Online Appendix

Journal of Service Management

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

Melanie Schwede

Chair of Marketing and Innovation Management, University of Goettingen, Germany

Maik Hammerschmidt

Chair of Marketing and Innovation Management, Retail Logistics & Innovation Lab, University of Goettingen, Germany

Stefan Morana

Chair of Digital Transformation and Information Systems, Saarland University, Germany

Table of Contents

Online Appendix A: Literature Review

Related work on voice assistants' language style

To date, empirical research on voice assistants in marketing covers a wide range of themes and offers insights into their various effects on consumers and firms. One important research field focuses on how to design voice assistants to improve consumer evaluations and adoption. This includes examining the (technical) capabilities of voice assistants (e.g., Bahmani *et al.*, 2022; Hernández Ortega and Lucia-Palacios, 2023; Yoganathan and Osburg, 2024) but also the content they deliver (e.g., Weith and Matt, 2023) and the properties of their voices (e.g., Efthymiou *et al.*, 2024; Mohsenin and Munz, 2024). Accordingly, researchers emphasize that designing such assistants requires consideration of elements from human-to-human communication (Gnewuch, Morana, *et al.*, 2022). Since voice assistants rely solely on verbal cues to provide information, different language styles – the framing of messages – become a critical factor, particularly the language styles that have been used almost exclusively for human-to-human communication.

Table 1 provides an overview of the relevant literature on voice assistants' language style. The table outlines key factors necessary for evaluating the findings of prior work: (1) voice assistants' language styles, (2) the underlying psychological mechanisms that influence consumer behavior, (3) the boundary conditions that determine the impact of language styles, and (4) the behavioral outcomes examined to capture voice assistant success. Several important take-aways result from the table.

First, 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 voice assistants. 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, to date, no empirical study in the field of voice assistants has examined language styles that use vivid expressions. This is a crucial shortcoming, as success in screenless commerce may depend on the ability of voice assistants to evoke mental images, which emphasizes the need to investigate the potential of figurative language styles in this context.

Second, prior research has investigated different psychological mechanisms related to either message-related (Lan *et al.*, 2024) or source-related (e.g., Whang and Im, 2021) evaluations. However, except of Chattaraman *et al*. (2019), limited research has examined both levels (message and source) of psychological mechanisms simultaneously. It follows that prior studies have focused primarily on trust or relational properties as mechanisms triggered by language styles, rather than how information processing can be facilitated. Notably, no study has explored visual fluency or how easily consumers can mentally visualize a product. Additionally, prior studies did not consider whether trade-offs may occur between the message- and source-related psychological mechanisms and what consequences this may have for voice commerce.

Third, prior research has examined boundary conditions related to the product or service (Lan *et al.*, 2024; Rhee and Choi, 2020), the modality of the assistant (Whang and Im, 2021), or message content (Rhee and Choi, 2020). However, except of Chattaraman *et al.*'s (2019) study on the competence of older consumers, no study so far has addressed consumer-related boundary conditions such as consumption goals.

Fourth, prior research has mainly focused on perceptual outcomes, such as product evaluation (e.g., Whang and Im, 2021) and reuse intention (e.g., Lan *et al.*, 2024). While these are valuable, studies are needed that consider the actual adoption of a recommendation, which includes the willingness to purchase a recommended product or book a recommended service, as it is a key element of economic success of voice commerce.

As summarized in Table 1, no research so far has examined the use of figurative vs. literal language styles as a crucial instrument for enhancing visual fluency, although it has been noted as a key challenge in voice commerce (Gnewuch, Ruoff, *et al.*, 2022). We therefore explore the impact of both language styles on recommendation adoption. Additionally, we explore whether improving message-related evaluations (i.e., visual fluency) comes at the expense of sacrificing source-related evaluations (i.e., credibility). This potential trade-off has been neglected so far, although its identification is crucial for designing voice assistants in a way that they trigger positive evaluations and mitigate negative evaluations. Furthermore, we investigate how consumer-related context factors (i.e., consumption goals) shape the extent to which the different language styles exert the trade-off between visual fluency and credibility and, in turn, the adoption of recommendations.

Table 1. Summary of prior empirical research on the effects of voice assistant's language styles on consumer behavior

Study	Theory basis	Voice assistant language condition	Psychological mechanisms		Boundary condition(s)	Outcome variable	Key findings
			Message-related	Source-related			
Chattaraman *et al*. (2019)	Social response theory, cognitive load theory, social cognitive theory	Social- vs. task-oriented	Information overload	Interactivity, interface trust	Older consumers' Internet competence	Website patronage intention	Older consumers with high internet competence have better social experiences with social-oriented assistants. Older consumers with less internet competence benefit from task-oriented assistants, finding them less cognitively overwhelming.
Lan *et al*. (2024)	Construal-level theory	Abstract vs. concrete	Processing fluency	/	Service context	Continuous usage intention	Consumers find it easier to process abstract language from voice assistants in a hedonic service context and concrete language in a utilitarian service context.
Mari *et al*. (2024)	Technology acceptance theories, AI empathy	High vs. low empathy	/	Functional, relational, and social-emotional attributes	Social context	Task delegation, decision support seeking, and recommendation trust	Consumers are more likely to delegate tasks, seek decision support, and trust recommendations from voice assistants perceived as empathic. In a family context, they respond better to functional attributes within high AI empathy.
Rhee and Choi (2020)	Social role theory	Informal vs. formal	/	/	Product involvement, personalization	Attitude toward the product	Consumers like the products more from voice assistants using an informal (vs. formal) language, especially for products with low (vs. high) involvement.
Whang and Im (2021)	Parasocial interaction theory	Social- vs. task-oriented	/	Human likeness, parasocial relationship	Interaction modality	Product evaluation	Consumers perceive websites as more human-like and evaluate them more positively than voice assistants, regardless of the language style.

This research	Language expectancy theory, task-technology fit theory	Figurative vs. literal	Visual fluency	Credibility	Consumer's consumption goal	Recommendation adoption	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.

