for practitioners aiming to provide their customers with a better voice commerce experience.

The paper is organized as follows: In the next section, we provide an overview of related work on VAs, ways to design them, and similarity-attraction theory. We then derive our hypotheses, describe our methods, and test as well as discuss various effects of personality and gender matches on user trust in voice commerce. Finally, we discuss our study's implications, its limitations, and suggestions for future research.

2. Theoretical foundations and related work

2.1. Voice assistants and auditory cues

Recent developments in artificial intelligence (AI) have fueled the rise of VAs as an important class of IS artifacts. VAs exhibit greater interactivity and intelligence than previous generations of virtual assistants or traditional software applications [20] and can be considered a type of conversational agent. However, they differ from IS artifacts such as chatbots and recommendation agents in that their visual interface is non-existent or only limited; they interact solely through spoken language. Prominent examples of VAs include Amazon's Alexa or Google's Assistant on smart speakers and Apple's Siri on iPhones. As with mobile apps on smartphones, third-party companies can offer customers various services by building their own VAs for these new ecosystems using readily available web services (e.g., Amazon's Alexa Skills Kit). These VAs—also known as *skills* for Alexa or *actions* for Google's Assistant—can feature not only simple functions (e.g., giving information about stock prices or local gasoline prices) but also more complex voice commerce activities, such as comparing and ordering products while offering answers to contextual questions. For example, customers can use company-specific VAs on their smart speakers to find the perfect blend of Johnnie Walker, request a ride from Uber, or quickly (re)order a meal at Just Eat—all using only voice commands. Although companies have identified the vast potential of VAs, they are often unsure about how VAs should be designed to foster a pleasant customer experience.

Following the "computers are social actors" (CASA) paradigm, companies can integrate social cues into the design of VAs, as people assign human traits to computers when interacting with them [16, 21]. Social response theory goes a step further, stating that those who see human-like traits in computers apply social rules to them and therefore treat the computers as if they were human [16, 21]. For example, people attribute a personality to computers, then apply personality-based social rules such as similarity attraction to them [17, 22]. Following this reasoning, people may perceive VAs as having various human traits, such as gender and personality [4, 19], whereupon different user actions might follow. Understanding and leveraging these social cues and their consequences could enable companies to optimize the design of their VAs.

Research in the domain of IS and human–computer interaction (HCI) has identified several dimensions of social cues for the design of conversational agents: auditory, verbal, invisible, and visual [23]. Auditory cues are of particular interest in this study, as they are the defining characteristic of VAs, differentiating them from traditional e-commerce websites and text-based chatbots. As intelligent and interactive voice-based HCI has only become possible with recent improvements in natural language processing, auditory cues represent an under-researched design dimension for conversational agents [23, 24, 25]. There are two main subdimensions

of auditory cues: voice qualities (e.g., gender, pitch range, speech rate, volume) and vocalizations (e.g., vocal segregations such as "Uh-huh" or laughing) [23, 26, 27].

A counterpart's personality and gender are important to people because these characteristics are strong determinants of how one behaves in interactions [16, 18, 25]. Personality comprises someone's behaviors, cognitions, and emotions, which are derived from both biological and social factors [28, 29]. Research has homed in on these traits, focusing particularly on the big five personality traits with its extraversion/introversion dichotomy [29]. Based on auditory cues, people assign gender and personality traits to technologies [16, 22]. For example, research has shown that users perceive a computer as extraverted rather than introverted when it speaks in a voice that is faster, louder, and of higher frequency [17, 22, 30]. Different gender perceptions can be detected based on elements such as the frequency of the voice, with people perceiving voices at about 120 Hz as male and 225 Hz as female [22].

Recent developments in AI have enabled companies to change the voice of a VA while keeping the VA's appearance somehow human-like [31]. Though they can give their assistants a desired personality and gender, the consequences of doing so remain unclear; there is limited empirical research in the context of conversational agents addressing the role of both users and VA personality [4] along with the influence of auditory cues [25]. In our study, we focus on analyzing the extraversion/introversion personality dimension [32] and the gender of the VA, as these can be strongly shaped by auditory cues [22, 26, 33].

2.2. Similarity-attraction theory

The central premise of similarity-attraction theory (also referred to as the law of attraction) is that people are more likely to be attracted to those who are similar to them than those who are dissimilar [9]. A large body of research in the field of human–human interaction has shown that relevant factors include both surface-level demographic characteristics (e.g., age, gender, ethnicity) and personality traits (e.g., degree of extraversion). Hence, people who share similar demographic characteristics and personality traits are more attracted to and trusting of one another [10].

