Journal of the Academy of Marketing Science
https://doi.org/10.1007/s11747-022-00868-5

ORIGINAL EMPIRICAL RESEARCH

Voice bots on the frontline: Voice-based interfaces enhance flow-like consumer experiences & boost service outcomes

Naim Zierau[1] · Christian Hildebrand[2] · Anouk Bergner[2] · Francesc Busquet[2] · Anuschka Schmitt[1] · Jan Marco Leimeister[1,3]

Received: 15 February 2021 / Accepted: 16 April 2022

Abstract

Voice-based interfaces provide new opportunities for firms to interact with consumers along the customer journey. The current work demonstrates across four studies that voice-based (as opposed to text-based) interfaces promote more flow-like user experiences, resulting in more positively-valenced service experiences, and ultimately more favorable behavioral firm outcomes (i.e., contract renewal, conversion rates, and consumer sentiment). Moreover, we also provide evidence for two important boundary conditions that reduce such flow-like user experiences in voice-based interfaces (i.e., semantic disfluency and the amount of conversational turns). The findings of this research highlight how fundamental theories of human communication can be harnessed to create more experiential service experiences with positive downstream consequences for consumers and firms. These findings have important practical implications for firms that aim at leveraging the potential of voice-based interfaces to improve consumers' service experiences and the theory-driven "conversational design" of voice-based interfaces.

Keywords Voice-based interface · Speech interaction · Natural language processing · Artificial intelligence · Media richness · Flow · Service research

Introduction

The proliferation of voice-based interfaces driven by recent advances in artificial intelligence and natural language processing is radically transforming how consumers interact with firms along every touchpoint of the customer journey (Huang & Rust, 2018; Mende et al., 2019; van Doorn et al., 2017). Voice-based interfaces have emerged as a novel interaction paradigm, allowing firms to engage consumers in increasingly personal ways, almost as if they are talking to a human service provider (Hollebeek et al., 2021; Huang & Rust, 2021). From ordering a coffee with the Starbucks Barista bot (Perez, 2017) to speaking to Bank of America's virtual advisor Erica to conduct money wires (Fuscaldo, 2019). Voice-based interfaces increasingly support consumers in completing service requests more naturally compared to traditional text-based interface modalities (Singh et al., 2020; Wirtz et al., 2018). Indeed, recent industry reports indicate that the number of voice-enabled interfaces is rising exponentially (e.g., 8.4 billion voice-based interfaces by 2024, Smith, 2018).

Stephanie Noble and Martin Mende served as Guest Editors for this article.

✉ Christian Hildebrand
christian.hildebrand@unisg.ch

Naim Zierau
naim.zierau@unisg.ch

Anouk Bergner
anouk.bergner@unisg.ch

Francesc Busquet
francesc.busquet@unisg.ch

Anuschka Schmitt
anuschka.schmitt@unisg.ch

Jan Marco Leimeister
janmarco.leimeister@unisg.ch

[1] Institute of Information Management, University of St. Gallen, Müller-Friedberg-Strasse 8, CH-9000 St. Gallen, Switzerland

[2] Institute of Behavioral Science & Technology, University of St.Gallen, Torstrasse 25, CH-9000 St. Gallen, Switzerland

[3] Information Systems, Research Center for IS Design (ITeG), University of Kassel, Pfannkuchstraße 1, 34121, Kassel, Germany

Published online: 21 June 2022

Such voice-based interfaces combine automatic speech recognition with natural language processing to enable a fully voice-mediated service experience (Diederich et al., 2022; Seaborn et al., 2021). The fact that consumers execute service requests through voice-based interfaces highlights that they are both increasingly more comfortable interacting through voice-based interfaces and that firms expanded the use of voice-mediated service channels across touch points (e.g., 72% of US consumers have already interacted with voice-based interfaces in business settings according to a study by PwC; Hayes & Wagner, 2018). This burgeoning trend raises the question of how consumers respond to such voice-based interactions in customer service settings, how they perceive the firm providing such voice-based service exchanges, and what underlying psychological mechanism might explain these effects with which downstream effects on consumers and firms.

To answer these questions, the current work introduces a novel conceptualization of voice-based interfaces in marketing. Specifically, the current work examines how the affordances of voice-based interfaces affect customer experiences during self-service encounters, building on and integrating prior work on multiformat communication (Moffett et al., 2021) and media richness theory (Daft & Lengel, 1986). The findings of this research shed light on how affordances unique to voice-based interfaces (i.e., verbal cues, channel synchronicity) enhance flow-like user experiences and how these experiences, in turn, shape consumers' perception of their service experience, the firm, and downstream behavioral outcomes.

In what follows, we first discuss related work on voice-based interfaces, delineate the key conceptual properties of voice-based compared to text-based interfaces, and develop a set of hypotheses on how voice-based interfaces alter perceptions of flow, firm evaluation, and behavioral service outcomes. We then present the results of four studies that were designed to test our theorizing. Finally, we discuss the theoretical and practical implications for the effective conversational design of voice-based interfaces and the implications for customer service in an increasingly AI-driven economy.

Theoretical background and hypotheses

Related prior work on voice-based interfaces

In what follows, we will first review related prior work on voice-based interfaces and highlight how they relate to the current research. As summarized in Table 1, four key factors help to understand the unique insight of prior work with respect to (1) the nature or focus of the voice modality (voice primarily as input vs. output of the task), (2) the type of examined boundary conditions, (3) whether outcomes focused more on perception as opposed to behavioral consequences, and (4) the underlying psychological mechanism of the effect.