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

Online Appendix B: Demographic Data Across Studies

		Study 1		Study 2		Study 3	
		n	***%***	***n***	***%***	***n***	***%***
Gender	Female	62	47.33	224	70.66	133	42.36
	Male	65	49.62	90	28.39	179	57.00
	Diverse	4	3.05	2	0.63	1	0.32
	No information	0	0	1	0.32	1	0.32
Age	< 21	12	9.16	10	3.15	8	2.56
	21 - 30	82	62.60	239	75.40	58	18.47
	31 - 40	29	22.14	27	8.52	115	36.62
	41 - 50	3	2.29	13	4.10	64	20.38
	51 - 60	3	2.29	22	6.94	44	14.01
	>= 61	1	0.76	6	1.89	23	7.32
	No information	1	0.76	0	0	2	0.64
Education	No degree	0	0	0	0	0	0
	Intermediate school	11	8.40	4	1.26	38	12.10
	High school diploma	48	36.64	60	18.93	50	15.92
	Completed vocational training	10	7.63	22	6.94	74	23.57
	University degree	61	46.57	230	72.55	149	47.45
	No information	1	0.76	1	0.32	3	0.96
Occupation	Student (school)	1	0.76	1	0.32	2	0.64
	Student (university)	57	43.51	194	61.19	17	5.41
	Trainee (apprentice)	7	5.35	1	0.32	3	0.96
	Employed	56	42.75	115	36.27	252	80.25
	Not employed	9	6.87	1	0.32	12	3.82
	Homemaker	0	0	2	0.63	10	3.19
	Retiree	0	0	3	0.95	7	2.23
	No information	1	0.76	0	0	11	3.50
Prior experience with voice assistants		$M = 4.27$; $SD = 1.59$		$M = 3.33$; $SD = 1.64$		$M = 4.51$; $SD = 1.51$	

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

Online Appendix C: Measures of Constructs

Construct	Items	Item loadings Study 1	Study 2	Study 3
Manipulation checks				
Language embodiment adapted from Choi *et al*. (2019)	How would you rate the language style of the voice assistant? (1) very literal (use of functional expressions) … (7) very figurative (use of vivid expressions)	single item	single item	single item
Consumer's consumption goal adapted from Chen *et al*. (2016)	Buying the headphones, you are looking for in the scenario is ... (1) a rational-oriented consumption (i.e., something you buy to fulfill a necessary function or task) ... (7) a pleasure-oriented consumption (i.e., something that is fun or an experience)	-	single item	single item
Focal psychological mechanisms				
Visual fluency adapted from Landwehr *et al*. (2011)	Constructing a mental image of this restaurant / these headphones ...			
	(1) felt difficult … (7) felt easy	0.907	0.933	0.915
	(1) was exhausting … (7) was relaxing	0.905	0.924	0.906
	(1) took a long time … (7) happened instantly	0.896	0.912	0.893
Credibility adapted from Filieri (2015) & Flavián *et al*. (2023)	The voice assistant's recommendation …			
	… was fair.	0.806	0.851	-
	… was accurate.	0.795	0.890	-
	… was credible.	0.909	0.922	-
	… contained convincing arguments.	0.799	0.857	-
	… was trustworthy.	0.905	0.912	-
Outcome				
Recommendation adoption adapted from Bleier *et al*. (2019) & Flavián *et al*. (2023)	Based on the recommendation of the voice assistant …			
	... it is very likely that I would book a table in the restaurant. / ... it is very likely that I would buy the headphones.	0.951	0.958	0.943
	... I would definitely book a table in the restaurant. / ... I would definitely buy the headphones.	0.900	0.922	0.911
	... I would consider booking a table in the restaurant. / ... I would consider buying the headphones.	0.866	0.905	0.894
Additional psychological mechanisms				
Expectancy dis-/confirmation adapted from Garvey *et al*. (2023) & Burgoon and Walther (1990)	Think about what your expectations were for the language style of the voice assistant. How does the language style that you were provided compare to your expectations?			
	(1) I expected exactly this language style … (7) I expected a completely different language style	single item	-	0.869
	(1) The language style was very usual … (7) The language style was very unusual …	-	-	0.906
	(1) The language style was very normal… (7) The language style was very abnormal …	-	-	0.875
Arousal adapted from Berger and Milkman (2012) & Cascio Rizzo *et al*. (2024)	Due to the voice assistant's language style, I felt ...			
	(1) very passive … (7) very active	-	-	0.915
	(1) very mellow … (7) very fired up	-	-	dropped
	(1) very low energy … (7) very high energy	-	-	0.892

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

Online Appendix D: Supplementary Material of Study 1

D1. Power analysis and data cleaning

A power analysis using G*Power (Faul *et al.*, 2007) with a significance level of 0.05 determined a minimum sample size of 128 participants to achieve a statistical power of 0.80 for detecting a medium effect size ($f = 0.25$). As we anticipated that some participants did not attend actively in our study or might encounter (technical) difficulties with the experiment, the sample of the main study initially consisted of 153 participants. We excluded those who exhibited click-through behavior and those who did not correctly answer attention checks related to the scenario ("What do you want to request via the voice assistant?") or the survey ("Please mark scale point (5) to indicate that you read the questionnaire carefully"). In addition, we excluded those who did not complete the questionnaire conscientiously ("I answered the questionnaire conscientiously"), those who had technical difficulties with the audio file, and finally, those who completed the study in an unrealistic time (Haenel *et al.*, 2019); following the guidelines of Meyvis and Van Osselaer (2018) for data cleaning. 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.

D2. Stimulus material

- Scenario of Study 1 -	
Imagine you are going on holiday to the German Baltic coast by car. Since you don't feel like cooking dinner in your vacation apartment the first night, you decide to look for a restaurant to eat out. You don't know the area, you're driving, and you don't want to spend a lot of time searching for a suitable restaurant, so you ask your car's voice assistant. You have already told it your preferences for the type of restaurant, its rating, location, and price, and now you get a restaurant recommendation.	
- Voice assistant's recommendation | Audio output - Language embodiment	
Figurative	**Literal**
Nearby is the restaurant Ennas Strandhus, which has been rated 4.5 out of 5 based on 360 reviews. The restaurant looks like a royal palace. The location on the beach will blow you away and the views of the endless sea are breathtaking. The food is a delicious taste experience from the region, and the service is like that of a royal reception. In summary: paradise disguised as a restaurant.	Nearby is the restaurant Ennas Strandhus, which has been rated 4.5 out of 5 based on 360 reviews. The interior of the restaurant is beautifully decorated. It is located very close to the beach and offers an unobstructed view of the sea. The food is very delicious and local, and the service is very professional. All in all, it is a very good choice for your restaurant visit.