Similar results have been found for interactions with IS such as recommendation agents or conversational agents. For example, recommendation agents that match the ethnicity of their users are perceived as more sociable, enjoyable, and useful than mismatched agents [12]. In the context of embodied conversational agents in eHealth applications, ter Stal et al. [34] found that participants preferred an agent image that looked similar to them in terms of age and gender. Moreover, in the context of websites, Lee and Nass [17] showed that people felt more social presence when listening to book descriptions in voices matching their own personality. Furthermore, an IS having personality traits similar to those of the user (e.g., dominant/submissive) also increases trust and the likelihood of accepting its advice [35, 36].

While research provides valuable insights into the positive effects of similarity in HCI, the opportunities to design VAs to be similar to their users are limited to their voices (what and how to say it); providing visual representations (e.g., a gendered avatar) is impossible. Therefore, there is little research on how VAs can be designed to match the gender and personality of the user and the effects of doing so.

3. Research model and hypotheses

In traditional sales interactions, customers evaluate the salesperson's abilities and similarities to themselves [37]. This can lead, for example, to a greater level of trust, which is a key factor in the customer–salesperson relationship [38]. Trust has also received much attention in the domain of HCI [37] and has been identified as an important factor in e-commerce [8]. At the same time, research has shown that people evaluate technologies in terms of their human-like traits, such as gender and personality [21], leading people to apply social rules (e.g., similarity attraction) to them [22, 33]. Therefore, in our study we focus on how perceived similarity between humans and VAs in terms of personality (extraverted vs. introverted) and gender (male vs. female) affect users' trust in VAs. More specifically, we hypothesize that VAs that match the user's personality and gender will inspire greater trust.

Qualitative research has shown that voice commerce faces challenges related to various dimensions of trust [3]. On this basis, and given that trust is a multidimensional construct, we analyze how matching the user affects three core trust dimensions (i.e., integrity, competence, and benevolence) [39], in line with recent research [40, 41]. First, integrity refers to the user's beliefs that the VA adheres to the principles accepted by the user (e.g., being honest). Second, competence refers to the VA's capability to effectively perform a designated task. Third, benevolence refers to the belief that the VA's motivation is to act in the best interest of the user [39, 41]. Figure 1 depicts our research model, which is explained in greater detail in the following sections.
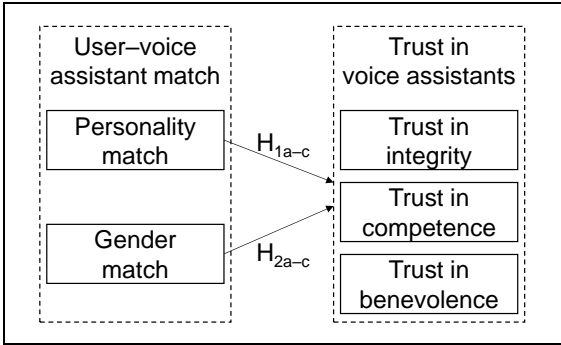
Figure 1. Research model

3.1. The effect of a personality match on trust in voice assistants

Several studies have shown that the design of a technology's auditory cues can lead users to attribute a personality to it [17, 22, 30]. As many people already assign personalities to existing VAs, such as Amazon's Alexa or Apple's Siri [4, 19], it is evident that personality characteristics are crucial in interactive settings. Therefore, it is important to understand their influence on user perception [30]. Similarity-attraction theory states that people who share personality traits are attracted to one another [9]. Based on this theory, research on websites has shown that users who hear a synthesized voice are more likely to be attracted if the voice matches their own personality (extravert or introvert) [22]. Furthermore, people who experience a personality match when listening to a machine-generated voice have an enhanced sense of the technology's social presence [17], which is an important basis for building trust in e-commerce services [8]. Taken together, similarity-attraction theory and previous HCI research lead us to propose the following hypothesis:

H₁: A personality match between the VA and the user has a positive effect on the user's trust in the VA's (a) integrity, (b) competence, and (c) benevolence.

3.2. The effect of a gender match on trust in voice assistants

In human–human communication, gender is one of the most studied auditory cues [14]. As people often assign a gender to technologies, it is clearly an important cue to consider when designing a VA [21]. However, there is no consensus on the role of matching the counterpart's gender. These mixed results make it particularly interesting for analysis in the context of VAs. Ter Stal et al. [34] found a significant correlation between the gender of respondents and their selected embodied agent design. Moreover, in the context of

virtual agents in real-world job interviews, a recent study showed a positive influence of gender similarity on trustworthiness [42]. Consistent with our previous reasoning based on similarity-attraction theory, we propose that interacting with a VA of the same gender whose voice contains similar auditory cues to one's own voice should create perceptions of similarity that increase trust. Therefore, we hypothesize the following:

H₂: A gender match between the VA and the user has a positive effect on the user's trust in the VA's (a) integrity, (b) competence, and (c) benevolence.

4. Method

To address our research question and test our hypotheses, we conducted a scenario-based experiment with four different VAs.