First, the majority of prior work either examined voice-based interface modalities by conceptualizing *voice as an input* as the focal independent variable or by examining *voice as an output* of the task, i.e., either assessing the role of the user's own speech during a task (*voice as input*) or how an interface responds back (*voice as output*). For example, Klesse et al. (2015) examined consumers' voice as an input in how selecting a product through spoken as opposed to written responses leads consumers to select more indulgent product options (see also Son & Oh, 2018 showing similar effects on hedonic product choices using a voice-based interface). In contrast, prior work focusing on voice as an output examined predominantly questions related to how much people trust a recommendation that is delivered verbally (vs. in written form) by a voice-based interface (Qiu & Benbesat, 2005). Only few studies examined a fully conversational interaction that required both the voice input by the user and a verbal response by the system (see Rzepka et al., 2021). Second, the type of boundary conditions typically examined either appearance-related factors (e.g., predominantly the presence of avatars; see Hess et al., 2009; Qiu & Benbasat, 2009) or broad consumer characteristics (e.g., consumers' general privacy concerns or emotional attachment tendencies; see Pagani et al., 2019; Son & Oh, 2018). Except for emerging work examining the task characteristics as possible boundary conditions (Rzepka et al., 2021), no prior work we are aware of examined the unique *conversational design* characteristics during an interaction such as the specific semantic properties of the language or how the conversation itself is structured. Third, the type of outcomes studied in earlier work focused predominantly on perception outcomes (such as trust or satisfaction measures; Hess et al., 2009; Pagani et al., 2019; Qiu & Benbasat, 2009; Qiu & Benbesat, 2005) as opposed to the evaluation of the firm or downstream behavioral outcomes (see Son & Oh, 2018 for an important exception). Finally, the type of mechanisms examined in earlier work illuminated important processes such as the extent of perceived social presence (Hess et al., 2009; Qiu & Benbasat, 2009) as well as fundamental cognitive processes (such as the extent of deliberation or how efficient the process was perceived overall; Berger et al., 2021; Rzepka et al., 2021). However, whether affordances unique to a voice-based interaction also change how consumers experience a service interaction, such as whether the interaction is perceived as more immersive or absorptive, and how these effects in turn shape downstream firm-related behavior is unclear.

As summarized in the review of Table 1, the current work examines a fully conversational service interaction experience (voice both as input and output), examines key

Table 1 Review of relevant literature on voice-based interface modalities

Study	Task	Voice modality focus	Psychological mechanism	Boundary condition(s)	Outcomes		Key findings
					Perception	Behavior	
Qiu and Benbesat (2005)	Receive product recommendation	Output (Website help interface)		Interface design (Avatar presence)	Interface trust		Voice output increases cognitive and emotional trust toward the interface. Avatar presence strengthens this effect
Hess et al. (2009)	Receive product recommendation	Output (Website help interface)	Social Presence	Interface design (Agent extraversion)	Interface trust		Agent extraversion strengthens the positive effect of voice output on social presence
Qiu and Benbasat (2009)	Receive product recommendation	Output (Website help interface)	Social Presence	Interface design (Avatar presence)	Interface trust, Usage intention		Voice output increases perceptions of social presence towards the interface. Avatar presence strengthens this effect
Klesse et al. (2015)	Order product	Input (Vending machine)				Product choice	Spoken preference expression prompts more indulgent product choices
Pagani et al. (2019)	Search product information	Input (Search engine)		User characteristics (Privacy concerns)	Platform engagement, Brand trust		Voice input reduces platform engagement and ultimately brand trust. Higher privacy concerns strengthen this effect
King et al. (2021)	Search product information	Input (Search engine)	Action Orientation		Intention to purchase		Voice input decreases purchase intention. This effect is mediated by reduced action orientation
Berger et al. (2021)	Provide customer review	Input (Non-conversational interface)	Deliberation		Firm interest	Emotionality of expression	Voice input leads to more emotional attitudes, which is driven by reduced deliberation

Table 1 (continued)

Study	Task	Voice modality focus	Psychological mechanism	Boundary condition(s)	Outcomes		Key findings
					Perception	Behavior	
Son and Oh (2018)	Access video-on-demand content	Input & Output (Pre vs. post adoption of smart speakers)		User characteristics (Emotional attachment)		Purchase, Content consumption	The adoption of smart speakers increases content consumption and leads to more hedonic product choices. The effects on product choice are strengthened by emotional attachment towards the interface
Rzepka et al. (2021)	Receive service recommendation	Input & Output (Conversational agent accessed on smart speaker vs. mobile device)	Efficiency, Enjoyment, Cognitive Effort	Task characteristics (Goal directedness)	Service satisfaction		Smart speakers enhance efficiency and enjoyment, reduce cognitive effort and, thus, increase satisfaction. These effects are strengthened by the goal directedness of the task
This research	File insurance claim	Input & Output (Voice-based vs. text-based claim filing interface)	Interface Flow	Conversational design (Semantic fluency, Conversational turns)	Service experience, Firm evaluation	Contract renewal, Consumer sentiment, Conversion rates	Voice-based interfaces lead to enhanced interface flow and in turn improved perceptual and behavioral service outcomes. This effect is moderated by the conversational design, such that higher semantic fluency and reduced conversational turns strengthen the positive effect of voice-based interfaces on interface flow

Table 2 Key conceptual properties of text-mediated vs. voice-mediated communication

Communication format	Text-mediated communication	Voice-mediated communication
Cue Characteristics	**Textual** - Formal structure and grammar - Precise syntax - Limited symbol set - No prosodic or temporal cues	**Verbal** - Informal structure - Flexible syntax - Extensive symbol set - Prosodic and temporal cues
Channel Characteristics	**Low Synchronicity** - Slower-paced - Asynchronous (sequential processing) **High Revisability** - Possibility to assess, deliberate, and rehearse	**High Synchronicity** - Faster-paced - Synchronous (parallel processing) **Low Revisability** - Little to no possibility to rehearse

conversational design characteristics (how the conversation is structured and the adaptation of the language during the interaction), examining both perceptual and behavioral outcomes relevant for consumers as well as firms, and a novel mechanism illuminating the extent of absorption during the task. In what follows, we will first review the key properties of voice vs. text-based communication followed by our conceptual model and hypotheses.

Key conceptual properties of voice-mediated communication

Our literature review revealed that voice-based interfaces have been associated with a more intuitive, positively valenced experience (Rzepka et al., 2021) that can impact preference construction (Klesse et al., 2015) and the type of content that consumers reveal about themselves (Berger et al., 2021). These effects raise the question about the underlying properties that make speaking versus writing truly unique. We build on Moffett et al.'s (2021) framework of multiformat communication to carve out the unique nature of voice-mediated communication and how to separate the "cue characteristics" of verbal speech (more informal structure, flexible syntax, prosodic and temporal communication cues) and the channel characteristics (greater synchronicity due to fast-paced and more parallel processing with low revisability during a conversation, see Table 2).

Cue characteristics Verbal cues refer to the specific vocal features of spoken language (e.g., tone, pitch, inflection, accent; Hildebrand et al., 2020; Moffett et al., 2021; Walther, 2005), while textual cues are linked to the written or typed language, such as spelling and punctuation (Moffett et al., 2021). Prior work on media richness theory suggests that the verbal cues in voice-based communication allow individuals to communicate more intuitively and naturally compared to text-based exchanges, leading to a sensorially richer experience (Daft & Lengel, 1986). For example, pauses, changes in speed, and differences in intonation evoke a more immersive experience during a voice-based as opposed to a text-based interaction (Redeker, 1984). This makes voice a richer and more vivid medium, leading to enhanced levels of involvement in a task (Daft & Lengel, 1986; Dennis & Kinney, 1998; Moffett et al., 2021).