Note: Stimulus material is translated from German into English

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

We used a scenario-based approach to ensure the interactions were identical except for the respective manipulations. This allowed us to control for confounding influences to achieve high internal validity. We chose the restaurant scenario because a significant percentage of consumers already make restaurant reservations with their VA (16%) or are willing to receive restaurant recommendations from their VA when asking for it (32%)

(PWC, 2018). Furthermore, restaurants represent a product category many consumers are familiar with or can easily relate to. We informed the participants that VAs are agents based on AI.

D3. Reliability and validity assessment

	M	*SD*	CR	CA	AVE	(1)	(2)	(3)
(1) Visual fluency	4.77	1.33	0.929	0.886	0.814	**0.902**		
(2) Credibility	4.56	1.19	0.925	0.898	0.713	0.214*	**0.845**	
(3) Recommendation adoption	4.77	1.31	0.932	0.888	0.820	0.283**	0.681***	**0.906**

Notes: N_{Study1} = 131; *M* = mean; *SD* = standard deviation; CR = composite reliability; CA = Cronbach's alpha; AVE = average variance extracted. Bolded diagonal elements are the square root of AVE; the values below represent the correlation between constructs; all HTMT ratios are below the threshold of 0.85;
*$p \leqslant 0.05$; **$p \leqslant 0.01$; ***$p \leqslant 0.001$

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

This study assesses the reliability and validity of the three multi-item constructs using established methods. Composite reliability (CR) and Cronbach's alpha (CA) values exceed the threshold of 0.7 for all constructs, confirming construct-level reliability (Hair *et al.*, 2019; Hulland *et al.*, 2018). The Fornell and Larcker (1981) test supports convergent validity, as the average variance extracted (AVE) for each construct surpasses 0.50. Additionally, discriminant validity is confirmed, as the square root of each construct's AVE (bolded diagonal values in the table) exceeds its correlations with other constructs (Fornell and Larcker, 1981). The heterotrait–monotrait (HTMT) method further demonstrates discriminant validity, with all HTMT ratios falling below the conservative threshold of 0.85 (Hair *et al.*, 2019; Henseler *et al.*, 2015). These results establish the reliability and validity of the tested constructs.

D4. Common method variance

Common Method Variance (CMV) refers to systematic, non-substantive influences on observed responses that arise from the measurement method, potentially distorting relationships among variables of interest (Hulland *et al.*, 2018). As a recognized threat to the validity of empirical research, CMV necessitates the adoption of both procedural (a priori) and statistical (post hoc) remedies to mitigate common method bias (CMB; Podsakoff *et al.*, 2024).

A Priori Methods: Procedural Remedies

To address potential CMV arising from social desirability, consistency motifs, and implicit theories, we implemented a comprehensive set of procedural remedies in line with the recommendations of Podsakoff *et al.* (2024). Specifically, we ensured participant anonymity to mitigate social desirability bias and provided financial incentives, such as raffle vouchers or monetary compensation, to motivate thoughtful and accurate responses.

Participants were informed that no specific knowledge or expertise was required and that their personal opinions were central to the study. This measure aimed to reduce the likelihood of socially desirable response patterns. To further minimize response editing influenced by implicit theories or consistency biases, we

separated the experimental task and subsequent survey, also ensuring that each measured construct was presented on distinct pages.

In addition, we experimentally manipulated the independent variable, language embodiment, to reduce reliance on subjective self-report measures and mitigate the risk of CMB. Pretests were conducted to ensure the clarity of instructions and survey items, thereby reducing ambiguity that might contribute to CMB. Finally, we varied scale formats throughout the survey, as detailed in Online Appendix C, to further minimize biases associated with social desirability.

Post Hoc Methods: Statistical Controls

Recognizing that a priori methods may not fully eliminate CMV, we implemented statistical controls using the full collinearity and marker variable approaches.

First, following Kock and Lynn (2012), we assessed vertical collinearity (predictor-predictor collinearity in regression models) and lateral collinearity (predictor-criterion relationships). The later approach detects multicollinearity introduced by CMV, where artificial correlations between predictors and criteria emerge due to shared method variance rather than substantive relationships. The variance inflation factors (VIFs) for assessing vertical and lateral collinearity are all below the recommended threshold of 3.3 (Kock and Lynn, 2012), indicating no multicollinearity concerns or evidence of CMB.

Second, we applied the marker variable approach by Lindell and Whitney (2001), using a variable that is conceptually unrelated to the research constructs but shares measurement characteristics (Podsakoff *et al.*, 2024). Therefore, we used post hoc the item "I enjoy taking responsibility in situations that require complex thinking", adapted from Amabile *et al*. (1994), which measures general intrinsic work engagement. We believe that, in our context, this item should have no theoretical correlation with our research constructs. Consistent with this expectation, this variable demonstrates non-significant correlations with the constructs of the research model (all $r \leq 0.08$, $p \geq 0.36$). When incorporated as a covariate in analysis of variance (ANOVA) tests and mediation testing, the marker variable does not alter the substantive results, verifying that it is unlikely our research is affected by CMB. By implementing both procedural and statistical remedies, we address the potential impact of CMV on our findings. These measures suggest that the substantive relationships in our model are unlikely to be influenced by CMB, supporting the validity of our results.

D5. Results of the pre-study

Our pre-study comprises a final sample of 33 participants from a European country through the crowdsourcing platform Prolific (45.45% female, M_{age} = 26.12 years, SD = 6.09), which are randomly and evenly distributed across the two scenarios. The results indicated that the manipulation was successful ($M_{\text{figurative}}$ = 6.12, SE = 0.22; M_{literal} = 3.56, SE = 0.43; $t(31)$ = -5.38, $p < 0.001$). Participants who heard the product recommendation in figurative language rated it as significantly more figurative compared to those who heard the recommendation in literal language. Moreover, the participants perceived the scenarios as realistic (M = 5.55, SD = 1.25) and Cronbach's α of the constructs are above the recommended threshold of 0.7.

D6. Descriptive statistics

	Visual fluency	Credibility	Recommendation adoption
Figurative language	5.01 (1.37)	4.12 (1.29)	4.55 (1.39)
Literal language	4.50 (1.25)	5.03 (0.87)	5.00 (1.19)
Total sample	4.77 (1.33)	4.56 (1.19)	4.77 (1.31)

Notes: $N_{Study1} = 131$; Means with standard deviations in parentheses. Measured on 7-point Likert scales.

