Journal of Service Management

Fluent voice, credible choice – how language embodiment and consumption goals shape voice commerce success

Melanie Schwede and Maik Hammerschmidt
Chair of Marketing and Innovation Management, Faculty of Business and Economics, University of Goettingen, Goettingen, Germany, and
Stefan Morana
Chair of Digital Transformation and Information Systems, Faculty of Empirical Humanities and Economics, Saarland University, Saarbruecken, Germany

Received 30 January 2024
Revised 31 July 2024
13 December 2024
28 January 2025
Accepted 29 January 2025

Abstract

Purpose – Voice assistants (VAs) could be a game changer in conversational commerce. However, consumers often reject purchase recommendations by VAs due to disfluency and low credibility of VAs' messages. To overcome these barriers and increase recommendation adoption, this study examines which language style leads to voice commerce success and how consumer-related boundary conditions shape this effect.

Design/methodology/approach – In three experiments, the authors collected data from consumers listening to VA recommendations in different styles and commerce settings. The first study examines the effect of figurative (vs literal) language on recommendation adoption, while the second study focuses on the moderating impact of consumers' consumption goals (hedonic vs utilitarian). The third study delves into the underlying mechanisms leading to enhanced visual fluency in voice interactions.

Findings – The experiments demonstrate that figurative language increases visual fluency, with no differences across consumption goals (hedonic vs utilitarian). However, figurative language enhances credibility in a hedonic context while reducing it in a utilitarian one. Visual fluency and credibility mediate the effect of figurative language on recommendation adoption. The results further reveal the crucial role of arousal for triggering visual fluency in response to figurative language.

Originality/value – This research examines how and when using language styles from human-to-human communication context can enhance service interactions in the voice-driven marketplace. It shows that a VA's figurative language can lead to a trade-off between visual fluency and credibility, but a match with the right consumer goal eliminates it. Therefore, this research provides new insights on when embodied voice systems either limit or boost voice commerce success.

Keywords Voice assistant, Voice commerce, Language style, Consumption goal, Fluency, Credibility

Paper type Research paper

We would like to thank editor Jay Kandampully and special issue editor Arne de Keyser, as well as the two anonymous reviewers, for their constructive guidance and thoughtful feedback during the development of this article. We are also grateful for the opportunity to present an earlier version of this article at the Virtual Thought Leader Workshop for the Journal of Service Management special issue "Rise of Voice Conversation Capabilities in Smart Service Systems". We thank Sina Ueberländer for her assistance in implementing the experimental setup and data collection for Study 2 and Imke Eisenträger for her assistance in implementing the pre-test for Study 1. In addition, we greatly appreciate Simon Trang's thoughtful comments on our statistical analyses. Finally, we would like to thank Thorsten Hennig-Thurau, Gianfranco Walsh, and all participants of the 2024 Research Bootcamp on Marketing for their valuable feedback.

Journal of Service Management
Emerald Publishing Limited
e-ISSN: 1757-5826
p-ISSN: 1757-5818
DOI 10.1108/JOSM-01-2024-0051

Introduction

The development of voice technologies, such as ChatGPT Voice, Google Gemini, and Alexa's LLM treatment, transforms how consumers interact with digital devices. These technologies incorporate the latest advances in Artificial Intelligence (AI) and natural language processing, making conversations with VAs more natural and intuitive (Schmitt *et al.*, 2023). VAs are AI-based artifacts interacting with consumers through voice commands and responses (Diederich *et al.*, 2022; Zierau *et al.*, 2023). VAs are embedded in different mobile and interactive devices, such as smartphones (e.g. ChatGPT Voice), tablets or laptops (e.g. Cortana, Google Gemini), smart speakers at home (e.g. Amazon Alexa, Google Assistant), or vehicles (e.g. Volkswagen's IDA; Coker and Thakur, 2023; Guha *et al.*, 2023). This widespread integration allows consumers to utilize VAs for simple tasks, such as turning on lights, and more complex tasks, such as information search or shopping (Melumad, 2023). Because more than 70% of consumers prefer to use voice commands instead of relying on their keyboards when searching online (Voicebot.ai, 2023), firms increasingly utilize voice technology for commercial purposes (Forbes, 2023). VAs have the potential to facilitate purchase decisions by providing personalized product or service recommendations (Blut *et al.*, 2023). However, consumers' adoption of voice recommendations lags behind (Coker and Thakur, 2023). Even when consumers frequently use a VA for search requests, they often quit the interaction instead of adopting the recommendation from the VA, such as purchasing a suggested product or booking a table at a recommended restaurant. Two challenges to the adoption of purchase-related recommendations stand out: difficulties in processing VAs' recommendations (Melumad and Meyer, 2024; Schindler *et al.*, 2023; Schwede *et al.*, 2022) and credibility concerns regarding these VAs' messages (Flavián *et al.*, 2023; Luo *et al.*, 2019).

Processing recommendations from VAs can be challenging for consumers due to the ephemeral nature of voice, particularly when they lack access to screens (Zierau *et al.*, 2023). To get consumers to adopt recommendations by VAs and make a purchase, firms need to design voice recommendations that are easy to understand without a screen. Therefore, visual fluency, or the ease of mentally visualizing the recommended product, is crucial (Landwehr *et al.*, 2011). One way to improve visual fluency is to use language styles that stimulate the imagination of consumers. Language embodiment is a valuable technique in this respect that uses principles of human dialogue to create the appeal of an embodied system (Bergner *et al.*, 2023). For example, using a figurative language style, which contains vivid expressions (Choi *et al.*, 2019), can facilitate consumer's mental imagery and hence visual fluency. However, in voice commerce, literal language, as opposed to figurative language, is so far the norm, in which words are used in their exact, basic meaning (Choi *et al.*, 2019). According to language expectancy theory (LET), if consumers do not expect figurative language to appear in typical voice commerce conversations, using this language style could be detrimental to the credibility of the assistant's recommendations. Our research examines this potential trade-off of improving visual fluency at the expense of lowered credibility when using figurative language. In addition, we investigate how these psychological mechanisms influence purchase in terms of adopting the recommendation.

RQ1. How does language embodiment (figurative vs literal language) of VAs influence recommendation adoption through visual fluency and credibility?

To facilitate purchases, it is imperative for firms to identify conditions in which this trade-off can be mitigated. This research, therefore, investigates how consumers' consumption goals, an important boundary condition in VA interactions (Schmitt *et al.*, 2023), influence the effectiveness of figurative language in VA communication. Consumers pursue either a hedonic (affective motivation) or utilitarian (functional motivation) consumption goal, depending on their motivation when purchasing or using a product or service (Batra and Ahtola, 1991; Botti and McGill, 2011). While products and services often have hedonic and utilitarian attributes, a consumer's consumption goal determines which attributes dominate for purchase decisions (Voss *et al.*, 2003). For instance, a restaurant may be chosen for a private meeting with friends

or a business dinner, while headphones may be selected for leisure or work purposes. This associates leisure activities with hedonic consumption and work activities with utilitarian consumption (Wien and Peluso, 2021). According to the task-technology fit theory (TTF theory; Goodhue and Thompson, 1995), a match between the task (consumer's consumption goal) and technology (VA's language embodiment) leads to positive consumer evaluations. This research explores how these interactions affect voice commerce success. Hence, the second research question asks:

RQ2. How does the consumption goal, hedonic vs utilitarian, shape the effects of language embodiment of VAs?

This study contributes to research on the impact of VAs at the smart service frontline. First, our study introduces figurative language as a novel language style for VAs that facilitates information processing despite the lack of a visual interface. By doing so, this study expands research regarding embodied systems (e.g. Bergner *et al.*, 2023), emphasizing the significance of vivid messaging in voice commerce.

Second, we highlight that using figurative language in voice commerce can be a "double-edged" sword. While our findings indicate that the vividness of figurative language increases visual fluency, they also show that this benefit comes at the price of decreasing credibility in the context of utilitarian consumption goals. While LET is suitable for explaining the decrease in credibility, it does not adequately explain the increase in visual fluency. Our results suggest that arousal, rather than expectancy disconfirmation, is the mechanism responsible for the increase in visual fluency. These results advance LET by adding new insights that allow it to sufficiently explain the effects of unexpected language styles. The study further highlights that the trade-off between visual fluency and credibility fully mediates the relationship between language embodiment and recommendation adoption.

Third, we shed light on the relevance and applicability of TTF theory in voice commerce. Interestingly, our research indicates that while the impact of figurative language on credibility is influenced by consumption goals (i.e. hedonic vs utilitarian), for the impact on visual fluency, these goals do not play a role. Hence, our results demonstrate that a fit between the technology (language style of the VA) and the task (consumer's consumption goal) does not always result in improved consumer evaluations. This finding suggests that the benefits of matching technology and tasks in voice commerce are less universal than previously thought.

For practitioners, this study offers new insights into the role of language embodiment as a new instrument in voice commerce. It provides a nuanced understanding of the impact of using unexpected, innovative language styles and the interplay between language embodiment and consumers' consumption goals. These insights can guide firms in designing VAs that enhance sales by implementing the right language style for a particular consumer segment.

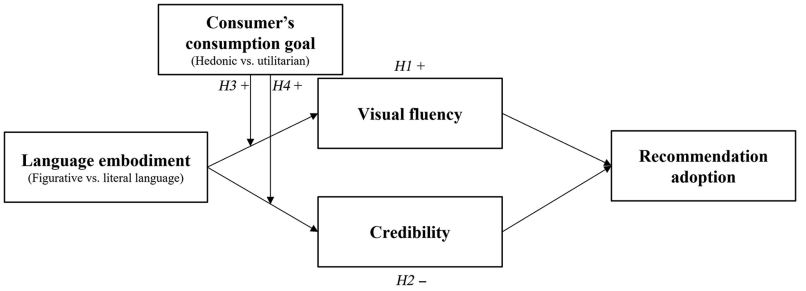
Theoretical background and hypotheses

This section first derives the relevance of visual fluency and credibility of VAs and presents extant research on VA's language style as a key factor for shaping these psychological mechanisms. Then we elaborate on how these mechanisms mediate the relationship between language embodiment and recommendation adoption and discuss the role of consumers' consumption goals. Figure 1 presents our research model along with the associated hypotheses.