Channel characteristics The literature on the interactivity of media channels suggests that the speed and immediacy of an interaction with a medium can lead to an increase in its perceived interactivity (Steuer, 1992). Voice-based as opposed to text-based exchanges unfold more synchronously. This enhanced channel synchronicity in voice-mediated interactions has been shown to lead to enhanced absorption in a task (Agarwal & Karahanna, 2000; Haeckel, 1998; Moffett et al., 2021). Contrary to written communication that allows consumers to re-read and re-type responses (so-called channel revisability, Moffett et al., 2021), spoken language must be encoded and decoded in a parallel process, thus making voice-based interactions substantially more immediate (Rubin et al., 2000). Voice-based interfaces allow users to have a more immediate influence on the content and the process of communication, promoting the exchange of social information at a faster rate and even impacting early relationship development (Moffett et al., 2021).

Voice-based interfaces and flow

How do these conceptual properties of voice-based communication (e.g., faster-paced, more parallel, synchronous processing of information and a rich sensory experience) impact consumers' task experience? One construct to describe such positively-valenced, immersive user experiences is the concept of *flow* (Agarwal & Karahanna, 2000; Csikszentmihalyi, 1975; Hoffman & Novak, 1996). Flow represents a pleasant, positively-valenced experience (Agarwal & Karahanna, 2000; Ghani et al., 1991; Trevino & Webster, 1992; Webster et al., 1993) that is characterized by a high level of immersion, positive affect, and optimal level of challenge (i.e., balancing boredom and excessively challenging experiences). Greater levels of flow have been shown to lead to more playful, intuitive, and effortless interaction experiences during web navigation (Hoffman & Novak, 1996) and a sense of total engagement and immersion with a system (Agarwal & Karahanna, 2000). Flow-like user experiences can even arise from the properties of the medium itself (Brannon Barhorst et al., 2021; Kim & Ko, 2019). For instance, Brannon Barhorst et al. (2021) demonstrate that the unique affordances of virtual reality applications facilitate a state of flow through enhanced levels of interactivity and vividness of the experience. Due to the sensory-rich, faster-paced, and extensive symbol set of voice-mediated communication, it is plausible that these distinct features of verbal communication similarly affect the experience of flow in voice as opposed to text-based service interactions.

Flow is characterized by "a seamless sequence of responses facilitated by machine interactivity" and occurs when consumers are immersed in a computer-mediated activity (Hoffman & Novak, 2009). Thus, we expect that the extent of interactivity and vividness afforded by the medium will impact the extent of interface flow that users experience (Novak et al., 2000). Contributing to the expected impact of enhanced interactivity of the communication medium, prior work provides initial evidence that two-way interactions and synchronicity might impact flow experiences (Guan et al., 2021). Given that voice-based (as opposed to text-based) communication facilitates more immediate and synchronous interactions (Moffett et al., 2021), we expect that the greater richness of verbal cues and greater extent of synchronicity in voice-based interactions enhances consumers' experience of interface flow, i.e., the extent of flow that consumers experience in connection with the interface. Thus, due to the unique cue and channel characteristics of voice-mediated communication, we formally hypothesize that using a voice-based interface (vs. text-based interface) enhances consumers' experience of interface flow.

H1 Interacting with a voice-based compared to a text-based-interface leads to enhanced levels of interface flow.

The impact of interface flow on service experience & firm outcomes

What might be the impact of enhanced levels of interface flow on service and firm outcomes? Prior work suggests that flow is a positively-valenced state that is intrinsically enjoyable (Agarwal & Karahanna, 2000; Hoffman & Novak, 1996; Johnson et al., 2003) and can even carry over to subsequent tasks, making them appear more experiential and enjoyable (Harmeling et al., 2017; Mathwick & Rigdon, 2004). Due to the increased sensory involvement and high synchronicity of the voice modality, we hypothesize that greater levels of flow during interface use carry over to consumers' overall experience with the service. This is consistent with Zomerdijk and Voss' (2010) work on designing "experience-centric services" that highlights the importance of the interface modality to provide more experiential service deliveries.

Experiences that go beyond customers' expectations have been shown to impact also the firm providing the service (Harmeling et al., 2017). Such experiences can move a customer from seeing interaction as merely transactional to building a positive consumer-firm relationship (Zomerdijk & Voss, 2010). This is consistent with prior research showing that interactive and multisensory service experiences can enhance firm perceptions along the customer journey (Lemon & Verhoef, 2016). In short, we expect that the enhanced experience of interface flow when using a voice compared to a text-based interface will in turn lead to an overall more experiential and positively-valenced service and firm evaluation.

H2 Greater levels of interface flow when interacting with a voice-based as compared to a text-based interface have a positive effect on (1) a more experiential service delivery and (2) a more positively-valenced evaluation of the service firm.

These changes in perception might also impact consumers more directly. As the cue and channel characteristics of voice-mediated interfaces may drive a more experiential service delivery (H_2), it is plausible that they might also in turn help form a closer connection between the consumer and the firm (Harmeling et al., 2017; Schouten et al., 2007). While the experience of interface flow is transitory, the fact that it increases the enjoyment of an activity for its own sake suggests that it may have a more lasting impact on behavior related to future interactions with the same firm, making individuals wanting to experience a state of flow again and again (Schouten et al., 2007). Thus, we hypothesize that the enhanced level of interface flow that consumers experience when interacting with a voice-based interface will positively impact behavioral outcomes in connection with the firm. Specifically, we hypothesize that greater levels of interface flow in turn enhance consumers' desire to stay committed to their relationship with the firm in the future, to engage in future interactions with the service firm, and to advocate the firm and its services to others.

H3 More positive perceptions of the firm and the experiential service delivery (H2), in turn, enhance consumers' contract renewal decision, consumer sentiment, and conversion rates to engage in further service exchanges with the same firm again.