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

D7. Results with covariates

When controlling for prior experience with voice assistants (measured using a single item: "How would you rate your experience with using voice assistants?" on a 7-point scale ranging from (1) very low to (7) very high, adapted from Dabholkar (1996); $M = 4.27$, $SD = 1.59$) and audio repetition (measured using a single item: "How many times have you listened to the VA message (audio output)? Please select the appropriate number"; $M = 1.22$, $SD = 0.42$), the results of the tested research model remain robust.

In detail, the first analysis of covariance (ANCOVA) results indicate that language embodiment significantly affects visual fluency ($F(1, 127) = 4.94$, $p < 0.05$, $d = 0.39$). In addition, neither prior experience with voice assistants ($F(1, 127) = 0.10$, $p = 0.752$) nor audio repetition ($F(1, 127) = 1.29$, $p = 0.258$) affects visual fluency.

Furthermore, the second ANCOVA reveals that language embodiment has a significant effect on credibility ($F(1, 127) = 21.74$, $p < 0.001$, $d = 0.83$), while neither prior experience with voice assistants ($F(1, 127) = 0.13$, $p = 0.720$) nor audio repetition ($F(1, 127) = 0.21$, $p = 0.644$) shows any effect.

The mediation analysis including the two covariates indicates a significant indirect effect of figurative (vs. literal) language on recommendation adoption through both visual fluency ($\beta = 0.06$, $SE = 0.04$, $CI_{95\%} = [0.002, 0.182]$) and credibility ($\beta = -0.68$, $SE = 0.16$, $CI_{95\%} = [-1.042, -0.395]$). There is no significant direct effect of language embodiment on recommendation adoption ($\beta = 0.17$, $SE = 0.19$, $p = 0.366$). In addition, there is also no direct effect of prior experience with voice assistants ($\beta = -0.05$, $SE = 0.05$, $p = 0.336$) nor audio repetition ($\beta = -0.04$, $SE = 0.20$, $p = 0.840$) on recommendation adoption.

Online Appendix E: Supplementary Material of Study 2

E1. Power analysis and data cleaning

Considering the effect sizes found in Study 1, suggesting also small effect sizes, we decided to calculate the sample size for the second study a priori with a smaller effect size (f = 0.17) than in Study 1, a significance level of 0.05 and a statistical power of 0.80. Therefore, the total sample size should be 274 participants. Following our argumentation of the first study, our study initially included 356 participants. We used the same data cleaning process as in Study 1. 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.

E2. Stimulus material

- Scenario of Study 2 and 3 – Consumer's consumption goal	
Hedonic	**Utilitarian**
Imagine you are looking for a new pair of over-ear headphones because your current pair has broken. You need the headphones for personal use, to listen to music or podcasts. This is one of your favorite leisure activities and gives you a lot of pleasure. Headphones therefore have an emotional value to you because you can totally relax while listening to music. The number of headphones on the market is very large and therefore quite overwhelming. You don't want to spend a lot of time searching, so you start an interaction with your voice assistant. You have already told it your preferences for type, color, reviews, and price, and now you ask it to recommend a new pair of over-ear headphones.	Imagine you are looking for a new pair of over-ear headphones because your current pair has broken. You need the headphones for your work, to answer customer inquiries, and to attend online meetings. The headphones are a device that helps you in doing your job. Headphones therefore have a functional value to you because they allow you to do your job efficiently. The number of headphones on the market is very large and therefore quite overwhelming. You don't want to spend a lot of time searching, so you start an interaction with your voice assistant. You have already told it your preferences for type, color, reviews, and price, and now you ask it to recommend a new pair of over-ear headphones.
- Voice assistant's recommendation | Audio output - Language embodiment	
Figurative	**Literal**
Based on your preferences, I recommend the new over-ear headphones from the brand Osmo. The wearing comfort is unbeatable, as the elegant ear cups hug your ears like silk. They are also as light as a feather. The gentle curves create an appealing design and the metallic silver shimmers in the sunlight. The sound quality will blow your mind - a breathtaking sound with which you can fully immerse yourself and that will make your heart beat faster. What's more, these headphones are currently in great demand, not least because the price won't burn a big hole in your wallet. Shall I add these headphones to the shopping cart?	Based on your preferences, I recommend the new over-ear headphones from the brand Osmo. They are comfortable to wear, as the comfortable ear cups automatically adapt to your ears. They are also very light. The rounded shape provides an attractive design and the silver color makes a very good impression. The sound quality is of high quality - a great sound with which you will no longer notice disturbing noises in your surroundings and which will impress you. What's more, these headphones are currently very popular, not least because the price is not very high. Shall I add these headphones to the shopping cart?

Note: Stimulus material is translated from German into English

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

We chose the headphone scenario because consumers are already using voice assistants to purchase electronics (22%) or are willing to receive recommendations from a voice assistant when asking about electronics (26%) (PWC, 2018). Additionally, many consumers are familiar with headphones, which can serve both hedonic and utilitarian purposes and have been used for both purposes in prior studies (Wien and Peluso, 2021). We informed the participants that VAs are agents based on AI.

E3. Reliability and validity assessment

	M	*SD*	CR	CA	AVE	(1)	(2)	(3)
(1) Visual fluency	4.72	1.54	0.945	0.913	0.852	**0.923**		
(2) Credibility	4.42	1.42	0.948	0.931	0.786	0.233***	**0.887**	
(3) Recommendation adoption	3.66	1.63	0.949	0.920	0.862	0.315***	0.770***	**0.929**

Notes: N_{Study2} = 317; *M* = mean; *SD* = standard deviation; CR = composite reliability; CA = Cronbach's alpha; AVE = average variance extracted. Bolded diagonal elements are the square root of AVE; the values below represent the correlation between constructs; all HTMT ratios are below the conservative threshold of 0.85; *$p \leqslant 0.05$; **$p \leqslant 0.01$; ***$p \leqslant 0.001$

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

The same analytical methods as in Study 1 are applied in Study 2, confirming reliability and validity for all three multi-item constructs, with CR, CA, and AVE values meeting thresholds and Fornell-Larcker and HTMT criteria validating discriminant validity.

E4. Common method variance

As in Study 1, we implemented the same procedural and statistical remedies in Study 2 to mitigate CMB. For a comprehensive overview and detailed description of the a priori (procedural) and post hoc (statistical) methods, we refer to Online Appendix D4. Below, we focus solely on the results of the post hoc methods.