4.1. Experimental design

The experiment used a 2 (personality: extraverted vs. introverted) × 2 (gender: male vs. female) between-subjects design. Following previous research on text-based chatbots [13], we used the conversational design tool Botsociety [43] to create our four distinct VAs for the experiment. In doing so, we avoided using widely known voices, thus minimizing associations with specific VA brands and reducing potential biases linked to participants' prior experiences. Using the Amazon Polly service available in Botsociety, we could also individually configure the VA's auditory cues using speech synthesis markup language (SSML) for the whole interaction.

First, to create two distinct genders across the VAs, we selected the German voices *Hans* (a male, lower-pitched voice) and *Vicki* (a female, higher-pitched voice). Neither voice is used by default in any of the VAs available on the market.

Second, we created two different personalities. Previous research has shown that one can manipulate personality perceptions by changing only auditory cues [30]. For example, people ascribe a more extraverted personality to a voice that is faster, higher pitched, and louder [22, 26]. As auditory cues are a characteristic distinguishing VAs from other technologies, we included a combined set of these cues across the dialogue to render the interaction more extraverted or introverted. We focused on adjustments in the auditory cue subdimension of voice qualities, as it enabled us to make necessary adjustments for the whole interaction flow; changes in vocalizations would have required modifying individual messages of the VA. For example, we configured the speech rate for the female extraverted version to a higher tempo (151 words per minute) compared to the female introverted version (103 words

per minute). Table 1 provides an overview of our four experimental conditions.

Table 1. Manipulation of VA personality and gender

Group	Voice configuration				N
	Gender	Pitch	Tempo	Volume	
1 Extra	Male	High	Fast	High	94
2 Extra	Female	High	Fast	High	90
3 Intro	Male	Low	Slow	Low	102
4 Intro	Female	Low	Slow	Low	94

To ensure comparability between the prototypes, we developed a more guided dialogue flow for the experiment. This flow steered participants through a purchasing process using the same 10 interaction points (e.g., choice of book cover, payment method, delivery option) and the same script for all four experiment groups (following similar research, e.g., [40, 44]). In a first pre-study, participants ($N = 8$; 37.5% female, M_{age} = 26 years) were asked to role-play the situation and provide direct qualitative feedback on the realism of the human responses in the script. The results confirmed that the setup with short human responses is realistic in VA–human interaction.

Moreover, to ensure the validity and reliability of our experimental setting and questionnaire, we conducted a second pre-study ($N = 36$; 33.3% female, $M_{age} = 26.14$ years) that included a qualitative feedback free-form field. After making minor adjustments, we conducted the main study.

4.2. Procedure

The online experiment consisted of the following steps. At the beginning, participants were introduced to the topic of VAs, including their functionalities, hardware alternatives (e.g., VAs on smart speakers vs. on smartphones), and current main use cases. Participants then had to perform a sound check and optimally adjust the audio settings on the device they were using. They were instructed not to make any further changes to the audio settings during the experiment. Afterward, participants were tasked to read the experiment's scenario instructions and imagine buying a book via VA. We chose purchasing books because it seems to be a realistic scenario in voice commerce [40] and has been used successfully in similar experimental setups [17, 22]. Next, participants were randomly assigned to one of the four experimental conditions. In the experiment, they were directed to listen to an audio conversation between a VA and a human while putting themselves in the shoes of the human user. The conversation lasted between 104 and 154 seconds. After listening to the conversation, participants were asked to answer a control question about the interaction. Finally, they completed a post-experiment questionnaire that allowed us to measure our constructs.

4.3. Participants

We recruited our participants via social media, online distribution lists, and personal and university networks. Participation was rewarded with a raffle entry for one of two €25 Amazon vouchers. To ensure reliable results, we collected data from an initial large sample of 527 participants. We cleaned the data set for participants who were not on the experiment site long enough to listen to the entire conversation (97), had technical problems listening to the sound check or the conversation itself (7), or answered the control question incorrectly (11). Of the 412 who successfully completed the experiment, 32 failed to pass further attention or comprehension checks. The final sample then consisted of 380 participants (73.7% female, distributed almost equally among the four groups). The age of the subjects ranged from 17 to 72 years, with an average age of 32.7 years.

4.4. Measures and data analysis

We used validated scales from prior research and adapted them to our research environment (see Table 2 for an item overview). Participants were asked to rate the personality of the VA as well as their own using seven established personality adjective items each for their impression of the level of extraversion and introversion [30, 32]. Personality was measured on a 7-point Likert scale ranging from 1 (*extremely inaccurate*) to 7 (*extremely accurate*). As in similar research [40, 41], trust was assessed on three dimensions, with two items measuring trust in the VA's integrity and four items each assessing trust in its benevolence and competence [39, 45]. These were also measured on 7-point Likert scales, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Finally, we asked participants about their gender, age, disposition to trust [46], need for human interaction [47], product involvement for books [48], and experience with VAs. There were no significant differences with respect to these control variables between the experimental groups (all $p > .05$).