Boundary conditions: The role of semantic fluency & conversational turns

Flow has been found to only emerge when a user's skill matches the difficulty of the task (Hoffman & Novak, 1996). More specifically, there needs to be an optimal match between the capabilities of the user and the opportunities provided by the interface to execute a task. While the literature vastly disagrees on the number of possible categories of combinations between different levels of user skill and task complexity (Clarke & Haworth, 1994; Csikszentmihalyi, 1988; Csikszentmihalyi & LeFevre, 1989) the literature largely agrees that for flow to emerge, there needs to be *some* level of challenge above a certain threshold (see prior work on optimal stimulation theory; Holbrook & Gardner, 1993; Steenkamp & Baumgartner, 1992). Thus, if the task is too challenging, users get anxious or frustrated, while if the task is not challenging enough, they may get bored. Thus, how a certain task is designed has an important impact on the extent to which flow can emerge.

This suggests that a key boundary condition to experience flow in conversational interfaces is linked to how the conversation is governed or designed. During interactions with a conversational interface, the user engages in information processing based on *what* and *how* the interface communicates (Clark & Brennan, 1991; Pogacar et al., 2018). Thus, the way information is processed depends on both (1) how the language used by the interface is processed (what we refer to as the *semantic fluency* of a conversation), and (2) how the messages are structured as a whole (what we refer to as the extent of *conversational turns* during a conversation).

Semantic fluency Research in cognitive psychology and linguistics has demonstrated the important connection between linguistic properties of a conversation and the ease with which it is processed (Fernández-Sabiote & López-López, 2020; Song & Schwarz, 2008). Specifically, research has shown that simpler language leads to more fluent processing of messages (Thompson & Ince, 2013). We therefore expect that if an interface employs simpler language, users will process it with greater fluency. The differences in channel synchronicity between text-based and voice-based communication suggests that semantic fluency may differentially affect each channel. Indeed, research on interpersonal communication suggests that a simpler conversational style better suits the properties of a dynamic dialogue as opposed to those of written text (Redeker, 1984). This is primarily due to the extent of synchronicity of the communication channel. When using a text-based interface, consumers can re-read a message with lower semantic fluency to aid message processing due to the asynchronous nature of the channel. Conversely, when interacting with a voice-based interface, the synchronous nature of the channel makes processing low semantic fluency messages more difficult, requiring greater cognitive processing. Based on our theorizing about the informal structure, flexible syntax, and more extensive symbol set with low revisability in verbal compared to text-based communication (see Table 2 for a summary) as well as previous research showing that messages in spoken as opposed to written communication require greater cognitive processing (e.g., Van Zeeland and Schmitt 2013), we expect that the cognitive processing of messages with lower semantic fluency is more pronounced in voice compared to text interfaces and will therefore affect flow more negatively in voice (as opposed to text-based) interfaces.

H4a The positive effects of voice-based interfaces on interface flow are increased (decreased) under conditions of high (low) semantic fluency.

Conversational turns The second boundary condition assesses to which extent consumers' experience of interface flow is disrupted by increasing the number of conversational turns during the interaction with a voice-based interface. Individual messages are linked together in a conversation by conversational partners taking turns, which can vary in length depending on the context of the conversation (Levinson, 2016). Thus, when interacting with a voice-based interface, users are forced to adjust to the pace (i.e., speaking rate) of the interface and remain attentive to what is being said throughout the conversation (Redeker, 1984). In contrast, when using a text-based interface, consumers can deliberately choose their own pace of processing information and rehearse or reflect as they wish on what was being said (Moffett et al., 2021). This suggests that the extent of conversational turns can either disrupt or enhance interface flow in voice-based interfaces. We expect that a greater number of conversational turns to complete a service task will reduce channel synchronicity and provoke a less absorptive experience, as the interface flow is interrupted more often. Indeed, as flow is experienced as a high level of continuous immersion in an activity (Csikszentmihalyi, 1975), we expect a higher number of conversational turns to cause a reduced absorption in a task. Thus, we expect that a higher number of turns creates a feeling of disjointedness between individual messages. Taken together, we predict that an increase in conversational turns will negatively impact interface flow in voice-based (vs. text-based) interfaces.

H4b The positive effects of voice-based interfaces on interface flow decrease (increase), with a greater (lower) number of conversational turns during the interaction.

Our overarching hypotheses are summarized in Fig. 1. In what follows, we provide a short description of the

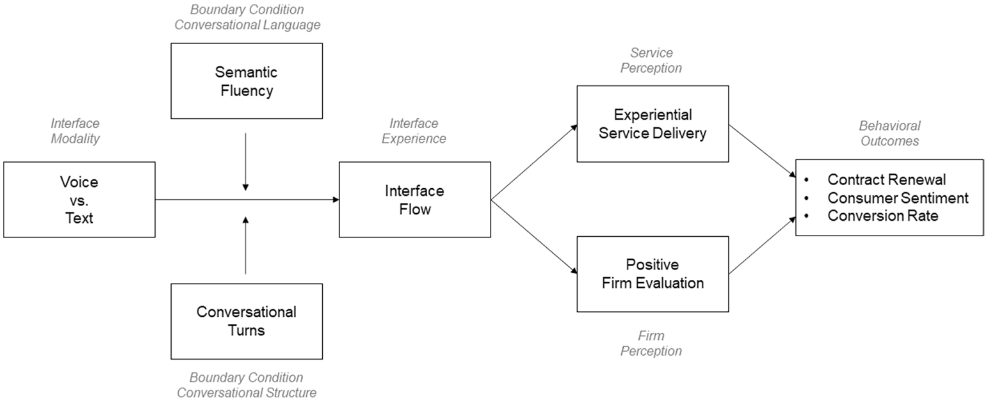

Fig. 1 Conceptual model

experimental setup and technical implementation, followed by our empirical studies.

Experimental paradigm and context

To test our hypotheses and to ensure maximum experimental control, we developed a custom-made voice-based interface (using Python and Google WaveNet; see Web Appendix A1 for details). The appearance of the interface was kept neutral to avoid any confounds that could have produced spurious effects on our key measures of interest (appearance of the avatar or any other visual design cues) and to isolate the hypothesized interface modality effects we predict. We provide a detailed description in the Web Appendix explaining the technical infrastructure for both research transparency reasons as well as to provide researchers with the opportunity to build on and expand on our current technical implementation. We predefined the dialogue flow based on the context of each study and ensured that the order of all questions was identical between conditions across experiments. We used Google's state-of-the-art text-to-speech generator WaveNet for all studies (Oord et al., 2016) and participants went through a fully voice-based service interaction with individual prompts to complete the service task (see also Web Appendix A1 for further details). To rule out potential confounds related to the gender of the voice bot, we used both female (Study 1 and Study 4) and male voices across studies (Study 2 and Study 3). Finally, in determining our sample sizes, we followed prior work that used a similar experimental paradigm with intermediate effect sizes (e.g., King et al. (2021) sampled around 125 participants per condition comparing text-based vs. voice-based user search).