First, we assessed both vertical and lateral collinearity. The VIFs for these assessments are all below the threshold of 3.3, indicating no concerns of multicollinearity or evidence of CMB.

Second, we used post hoc the marker variable "I really enjoy tasks where I have to find new solutions to problems", another item adapted from Amabile *et al*. (1994) that measures general intrinsic work engagement. Our results show that the marker variable exhibits weak and non-significant correlations with the constructs in our research model (all $r \leq 0.06$, $p \geq 0.26$). When included as a covariate in our ANOVA tests and mediation testing, the marker variable does not alter the substantive results. These findings confirm that CMB is unlikely to influence our research outcomes.

By employing both procedural and statistical remedies, we effectively address the potential impact of CMV on the findings of Study 2. These measures reinforce that the substantive relationships in our model are not influenced by CMB, thereby supporting the validity of our results.

E5. Results of the pre-study

Our pre-study compromises a final sample of 72 participants from a European university (65.28% female, M_{age} = 29.63 years, SD = 10.43), which are randomly and evenly distributed across the four scenarios. The results indicated that the manipulation of language embodiment was successful ($M_{\text{figurative}}$ = 5.71, SE = 0.30; M_{literal} = 3.76, SE = 0.28; $t(70)$ = -4.70, $p < 0.001$), meaning that participants who heard the recommendation in a figurative language evaluated this style accordingly as more figurative than those who heard the product recommendation in literal language. The manipulation of consumers consumption goal was also successful ($M_{hedonic}$ = 5.28, SE = 0.28; $M_{utilitarian}$ = 3.30, SE = 0.37; $t(70)$ = -4.29, $p < 0.001$), meaning that participants who were in the hedonic scenario also imagined that they had a hedonic consumption goal than those who were in the utilitarian scenario. Moreover, the participants perceived the scenarios as realistic (M = 5.58, SD = 1.33) and Cronbach's α of the constructs are above the recommended level of 0.7.

E6. Overview of the ANOVA results for visual fluency

	Dependent variable: visual fluency	
Main effect: Language embodiment	$F(1, 313) = 6.99$, $p < 0.01$	
	$M_{figurative}$ = 4.95, SE = 0.12; $M_{literal}$ = 4.50, SE = 0.12; t = 2.64, $p < 0.01$	
Main effect: Consumer's consumption goal	$F(1, 313) = 2.02$, $p = 0.156$	
	$M_{hedonic}$ = 4.60, SE = 0.12; $M_{utilitarian}$ = 4.85, SE = 0.12; t = -1.42, $p = 0.156$	
Interaction effect: Language embodiment x consumer's consumption goal	$F(1, 313) = 1.78$, $p = 0.183$	
	$M_{figurative \times hedonic}$ = 4.94, SE = 0.17; $M_{figurative \times utilitarian}$ = 4.96, SE = 0.17; t = -0.06, $p = 0.952$	$M_{literal \times hedonic}$ = 4.26, SE = 0.17; $M_{literal \times utilitarian}$ = 4.74, SE = 0.17; t = -1.96, $p = 0.051$
	$M_{figurative \times hedonic}$ = 4.94, SE = 0.17; $M_{literal \times hedonic}$ = 4.26, SE = 0.17; t = 2.82, $p < 0.01$	$M_{figurative \times utilitarian}$ = 4.96, SE = 0.17; $M_{literal \times utilitarian}$ = 4.74, SE = 0.17; t = 0.92, $p = 0.357$

Note: N_{Study2} = 317

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

E7. Overview of the ANOVA results for credibility

	Dependent variable: credibility	
Main effect: Language embodiment	$F(1, 313) = 10.16$, $p < 0.01$	
	$M_{figurative} = 4.17$, $SE = 0.11$; $M_{literal} = 4.67$, $SE = 0.11$; $t = -3.19$, $p < 0.01$	
Main effect: Consumer's consumption goal	$F(1, 313) = 2.16$, $p = 0.143$	
	$M_{hedonic} = 4.53$, $SE = 0.11$; $M_{utilitarian} = 4.30$, $SE = 0.11$; $t = 1.47$, $p = 0.143$	
Interaction effect: Language embodiment x consumer's consumption goal	$F(1, 313) = 3.11$, $p = .079$	
	$M_{figurative \times hedonic} = 4.42$, $SE = 0.16$; $M_{figurative \times utilitarian} = 3.92$, $SE = 0.16$; $t = 2.27$, $p < 0.05$	$M_{literal \times hedonic} = 4.64$, $SE = 0.15$; $M_{literal \times utilitarian} = 4.69$, $SE = 0.16$; $t = -0.21$, $p = 0.835$
	$M_{figurative \times hedonic} = 4.42$, $SE = 0.16$; $M_{literal \times hedonic} = 4.64$, $SE = 0.15$; $t = -1.01$, $p = 0.314$	$M_{figurative \times utilitarian} = 3.92$, $SE = 0.16$; $M_{literal \times utilitarian} = 4.69$, $SE = 0.16$; $t = -3.50$, $p < 0.01$
Note: $N_{Study2} = 317$		

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

E8. Descriptive statistics

Experimental condition	Visual fluency			Credibility			Recommendation adoption		
	HC goal	UC goal	All goals	HC goal	UC goal	All goals	HC goal	UC goal	All goals
Figurative language	4.94 (1.52)	4.96 (1.47)	4.95 (1.49)	4.42 (1.49)	3.92 (1.53)	4.16 (1.53)	3.60 (1.70)	3.30 (1.53)	3.45 (1.62)
Literal language	4.26 (1.68)	4.74 (1.38)	4.49 (1.55)	4.64 (1.29)	4.69 (1.25)	4.67 (1.26)	3.81 (1.64)	3.94 (1.61)	3.87 (1.62)
Total sample	4.60 (1.64)	4.85 (1.43)	4.72 (1.54)	4.53 (1.39)	4.30 (1.45)	4.42 (1.42)	3.70 (1.67)	3.62 (1.60)	3.66 (1.63)

Notes: $N_{Study2} = 317$; Means with standard deviations in parentheses; Measured on 7-point Likert scales; HC Goal = Hedonic consumption goal; UC goal = Utilitarian consumption goal

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

E9. Results with covariates

When controlling for prior experience with voice assistants (M = 3.33; SD = 1.64) and audio repetition (M = 1.08, SD = 0.34), the tested research model remains robust. Specifically, the first ANCOVA results reveal a significant main effect of language embodiment on visual fluency ($F(1, 311) = 7.27$, $p < 0.01$, $d = 0.30$). The main effect of consumption goal is not significant ($F(1, 311) = 2.13$, $p = 0.146$), and there is no significant interaction between language embodiment and consumption goal ($F(1, 311) = 1.87$, $p = 0.173$). Planned contrasts indicate that when consumers pursue a hedonic (vs. utilitarian) consumption goal, figurative language used by a voice assistant does not significantly influence visual fluency (t = -0.06, $p = 0.948$). Neither prior experience with voice assistants ($F(1, 311) = 0.38$, $p = 0.538$) nor audio repetition ($F(1, 311) = 0.73$, $p = 0.395$) significantly affect visual fluency.