We found satisfactory results when assessing the reliability and validity of our measures (see Table 2). After dropping one item of extraversion, all factor loadings were higher than .6 [49]. Cronbach's alpha and composite reliability scores were above .7 [50]. The average variance extracted (AVE) of each construct was

Table 2. Constructs, items, Cronbach's α, CR, AVE, and factor loadings

Constructs	Items	Loadings	Sources
Extraversion of VA / user (α = .911 / .839, CR = .931 / .877, AVE = .693 / .547)	In the purchase process, the voice assistant appeared ... / Generally, I am …		[30, 32]
	… outgoing.	.816 / .698	
	… vivacious.	.868 / .749	
	… enthusiastic.	.866 / .794	
	… cheerful.	.845 / .804	
	… confident. (dropped)	*.690 / .544*	
	… extraverted.	.723 / .674	
	… jovial.	.860 / .728	
Trust in integrity (α = .834, CR = .923, AVE = .858)	In the purchase process, the voice assistant appeared ...		[39, 45]
	... honest with me.	.930	
	... sincere and genuine.	.923	
Trust in competence (α = .844, CR = .894, AVE = .679)	In the purchase process, the voice assistant was ...		
	… competent.	.888	
	… very effective.	.771	
	… very knowledgeable about the products.	.796	
	… capable of providing me suitable recommendations.	.837	
Trust in benevolence (α = .921, CR = .944, AVE = .808)	The voice assistant gave the impression …		
	... that its actions were in my best interest.	.930	
	... of doing its best to help me with the purchase.	.888	
	... of being interested in my needs and not someone else's.	.864	
	... of acting in my best interest.	.912	
Note. α = Cronbach's alpha; CR = composite reliability; AVE = average variance extracted.			

above .5 [51]. We compared all square roots of the AVEs for each construct with its correlations with others constructs (Fornell–Larcker criterion) and found satisfying discriminant validity [52]. As the ability of this test to reliably detect problems of discriminant validity has recently been questioned [53], we also applied the heterotrait–monotrait ratio of correlations (HTMT) criterion; the values were below the recommended threshold of .9 [54].

To check whether our manipulation of the personality of the VAs was successful, we compared users' perceptions of VA extraversion between the extraverted and introverted designs. The results confirmed that VA extraversion was perceived to be significantly higher in the extraverted conditions than in the introverted ones ($p < .001$).

To calculate the personality match between the VA and the participant, we used the R package psy to compute a dyadic personality similarity score using pairwise intraclass correlations between the user's extraversion and that of the VA. The values of the intraclass correlation coefficient (ICC) range from −1.0 to 1.0, with 1.0 indicating a perfect match. Consistent with the approach described in the literature [36, 55], we dichotomized the dyadic similarity scores into two groups by a median split and coded a match as 1 and a mismatch as 0. Finally, the gender match was also assessed as a binary variable. More specifically, we created a variable where 1 indicated a match—assistant and human were both male or both female—and 0 designated a mismatch.

5. Results

5.1. Descriptive results

Table 3 provides the descriptive statistics for the three trust dimensions in the four experimental conditions.

Table 3. Descriptive statistics

Match		Trust		
		Integrity	Competence	Benevolence
Personality	0	4.38 (1.59)	5.04 (1.36)	4.56 (1.60)
	1	5.02 (1.25)	5.56 (1.12)	5.16 (1.37)
Gender	0	4.65 (1.48)	5.22 (1.24)	4.78 (1.52)
	1	4.75 (1.45)	5.40 (1.30)	4.95 (1.51)
Note. 0 = mismatch; 1 = match; means with standard deviations in parentheses; items were measured on 7-point Likert scales.				

5.2. Structural model and hypothesis testing

To test our proposed relationships in the structural model and perform additional analyses, we used the PLS-SEM approach in SmartPLS version 3 and a bootstrap resampling method with 5,000 samples.

The results support the proposed positive effects of a personality match on trust in integrity ($β = .216$, $p < .001$), competence ($β = .235$, $p < .001$), and benevolence ($β = .196$, $p < .001$). This provides support for $H_{1a–c}$. Contrary to our expectations, the paths from gender

match on trust in integrity (β = .036, p = .463), competence (β = .070, p = .182), and benevolence (β = .059, p = .236) have no significant effects, leading us to reject H_{2a-c}.

Furthermore, we found some significant effects of the control variables on the latent variables: First, male users placed less trust in the VA's benevolence (β = −.138, p < .05). Second, older participants tended to perceive less trust in the VA's integrity (β = −.154, p < .01) and benevolence (β = −.139, p < .01).