Visual fluency and credibility in human-voice assistant interactions

Our research examines non-human assistants that provide purchase recommendations solely through spoken communication, lacking visual support. This reliance on voice comes with two key challenges.

First, without visual elements (e.g. an illustrated product) and with spoken words being transient, consumers face challenges in processing recommendations (Zierau *et al.*, 2023). To



Source(s): The above figure was created by the authors

Figure 1. Research model

overcome this, VAs should communicate in a way that makes it easy for consumers to construct mental images of the recommended products or services. This ease of mental visualization is referred to as *visual fluency*, which describes the cognitive effortlessness with which consumers form mental representations of objects (Landwehr *et al.*, 2011), such as a recommended product. For example, when consumers receive a recommendation for a restaurant or headphones, they assess how easily they can visualize the described features and whether the process feels relaxing or exhausting, happens instantly or takes a long time (Landwehr *et al.*, 2011). Notably, consumers generally perceive information that is easier to process as more favorable, a pattern that extends to commerce (Kim and McFarland, 2024).

Second, communication must establish credibility, as consumers need to trust that VAs are acting in their best interest (Kim and Duhachek, 2020; Luo *et al.*, 2019). Credibility refers to the perceived reliability and trustworthiness of a source's message (Filieri, 2015; Flavián *et al.*, 2023), such as a VA's recommendation. Credibility is particularly crucial when consumers lack access to complete information and must rely on the source's judgment. This reliance is especially prominent in voice commerce, where consumers evaluate whether a VA's recommendation is fair, accurate, trustworthy, and convincing. Such assessments are critical for adopting recommendations as consumers often view VAs as algorithm-driven systems that lack human expertise (Luo *et al.*, 2019).

Voice assistants' language style

Since VAs rely solely on verbal cues to provide information, examining different language styles – the framing of messages – becomes a critical factor for shaping visual fluency and credibility. In this regard, researchers emphasize that designing such assistants requires careful consideration of elements from human dialogue (Gnewuch *et al.*, 2022a). Therefore, prior research has examined four different pairs of language styles: social- vs task-oriented, informal vs formal, high vs low empathy, and abstract vs concrete. The social-oriented, informal, and highly empathic language styles promote social-emotional exchange by using friendship-like, caring expressions. In contrast, the task-oriented, formal, and low empathic language styles are more goal-oriented and, therefore, uses expressions in a professional manner (Chattaraman *et al.*, 2019; Mari *et al.*, 2024; Rhee and Choi, 2020; Whang and Im, 2021). These language styles primarily aim to build relationships between consumers and VAs. In comparison, the abstract vs concrete language style distinction focuses mainly on information dissemination. Abstract language, such as "intelligence" or "brilliance" seeks to inform about intangible qualities and concepts while concrete language, such as "she scored 95 out of 100 on the test", provides highly specific, tangible details (Lan *et al.*, 2024).

However, although the use of vivid expressions like figurative language in human dialogue is common, no empirical study in the field of VAs has examined such language styles. This is a

crucial shortcoming, as success in screenless commerce may depend on the ability of VAs to evoke mental images through figurative language. However, mimicking such human-like language in machines may come at the expense of credibility evaluations due to expectancy disconfirmation. This potential trade-off has been neglected in prior research on VA language styles, even though identifying if and when it occurs is essential for designing voice assistants that trigger positive evaluations while mitigating negative ones.

Online Appendix A provides a more in-depth review of prior research on the language styles of VAs and their impact on consumer behavior.

Language embodiment and language expectancy theory

To address the challenges of visual disfluency and low credibility, language embodiment is a promising approach. This is because language embodiment entails the use of human dialogue principles by non-human entities, such as typically human-like language styles, to create the appeal of a system that mimics a human (i.e. an "embodied system"; Bergner *et al.*, 2023, p. 744). Figurative language is a style typically used in human conversations (Luri *et al.*, 2024). It involves words or phrases with a meaning that is different from the literal interpretation, often relying on metaphorical or symbolic language to create vivid mental images (Sopory and Dillard, 2002). Metaphors (e.g. time is a thief), comparisons (e.g. as brave as a lion), hyperbole (e.g. I have a million things to do), or idioms (e.g. it is raining cats and dogs) are typical elements of figurative language. In contrast to figurative language, literal language refers to using words in their basic, direct sense (Kronrod and Danziger, 2013). If someone uses the expression "it is raining cats and dogs" and it is literally true, it would mean that cats and dogs are falling from the sky. Therefore, literal language is essential for factual communication, while figurative language can evoke emotional associations (Wu *et al.*, 2017). Humans naturally use figurative phrases in their everyday conversations with other people to give their messages an expression that draws on associative memory networks and can therefore be understood quickly (Choi *et al.*, 2019; Sopory and Dillard, 2002). Therefore, this language style is highly suitable to embody a VA.

To theoretically derive the effect of language embodiment of VAs on recommendation adoption through visual fluency, we rely on language expectancy theory (LET). LET provides a fundamental framework for comprehending the dynamics of persuasive communication (Burgoon *et al.*, 2002; Miller and Burgoon, 1979). LET is based on the idea that individuals hold expectations about appropriate communication, influenced by social norms (Burgoon *et al.*, 2002). However, as in any rule-governed system, also in language systems individuals can choose to follow or violate norms (Averbeck and Miller, 2014). Commonly, communicators follow the rules, which confirms and reinforces their normative status. However, communicators can also interact in ways that violate expectations leading to expectancy disconfirmation (Burgoon, 1993). Since machines do not commonly use figurative language but have been using literal language for a long time (Choi *et al.*, 2019), the use of figurative language is perceived as a norm violation, so it disconfirms consumer expectations. Due to expectancy disconfirmation in response to a figurative message, consumers experience the message more intensively (Burgoon *et al.*, 2002). It means that consumers are aroused, which provides the level of energy needed for unlocking the processing capacity to construct mental images of a recommended service or product (Yin *et al.*, 2017). Consequently, figurative language is more effective than literal language in supporting consumers' visual processing, thereby increasing visual fluency. High fluency has been shown to have strong implications for consumer behavior as it creates a sense of familiarity through subjective ease of processing (Landwehr *et al.*, 2011; Schwarz *et al.*, 2021), thereby reducing the risk of adopting purchase recommendations. Thus, we assume that visual fluency mediates the relationship between figurative (vs literal) language and recommendation adoption. Accordingly, we formulate the following hypothesis:

H1. The effect of language embodiment on recommendation adoption is mediated by visual fluency, such that visual fluency and in turn recommendation adoption is higher for a VA's figurative than literal language.

According to LET, consumers' perceptions of expectancy disconfirmation can lead to changes in credibility (Averbeck and Miller, 2014). While figurative language has been shown to have positive disconfirmation effects in face-to-face communications that enhance credibility (Chang and Yen, 2013), these findings do not apply uniformly across all contexts. For instance, Wu *et al.* (2017) found that the use of figurative language by non-experts in online reviews decreases persuasion because it is negatively evaluated by consumers. Similarly, Choi *et al.* (2019) demonstrate that consumers prefer literal language when interacting with AI assistants. This preference stems from the perception that these assistants are neither autonomous nor experts for product recommendations (Luo *et al.*, 2019). Consequently, any deviation from literal language through VAs, such as the use of figurative language, may lead to mistrust. According to LET, in this case consumers will evaluate the unexpected figurative language negatively and initially reject it, resulting in a decrease in perceived credibility. As the credibility of an AI assistant is closely related to the adoption of its recommendations (Yen and Chiang, 2021), we expect the following:

H2. The effect of language embodiment on recommendation adoption is mediated by credibility, such that credibility and in turn recommendation adoption is lower for a VA's figurative than literal language.

Consumer-related context factors and task-technology fit theory

The impact of language embodiment may vary depending on consumers' preferences. Therefore, task-technology fit theory (TTF theory) emphasizes the match between the characteristics of a task that consumers prefer to perform and the characteristics of the technology they use to perform it (Goodhue and Thompson, 1995). The lower the fit between task and technology, the lower the consumer's evaluation of the technology. In contrast, a high fit positively affects a consumer's evaluation of the technology and its performance (Goodhue and Thompson, 1995). However, tasks can have various characteristics, such as the goal that motivates performing the task (Rzepka *et al.*, 2022). With regard to the task of buying a specific product or service, consumers may pursue either a hedonic or utilitarian consumption goal. Thus, the way a technology communicates, such as its language style, and the consumer's predominant goal should match.

Hedonic consumption is associated with emotional and multisensory stimulation (e.g. enjoyment), while utilitarian consumption relates to factual and rational stimulation (e.g. efficiency; Longoni and Cian, 2022). We suggest that consumers with a hedonic consumption goal, who value enjoyment and fun, and VAs that use figurative language might be a close match (Batra and Ahtola, 1991; Hirschman and Holbrook, 1982). This is because figurative language employs emotional elements such as metaphors or hyperboles, which may stimulate vivid experiences that fit to the multi-sensory needs of consumers who pursue hedonic goals (Castelo *et al.*, 2019; Kronrod and Danziger, 2013; Ruan and Mezei, 2022). Compared to hedonic consumption goals, consumers with utilitarian consumption goals prioritize objective and functional aspects (Dhar and Wertenbroch, 2000). Consumers perceive utilitarian consumption as necessary and refer to such purchases as "work" that brings satisfaction mainly upon completion (e.g. Babin *et al.*, 1994). We therefore assume that a utilitarian consumption goal and a literal language style fit best, as they are both inherently straightforward. It is important to note that the very same product can be perceived differently depending on whether a hedonic or utilitarian goal prevails, and these goals are not mutually exclusive (Batra and Ahtola, 1991). If a fit between the language embodiment (technology characteristic) and the consumption goal (task characteristic) is present, it may trigger "feeling right" experiences (Goodhue and Thompson, 1995).