The second ANCOVA results show a significant main effect of language embodiment on credibility ($F(1, 311) = 9.62$, $p < 0.01$, $d = 0.36$), while the main effect of consumption goal is not significant ($F(1, 311) = 1.86$, $p = 0.173$). However, there is a marginally significant interaction between language embodiment and consumption goal ($F(1, 311) = 3.04$, $p = .082$, $d = 0.20$). Planned contrasts reveal that figurative language increases perceived credibility when consumers pursue a hedonic (vs. utilitarian) consumption goal (t = 2.06, $p < 0.05$, $d = 0.34$). Neither prior experience with voice assistants ($F(1, 311) = 1.95$, $p = 0.163$) nor audio repetition ($F(1, 311) = 0.00$, $p = 0.946$) significantly affect credibility.

The mediation analysis, accounting for the two covariates, reveals a positive indirect effect of figurative (vs. literal) language on recommendation adoption via visual fluency (β = 0.07, SE = 0.03, $CI_{95\%}$ = [0.018, 0.155]) and a negative indirect effect via credibility (β = -0.41, SE = 0.14, $CI_{95\%}$ = [-0.689, -0.151]). There is no significant direct effect of language embodiment on recommendation adoption (β = -0.07, SE = 0.12, $p = 0.553$). Additionally, neither prior experience with voice assistants (β = 0.02, SE = 0.03, $p = 0.483$) nor audio repetition (β = 0.02, SE = 0.17, $p = 0.908$) affect recommendation adoption. Finally, the results indicate no indirect effect of figurative language on recommendation adoption via visual fluency when consumers pursue a hedonic (vs. utilitarian) consumption goal (β = -0.002, SE = 0.05, $CI_{95\%}$ = [-0.101, 0.085]), but a significant positive indirect effect via credibility (β = 0.41, SE = 0.20, $CI_{95\%}$ = [0.004, 0.797]).

Online Appendix F: Supplementary Material of Study 3

F1. Power analysis and data cleaning

Considering the small effect sizes in Studies 1 and 2, we calculated the required sample size for Study 3 a priori, using the same small effect size (f = 0.17) as in Study 2, with a significance level of 0.05 and a power of 0.80. This resulted in a minimum required sample size of 274 participants. To account for potential data loss, we initially recruited 360 participants. We applied the same data cleaning process as in Studies 1 and 2. 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.

F2. Reliability and validity assessment

	M	***SD***	**CR**	**CA**	**AVE**	**(1)**	**(2)**	**(3)**	**(4)**
(1) Expectancy disconfirmation	3.66	1.54	0.914	0.859	0.780	**0.883**			
(2) Arousal	4.25	1.19	0.910	0.805	0.835	-0.076	**0.914**		
(3) Visual fluency	5.04	1.33	0.931	0.889	0.818	-0.123*	0.504***	**0.904**	
(5) Recommendation adoption	4.16	1.42	0.939	0.903	0.838	-0.305***	0.476***	0.386***	**0.915**

Notes: N_{Study3} = 314; *M* = mean; *SD* = standard deviation; CR = composite reliability; CA = Cronbach's alpha; AVE = average variance extracted. Bolded diagonal elements are the square root of AVE; the values below represent the correlation between constructs; all HTMT ratios are below the conservative threshold of 0.85; *$p \leqslant 0.05$; **$p \leqslant 0.01$; ***$p \leqslant 0.001$

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

The same analytical methods as in Study 1 confirm reliability and validity for all four multi-item constructs of Study 3, with CR, CA, and AVE values meeting thresholds and Fornell-Larcker and HTMT criteria supporting discriminant validity.

F3. Common method variance

As in the previous studies, we implemented the same procedural and statistical remedies in Study 3 to address potential CMB. Online Appendix D4 outlines the procedural remedies, while we summarize the statistical results here.

First, vertical and lateral collinearity checks indicate VIF values below 3.3, showing no multicollinearity issues or signs of CMB. Second, following Simmering *et al*. (2015), we used a priori the marker variable "I really like the color blue" (7-point Likert scale), which should be conceptually unrelated to our constructs. The marker variable exhibits weak correlations with all constructs ($r \leq 0.10$, $p \geq 0.09$), with no significant correlations present except for a correlation with recommendation adoption at the 10% level. However, including the marker variable as a covariate in ANOVA tests and mediation testing, the marker variable does not alter the substantive results.

Overall, these results suggest that CMB does not influence our findings. By employing procedural and statistical remedies, we ensure the robustness and validity of the results in Study 3.

F4. Hypotheses testing of H1 and H3 in Study 3

In Study 3, we re-tested hypotheses H1 and H3. First, a two-way ANOVA was conducted with language embodiment and consumption goal as independent variables and visual fluency as the dependent variable. The results reveal a significant main effect of language embodiment on visual fluency ($F(1, 310) = 4.46$, $p < 0.05$, $d = 0.24$). Figurative language leads to significantly higher visual fluency than literal language ($M_{figurative} = 5.20$, $SE = 0.11$; $M_{literal} = 4.89$, $SE = 0.10$), with a small effect size. The main effect of consumption goal is not significant ($F(1, 310) = 0.00$, $p = 0.988$), nor is the interaction of language embodiment and consumption goal ($F(1, 310) = 2.00$, $p = 0.158$). Planned contrasts further reveal that when consumers pursued a hedonic (vs. utilitarian) consumption goal, the use of figurative language by a voice assistant does not significantly enhance visual fluency ($M_{figurative \times hedonic} = 5.10$, $SE = 0.15$; $M_{figurative \times utilitarian} = 5.31$, $SE = 0.15$; $t = -0.97$, $p = 0.332$).