We focused on a service context with important economic and societal implications. Specifically, we developed a voice-based interface to submit and handle insurance claims for two major reasons. First, the claim filing process in the insurance industry has been heavily standardized to simplify claim management (from claiming the loss of a mobile phone to reporting a car damage) (Singh et al., 2019). Second, the claim handling process is often perceived as a major source of distress for individuals that negatively affects their experience across service touchpoints (Riikkinen et al., 2018). According to recent industry reports, the trend to phase out human service agents in favor of using automated file claiming systems has also led to reduced customer satisfaction rates, enhanced client frustration, and an experienced loss of "human touch" (Blake, 2018), rendering the use of voice-based interfaces as a novel intervention in the current service automation landscape.

Overview of studies

We present evidence from four studies that were designed to test our theorizing. In Study 1, we test our baseline hypothesis that a voice-based compared to a text-based interface promotes a more flow-like user experience, which in turn translates into a more positive service perception. Study 2 demonstrates an important boundary condition to the effect of voice-based interfaces, showing that reduced levels of semantic fluency can provoke negative user experiences which in turn also impact consumers' inclination to extend their relationship with the service firm. Study 3 further demonstrates a second key boundary condition, showing that a greater number of conversational turns in a voice-based interface reduces consumers' experience of interface flow and perception of the service firm. Study 3 also rules out that the current findings can be explained by a general positive affect account. Finally, Study 4 further demonstrates the impact of voice-based interfaces on behavioral outcomes (conversation rates and consumer sentiment) and provides further evidence on the unique role of interface flow to explain these effects (consistent with optimal stimulation theory and the matching of task affordances and skills; Holbrook & Gardner, 1993).

Study 1

Study 1 was designed to provide a first test of our baseline hypothesis that a voice-based compared to a text-based interface alters perceptions of interface flow. Furthermore, the current study also tests whether these changes in interface flow promote more experiential service delivery. We also further explore whether these effects are conditional on consumers' prior experience with voice technology.

Design and procedure

A total of 184 participants ($M_{\text{Age}} = 32.97$, $SD_{\text{Age}} = 11.05$, 50% females; for an overview of the demographics across all studies see Web Appendix B1) were recruited from a nationwide consumer panel (Prolific[1]) and randomly assigned to a two-cell between-subject experiment (text-based vs. voice-based interface). In both conditions, the experimental task comprised the filing of an insurance claim with a fictitious car insurance company called TER Insurance. Before the task, participants were presented with a scenario that described the events of an accident including the damage that occurred to the car (see Web Appendix A2.1 for a detailed description of the scenario). In the claim filing task, we asked participants to report all information concerning the incident via either the text- or voice-based interface (dependent on condition). The content and the sequence of questions were identical across conditions (i.e., the sequence of six conversational turns, content, and all other features of the file claiming process were held constant across conditions). Participants were screened to use a laptop or desktop computer (non-mobile devices) to avoid any interface modality confounds (such as differences in writing speed on mobile versus non-mobile devices). We used a female voice in this study but find consistent results across studies regardless of the gender of the voice-based interface (e.g., using a male voice in Study 2).

Measurement Immediately after completing the insurance claim filing task, we assessed participants' perception of interface flow when using the service interface with three items (scale adapted from Unger & Kernan, 1983, sample item: "This interface totally absorbed me."; 7-point Likert scale, from 1: "Strongly disagree" to 7: "Strongly agree"; $\alpha_{\text{Flow}} = 0.93$, see Web Appendix A3.1). Next, we measured consumers' extent of experiential service delivery using a four-item scale (scale adapted from Unger & Kernan, 1983; sample item: "This interface offers novel experiences."; 7 point-Likert scale, from 1: "Strongly disagree" to 7: "Strongly agree"; $\alpha_{\text{Experiential}} = 0.94$, see Web Appendix A3.2). Finally, participants answered how frequently they use voice-based interfaces in general (see Web Appendix A3.5) and self-reported their domain-specific insurance expertise (see Web Appendix A3.4).

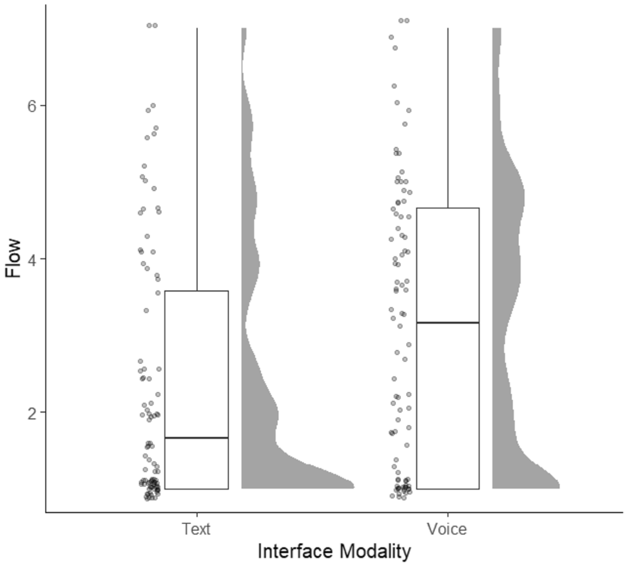

Fig. 2 Voice-based interfaces promote flow-like consumer experiences

Results

Participants who used the voice-based interface experienced significantly higher levels of interface flow compared to participants using the text-based interface ($M_{\text{VoiceInterface}} = 3.09$, $M_{\text{TextInterface}} = 2.31$, $t = 2.996$, $p < 0.01$; Cohen's $d = 0.44$; see Fig. 2). Similarly, we also found that participants using the voice-based interface reported a significantly more experiential service delivery compared to participants using the text-based interface ($M_{\text{VoiceInterface}} = 3.78$; $M_{\text{TextInterface}} = 2.90$, $t = 3.239$, $p < 0.01$; Cohen's $d = 0.48$).

To provide a direct test of our theorizing on the effect of interface modality on consumers' perception of experiential service delivery, we estimated a simple mediation model (5,000 bootstrap samples) with the interface modality condition as the independent variable, interface flow as the mediator, and experiential service delivery as the dependent variable. In support of our theorizing, the voice-based interface led to a significant increase in interface flow ($\beta_{\text{Interface}} = 0.78$,

[1] Peer et al., (2017, 2021) compared the data quality of different crowdsourcing platforms for behavioral research, showing that Prolific scored highest on response quality and sample diversity. Prolific was used to recruit participants and manage payments, and routed to our own server to complete the experimental task.