Prior research has shown that a "feeling right" experience due to perceiving a close match between elements is associated with a sense of ease in processing (e.g. Lee *et al.*, 2010; Lee and Aaker, 2004). Furthermore, Lee and Labroo (2004) show that when information fits the consumer's motivational stance it is more fluid to process – a phenomenon referred to as "fit fluency" (Graf *et al.*, 2018, p. 397). Similarly, the study of Schwede *et al.* (2022) demonstrates that information from VAs are more difficult to process, however, Schindler *et al.* (2023) show that this difficulty varies depending on the product type, indicating a matching effect. Therefore, we propose that matching VAs' figurative language with consumers who are in a hedonic (vs utilitarian) consumption mood triggers visual fluency, making the construction of mental images of a product or service less effortful (Novemsky *et al.*, 2007). As argued above, triggering the experience that imagining the recommended product is easy (i.e. visual fluency) enhances purchase decisions. Therefore, we hypothesize:

H3. If a consumer pursues a hedonic (vs utilitarian) consumption goal, a VA's figurative language increases visual fluency which in turn enhances recommendation adoption.

In the realm of interactions with VAs, the perceived accuracy and trustworthiness of the assistant are key factors in evaluating its communication credibility (Flavián *et al.*, 2023). Nonetheless, consumers believe that AI assistants do not act autonomously, making consumers less likely to trust its recommendations (Kim and Duhachek, 2020). However, according to TTF theory, with a fit between the language style and consumer's consumption goal, a consumer may attribute a high level of capability to the technology to autonomously accomplish certain tasks. Therefore, in case of a close match with a hedonic (vs utilitarian) consumption goal, even figurative language can be perceived as credible although it deviates from the expected language style. As established above, if consumers believe that a VA's recommendation is credible, the likelihood of adopting this recommendation increases (Flavián *et al.*, 2023). We therefore assume that a fit between figurative language and a hedonic (vs utilitarian) consumption goal increases credibility. Hence:

H4. If a consumer pursues a hedonic (vs utilitarian) consumption goal, a VA's figurative language increases credibility which in turn enhances recommendation adoption.

Study 1

Study 1 tests the effect of language embodiment (figurative vs literal language) on recommendation adoption using a two-cell between-subject experiment. The study also considers the expected mediating role of perceived visual fluency and credibility of the VA's recommendation. In addition, in a pre-study, we will examine expectancy dis-/confirmation of figurative and literal language to support the baseline assumption in our hypotheses regarding the effects of figurative language on visual fluency and credibility.

Method

Participants. We recruited participants from a European country through the crowdsourcing platform Prolific. We financially compensated participants based on the legal minimum wage. The participants did not have to meet any special requirements to take part in the study. The final sample comprises 131 participants (47.33% female, M_{age} = 27.82 years, SD = 7.71), which are randomly and evenly distributed across the two scenarios. There are no significant differences between scenarios regarding participants' prior experience with VAs (i.e. the extent of prior usage of VAs), audio repetition (i.e. the frequency of listening to the VA message), and socio-demographics (all $p > 0.1$). See Online Appendix B for an overview of participants' socio-demographics across all studies and Online Appendix D1 for power analysis and data cleaning procedures.

Experimental procedure and task. In both conditions, participants listened to a restaurant recommendation they had requested from a car's VA (audio files are available on request). The

VA used either figurative or literal language when providing the restaurant recommendation. We adopted the language style manipulation by Choi *et al.* (2019) as well as Kronrod and Danziger (2013). We used metaphors (e.g. "the location on the beach will blow you away"), comparisons (e.g. "the service is like that of a royal reception"), and hyperboles (e.g. "the restaurant looks like a royal palace") for the manipulation of the figurative language. In contrast, we used words in their exact, basic meaning without metaphorical or vivid aspects to manipulate the literal language (e.g. "the interior of the restaurant is beautifully decorated" or "the service is very professional"). The VA's recommendation included relevant service features commonly found in restaurant recommendations (Choi *et al.*, 2019). As a result, both audio outputs are 32 s in duration. We used the Neural2-F voice from Google Dialogflow for the audio output. The choice of a female voice was motivated by consumers' preference for female voices (Tolmeijer *et al.*, 2021), as consumers perceive female voices as friendlier and more pleasant than male voices in service settings (Mitchell *et al.*, 2011). The voice, speaking rate, pitch, and volume were identical for both audio outputs. Participants could listen to the audio output as many times as they desired. For a complete overview of the scenario description and language embodiment manipulations, please refer to Online Appendix D2. The participants of the main study perceived the scenarios as realistic ($M = 5.11$, $SD = 1.69$; "The scenario is realistic"; Bagozzi *et al.*, 2016).

Measures. We used previously validated measures for all constructs and slightly adapted them to the context of this study. After the experiment, participants answered questions about their perceptions of language embodiment, visual fluency, credibility, and recommendation adoption. All items, except for socio-demographics, were measured on 7-point Likert scales anchored by 1 = strongly disagree to 7 = strongly agree, unless otherwise stated. For a comprehensive list of all constructs and corresponding items, please refer to Online Appendix C. Online Appendix D3 lists the reliability and validity measures. The tables show that the factor loadings, composite reliability, Cronbach's α, convergent, and discriminant validity are above the recommended thresholds (Fornell and Larcker, 1981; Hair *et al.*, 2019). Moreover, Online Appendix D4 provides a detailed account of the procedural and statistical remedies implemented to address common method variance (CMV). Based on these measures, we confidently conclude that common method bias (CMB) is not a significant concern in this study.

Results

Pre-study results. Before testing our hypotheses, we conducted a pre-study ($N = 33$; 45.45% female, $M_{age} = 26.12$ years, $SD = 6.09$) to assess the validity of our baseline assumption that figurative language leads to expectancy disconfirmation. The results confirm that participants who received the figurative language expected a different language style ($M_{figurative} = 5.76$, $SE = 0.39$), while those who received the literal language expected this type of language ($M_{literal} = 3.13$, $SE = 0.46$). This difference is significant ($t(31) = -4.43$, $p < 0.001$), and the results remained robust after controlling for participants' experience with VAs and audio repetition. Online Appendix C lists the item of the expectancy dis-/confirmation scale while Online Appendix D5 shows the pre-study results that confirm the validity of our manipulations, the reliability of the constructs, and the scenario realism check.

Manipulation check for the main study. The manipulation check for perceived language embodiment in terms of figurative versus literal language was successful ($M_{figurative} = 6.03$, $SE = 0.15$; $M_{literal} = 2.35$, $SE = 0.17$; $t(129) = -16.33$, $p < 0.001$).

Hypotheses testing for the main study. To initially test the effect of language embodiment on visual fluency and credibility, we used two one-way analyses of variance (ANOVA) with language embodiment as the independent variable and visual fluency and credibility as the dependent variables. The results indicate that language embodiment significantly affects visual fluency ($F(1, 129) = 4.88$, $p < 0.05$, $d = 0.39$). Specifically, figurative language leads to significantly higher visual fluency than literal language ($M_{figurative} = 5.01$, $SE = 0.16$;

$M_{literal} = 4.50$, $SE = 0.17$); with a small effect size according to Cohen's (1992) classification. Furthermore, language embodiment has a significant effect on credibility ($F(1, 129) = 22.46$, $p < 0.001$, $d = 0.83$), whereas figurative language leads to significantly lower perceived credibility than literal language ($M_{figurative} = 4.12$, $SE = 0.13$; $M_{literal} = 5.03$, $SE = 0.14$); with a large effect size. Online Appendix D6 presents descriptive statistics, including means and standard deviations, for all constructs in both the figurative and literal language conditions.

We tested mediation using a bootstrapping procedure with 5,000 iterations to estimate indirect effects, generating bias-corrected confidence intervals (Zhao *et al.*, 2010). The results indicate significant indirect effects of figurative (vs literal) language on recommendation adoption through both visual fluency ($\beta = 0.06$, $SE = 0.04$, $CI_{95\%} = [0.003, 0.201]$) and credibility ($\beta = -0.68$, $SE = 0.16$, $CI_{95\%} = [-1.019, -0.401]$). The results support H1 and H2 as the confidence intervals do not include zero. Moreover, as expected, the results show a positive indirect effect through visual fluency and a negative indirect effect through credibility. Additionally, there is no significant direct effect of language embodiment on recommendation adoption ($p = 0.359$), suggesting indirect-only mediation (Zhao *et al.*, 2010). Moreover, the mediation path of credibility is significantly stronger than that of visual fluency ($\beta = -0.62$, $SE = 0.18$, $CI_{95\%} = [-0.977, -0.292]$). All results remain robust after controlling for prior experience with VAs and audio repetition [1]. Prior experience may shape consumers' perceptions and behaviors through increased familiarity with VAs (Dabholkar, 1996), while audio repetition may influence how consumers process and interpret a VA's recommendation through increased familiarity with the content (Mohsenin and Munz, 2024; see Online Appendix D7 for detailed items and results).

Discussion. The pre-study results indicate that participants do not expect figurative language (compared to literal language) in VA interactions. In the main study, figurative language, compared to literal language, enhances visual fluency but diminishes credibility, ultimately influencing recommendation adoption.

Study 2

Study 2 has two main goals. First, it seeks to enhance the generalizability of the findings from Study 1 by using electronics as an alternative product context. Second, it examines whether consumers' consumption goals, hedonic vs utilitarian, influence the impact of figurative language. Study 1 demonstrated opposing effects of figurative language on visual fluency and credibility indicating that a trade-off between both evaluations exists. Therefore, Study 2 provides evidence on whether this undesirable trade-off can be mitigated under certain boundary conditions in terms of the consumer's consumption goal. For Study 2, we conduct a 2 (figurative vs literal language) × 2 (hedonic vs utilitarian consumption goal) between-subject experiment.

Method

Participants. We recruited participants from a European university via mailing lists and social media to assure robustness of results across different data collection modes. We did not require any prerequisites for participants to take part in the study. After completing the survey, participants could take part in a raffle for online shopping vouchers. The final sample consists of 317 participants (70.66% female, $M_{age} = 29.56$ years, $SD = 10.42$), which are randomly and evenly distributed across the four scenarios. There are no significant differences between scenarios regarding participants' prior experience with VAs, audio repetition, and socio-demographics (all $p > 0.1$). Online Appendix E1 details the power analysis and data cleaning process.