Given that most current VAs have a "female-by-default" design [56], participants' pre-existing expectations about the voice's gender may have influenced our results and could provide an explanation for the non-significant effects of gender matching. Therefore, we conducted a follow-up multigroup analysis to compare the effect of a gender match on trust between the male and female VAs. The results show a significant difference in the effects of a gender match on trust in benevolence (β = −.293, p < .05), while the differences in the effects on trust in competence and integrity were non-significant. More specifically, in the female VA condition, we found a significant positive effect of a gender match on trust in benevolence (β = .172, p < .05), while this effect was negative but not significant in the male VA condition (β = −.119, p = .129). This finding indicates that the effect of a gender match depends on the VA's gender.

6. Discussion

In this study, we investigated how personality and gender matches between VAs and users—created via auditory cues alone—influence user trust in voice commerce.

The results of our experiment show that users trust a VA more (in terms of its integrity, competence, and benevolence) when its personality is similar to their own. In contrast, matching VAs to the user's gender had no significant effect on trust. In the following, we discuss the implications of our findings, acknowledge the limitations of our study, and identify further research opportunities.

6.1. Theoretical contributions

Our study makes three contributions. First, we contribute to IS literature on the design of VAs. The general assumption based on similarity-attraction theory is that people prefer interacting with people and technologies that exhibit characteristics similar to their own [9]. Our study offers a more nuanced view of this assumption by revealing that some types of matches are more effective than others: While a match in personality increases trust, the effect of a gender match is non-

significant. This lack of gender effect may be because most commercial VAs are designed to be female or have a female voice [56], leading users to have preconceived expectations about a VA's voice [57]. This could also explain the results of our post hoc analysis indicating that a gender match has a positive effect on trust in benevolence for a female VA but not for a male VA.

Second, our results extend prior research on social responses to VAs [e.g., 13, 22, 33] by demonstrating how different personality attributions can be created based on auditory cues alone. While previous research has shown that such attributions can be created using a different language style (e.g., more assertive statements) [55], we show that distinct personality impressions can be triggered in human–VA interactions by adjusting only auditory cues—namely voice pitch, tempo, and volume—without changing the content of messages and regardless of the VA's gender. Thus, in contrast to previous research, which has mainly focused on visual or verbal cues (what to say), we investigated the effect of designing different personalities with auditory cues (how to say it).

Third, our findings contribute to previous research that has analyzed existing barriers to voice commerce adoption and how to overcome them (e.g., [3, 40, 58]). We also add to research on trust that has recently called for studies analyzing trust in VAs on multiple dimensions (i.e., integrity, competence, benevolence) [6]. More specifically, we found some differential effects on the dimensions of trust. In the treatment conditions with female VAs, a gender match had a positive effect on users' trust in the VA's benevolence but not its competence and integrity.

6.2. Practical implications

Voice commerce offers customers the opportunity to experience intuitive back-and-forth dialogues in e-commerce by interacting with a VA using natural language. Given the importance of this new customer touchpoint for companies [59], our study also offers useful insights for practitioners.

First, VA developers and designers should not focus solely on technical components and the formulation of VA responses; there is also the opportunity to customize a VA's auditory cues. SSML provides a good approach to create unique personalities that can match the users' personalities and is available for all VAs (e.g., the Amazon Polly service for Amazon's Alexa). For example, VA designers can individually configure the pitch, volume, and tempo of a VA's voice. Currently, companies tend to design verbal cues for each VA response (e.g., adding small talk, different language styles). This seems impractical for longer conversations because each message would

have to be customized to the desired VA personality. Hence, compared to other design approaches, customizing auditory cues appears to be more promising for companies, as they can use SSML to modify the personality relatively easily for the entire conversation and also iteratively adapt to a certain user personality or mood in an ongoing conversation.

Second, companies should avoid pursuing a one-size-fits-all strategy when designing VAs. Instead, as corroborated by our results, companies could match the VA to their users' personalities. To implement such a strategy, companies that use VAs in voice commerce or plan to do so need to know their users. Because users are often unwilling to share sensitive personality-related information with companies, automated approaches could be used to infer gender and personality. Companies could analyze the user's voice to recognize their characteristics (e.g., extraverts speak faster and louder than introverts) [22] and then customize the voice of the assistant. However, major adaptations, such as changing the VA's gender, should be avoided during conversations, as they could cause undesirable disruptions to the customer experience. Furthermore, VA developers and designers should consider going beyond the default gender of a voice. As research has shown that gender attributions to technology can exacerbate gender stereotypes [16, 57], designers should critically reflect on the use of female voices as the default.

In summary, companies should take advantage of this new touchpoint and provide a personalized design according to their target groups or specific customers. As mentioned previously, VAs lack users' trust—for example, in their benevolence and competence [3]. By enabling a match between the personality of the VA and that of the user, companies can take a big step toward overcoming a key challenge in voice commerce.