$SE = 0.26$, $t = 3.002$, $p < 0.01$). In turn, greater levels of interface flow led to a more experiential service delivery ($\beta_{Flow} = 0.77$, $SE = 0.05$, $t = 14.649$, $p < 0.001$), rendering the residual direct effect non-significant ($\beta_{Interface} = 0.28$, $SE = 0.19$, $t = 1.469$ $p = 0.14$), and a significant indirect effect with a confidence interval excluding zero ($\beta_{Indirect} = 0.60$, 95% *CI*: [0.19, 0.99]), indicating full mediation. These results were robust even after controlling for differences in participants' age, gender, income level, car ownership, domain expertise, and previous experience with voice technology (i.e., lower information criteria contrasting our main model to the alternative control model; $BIC_{MainModel}$: 621.11, $BIC_{ControlModel}$: 637.80). See Web Appendix B2 for additional statistical analyses including control variables.

Discussion

The findings of Study 1 provide initial evidence for the baseline hypothesis that voice-based interfaces evoke a more experiential service delivery compared to text-based interfaces and that these effects are driven by the flow-enhancing effect of voice-based interfaces.

Study 2

The objectives of Study 2 were threefold. First, the current study examines a theoretically important and practically relevant boundary condition. Study 2 tested to which extent the semantic fluency of a query can either enhance or reduce a users' perception of interface flow when using a voice-based interface. Second, we further test whether the use of a voice-based interface impacts only perceptual outcomes or also behavioral outcomes (contract renewal with the current service provider). Third, the current study used a male voice to rule out that the current findings are a function of the digital assistants' gender (female voice in Study 1 vs. male voice in Study 2).

Design and procedure

A total of 612 participants ($M_{Age} = 37.19$, $SD_{Age} = 13.34$, 45.4% males) were recruited from a nationwide consumer panel (Prolific) and randomly assigned to a 2 (interface modality: voice-based vs text-based interface) × 2 (semantic fluency: high vs. low) between-subjects design, using the same experimental paradigm and scenario as in Study 1. We used the same experimental paradigm as in Study 1 and only added short filler questions to provide a more realistic experimental setting (e.g., whether another insurance is already involved, and whether a technical expert was consulted; total of ten turns per interaction, see Web Appendix 2.2). We altered semantic fluency of each turn based on prior work in linguistics and communication research by replacing simple through more complex words (Khawaja et al., 2010; Redeker, 1984). For example, the high semantic fluency condition used easily to process phrases (e.g., "Please tell us more about the damage that you want to report."), the low semantic fluency condition used more difficult to process phrases (e.g., "Please elucidate on the nature of the damage that you wish to declare."). To avoid any confounding effects in terms of the length of a query and to isolate the effect of the semantic properties of the conversation, we held the number of words per turn constant between conditions ($M_{HighFluency} = 16.91$, $M_{LowFluency} = 17.09$, $t = -0.04$, $p = 0.97$).

To study whether a voice-based interface might also impact behavioral outcomes, users were informed after completing the claim filing task that their contract is about to expire and that they could extend their contract if they wish to do so. All participants were then asked directly by the respective interface to either renew their contract or not (see Web Appendix A2.2 for further information on the scenario).

Manipulation check We used objective readability scores to assess the discriminatory power of the semantic fluency manipulation, using the Flesch reading ease score (FRE; higher is better / more fluent), the New Dale-Chall readability formula (NDR; lower is better / more fluent), and the Automated Readability Index (ARI; lower is better / more fluent). All measures confirmed that semantic fluency was indeed higher in the high versus low semantic fluency condition ($FRE_{HighFluency} = 80.03$ vs. $FRE_{LowFluency} = 56.44$; $NDR_{HighFluency} = 6.63$ vs. $NDC_{LowFluency} = 9.98$; $ARI_{HighFluency} = 3.15$ vs. $ARI_{LowFluency} = 6.69$).[2]

Measurement We measured consumers' experience of interface flow using the same items as in Study 1 ($\alpha_{Flow} = 0.92$), along with the same items to assess consumers' perception of experiential service delivery ($\alpha_{Experiential} = 0.91$). Moreover, we measured their perception of the service firm using a two-item scale (scale adapted from Lee & Lin, 2005; sample item: "Overall, I am satisfied with the online experience with TER insurance"; 7 point-Likert scale, from 1: "Strongly disagree" to 7: "Strongly agree"; $\alpha_{FirmEval} = 0.96$, see Web Appendix A3.3). Finally, we measured consumers' decision to extend their contract with TER insurance as 1 when extending and 0 when terminating their contract.

[2] To compute the FRE, the ARI, and the DCR we used the Spacy package in Python. In the web appendix (A3.6) we provide the corresponding formulas to compute these metrics. Please refer to DuBay (2004) for a review and methodological details.

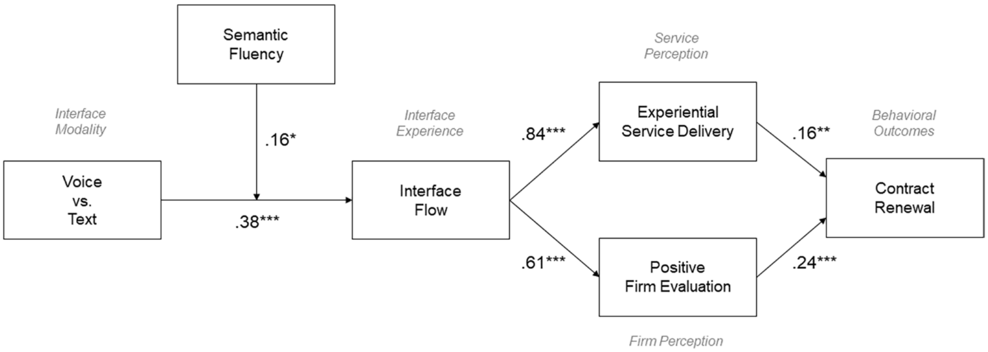

Fig. 3 Path model results (Study 2)