Experimental procedure and task. In all conditions, we instructed participants to listen to a recommendation for over-ear headphones they had previously requested from their smart speaker. In addition to this information, the scenario description included the manipulation of

the consumption goal. We manipulated the consumption goal for the same product, based on Wien and Peluso's (2021) work. The manipulation of the hedonic consumption goal underlined the need for leisure use by stating that listening to music or podcasts is one of the participant's favorite activities, thus giving the headphones a high emotional value (e.g. "You need the headphones for personal use, to listen to music or podcasts"). In contrast, manipulating the utilitarian consumption goal emphasized the headphones' role in supporting work tasks and their highly functional value for the participant (e.g. "You need the headphones for your work, to answer customer inquiries and attend online meetings").

After the scenario description, participants listened to an audio output (audio files are available on request) in which the VA provided them with a headphone recommendation using either figurative or literal language. We adopted the language style manipulation from Choi *et al.* (2019) as well as Kronrod and Danziger's (2013) studies. As in Study 1, for the manipulation of language, we used metaphors, comparisons, and hyperboles for figurative language, and words in their basic, literal meaning for literal language. The VA's recommendation included relevant product features commonly used in headphone recommendations, adapted from prior work by Kronrod and Danziger (2013). As a result, both audio outputs have a duration of 46 s. The voice conditions used in Study 2 were the same as those in Study 1, including the same female voice, speaking rate, pitch, and volume. For a complete overview of the stimulus material, please refer to Online Appendix E2. The participants of Study 2 perceived the scenarios as realistic ($M = 5.38$, $SD = 1.41$).

Measures. In addition to the measures considered in Study 1, we included a manipulation check for the consumption goal. All other measurements were identical to those in Study 1. Online Appendix E3 lists the measures for reliability and validity of Study 2. As in Study 1, the factor loadings, the composite reliability, Cronbach's α, and the convergent validity meet the recommended thresholds. Also, high discriminant validity is established (Fornell and Larcker, 1981; Hair *et al.*, 2019). Online Appendix E4 summarizes the remedies used to address CMV, confirming that CMB is not a significant concern in Study 2. We conducted a pre-study for Study 2 to validate our manipulations, ensure the reliability of the constructs, and confirm the realism of the scenarios. Online Appendix E5 details our findings of the pre-study.

Results

Manipulation check. The two manipulation checks for perceived language embodiment ($M_{\text{figurative}} = 5.87$, $SE = 0.10$; $M_{\text{literal}} = 3.08$, $SE = 0.14$; $t(315) = -15.92$, $p < 0.001$) and consumption goal ($M_{\text{hedonic}} = 4.84$, SE = 0.15; $M_{\text{utilitarian}} = 2.48$, $SE = 0.16$; $t(315) = -10.78$, $p < 0.001$) were successful.

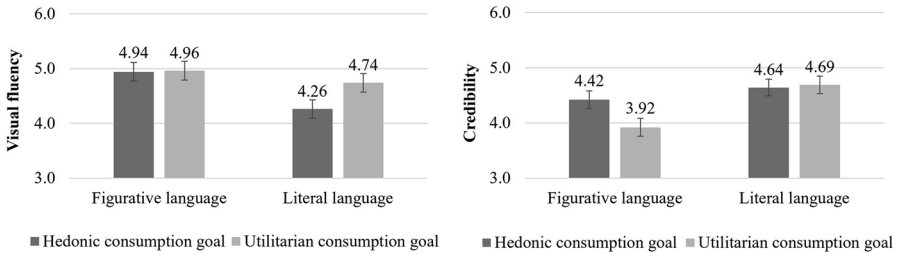
Hypotheses testing. We conducted two two-way ANOVAs, with language embodiment and consumption goal as independent variables, and visual fluency and credibility as dependent variables respectively. The results of the first ANOVA indicate that language embodiment has a significant main effect on visual fluency ($F(1, 313) = 6.99$, $p < 0.01$, $d = 0.30$), with figurative language resulting in significantly higher visual fluency than literal language ($M_{\text{figurative}} = 4.95$, $SE = 0.12$; $M_{\text{literal}} = 4.50$, $SE = 0.12$); with a small effect size. The main effect of consumption goal is not significant ($F(1, 313) = 2.02$, $p = 0.156$). Finally, there is no interaction of language embodiment and consumption goal ($F(1, 313) = 1.78$, $p = 0.183$). According to this, findings of planned contrasts reveal that when a consumer pursues a hedonic (vs utilitarian) consumption goal, a VA's figurative language does not increase visual fluency ($M_{\text{figurative}\times\text{hedonic}} = 4.94$, $SE = 0.17$; $M_{\text{figurative}\times\text{utilitarian}} = 4.96$, $SE = 0.17$; $t = -0.06$, $p = 0.952$). Therefore, the results suggest that figurative language generally leads to higher fluency, regardless of the consumption goal.

The second ANOVA results show a significant main effect of language embodiment on credibility ($F(1, 313) = 10.16$, $p < 0.01$, $d = 0.36$), where figurative language leads to significantly lower credibility than literal language ($M_{\text{figurative}} = 4.17$, $SE = 0.11$; $M_{\text{literal}} = 4.67$, $SE = 0.11$); with a small effect size. The main effect of the consumption

goal is not significant ($F(1, 313) = 2.16$, $p = 0.143$). Finally, there is a statistically significant interaction between language embodiment and consumption goal at the 10% level ($F(1, 313) = 3.11$, $p = 0.079$, $d = 0.20$). Planned contrasts indicate that a VA's figurative language increases perceived credibility when consumers pursue a hedonic consumption goal compared to a utilitarian goal ($M_{\text{figurative×hedonic}} = 4.42$, $SE = 0.16$; $M_{\text{figurative×utilitarian}} = 3.92$, $SE = 0.16$; $t = 2.27$, $p < 0.05$, $d = 0.33$); with a small effect size. Interestingly, within the hedonic consumption context, both figurative and literal language are perceived as equally credible ($M_{\text{figurative×hedonic}} = 4.42$, $SE = 0.16$; $M_{\text{literal×hedonic}} = 4.64$, $SE = 0.15$; $t = -1.01$, $p = 0.314$). In contrast, within the utilitarian consumption context, literal language is significantly more credible than figurative language ($M_{\text{figurative×utilitarian}} = 3.92$, $SE = 0.16$; $M_{\text{literal×utilitarian}} = 4.69$, $SE = 0.16$; $t = -3.50$, $p < 0.01$, $d = 0.55$); with a medium effect size. In conclusion, consumers perceive a VA's message using figurative language as credible only in hedonic consumption contexts, while literal language is consistently advantageous for credibility. Figure 2 illustrates the predictive margins of the interaction effect for Study 2 and an overview of the ANOVA results is in Online Appendix E6 and E7. Additionally, Online Appendix E8, presents descriptive statistics for Study 2. The results hold after controlling for prior experience with VAs and audio repetition (see Online Appendix E9).

Consistent with Study 1, mediation analysis using a bootstrapping procedure with 5,000 iterations and bias-corrected confidence intervals reveals a positive indirect effect of figurative (vs literal) language on recommendation adoption via visual fluency ($\beta = 0.07$, $SE = 0.03$, $CI_{95\%} = [0.018, 0.159]$) and a negative indirect effect via credibility ($\beta = -0.42$, $SE = 0.14$, $CI_{95\%} = [-0.694, -0.156]$). Therefore, Study 2 supports H1 and H2. Likewise, we find no significant direct effect of language embodiment on recommendation adoption ($p = 0.530$), indicating an indirect-only mediation (Zhao *et al.*, 2010). As in Study 1, the mediation path for credibility is significantly higher than for visual fluency ($\beta = -0.35$, $SE = 0.15$, $CI_{95\%} = [-0.641, -0.056]$). While not supporting H3, the results show a non-significant indirect effect of figurative language on recommendation adoption via visual fluency when a consumer pursues a hedonic (vs utilitarian) consumption goal ($\beta = -0.003$, $SE = 0.04$, $CI_{95\%} = [-0.092, 0.089]$), but a significant positive indirect effect via credibility ($\beta = 0.41$, $SE = 0.20$, $CI_{95\%} = [0.031, 0.806]$); supporting H4. In summary, both Studies 1 and 2 confirm our hypotheses, except for H3.

Discussion. The results of Study 2 provide evidence, even in a different commerce setting, that figurative language leads to higher visual fluency but lower credibility, and both influence recommendation adoption. Contrary to our hypothesis, figurative language increases visual fluency regardless of the consumption goal. However, the VA's use of figurative language enhances perceived credibility only when consumers pursue a hedonic consumption goal compared to a utilitarian goal. Thus, figurative language can be advantageous in a hedonic consumption context but disadvantageous in a utilitarian consumption context due to its lower credibility.



Source(s): The above figure was created by the authors

Figure 2. Interaction of language embodiment × consumer's consumption goal on visual fluency and credibility

Study 3

The first two studies support the hypothesized process of the impact of language embodiment and consumer's consumption goal on recommendation adoption. However, the precise mechanism that explains why language embodiment enhances visual fluency warrants deeper empirical elaboration. In our development of H1 we assumed that the positive effect of figurative language on visual fluency could occur because consumers do not expect VAs to use figurative language, leading to expectancy disconfirmation and, in turn, increased arousal as consumers experience the figurative message more intensely due to this expectation violation. Greater arousal then leads to increased fluency. Study 3 explicitly tests this path for enhancing visual fluency. In addition, Study 3 examines the possibility that arousal may not be affected by the expectancy disconfirmation induced by the unusual figurative language, but that arousal directly increases in response to figurative language. This could be reasonable due to the emotional nature of vivid figurative language, which may immediately increase arousal (Berger *et al.*, 2023). To test both possible mechanisms, we run another 2 (figurative vs literal language) × 2 (hedonic vs utilitarian consumption goal) between-subject experiment using the same experimental design as in Study 2.

Method

Participants. We recruited participants from a European country through a crowdsourcing platform. We financially compensated participants based on the legal minimum wage. The participants did not have to meet any special requirements to take part in the study. The final sample consists of 314 participants (42.36% female, $M_{age} = 40.41$ years, $SD = 11.81$), which were randomly and evenly distributed across the four scenarios. There are no significant differences between scenarios in terms of participants' prior experience with VAs, audio repetition, and socio-demographics (all $p > 0.1$). Online Appendix F1 outlines the power analysis and data cleaning process.