6.3. Limitations and future research

There are limitations to our study that offer future research opportunities. First, consistent with previous research on conversational agents [13, 58, 60], we chose a scenario-based experiment to maximize internal validity and control for confounding influences. Therefore, future research should seek to increase external validity by confirming our findings in real-life user–VA interactions (e.g., field experiments). This would also address the current limitation whereby each participant had to imagine being in the role of the same user, whose voice may not necessarily reflect the participant's actual voice. In addition, analyzing real voice input could provide further valuable insights, including the user's current emotional state during the interaction [56]. Such insights could then be leveraged to create adaptive VA designs that automatically adjust to match user characteristics (e.g., personality) and mood (e.g., frustration or relaxation in the voice).

Second, we considered only binary genders and females were overrepresented in our sample. Although participants were equally distributed among the four experimental conditions, future studies should strive for even gender distributions that also include non-binary genders to increase the generalizability.

Third, we focused on the match of two important individual characteristics: a personality trait (i.e., extraversion) and gender of voice [14, 18]. As we found interesting effects of personality matches, it may be worth analyzing additional personality traits (e.g., agreeableness) and other types of personalization (e.g., based on context or customer profile). Furthermore, as we did not find a significant effect of gender match, implementing a gender-ambiguous voice like the first genderless voice Q [61] could be a viable alternative for companies. More research is needed to investigate the effects of such voices; this could also have important implications from an ethical perspective.

A fourth limitation of the current study is that we examined initial trust perceptions, similarly to previous research [6, 7, 40]. However, if users interact with a particular VA over a longer time frame, they might perceive its voice to be more familiar—regardless of whether it matches their personality—and change the way they form trust because they know how to interact with the VA. Future studies could therefore test our research model in a longitudinal design to see how trust develops. To get a more complete picture, researchers could study other forms of trust (e.g., emotional trust) [62] and their effects on actual behavior.

Fifth, the operationalization of social cues and how users process them and build trust could be analyzed from different perspectives of information processing (e.g., in light of the elaboration likelihood model; [63]). It is possible that trust could be built not only through the way content is communicated but also through the content itself. In the future, it may be useful to compare the effectiveness of the two strategies (what to say vs. how to say it).

7. References

[1] Voicebot, Smart Speaker Consumer Adoption Report: April 2021.

[2] Abramovich, G., Adobe Study: Voice Commerce Rises As Top Priority, 2019.

[3] Rzepka, C., B. Berger, and T. Hess, "Why Another Customer Channel? Consumers' Perceived Benefits and Costs of Voice Commerce", in Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020.

[4] Cowan, B.R., N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Earley, and N. Bandeira, "What can I help

you with?: Infrequent Users' Experience of Intelligent Personal Assistants", in Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, Vienna, Austria. 2017.

[5] Voicebot and Voysis, Voice Shopping Consumer Report: June 2018.

[6] Mari, A. and R. Algesheimer, "The Role of Trusting Beliefs in Voice Assistants during Voice Shopping", in Proceedings of the 54th Hawaii International Conference on System Sciences. 2021.

[7] Benbasat, I. and W. Wang, "Trust in and Adoption of Online Recommendation Agents", Journal of the Association for Information Systems, 6(3), 2005, pp. 72–101.

[8] Gefen, D. and D. Straub, "Managing User Trust in B2C e-Services", e-Service Journal, 2(2), 2003, pp. 7–24.

[9] Byrne, D., The attraction paradigm, Academic Press, New York, 1971.

[10] Byrne, D., W. Griffitt, and D. Stefaniak, "Attraction and similarity of personality characteristics", Journal of Personality and Social Psychology, 5(1), 1967, pp. 82–90.

[11] Diederich, S., A.B. Brendel, and L.M. Kolbe, "Designing Anthropomorphic Enterprise Conversational Agents", Business & Information Systems Engineering, 62(3), 2020, pp. 193–209.

[12] Qiu, L. and I. Benbasat, "A study of demographic embodiments of product recommendation agents in electronic commerce", International Journal of Human-Computer Studies, 68(10), 2010, pp. 669–688.

[13] Ahmad, R., D. Siemon, and S. Robra-Bissantz, "Communicating with Machines: Conversational Agents with Personality and the Role of Extraversion", in Proceedings of the 54th Hawaii International Conference on System Sciences. 2021.

[14] Zoghaib, A., "Persuasion of voices: The effects of a speaker's voice characteristics and gender on consumers' responses", Recherche et Applications en Marketing (English Edition), 34(3), 2019, pp. 83–110.

[15] Liu, x., F. Liu, Y. Li, E. Lim, and C.-W. Tan, "Modulating Your Voice: The Role of Paralinguistic Cues in Improving Podcasters' Competitiveness", in Twenty-Third Pacific Asia Conference on Information Systems. 2020.