Results

Interface flow A two-way ANOVA revealed a significant main effect of condition on interface flow ($F(1, 608) = 46.84$, $p < 0.001$, $\eta^2 = 0.07$), demonstrating a significantly higher experience of interface flow in the voice as opposed to text-based interface ($M_{\text{VoiceInterface}} = 3.59$, $M_{\text{TextInterface}} = 2.72$, $t = 6.693$, $p < 0.001$). The semantic fluency main effect was non-significant ($F(1, 608) = 0.24$, $p = 0.63$, $\eta^2 = 0.00$) while the interaction between both experimental factors was statistically significant ($F(1, 608) = 6.72$, $p < 0.01$, $\eta^2 = 0.01$). Follow-up contrasts with Holm correction for family-wise errors revealed that, consistent with our theorizing, participants' experience of interface flow in the voice-based interface condition was reduced in the low semantic fluency condition ($M_{\text{Voice\_HighFluency}} = 3.77$, $M_{\text{Voice\_LowFluency}} = 3.38$, $t = 2.189$, $p = 0.057$) while semantic fluency had no effect on participants in the text-based interface condition ($M_{\text{Text\_HighFluency}} = 2.59$, $M_{\text{Text\_LowFluency}} = 2.86$, $t = 1.470$, $p = 0.14$).

Service experience Mirroring the previous effect when assessing the impact on consumers' overall evaluation of the service, a two-way ANOVA revealed a significant increase of a more experiential service delivery in the voice as opposed to text-based interface ($F(1, 608) = 76.87$, $p < 0.001$, $\eta^2 = 0.11$; $M_{\text{VoiceInterface}} = 3.58$, $M_{\text{TextInterface}} = 2.51$), a non-significant main effect of semantic fluency ($F(1, 608) = 0.171$, $p < 0.78$, $\eta^2 = 0.00$), and significant interaction between both factors ($F(1, 608) = 6.365$, $p = 0.01$, $\eta^2 = 0.009$). Follow-up contrasts with Holm correction confirmed that participants in the voice-based interface perceived the service more negatively in the low semantic fluency condition ($M_{\text{Voice\_HighFluency}} = 3.76$, $M_{\text{Voice\_LowFluency}} = 3.40$, $t = 2.189$, $p = 0.07$) while semantic fluency had no effect in the text-based interface condition ($M_{\text{Text\_HighFluency}} = 2.39$, $M_{\text{Text\_LowFluency}} = 2.65$, $t = 1.470$, $p = 0.14$).

Path model & serial mediation Finally, we estimated a path model using the lavaan package in R with robust standard errors to assess the overall system of hypotheses and the impact of interface flow and service experience on firm evaluation and consumer choice (contract renewal). The path model results are summarized in Fig. 3. These findings demonstrate that the increase of interface flow evoked by the voice-based interface ($\beta_{\text{Voice}} = 0.38$, $z = 6.504$, $p < 0.001$) in turn led to a significant increase in consumers' perception of experiential service delivery ($\beta_{\text{Flow}} = 0.84$, $z = 21.180$, $p < 0.001$) and an overall more positive perception of the firm ($\beta_{\text{Flow}} = 0.61$, $z = 15.320$, $p < 0.001$). Focusing on the impact on contract renewal, we found that the positive impact of the voice-based interface condition on service and firm perception, in turn, led to a significant increase in contract renewal rates ($\beta_{\text{ExpService}} = 0.16$, $z = 2.879$, $p < 0.01$; $\beta_{\text{FirmEval}} = 0.24$, $z = 4.615$, $p < 0.001$). These path model results were confirmed by a serial mediation model (Hayes 2017; model 81 with 5000 bootstrap samples) estimating the effect of the voice (vs. text) condition on contract renewal rate via flow (proximal mediator) and firm evaluation as well as service experience ratings (distal mediators), with a significant indirect effect excluding zero ($\beta_{\text{Indirect}} = 0.32$, 95% *CI*: [0.16, 0.49]).

Discussion

The results of Study 2 replicate the finding that voice-based interfaces impact consumers' experience of interface flow. The findings reveal further that these enhanced levels of interface flow drive both important perceptual outcomes (how consumers perceive the service firm and the service experience itself) as well as behavioral outcomes (consumers' decision to renew their contract with the same firm). We also provide evidence for an important boundary condition, demonstrating that these positive effects are reduced with lower levels of semantic fluency. While one might expect the main effect of semantic fluency across the board (such that reduced semantic fluency might negatively impact both voice and text interfaces alike), the fact that consumers can easily re-read (initially) more difficult to comprehend phrases makes consumers arguably more immune to the negative effect of reduced semantic fluency.

Study 3

The objectives of Study 3 were twofold. First, the current study was designed to further illuminate the conditions under which voice-based interfaces might enhance versus reduce perceptions of interface flow. Specifically, the current study tests the boundary condition of conversational turns during a voice-based interaction. Second, the current study also measured a set of alternative explanations. Given that flow-like experiences are by definition more positively-valenced experiences, we further assess the possibility that the current findings can be explained merely via a general positive affect account (as opposed to a *specific* positive affective experience, i.e. interface flow). We also further assess whether the mere presence of a voice-based interface might enhance generic attributions of perceived care (given that a voice-based interface actively "talks" to consumers).

Design and procedure

A total of 610 participants ($M_{\text{Age}} = 35.59$, $SD_{\text{Age}} = 12.70$, 46.9% males) were recruited from a nation-wide consumer panel (Prolific) and randomly assigned to a 2 (interface modality: voice-based vs text-based interface) × 2 (conversational turns: high vs. low) between subjects' design. The interface modality conditions mirrored the paradigm used in Study 1. We manipulated the extent of conversational turns building on prior work on turn-taking (Levinson, 2016; Wiemann & Knapp, 1975). Specifically, in the low conversational turns condition, participants answered corresponding questions within one turn (e.g., "Was it a traffic accident, glass breakage, theft or burglary? How did this damage occur? Please describe the course of events as accurate as possible.") whereas in the high conversational turns condition these questions were broken up into separate sub-questions (e.g., Question 1: "Was it a traffic accident, glass breakage, theft, or burglary?", Question 2: "How did this damage occur? Please describe the course of events as accurately as possible."; see Web Appendix A2.3 for further details on the scenario). We predict that separating corresponding questions in a voice- compared to text-based interface interaction will be more detrimental as compared to keeping larger question blocks to consumers' experience of interface flow, given the reduced synchronicity of the interaction (see Table 2). However, increasing the number of conversational turns in the text-based interface might even enhance perceptions of interface flow as text-based communication often unfolds in shorter, more rapid turns (think of the common short message exchanges using social messenger platforms).