Experimental procedure and task. The experimental procedure and design were like in Study 2. As in Study 2, participants asked the VA to provide a headphone recommendation. The participants were either in a hedonic consumption mood (searching for headphones for leisure activities) or in a utilitarian consumption mood (searching for headphones for working activities) and received the headphone recommendation from the VA who was using figurative (vivid expressions) or literal (functional expressions) language. Like in the prior study, the participants perceived the scenarios as realistic ($M = 5.44$, $SD = 1.33$).

Measures. We used the measures for language embodiment, consumer's consumption goal, visual fluency, and recommendation adoption from the prior studies. In addition to these prior measures, we assessed expectancy dis-/confirmation and arousal. All items were taken from established scales and are listed in Online Appendix C. As seen in Online Appendix F2, the results support the reliability and validity of the measurements of Study 3. Online Appendix F3 details the CMV remedies, confirming no significant CMB in Study 3.

Results

Manipulation Check. The manipulation check of this study was successful for perceived language embodiment in terms of figurative vs literal language ($M_{figurative} = 5.09$, $SE = 0.14$; $M_{literal} = 3.34$, $SE = 0.13$; $t(312) = -9.21$, $p < 0.001$). Additionally, the manipulation check for the perceived consumption goal in terms of hedonic vs utilitarian ($M_{hedonic} = 5.00$, SE = 0.14; $M_{utilitarian} = 3.70$, $SE = 0.14$; $t(312) = -6.60$, $p < 0.001$) was successful.

Testing the underlying mechanisms of expectancy disconfirmation and arousal [2]. The first ANOVA results show that figurative (vs literal) language significantly leads to higher expectancy disconfirmation ($M_{figurative} = 4.22$, $SE = 0.12$; $M_{literal} = 3.16$, $SE = 0.11$; $F(1, 310) = 41.84$, $p < 0.001$, $d = 0.73$). The consumption goal has no main effect on expectancy disconfirmation ($F(1, 310) = 0.06$, $p = 0.807$) and there is no interaction of language embodiment and consumer's consumption goal on expectancy disconfirmation ($F(1, 310) = 0.00$, $p = 0.964$). The results

indicate that independently of the consumer's consumption goal, an expectancy disconfirmation occurs for figurative language in comparison to literal language.

Moreover, the second ANOVA results suggest that figurative (vs literal) language significantly increases arousal ($M_{figurative} = 4.48$, $SE = 0.10$; $M_{literal} = 4.05$, $SE = 0.09$; $F(1, 310) = 10.33$, $p < 0.01$, $d = 0.36$). The consumption goal has a statistically significant effect on arousal at the 10% level ($M_{hedonic} = 4.37$, $SE = 0.09$; $M_{utilitarian} = 4.15$, $SE = 0.09$; $F(1, 310) = 2.74$, $p = 0.099$, $d = 0.19$) and there is no significant interaction of language embodiment and consumer's consumption goal on arousal ($F(1, 310) = 0.35$, $p = 0.554$).

Mediation analysis (5,000 bootstrap samples; bias-corrected confidence intervals) reveals a significant negative indirect effect of language embodiment on recommendation adoption via expectancy disconfirmation, arousal, and visual fluency ($\beta = -0.01$, $SE = 0.01$, $CI_{95\%} = [-0.033, -0.003]$). Interestingly, results show that there is a significant positive indirect effect of figurative language on recommendation adoption which is sequentially mediated by greater arousal and increased visual fluency ($\beta = 0.06$, $SE = 0.02$, $CI_{95\%} = [0.021, 0.146]$). Notably, the serial mediation for arousal is significantly higher than the one for expectancy disconfirmation ($\beta = 0.04$, $SE = 0.02$, $CI_{95\%} = [0.015, 0.095]$). In addition, there is no significant direct effect of language embodiment on recommendation adoption ($p = 0.141$), suggesting indirect-only mediation (Zhao *et al.*, 2010). Finally, the indirect effect of the interaction of figurative language and hedonic (vs utilitarian) consumption goal on recommendation adoption, mediated through expectancy disconfirmation, arousal, and visual fluency, is not significant ($\beta = -0.00$, $SE = 0.00$, $CI_{95\%} = [-0.012, 0.004]$). Likewise, the mediation through arousal and visual fluency alone is also not significant ($\beta = 0.02$, $SE = 0.02$, $CI_{95\%} = [-0.016, 0.081]$). Table 1 provides an overview of the bootstrapping results and Online Appendix *F5* provides descriptive statistics for all constructs in the four experimental conditions. Interestingly, all results, except for the significant mediation effect of expectancy disconfirmation, hold after including covariates (see Online Appendix F6 for details).

Discussion. The results suggest that arousal is the underlying mechanism through which language embodiment enhances visual fluency. This finding indicates that expectancy disconfirmation is not the reason behind the increase in arousal and subsequently visual fluency, instead, this mechanism tends to weaken the relationship between language embodiment and recommendation adoption. This contradicts the initial assumption that the surprising effect of figurative language due to expectancy disconfirmation is the mechanism leading to enhanced fluency.

Table 1. Mediation testing of study 3

			CI 95%		
Path	*Coeff.*	*SE*	*LLCI*	*ULCI*	Mediation
Language embodiment → expectancy disconfirmation → arousal → visual fluency → recommendation adoption	−0.01	0.01	−0.033	−0.003	✓
Language embodiment → arousal → visual fluency → recommendation adoption	0.06	0.02	0.021	0.146	✓
Figurative language × consumer's consumption goal → expectancy disconfirmation → arousal → visual fluency → recommendation adoption	−0.00	0.00	−0.012	0.004	×
Figurative language × consumer's consumption goal → arousal → visual fluency → recommendation adoption	0.02	0.02	−0.016	0.081	×

Note(s): $N_{Study3} = 314$, number of bootstrap iterations = 5,000; *Coeff.* = coefficient; *SE* = standard error; *LLCI* = 95% bias-corrected lower-level confidence interval; *ULCI* = 95% bias-corrected upper-level confidence interval

Source(s): Created by the authors

General discussion

This research explores how the challenge of visual disfluency in screenless voice commerce can be addressed through language embodiment, leveraging principles of human dialogue such as figurative language. In addition, we examine a potential trade-off between visual fluency and credibility when using figurative language, assuming that visual fluency and credibility mediate the relationship between language embodiment and recommendation adoption. Our study further explores how this trade-off might vary under certain boundary conditions, particularly with respect to consumer-related context factors such as consumption goals. In addition, we provide deeper insights into the underlying mechanisms of the language embodiment effect on visual fluency. In the final chapter, we discuss the findings and their implications for theory and practice. We also highlight the limitations of our research and suggest directions for future studies.

Theoretical implications

Our study offers significant theoretical contributions to voice commerce and consumer behavior research, addressing two key research questions: how language embodiment influences recommendation adoption through visual fluency and credibility, and how consumption goals shape this influence.

First, based on our research, we demonstrate that successful voice commerce depends not only on message-related aspects, such as visual fluency, but also on source-related aspects, such as credibility. Credibility is particularly crucial, but our findings suggest that improving visual fluency is also important for screenless voice commerce to be effective. While figurative language enhances visual fluency across contexts, its impact on credibility varies depending on the consumer's consumption goal, revealing a trade-off that has important implications for the understanding of language use in voice-assisted commerce (De Keyser and Kunz, 2022). Importantly, unlike prior studies on relational benefits, our research addresses the unique challenge that arise if visual representation (e.g. via display) is missing in VA recommendations. This provides a fresh perspective in the burgeoning field of conversational commerce (De Keyser *et al.*, 2019; Grewal *et al.*, 2022; Lim *et al.*, 2022; Puntoni *et al.*, 2021).

Second, we discover that figurative language leads to expectancy violations because participants currently do not often encounter this language style when interacting with VAs. However, our findings highlight that figurative language enhances visual fluency through arousal, rather than expectancy disconfirmation. This suggests that its emotional expressiveness plays a central role in how consumers process recommendations. This mechanism underscores the robust effect of figurative language on visual fluency, which remains consistent regardless of the consumption goal. As the strong positive effect of figurative language via arousal operates independently from an expectancy disconfirmation effect (which might decline with increasing experience and resulting expectation adjustments; Falk *et al.*, 2010), it should also be stable.

Third, our research highlights that unexpected figurative language undermines credibility in the context of voice commerce. This is notable because some studies suggest that unexpected language can enhance credibility, especially when consumers perceive the recommender as possessing unexpectedly advanced skills (Sopory and Dillard, 2002), such as a VA using human-like capabilities like figurative language. However, our findings suggest that there are additional boundary conditions that need to be considered when investigating unexpected language styles. We provide initial evidence that figurative language is more effective in hedonic consumption contexts than in utilitarian ones for creating a positive credibility effect of expectancy disconfirmation, thus adding to the existing research on LET and machines (e.g. Burgoon *et al.*, 2016; Gnewuch *et al.*, 2022a). As a conclusion, our results demonstrate that while expectancy disconfirmation plays some role, its influence and operation are more complex than initially thought.

Fourth, our findings significantly contribute to TTF theory by demonstrating that a fit between technology (i.e. the language style of the VA) and task (i.e. the consumer's consumption goal) does not always lead to improved consumer evaluations. Our second and third study find that the consumption goal does not influence the effect of figurative language on visual fluency. Therefore, the technology and its capabilities seem to overshadow a technology-task-fit effect. This suggests that specific capabilities of the VA robustly drive certain psychological mechanisms (i.e. visual fluency) so that they are advantageous in any context, independent of the characteristics of the consumers using the assistant. Conversely, for enhancing credibility, the fit between figurative language and the hedonic (vs utilitarian) consumption goal is crucial. Therefore, our study also enriches prior research on the fit between language styles and certain boundary conditions (e.g. Lan *et al.*, 2024; Rhee and Choi, 2020). Moreover, our study highlights the importance of consumer's consumption goals, demonstrating that the product can be the same, but the consumer's motivation for purchase may vary. It is, therefore, crucial to consider not only different products or services, but also the consumer's motivation behind the sought-after product.

Fifth, our research contributes to the literature on embodied systems by demonstrating that machines like VAs can enhance voice commerce success through the use of advanced skills in language style, such as figurative language, particularly under certain boundary conditions like the consumer's consumption goal. However, we also show that adding human-like skills is not always advantageous. Consumers may prefer more machine-like interactions, especially in utilitarian contexts. This aligns with the findings of Longoni and Cian (2022) and connects to the research stream that investigates when the humanization of machines is appropriate and when it is not (e.g. Mozafari *et al.*, 2022).