[16] Nass, C. and Y. Moon, "Machines and Mindlessness: Social Reponses to Computers", Journal of Social Issues, 56(1), 2000, pp. 81–103.

[17] Lee, K.-M. and C. Nass, "Social-Psychological Origins of Feelings of Presence: Creating Social Presence With Machine-Generated Voices", Media Psychology, 7(1), 2005, pp. 31–45.

[18] Kaplan, A.D., T. Sanders, and P.A. Hancock, "The Relationship Between Extroversion and the Tendency to Anthropomorphize Robots: A Bayesian Analysis", Frontiers in Robotics and AI, 5, 2019.

[19] Purington, A., J.G. Taft, S. Sannon, N.N. Bazarova, and S.H. Taylor, "Alexa is my new BFF", in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, Colorado, USA. 2017.

[20] Maedche, A., C. Legner, A. Benlian, B. Berger, H. Gimpel, T. Hess, O. Hinz, S. Morana, and M. Söllner, "AI-Based Digital Assistants: Opportunities, Threats, and Research Perspectives", Business & Information Systems Engineering, 61(4), 2019, pp. 535–544.

[21] Nass, C., J. Steuer, and E.R. Tauber, "Computers are Social Actors", in Proceedings of the ACM CHI Conference on Human Factors in Computing Systems, Boston, USA. 1994.

[22] Nass, C. and K.M. Lee, "Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction", Journal of Experimental Psychology: Applied, 7(3), 2001, pp. 171–181.

[23] Feine, J., U. Gnewuch, S. Morana, and A. Maedche, "A Taxonomy of Social Cues for Conversational Agents", International Journal of Human-Computer Studies, 132, 2019, pp. 138–161.

[24] Seeger, A.-M., J. Pfeiffer, and A. Heinzl, "Designing Anthropomorphic Conversational Agents: Development and Empirical Evaluation of a Design Framework", in Proceedings of the 39th International Conference of Information Systems (ICIS), San Francisco, CA, USA. 2018.

[25] van Pinxteren, M.M., M. Pluymaekers, and J.G. Lemmink, "Human-like communication in conversational agents: a literature review and research agenda", Journal of Service Management, 31(2), 2020, pp. 203–225.

[26] Lee, K.M. and C. Nass, "Designing social presence of social actors in human computer interaction", in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA. 2003.

[27] Nass, C. and L. Gong, "Speech interfaces from an evolutionary perspective", Communications of the ACM, 43(9), 2000, pp. 36–43.

[28] Hall, C.S., G. Lindzey, and J.B. Campbell, Theories of Personality, John Wiley & Sons Inc., Hoboken, NJ, 1957.

[29] Robert, L.P., "Personality in the Human Robot Interaction Literature: A Review and Brief Critique", in Proceedings of the 24th Americas Conference on Information Systems, New Orleans, LA. 2018.

[30] Lee, K.M., W. Peng, S.-A. Jin, and C. Yan, "Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction", Journal of Communication, 56(4), 2006, pp. 754–772.

[31] Guzman, A.L., "Voices in and of the machine: Source orientation toward mobile virtual assistants", Computers in Human Behavior, 90, 2019, pp. 343–350.

[32] Wiggins, J.S., "A psychological taxonomy of trait-descriptive terms: The interpersonal domain", Journal of Personality and Social Psychology, 37(3), 1979, pp. 395–412.

[33] Isbister, K. and C. Nass, "Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics", International Journal of Human-Computer Studies, 53, 2000, pp. 251–267.

[34] ter Stal, S., M. Tabak, H. op den Akker, T. Beinema, and H. Hermens, "Who Do You Prefer? The Effect of Age, Gender and Role on Users' First Impressions of Embodied Conversational Agents in eHealth",

International Journal of Human-Computer Interaction, 36(9), 2020, pp. 881–892.

[35] Al-Natour, S., I. Benbasat, and R. Cenfetelli, "The Adoption of Online Shopping Assistants: Perceived Similarity as an Antecedent to Evaluative Beliefs", Journal of the Association for Information Systems, 12(5), 2011, pp. 347–374.

[36] Gnewuch, U., Y. Meng, and A. Maedche, "The Effect of Perceived Similarity in Dominance on Customer Self-Disclosure to Chatbots in Conversational Commerce", in Proceedings of the 28th European Conference on Information Systems. 2020.

[37] Bickmore, T.W. and R.W. Picard, "Establishing and maintaining long-term human-computer relationships", ACM Transactions on Computer-Human Interaction, 12(2), 2005, pp. 293–327.

[38] Rajaobelina, L. and J. Bergeron, "Antecedents and consequences of buyer-seller relationship quality in the financial services industry", International Journal of Bank Marketing, 27(5), 2009, pp. 359–380.