Second, we conducted a mediation analysis using a bootstrapping procedure with 5,000 iterations and bias-corrected confidence intervals. The analysis reveals a positive indirect effect of figurative (vs. literal) language on recommendation adoption via visual fluency ($\beta = 0.13$, $SE = 0.07$, $CI_{95\%} = [0.013, 0.273]$). Thus, Study 3 supports H1. Unlike Studies 1 and 2, a significant direct effect of language embodiment on recommendation adoption is identified ($\beta = -0.32$, $SE = 0.05$, $p < 0.05$). Additionally, in no support for 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 ($\beta = -0.08$, $SE = 0.09$, $CI_{95\%} = [-0.290, 0.078]$). In summary, Study 3 confirms the robustness of the findings for H1 and H3, aligning with the results of Studies 1 and 2. Online Appendix F2 presents the correlations table.

F5. Descriptive statistics

Experimental condition	Expectancy disconfirmation			Arousal			Visual fluency		
	HC goal	UC goal	All goals	HC goal	UC goal	All goals	HC goal	UC goal	All goals
Figurative language	4.24 (1.43)	4.19 (1.67)	4.22 (1.55)	4.55 (1.05)	4.41 (1.16)	4.48 (1.11)	5.10 (1.32)	5.31 (1.33)	5.20 (1.32)
Literal language	3.17 (1.35)	3.14 (1.35)	3.16 (1.34)	4.20 (1.17)	3.90 (1.29)	4.05 (1.23)	4.99 (1.16)	4.78 (1.48)	4.89 (1.33)
Total sample	3.68 (1.48)	3.64 (1.60)	3.66 (1.54)	4.37 (1.12)	4.14 (1.25)	4.25 (1.19)	5.04 (1.24)	5.03 (1.43)	5.04 (1.33)

Experimental condition	Recommendation adoption		
	HC goal	UC goal	All goals
Figurative language	4.07 (1.45)	4.07 (1.49)	4.07 (1.46)
Literal language	4.25 (1.31)	4.26 (1.47)	4.25 (1.38)
Total sample	4.16 (1.37)	4.17 (1.48)	4.16 (1.42)

Notes: $N_{Study3} = 314$; Means with standard deviations in parentheses; Measured on 7-point Likert scales; HC goal = Hedonic consumption goal; UC goal = Utilitarian consumption goal

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

F6. Results with covariates

Testing the underlying mechanisms of expectancy disconfirmation and arousal, while controlling for prior experience with voice assistants (M = 4.51; SD = 1.51) and audio repetition (M = 1.25, SD = 0.65), reveals some changes in the prior results.

In detail, ANCOVA results show that figurative (vs. literal) language significantly increases expectancy disconfirmation ($M_{figurative}$ = 4.22, SE = 0.11; $M_{literal}$ = 3.16, SE = 0.11; F(1, 308) = 41.43, $p < 0.001$, d = 0.73). The consumption goal shows no significant effect (F(1, 308) = 0.07, p = 0.791), and there is no significant interaction between language embodiment and the consumption goal (F(1, 308) = 0.00 p = 0.950). Prior experience with voice assistants (F(1, 308) = 28.19, $p < 0.001$) significantly affects expectancy disconfirmation, while audio repetition (F(1, 308) = 0.80, p = 0.372) does not.

In terms of arousal, figurative (vs. literal) language significantly increases arousal ($M_{figurative}$ = 4.48, SE = 0.10; $M_{literal}$ = 4.05, SE = 0.09; F(1, 308) = 13.31, $p < 0.001$, d = 0.36). Additionally, the consumption goal had a significant effect on arousal ($M_{hedonic}$ = 4.37, SE = 0.09; $M_{utilitarian}$ = 4.14, SE = 0.09; F(1, 308) = 4.20, $p < 0.05$, d = 0.23). The interaction effect of language embodiment and consumption goal is not significant (F(1, 308) = 0.03, p = 0.869). Both covariates, prior experience with voice assistants (F(1, 308) = 21.95, $p < 0.001$) and audio repetition (F(1, 308) = 7.00, $p < 0.01$), have significant effects on arousal.

The mediation analysis with covariates shows no significant indirect effect of language embodiment on recommendation adoption via expectancy disconfirmation, arousal, or visual fluency (β = -0.01, SE = 0.01, $CI_{95\%}$ = [-0.020, 0.000]). However, a significant positive effect of figurative language on recommendation adoption exists, sequentially mediated by greater arousal and increased visual fluency (β = 0.04, SE = 0.02, $CI_{95\%}$ = [0.015, 0.106]). The direct effect of language embodiment on recommendation adoption is not significant (β = -0.20, SE = 0.14, p = 0.163). Prior experience with voice assistants (β = 0.11, SE = 0.05, $p < 0.05$) but not audio repetition (β = 0.10, SE = 0.10, p = 0.346) has a significant effect on recommendation adoption. Moreover, the results indicate no significant indirect effect of figurative language on recommendation adoption via expectancy disconfirmation, arousal, and visual fluency when consumers pursue a hedonic (vs. utilitarian) consumption goal (β = -0.00, SE = 0.00, $CI_{95\%}$ = [-0.007, 0.003]). Similarly, no significant indirect effect is observed via arousal and visual fluency alone under the same conditions (β = 0.03, SE = 0.03, $CI_{95\%}$ = [-0.005, 0.101]). The table below presents the results of the mediation testing.

Interestingly, when covariates are included, expectancy disconfirmation no longer plays a significant role in the overall model, leaving arousal as the sole mediator between language embodiment and visual fluency. As demonstrated in the main manuscript (without covariates), expectancy disconfirmation already showed a limited explanatory role in the relationship between language embodiment and recommendation adoption, and this finding reinforces those results. Furthermore, unlike in Studies 1 and 2, prior experience significantly affects all constructs in the model, highlighting the need for closer examination in future research. In addition, audio repetition significantly influences arousal, but does not affect other constructs. This indicates that while audio repetition may influence specific emotional responses, its overall impact is limited.