Practical implications

Firms can develop their own VAs to engage with consumers, such as on their customer service hotlines, or set up voice-activated apps on smart speakers like Amazon Alexa or Google Assistant (Bahmani *et al.*, 2022; Batat and Hammedi, 2023). It allows firms to customize features, including the language style of the VA (Amazon, 2024). Due to this flexibility, firms and VA designers can easily apply the following practical implications.

First, our study highlights the importance of both visual fluency and credibility in voice commerce. However, it indicates that credibility is superior to visual fluency in driving the adoption of purchase recommendations. Therefore, firms should design VAs with a focus on enhancing credibility to ensure the believability of the assistant's recommendations.

Second, when employing figurative language, firms should consider the consumer's consumption goal. With advances in AI, VAs are becoming increasingly adept at learning consumer preferences (De Keyser *et al.*, 2019), enabling firms to adjust the language style of their VAs to better align with consumers' goals. To ensure credibility, firms should use figurative language cautiously when the consumer's consumption goal is ambiguous. While figurative language generally enhances visual fluency and increases credibility in hedonic (vs utilitarian) contexts, it may reduce credibility in utilitarian ones. In contrast, VAs using literal language consistently enhance credibility but not visual fluency.

Finally, since our main findings are consistent across the studies, the practical implications apply to various commerce settings. For instance, in the first study, the VA recommended a suitable restaurant, and in the second and third study, it recommended an electronic product. Therefore, we provide practical insights for firms looking to deploy voice commerce and valuable insights for future VA applications.

In conclusion, transferring human conversational principles to machines is advantageous only under certain conditions. Although firms that implement voice commerce with a visual interface may have an advantage (Gnewuch *et al.*, 2022b; Weith and Matt, 2023), VAs with a display feature should still communicate recommendations as if they did not have a screen. Consumers use VAs for their hands- and eyes-free approach, not for the screen (Schmitt *et al.*, 2023). Therefore,

consumers may not always want or be able to look at a screen. Overall, it is important to emphasize the need for strategic conversational design for VAs.

Limitations and further research

Although this study offers valuable insights, it has limitations that suggest opportunities for future research. The first limitation pertains to the cultural context of our study. Our findings regarding expectancy disconfirmation, when VAs use figurative language, were derived from a Western-oriented cultural perspective. This focus limits the generalizability of the results to other cultural contexts. Prior research has shown that expectations can vary across cultural contexts (Baek *et al.*, 2022). Future research should aim to include a broader and more diverse range of cultures to determine the global applicability of expectancy disconfirmation in the context of VA interactions. Similarly, we used a female voice in our experiment, and there may be another effect when using a male or neutral voice (Mohsenin and Munz, 2024). In addition, the potential effect of the voice gender may also differ depending on the cultural context.

Additionally, our study design has a limitation related to temporal aspects. As our research utilized a cross-sectional approach, it cannot determine whether the impact of figurative language used by VAs endures over time. This limitation is especially significant when considering the possibility of habituation or a backfire effect, particularly if such language is perceived as bothersome, similar to repetitive advertising, which could lead to reduced credibility over time. Understanding the long-term dynamics of figurative language in consumer-VA interactions is crucial, as consumers may become accustomed to or increasingly irritated by it over time. Therefore, future research should consider longitudinal study designs to track changes over time and provide valuable insights into the durability and evolution of these effects.

Moreover, as humans naturally use a combination of figurative and literal language, this approach should also be tested in future studies. Therefore, it would be of great interest to investigate how a different mix of figurative and literal style affects consumer evaluations regarding expectancy disconfirmation, arousal, visual fluency, and credibility. Additionally, examining the impact of individual elements of figurative language, such as metaphors, hyperboles, idioms, or comparisons, on consumer evaluations would be valuable.

Our study highlights the importance of boundary conditions in this research field, indicating the need for further investigation into various consumer-related factors. Prior research has demonstrated that these factors, such as a consumer's ability to visualize (Petrova and Cialdini, 2005), preference for processing styles (Heirati *et al.*, 2024), or preference for different voices (Mohsenin and Munz, 2024), can significantly influence the effectiveness and impact of stimuli. Additionally, the preferred language style may vary depending on the consumer's relationship with the VA (e.g. partner or servant), as suggested by Novak and Hoffman (2019) and Grewal *et al.* (2022). Therefore, further research is needed to explore the conditions under which consumers prefer specific language styles.

In addition, we explored one possible technique using figurative language to enhance visual fluency. However, alternative techniques within the realm of sensory marketing can also influence mental imagery. Elements such as images, sounds, and statements can be crafted to trigger mental imagery by tapping into consumers' memories, as demonstrated in the work of Krishna *et al.* (2016) and Krishna and Schwarz (2014). With the increasing integration of displays into voice devices, leveraging these sensory inputs becomes even more feasible, opening up exciting new research avenues.

Lastly, our study's methodological approach poses certain limitations. Despite mitigating some challenges of the vignette-based experiment (Aguinis and Bradley, 2014) by including diverse commerce settings (restaurant and headphone recommendations), the lack of real-time interaction may limit external validity, as it cannot fully capture the dynamic nature of real-world consumer-VA interactions (Hill *et al.*, 2015). Future research should employ more interactive, real-time study settings to better reflect real-life complexities. Additionally, collecting data via a survey from a single key informant introduces potential CMB (Podsakoff

et al., 2024). While we implemented several *a priori* and post hoc measures to address this, the inherent limitations of such approaches remain. Nevertheless, studies such as ours, which use an experimental and questionnaire-based design with single informants, often serve as valuable means for opening new research fields and provide a solid basis for further research.

Notes

1. We thank the reviewers for suggesting the inclusion of this insightful detail of audio repetition. Across all studies, participants seldom replayed the VA message, with mean frequencies close to 1 (Study 1: $M = 1.22$, $SD = 0.42$; Study 2: $M = 1.08$, $SD = 0.34$; Study 3: $M = 1.25$, $SD = 0.65$). This consistency reflects natural consumer behavior and underscores the ecological validity of the study design.
2. In Study 3, we re-test H1 and H3, and the results confirm the findings from Study 1 and Study 2. Detailed results are provided in Appendix F4.

References

Aguinis, H. and Bradley, K.J. (2014), "Best practice recommendations for designing and implementing experimental vignette methodology studies", *Organizational Research Methods*, Vol. 17 No. 4, pp. 351-371, doi: 10.1177/1094428114547952.

Amazon (2024), "Alexa skills kit", available at: https://developer.amazon.com/en-US/alexa/alexa-skills-kit

Averbeck, J.M. and Miller, C. (2014), "Expanding language expectancy theory: the suasory effects of lexical complexity and syntactic complexity on effective message design", *Communication Studies*, Vol. 65 No. 1, pp. 72-95, doi: 10.1080/10510974.2013.775955.

Babin, B.J., Darden, W.R. and Griffin, M. (1994), "Work and/or fun: measuring hedonic and utilitarian shopping value", *Journal of Consumer Research*, Vol. 20 No. 4, pp. 644-656, doi: 10.1086/209376.

Baek, T.H., Kim, S., Yoon, S., Choi, Y.K., Choi, D. and Bang, H. (2022), "Emojis and assertive environmental messages in social media campaigns", *Internet Research*, Vol. 32 No. 3, pp. 988-1002, doi: 10.1108/INTR-11-2020-0638.

Bagozzi, R.P., Belanche, D., Casaló, L.V. and Flavián, C. (2016), "The role of anticipated emotions in purchase intentions", *Psychology and Marketing*, Vol. 33 No. 8, pp. 629-645, doi: 10.1002/mar.20905.

Bahmani, N., Bhatnagar, A. and Gauri, D. (2022), "Hey, Alexa! What attributes of Skills affect firm value?", *Journal of the Academy of Marketing Science*, Vol. 50 No. 6, pp. 1219-1235, doi: 10.1007/s11747-022-00851-0.

Batat, W. and Hammedi, W. (2023), "The extended reality technology (ERT) framework for designing customer and service experiences in phygital settings: a service research agenda", *Journal of Service Management*, Vol. 34 No. 1, pp. 10-33, doi: 10.1108/JOSM-08-2022-0289.

Batra, R. and Ahtola, O.T. (1991), "Measuring the hedonic and utilitarian sources of consumer attitudes", *Marketing Letters*, Vol. 2 No. 2, pp. 159-170, doi: 10.1007/BF00436035.

Berger, J., Moe, W.W. and Schweidel, D.A. (2023), "What holds attention? Linguistic drivers of engagement", *Journal of Marketing*, Vol. 87 No. 5, pp. 793-809, doi: 10.1177/00222429231152880.

Bergner, A.S., Hildebrand, C. and Häubl, G. (2023), "Machine talk: how verbal embodiment in conversational AI shapes consumer–brand relationships", *Journal of Consumer Research*, Vol. 50 No. 4, pp. 742-764, doi: 10.1093/jcr/ucad014.

Blut, M., Ghiassaleh, A. and Wang, C. (2023), "Testing the performance of online recommendation agents: a meta-analysis", *Journal of Retailing*, Vol. 99 No. 3, pp. 440-459, doi: 10.1016/j.jretai.2023.08.001.

Botti, S. and McGill, A.L. (2011), "The locus of choice: personal causality and satisfaction with hedonic and utilitarian decisions", *Journal of Consumer Research*, Vol. 37 No. 6, pp. 1065-1078, doi: 10.1086/656570.

Burgoon, J.K. (1993), "Interpersonal expectations, expectancy violations, and emotional communication", *Journal of Language and Social Psychology*, Vol. 12 Nos 1-2, pp. 30-48, doi: 10.1177/0261927X93121003.

Burgoon, M., Denning, V.P. and Roberts, L. (2002), "Language expectancy theory", in Dillard, J.P. and Pfau, M. (Eds), *The Persuasion Handbook: Developments in Theory and Practice*, Sage, pp. 117-136.

Burgoon, J.K., Bonito, J.A., Lowry, P.B., Humpherys, S.L., Moody, G.D., Gaskin, J.E. and Giboney, J.S. (2016), "Application of Expectancy Violations Theory to communication with and judgments about embodied agents during a decision-making task", *International Journal of Human-Computer Studies*, Vol. 91, pp. 24-36, doi: 10.1016/j.ijhcs.2016.02.002.