[39] McKnight, D.H., V. Choudhury, and C. Kacmar, "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology", Information Systems Research, 13(3), 2002, pp. 334–359.

[40] Reinkemeier, F., W. Toporowski, and K. Gutknecht, "Voice Apps in Voice Commerce: How Anthropomorphic Design Can Influence Users' Perceptions and Behavioral Intentions", in Proceedings of the 50th EMAC Conference. 2021.

[41] Wang, W. and I. Benbasat, "Empirical Assessment of Alternative Designs for Enhancing Different Types of Trusting Beliefs in Online Recommendation Agents", Journal of Management Information Systems, 33(3), 2016, pp. 744–775.

[42] Akbar, F., T. Grover, G. Mark, and M.X. Zhou, "The Effects of Virtual Agents' Characteristics on User Impressions and Language Use", in Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion, Tokyo, Japan. 2018.

[43] Botsociety, "Where teams design chatbots and voice", https://botsociety.io, accessed 9-7-2021.

[44] Gnewuch, U., S. Morana, C. Heckmann, and A. Maedche, "Designing Conversational Agents for Energy Feedback", in International Conference on Design Science Research in Information Systems and Technology (DESRIST). 2018.

[45] Qiu, L. and I. Benbasat, "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems", Journal of Management Information Systems, 25(4), 2009, pp. 145–181.

[46] Gefen, D., E. Karahanna, and D.W. Straub, "Inexperience and experience with online stores: The importance of tam and trust", IEEE Transactions on Engineering Management, 50(3), 2003, pp. 307–321.

[47] Dabholkar, P.A. and R.P. Bagozzi, "An Attitudinal Model of Technology-Based Self-Service: Moderating Effects of Consumer Traits and Situational Factors", Journal of the Academy of Marketing Science, 30(3), 2002, pp. 184–201.

[48] Malär, L., H. Krohmer, W.D. Hoyer, and B. Nyffenegger, "Emotional Brand Attachment and Brand Personality: The Relative Importance of the Actual and the Ideal Self", Journal of Marketing, 75(4), 2011, pp. 35–52.

[49] Gefen, D. and D. Straub, "A Practical Guide To Factorial Validity Using PLS-Graph: Tutorial And Annotated Example", Communications of the Association for Information Systems, 16, 2005, pp. 91–109.

[50] Nunnally, J.C. and I.H. Bernstein, "The Assessment of Reliability", Psychometric Theory, 3(1), 1994, pp. 248–292.

[51] Bagozzi, R.P. and Y. Yi, "On the evaluation of structural equation models", Journal of the Academy of Marketing Science, 16(1), 1988, pp. 74–94.

[52] Fornell, C. and D.F. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error", Journal of Marketing Research, 18(1), 1981, pp. 39–50.

[53] Henseler, J., C.M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling", Journal of the Academy of Marketing Science, 43(1), 2015, pp. 115–135.

[54] Teo, T.S.H., S.C. Srivastava, and L. Jiang, "Trust and Electronic Government Success: An Empirical Study", Journal of Management Information Systems, 25(3), 2009, pp. 99–132.

[55] Al-Natour, S., I. Benbasat, and R. Cenfetelli, "The Role of Design Characteristics in Shaping Perceptions of Similarity: The Case of Online Shopping Assistants", Journal of the Association for Information Systems, 7(12), 2006, pp. 821–861.

[56] Luo, X., S. Tong, Z. Fang, and Z. Qu, "Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases", Marketing Science, 38(6), 2019, pp. 913–1084.

[57] West, M., R. Kraut, and H.E. Chew, I'd blush if I could: Closing gender divides in digital skills through education, 2019.

[58] Whang, C. and H. Im, "I Like Your Suggestion! the role of humanlikeness and parasocial relationship on the website versus voice shopper's perception of recommendations", Psychology & Marketing, 38(4), 2021, pp. 581–595.

[59] Dawar, N. and N. Bendle, "Marketing in the Age of Alexa", Harvard Business Review, 05-06, 2018, pp. 80–86.

[60] Mozafari, N., W.H. Weiger, and M. Hammerschmidt, "Resolving the Chatbot Disclosure Dilemma: Leveraging Selective Self-Presentation to Mitigate the Negative Effect of Chatbot Disclosure", in Proceedings of the 54th Hawaii International Conference on System Sciences. 2021.

[61] Copenhagen Pride, Virtue, Equal AI, Koalition Interactive, and thirtysoundsgood, "Meet Q The First Genderless Voice", https://www.genderlessvoice.com, accessed 9-7-2021.

[62] Komiak, S.Y.X. and I. Benbasat, "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents", MIS Quarterly, 30(4), 2006, pp. 941–960.

[63] Petty, R.E. and J.T. Cacioppo, "The Elaboration Likelihood Model of Persuasion", Advances in experimental social psychology, 19, 1986, pp. 123–205.