Overview of mediation testing of Study 3 with covariates:

Path	*Coeff.*	*SE*	CI 95 %		Mediation
			LLCI	*ULCI*	
Language embodiment → expectancy disconfirmation → arousal → visual fluency → recommendation adoption	-0.01	0.01	-0.020	0.000	×
Language embodiment → arousal → visual fluency → recommendation adoption	0.04	0.02	0.015	0.106	✓
Figurative language ⨯ consumer's consumption goal → expectancy disconfirmation → arousal → visual fluency → recommendation adoption	-0.00	0.00	-0.007	0.003	×
Figurative language ⨯ consumer's consumption goal → arousal → visual fluency → recommendation adoption	0.03	0.03	-0.005	0.101	×

Notes: 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): The above table was created by the authors.

References

Amabile, T.M., Hill, K.G., Hennessey, B.A. and Tighe, E.M. (1994), "The Work Preference Inventory: Assessing Intrinsic and Extrinsic Motivational Orientations", *Journal of Personality and Social Psychology*, Vol. 66 No. 5, pp. 950–967, doi: https://doi.org/10.1037/0022-3514.66.5.950.

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.

Berger, J. and Milkman, K.L. (2012), "What Makes Online Content Viral?", *Journal of Marketing Research*, Vol. 49 No. 2, pp. 192–205, doi: 10.1509/jmr.10.0353.

Bleier, A., Harmeling, C.M. and Palmatier, R.W. (2019), "Creating Effective Online Customer Experiences", *Journal of Marketing*, Vol. 83 No. 2, pp. 98–119, doi: 10.1177/0022242918809930.

Burgoon, J.K. and Walther, J.B. (1990), "Nonverbal Expectancies and the Evaluative Consequences of Violations", *Human Communication Research*, Vol. 17 No. 2, pp. 232–265, doi: 10.1111/j.1468-2958.1990.tb00232.x.

Cascio Rizzo, G.L., Villarroel Ordenes, F., Pozharliev, R., De Angelis, M. and Costabile, M. (2024), "How High-Arousal Language Shapes Micro- Versus Macro-Influencers' Impact", *Journal of Marketing*, Vol. 88 No. 4, pp. 107–128, doi: 10.1177/00222429231207636.

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.

Chen, C.Y., Lee, L. and Yap, A.J. (2016), "Control Deprivation Motivates Acquisition of Utilitarian Products", *Journal of Consumer Research*, Vol. 43 No. 6, pp. 1031–1047, doi: 10.1093/jcr/ucw068.

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.

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.

Efthymiou, F., Hildebrand, C., De Bellis, E. and Hampton, W.H. (2024), "The Power of AI-Generated Voices: How Digital Vocal Tract Length Shapes Product Congruency and Ad Performance", *Journal of Interactive Marketing*, Vol. 59 No. 2, pp. 117–134, doi: 10.1177/10949968231194905.

Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007), "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences", *Behavior Research Methods*, Vol. 39 No. 2, pp. 175–191, doi: 10.3758/BF03193146.

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 & Marketing*, Vol. 40 No. 2, pp. 328–346, doi: 10.1002/mar.21765.

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.

Garvey, A.M., Kim, T. and Duhachek, A. (2023), "Bad News? Send an AI. Good News? Send a Human", *Journal of Marketing*, Vol. 87 No. 1, pp. 10–25, doi: 10.1177/00222429211066972.

Gnewuch, U., Morana, S., Adam, M.T.P. and Maedche, A. (2022), "Opposing Effects of Response Time in Human–Chatbot Interaction: The Moderating Role of Prior Experience", *Business & 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. (2022), "Multiexperience", *Business & Information Systems Engineering*, Vol. 64 No. 6, pp. 813–823, doi: 10.1007/s12599-022-00766-8.

Haenel, C.M., Wetzel, H.A. and Hammerschmidt, M. (2019), "The Perils of Service Contract Divestment: When and Why Customers Seek Revenge and How It Can Be Attenuated", *Journal of Service Research*, Vol. 22 No. 3, pp. 301–322, doi: 10.1177/1094670519835312.

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.

Henseler, J., Ringle, C.M. and Sarstedt, M. (2015), "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 115–135, doi: 10.1007/s11747-014-0403-8.

Hernández Ortega, B.I. and Lucia-Palacios, L. (2023), "Trust in word of voice communication: why consumers adhere to purchase recommendations made by smart voice assistants", *Marketing Intelligence & Planning*, Vol. 41 No. 8, pp. 1093–1120, doi: 10.1108/MIP-10-2022-0466.

Hulland, J., Baumgartner, H. and Smith, K.M. (2018), "Marketing survey research best practices: evidence and recommendations from a review of JAMS articles", *Journal of the Academy of Marketing Science*, Vol. 46 No. 1, pp. 92–108, doi: 10.1007/s11747-017-0532-y.

Kock, N. and Lynn, G. (2012), "Lateral Collinearity and Misleading Results in Variance-Based SEM: An Illustration and Recommendations", *Journal of the Association for Information Systems*, Vol. 13 No. 7, pp. 546–580, doi: 10.17705/1jais.00302.

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.

Lindell, M.K. and Whitney, D.J. (2001), "Accounting for common method variance in cross-sectional research designs.", *Journal of Applied Psychology*, Vol. 86 No. 1, pp. 114–121, doi: 10.1037/0021-9010.86.1.114.

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

Meyvis, T. and Van Osselaer, S.M.J. (2018), "Increasing the Power of Your Study by Increasing the Effect Size", *Journal of Consumer Research*, Vol. 44 No. 5, pp. 1157–1173, doi: 10.1093/jcr/ucx110.

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.

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.

PWC. (2018), "Consumer Intelligence Series - Prepare for the voice revolution", available at: https://www.pwc.com/us/en/advisory-services/publications/consumer-intelligence-series/voice-assistants.pdf.

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, p. 106359, doi: 10.1016/j.chb.2020.106359.

Simmering, M.J., Fuller, C.M., Richardson, H.A., Ocal, Y. and Atinc, G.M. (2015), "Marker Variable Choice, Reporting, and Interpretation in the Detection of Common Method Variance: A Review and Demonstration", *Organizational Research Methods*, Vol. 18 No. 3, pp. 473–511, doi: 10.1177/1094428114560023.

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 & 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.

Yoganathan, V. and Osburg, V.-S. (2024), "The mind in the machine: Estimating mind perception's effect on user satisfaction with voice-based conversational agents", *Journal of Business Research*, Vol. 175, p. 114573, doi: 10.1016/j.jbusres.2024.114573.