Castelo, N., Bos, M.W. and Lehmann, D.R. (2019), "Task-dependent algorithm aversion", *Journal of Marketing Research*, Vol. 56 No. 5, pp. 809-825, doi: 10.1177/0022243719851788.

Chang, C.-T. and Yen, C.-T. (2013), "Missing ingredients in metaphor advertising: the right formula of metaphor type, product type, and need for cognition", *Journal of Advertising*, Vol. 42 No. 1, pp. 80-94, doi: 10.1080/00913367.2012.749090.

Chattaraman, V., Kwon, W.-S., Gilbert, J.E. and Ross, K. (2019), "Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults", *Computers in Human Behavior*, Vol. 90, pp. 315-330, doi: 10.1016/j.chb.2018.08.048.

Choi, S., Liu, S.Q. and Mattila, A.S. (2019), "'How may i help you?' Says a robot: examining language styles in the service encounter", *International Journal of Hospitality Management*, Vol. 82, pp. 32-38, doi: 10.1016/j.ijhm.2019.03.026.

Cohen, J. (1992), "Quantitative methods in psychology: a power primer", *Psychological Bulletin*, Vol. 112 No. 1, pp. 155-159, doi: 10.1037//0033-2909.112.1.155.

Coker, K.K. and Thakur, R. (2023), "Alexa, may I adopt you? The role of voice assistant empathy and user-perceived risk in customer service delivery", *Journal of Services Marketing*, Vol. 38 No. 3, pp. 301-311, doi: 10.1108/JSM-07-2023-0284.

Dabholkar, P.A. (1996), "Consumer evaluations of new technology-based self-service options: an investigation of alternative models of service quality", *International Journal of Research in Marketing*, Vol. 13 No. 1, pp. 29-51, doi: 10.1016/0167-8116(95)00027-5.

De Keyser, A. and Kunz, W.H. (2022), "Living and working with service robots: a TCCM analysis and considerations for future research", *Journal of Service Management*, Vol. 33 No. 2, pp. 165-196, doi: 10.1108/JOSM-12-2021-0488.

De Keyser, A., Köcher, S., Alkire, L., Verbeeck, C. and Kandampully, J. (2019), "Frontline Service Technology infusion: conceptual archetypes and future research directions", *Journal of Service Management*, Vol. 30 No. 1, pp. 156-183, doi: 10.1108/JOSM-03-2018-0082.

Dhar, R. and Wertenbroch, K. (2000), "Consumer choice between hedonic and utilitarian goods", *Journal of Marketing Research*, Vol. 37 No. 1, pp. 60-71, doi: 10.1509/jmkr.37.1.60.18718.

Diederich, S., Brendel, A., Morana, S. and Kolbe, L. (2022), "On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research", *Journal of the Association for Information Systems*, Vol. 23 No. 1, pp. 96-138, doi: 10.17705/1jais.00724.

Falk, T., Hammerschmidt, M. and Schepers, J.J.L. (2010), "The service quality-satisfaction link revisited: exploring asymmetries and dynamics", *Journal of the Academy of Marketing Science*, Vol. 38 No. 3, pp. 288-302, doi: 10.1007/s11747-009-0152-2.

Filieri, R. (2015), "What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM", *Journal of Business Research*, Vol. 68 No. 6, pp. 1261-1270, doi: 10.1016/j.jbusres.2014.11.006.

Flavián, C., Akdim, K. and Casaló, L.V. (2023), "Effects of voice assistant recommendations on consumer behavior", *Psychology and Marketing*, Vol. 40 No. 2, pp. 328-346, doi: 10.1002/mar.21765.

Forbes (2023), "E-commerce evolution: predictions and trends shaping the future", available at: https://www.forbes.com/sites/forbesbusinesscouncil/2023/12/29/e-commerce-evolution-predictions-and-trends-shaping-the-future/

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50, doi: 10.1177/002224378101800104.

Gnewuch, U., Morana, S., Adam, M.T.P. and Maedche, A. (2022a), "Opposing effects of response time in human–chatbot interaction: the moderating role of prior experience", *Business and Information Systems Engineering*, Vol. 64 No. 6, pp. 773-791, doi: 10.1007/s12599-022-00755-x.

Gnewuch, U., Ruoff, M., Peukert, C. and Maedche, A. (2022b), "Multiexperience", *Business and Information Systems Engineering*, Vol. 64 No. 6, pp. 813-823, doi: 10.1007/s12599-022-00766-8.

Goodhue, D.L. and Thompson, R.L. (1995), "Task-technology fit and individual performance", *MIS Quarterly*, Vol. 19 No. 2, p. 213, doi: 10.2307/249689.

Graf, L.K.M., Mayer, S. and Landwehr, J.R. (2018), "Measuring processing fluency: one versus five items", *Journal of Consumer Psychology*, Vol. 28 No. 3, pp. 393-411, doi: 10.1002/jcpy.1021.

Grewal, D., Guha, A., Schweiger, E., Ludwig, S. and Wetzels, M. (2022), "How communications by AI-enabled voice assistants impact the customer journey", *Journal of Service Management*, Vol. 33 Nos 4/5, pp. 705-720, doi: 10.1108/JOSM-11-2021-0452.

Guha, A., Bressgott, T., Grewal, D., Mahr, D., Wetzels, M. and Schweiger, E. (2023), "How artificiality and intelligence affect voice assistant evaluations", *Journal of the Academy of Marketing Science*, Vol. 51 No. 4, pp. 843-866, doi: 10.1007/s11747-022-00874-7.

Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019), "When to use and how to report the results of PLS-SEM", *European Business Review*, Vol. 31 No. 1, pp. 2-24, doi: 10.1108/EBR-11-2018-0203.

Heirati, N., Pitardi, V. and Temerak, M.S. (2024), "When the recipe is more important than the ingredients: unveiling the complexity of consumer use of voice assistants", *Psychology and Marketing*, Vol. 41 No. 7, pp. 1489-1501, doi: 10.1002/mar.21992.

Hill, J., Randolph Ford, W. and Farreras, I.G. (2015), "Real conversations with artificial intelligence: a comparison between human–human online conversations and human–chatbot conversations", *Computers in Human Behavior*, Vol. 49, pp. 245-250, doi: 10.1016/j.chb.2015.02.026.

Hirschman, E.C. and Holbrook, M.B. (1982), "Hedonic consumption: emerging concepts, methods and propositions", *Journal of Marketing*, Vol. 46 No. 3, pp. 92-101, doi: 10.2307/1251707.

Kim, T.W. and Duhachek, A. (2020), "Artificial intelligence and persuasion: a construal-level account", *Psychological Science*, Vol. 31 No. 4, pp. 363-380, doi: 10.1177/0956797620904985.

Kim, Y. and McFarland, R.G. (2024), "Are you looking for something specific or just looking around? Adaptive selling on the basis of customer shopping goals in retail sales", *Journal of the Academy of Marketing Science*, Vol. 52 No. 6, pp. 1780-1804, doi: 10.1007/s11747-024-01015-y.

Krishna, A., Cian, L. and Sokolova, T. (2016), "The power of sensory marketing in advertising", *Current Opinion in Psychology*, Vol. 10, pp. 142-147, doi: 10.1016/j.copsyc.2016.01.007.

Krishna, A. and Schwarz, N. (2014), "Sensory marketing, embodiment, and grounded cognition: a review and introduction", *Journal of Consumer Psychology*, Vol. 24 No. 2, pp. 159-168, doi: 10.1016/j.jcps.2013.12.006.

Kronrod, A. and Danziger, S. (2013), "'Wii will rock you!' The use and effect of figurative language in consumer reviews of hedonic and utilitarian consumption", *Journal of Consumer Research*, Vol. 40 No. 4, pp. 726-739, doi: 10.1086/671998.

Lan, H., Tang, X., Ye, Y. and Zhang, H. (2024), "Abstract or concrete? The effects of language style and service context on continuous usage intention for AI voice assistants", *Humanities and Social Sciences Communications*, Vol. 11 No. 1, p. 99, doi: 10.1057/s41599-024-02600-w.

Landwehr, J.R., Labroo, A.A. and Herrmann, A. (2011), "Gut liking for the ordinary: incorporating design fluency improves automobile sales forecasts", *Marketing Science*, Vol. 30 No. 3, pp. 416-429, doi: 10.1287/mksc.1110.0633.

Lee, A.Y. and Aaker, J.L. (2004), "Bringing the frame into focus: the influence of regulatory fit on processing fluency and persuasion", *Journal of Personality and Social Psychology*, Vol. 86 No. 2, pp. 205-218, doi: 10.1037/0022-3514.86.2.205.

Lee, A.Y. and Labroo, A.A. (2004), "The effect of conceptual and perceptual fluency on brand evaluation", *Journal of Marketing Research*, Vol. 41 No. 2, pp. 151-165, doi: 10.1509/jmkr.41.2.151.28665.

Lee, A.Y., Keller, P.A. and Sternthal, B. (2010), "Value from regulatory construal fit: the persuasive impact of fit between consumer goals and message concreteness", *Journal of Consumer Research*, Vol. 36 No. 5, pp. 735-747, doi: 10.1086/605591.

Lim, W.M., Kumar, S., Verma, S. and Chaturvedi, R. (2022), "Alexa, what do we know about conversational commerce? Insights from a systematic literature review", *Psychology and Marketing*, Vol. 39 No. 6, pp. 1129-1155, doi: 10.1002/mar.21654.

Longoni, C. and Cian, L. (2022), "Artificial intelligence in utilitarian vs Hedonic contexts: the word-of-machine effect", *Journal of Marketing*, Vol. 86 No. 1, pp. 91-108, doi: 10.1177/0022242920957347.

Luo, X., Tong, S., Fang, Z. and Qu, Z. (2019), "Frontiers: machines vs Humans: the impact of artificial intelligence chatbot disclosure on customer purchases", *Marketing Science*, Vol. 38 No. 6, pp. 937-947, doi: 10.1287/mksc.2019.1192.

Luri, I., Schau, H.J. and Ghosh, B. (2024), "Metaphor-Enabled marketplace sentiment analysis", *Journal of Marketing Research*, Vol. 61 No. 3, pp. 496-516, doi: 10.1177/00222437231191526.

Mari, A., Mandelli, A. and Algesheimer, R. (2024), "Empathic voice assistants: enhancing consumer responses in voice commerce", *Journal of Business Research*, Vol. 175, 114566, doi: 10.1016/j.jbusres.2024.114566.

Melumad, S. (2023), "Vocalizing search: how voice technologies alter consumer search processes and satisfaction", *Journal of Consumer Research*, Vol. 50 No. 3, pp. 533-553, doi: 10.1093/jcr/ucad009.

Melumad, S. and Meyer, R.J. (2024), "How listening versus reading alters consumers' interpretations of news", *Journal of Marketing Research*, 00222437241280068, doi: 10.1177/00222437241280068.

Miller, M.D. and Burgoon, M. (1979), "The relationship between violations of expectations and the induction of resistance to persuasion", *Human Communication Research*, Vol. 5 No. 4, pp. 301-313, doi: 10.1111/j.1468-2958.1979.tb00642.x.

Mitchell, W.J., Ho, C.-C., Patel, H. and MacDorman, K.F. (2011), "Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management", *Computers in Human Behavior*, Vol. 27 No. 1, pp. 402-412, doi: 10.1016/j.chb.2010.09.002.

Mohsenin, S. and Munz, K.P. (2024), "Gender-ambiguous voices and social disfluency", *Psychological Science*, Vol. 35 No. 5, pp. 543-557, doi: 10.1177/09567976241238222.

Mozafari, N., Weiger, W.H. and Hammerschmidt, M. (2022), "Trust me, I'm a bot – repercussions of chatbot disclosure in different service frontline settings", *Journal of Service Management*, Vol. 33 No. 2, pp. 221-245, doi: 10.1108/JOSM-10-2020-0380.

Novak, T.P. and Hoffman, D.L. (2019), "Relationship journeys in the internet of things: a new framework for understanding interactions between consumers and smart objects", *Journal of the Academy of Marketing Science*, Vol. 47 No. 2, pp. 216-237, doi: 10.1007/s11747-018-0608-3.

Novemsky, N., Dhar, R., Schwarz, N. and Simonson, I. (2007), "Preference fluency in choice", *Journal of Marketing Research*, Vol. 44 No. 3, pp. 347-356, doi: 10.1509/jmkr.44.3.347.

Petrova, P.K. and Cialdini, R.B. (2005), "Fluency of consumption imagery and the backfire effects of imagery appeals", *Journal of Consumer Research*, Vol. 32 No. 3, pp. 442-452, doi: 10.1086/497556.

Podsakoff, P.M., Podsakoff, N.P., Williams, L.J., Huang, C. and Yang, J. (2024), "Common method bias: it's bad, it's complex, it's widespread, and it's not easy to fix", *Annual Review of Organizational Psychology and Organizational Behavior*, Vol. 11 No. 1, pp. 17-61, doi: 10.1146/annurev-orgpsych-110721-040030.

Puntoni, S., Reczek, R.W., Giesler, M. and Botti, S. (2021), "Consumers and artificial intelligence: an experiential perspective", *Journal of Marketing*, Vol. 85 No. 1, pp. 131-151, doi: 10.1177/0022242920953847.

Rhee, C.E. and Choi, J. (2020), "Effects of personalization and social role in voice shopping: an experimental study on product recommendation by a conversational voice agent", *Computers in Human Behavior*, Vol. 109, 106359, doi: 10.1016/j.chb.2020.106359.

Ruan, Y. and Mezei, J. (2022), "When do AI chatbots lead to higher customer satisfaction than human frontline employees in online shopping assistance? Considering product attribute type", *Journal of Retailing and Consumer Services*, Vol. 68, 103059, doi: 10.1016/j.jretconser.2022.103059.

Rzepka, C., Berger, B. and Hess, T. (2022), "Voice assistant vs Chatbot – examining the fit between conversational agents' interaction modalities and information search tasks", *Information Systems Frontiers*, Vol. 24 No. 3, pp. 839-856, doi: 10.1007/s10796-021-10226-5.

Schindler, D., Maiberger, T., Koschate-Fischer, N. and Hoyer, W.D. (2023), "How speaking versus writing to conversational agents shapes consumers' choice and choice satisfaction", *Journal of the Academy of Marketing Science*, Vol. 52 No. 3, pp. 634-652, doi: 10.1007/s11747-023-00987-7.

Schmitt, A., Zierau, N., Janson, A. and Leimeister, J.M. (2023), "The role of AI-based artifacts' voice capabilities for agency attribution", *Journal of the Association for Information Systems*, Vol. 24 No. 4, pp. 980-1004, doi: 10.17705/1jais.00827.

Schwarz, N., Jalbert, M., Noah, T. and Zhang, L. (2021), "Metacognitive experiences as information: processing fluency in consumer judgment and decision making", *Consumer Psychology Review*, Vol. 4 No. 1, pp. 4-25, doi: 10.1002/arcp.1067.

Schwede, M., Zierau, N., Janson, A., Hammerschmidt, M. and Leimeister, J.M. (2022), "'I will follow you!' – how recommendation modality impacts processing fluency and purchase intention", *Proceedings of the International Conference on Information Systems*, Vol. 55, Presented at the International Conference on Information Systems, Kopenhagen, Denmark, doi: 10.2139/ssrn.4239822.

Sopory, P. and Dillard, J.P. (2002), "The persuasive effects of metaphor: a meta-analysis", *Human Communication Research*, Vol. 28 No. 3, pp. 382-419, doi: 10.1111/j.1468-2958.2002.tb00813.x.

Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J.S., Leimeister, J.M.M. and Bernstein, A. (2021), "Female by default? – Exploring the effect of voice assistant gender and pitch on trait and trust attribution", *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, NY, USA, pp. 1-7, doi: 10.1145/3411763.3451623.

Voicebot.ai (2023), "Does ChatGPT mark the end of the voice assistant era or is it a false comparison?", available at: https://voicebot.ai/2023/10/20/does-chatgpt-mark-the-end-of-the-voice-assistant-era-or-is-it-a-false-comparison/

Voss, K.E., Spangenberg, E.R. and Grohmann, B. (2003), "Measuring the hedonic and utilitarian dimensions of consumer attitude", *Journal of Marketing Research*, Vol. 40 No. 3, pp. 310-320, doi: 10.1509/jmkr.40.3.310.19238.

Weith, H. and Matt, C. (2023), "Information provision measures for voice agent product recommendations— the effect of process explanations and process visualizations on fairness perceptions", *Electronic Markets*, Vol. 33 No. 1, p. 57, doi: 10.1007/s12525-023-00668-x.

Whang, C. and Im, H. (2021), "I like Your Suggestion!' the role of humanlikeness and parasocial relationship on the website versus voice shopper's perception of recommendations", *Psychology and Marketing*, Vol. 38 No. 4, pp. 581-595, doi: 10.1002/mar.21437.

Wien, A.H. and Peluso, A.M. (2021), "Influence of human versus AI recommenders: the roles of product type and cognitive processes", *Journal of Business Research*, Vol. 137, pp. 13-27, doi: 10.1016/j.jbusres.2021.08.016.

Wu, L., Shen, H., Fan, A. and Mattila, A.S. (2017), "The impact of language style on consumers' reactions to online reviews", *Tourism Management*, Vol. 59, pp. 590-596, doi: 10.1016/j.tourman.2016.09.006.

Yen, C. and Chiang, M.-C. (2021), "Trust me, if you can: a study on the factors that influence consumers' purchase intention triggered by chatbots based on brain image evidence and self-reported assessments", *Behaviour and Information Technology*, Vol. 40 No. 11, pp. 1177-1194, doi: 10.1080/0144929X.2020.1743362.

Yin, D., Bond, S.D. and Zhang, H. (2017), "Keep your cool or let it out: nonlinear effects of expressed arousal on perceptions of consumer reviews", *Journal of Marketing Research*, Vol. 54 No. 3, pp. 447-463, doi: 10.1509/jmr.13.0379.

Zhao, X., Lynch, J.G. and Chen, Q. (2010), "Reconsidering baron and kenny: myths and truths about mediation analysis", *Journal of Consumer Research*, Vol. 37 No. 2, pp. 197-206, doi: 10.1086/651257.

Zierau, N., Hildebrand, C., Bergner, A., Busquet, F., Schmitt, A. and Marco Leimeister, J. (2023), "Voice bots on the frontline: voice-based interfaces enhance flow-like consumer experiences and boost service outcomes", *Journal of the Academy of Marketing Science*, Vol. 51 No. 4, pp. 823-842, doi: 10.1007/s11747-022-00868-5.

Supplementary material

The supplementary material for this article can be found online.

About the authors

Melanie Schwede is a Research Associate and Ph.D. candidate at the Chair of Marketing and Innovation Management at the University of Goettingen, Germany. Her research focuses on the societal impact of AI-based information systems, particularly on how conversational agents and their design shape user experience and foster firm-beneficial outcomes. Her scholarly contributions have been published at leading academic conferences and in international peer-reviewed journals like *Electronic Markets and Journal of Service Management*. Melanie Schwede is the corresponding author and can be contacted at: melanie.schwede@uni-goettingen.de

Maik Hammerschmidt is Professor of Marketing and Innovation Management at the University of Goettingen, and Co-Director of the Retail Logistics and Innovation Lab, Germany. He earned his Ph.D. at the University of Mannheim, Germany. His research focuses primarily on the performance outcomes of artificial intelligence agents and immersive technologies in the service frontline, motivational information systems, and corporate social responsibility. His research has been published in the leading journals in the field like *Journal of Marketing, Journal of the Academy of Marketing Science, and Journal of Service Research*.

Stefan Morana is a Professor of Digital Transformation and Information Systems at Saarland University, Germany. He earned his Ph.D. at the University of Mannheim, Germany. His research focuses on the role of human-centered design of interactive systems, such as conversational interfaces, for digital transformation from the perspectives of the individual, organizations, and society. His research has been published in leading international journals, including *Information Systems Research, Journal of the Association for Information Systems, Business and Information Systems Engineering, and International Journal of Human-Computer Studies*.