Measurement Mirroring the preceding studies, we used the same items to assess consumers' experience of interface flow ($\alpha_{\text{Flow}} = 0.91$), experiential service delivery ($\alpha_{\text{ExpService}} = 0.94$), and firm evaluation ($\alpha_{\text{FirmEval}} = 0.94$). To further assess whether the current findings are driven by general positive affect, we measured PANAS (negative items reverse coded; $\alpha_{\text{PosAffect}} = 0.74$). Finally, we used a single item to assess respondents' feelings of perceived care on a 7-point Likert scale ("How caring did you perceive the interface you used to complete your claim filing procedure?"; see Web Appendix A3.7).

Results

Interface flow A two-way ANOVA revealed a significant main effect of the interface condition on interface flow ($F(1, 606) = 13.078$, $p < 0.001$, $\eta^2 = 0.02$), demonstrating a significantly higher experience of interface flow in the voice- as opposed to text-based interface modality condition ($M_{\text{VoiceInterface}} = 3.54$, $M_{\text{TextInterface}} = 3.03$, $t = 3.707$, $p < 0.001$). The conversational turns main effect was non-significant ($F(1, 606) = 0.858$, $p = 0.35$, $\eta^2 = 0.001$) while the interaction between both experimental factors was significant ($F(1, 606) = 4.206$, $p < 0.05$, $\eta^2 = 0.007$). Follow-up contrasts with Holm correction for family-wise errors revealed that while a smaller number of conversational turns enhance interface flow in voice-based interfaces, they reduce interface flow in text-based interfaces ($M_{\text{Voice\_LowTurns}} = 3.65$, $M_{\text{Text\_LowTurns}} = 2.85$, $t = 3.995$, $p < 0.001$). The presence of more conversational turns in the voice-based interface condition led to a directional reduction of interface flow ($M_{\text{Voice\_HighTurns}} = 3.43$, $M_{\text{Voice\_LowTurns}} = 3.65$, $t = 1.028$, $p = 0.47$), and was significant for participants that experienced at least some level of flow ($M_{\text{Voice\_HighTurns}} = 4.05$, $M_{\text{Voice\_LowTurns}} = 4.53$, $t = 2.437$, $p < 0.05$; trimmed mean minus half a standard-deviation, see Globerson, 1983).

Service experience Mirroring the previous effects on interface flow, a two-way ANOVA assessing the impact on consumers' service experience revealed a significant increase in experiential service delivery in the voice as opposed to text-based interface condition ($F(1, 606) = 37.135$, $p < 0.001$, $\eta^2 = 0.001$; $M_{\text{VoiceInterface}} = 3.78$, $M_{\text{TextInterface}} = 2.88$), a non-significant main effect of conversational turns ($F(1, 606) = 0.022$, $p = 0.88$, $\eta^2 = 0.00$), and a marginally significant interaction between both factors ($F(1, 606) = 3.402$, $p = 0.065$, $\eta^2 = 0.0007$). Follow-up contrasts with Holm correction confirmed that a smaller number of conversational turns enhanced consumers' service experience in voice-based interfaces, but reduced service experience in text-based interfaces ($M_{\text{Voice\_LowTurns}} = 3.97$, $M_{\text{Text\_LowTurns}} = 2.79$, $t = 5.552$, $p < 0.001$). As with flow, the presence of conversational turns in the voice-based interface

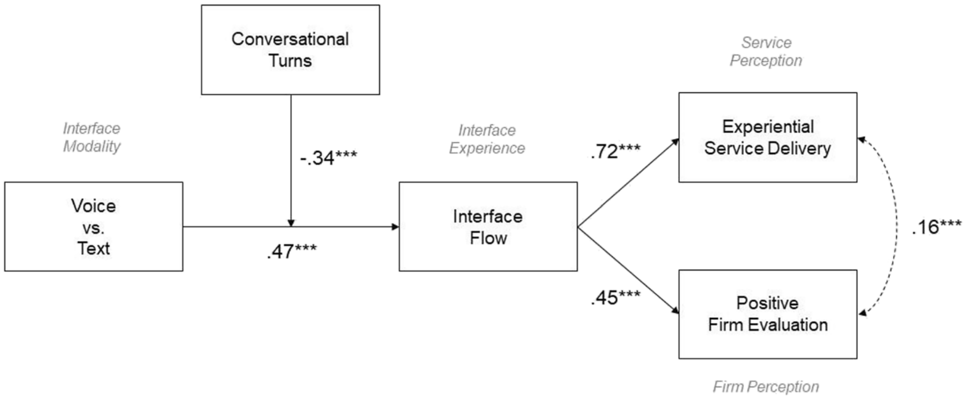

Fig. 4 Path model results (Study 3)

condition led to a directional reduction of interface flow ($M_{\text{Voice\_HighTurns}} = 3.61$, $M_{\text{Voice\_LowTurns}} = 3.97$, $t = 1.536$, $p = 0.25$), and was significant for participants that experienced at least some level of flow ($M_{\text{Voice\_HighTurns}} = 4.02$, $M_{\text{Voice\_LowTurns}} = 4.65$, $t = 2.468$, $p < 0.05$; trimmed mean minus half a standard-deviation, see Globerson, 1983).

Path model Next, we estimated a path model using the lavaan package in R with robust standard errors to assess the overall system of hypotheses. The path model results are summarized in Fig. 4. The findings demonstrate that the increase in interface flow evoked by the voice-based interface ($\beta_{\text{Voice}} = 0.47$, $z = 3.823$, $p < 0.001$), in turn, led to a significant increase in experiential service delivery ($\beta_{\text{Flow}} = 0.72$, $z = 28.420$, $p < 0.001$) and an overall more positive firm perception in line with our predictions ($\beta_{\text{Flow}} = 0.45$, $z = 12.556$, $p < 0.001$).

Alternative accounts Finally, we conducted a series of robustness checks. Specifically, we first tested whether interface modality might enhance general positive affect. Given that flow-like experiences are a specific type of positive affect, it is conceivable that voice-based interfaces are not specific to interface flow but drive general positive affect more broadly. Ruling out a general positive affect account, we found no main effect of modality on positive affect ($F(1, 606) = 1.121$, $p = 0.29$, $\eta^2 = 0.00$; $M_{\text{VoiceInterface}} = 4.64$, $M_{\text{TextInterface}} = 4.53$) and no interaction between both factors ($F(1, 606) = 1.834$, $p = 0.18$). Finally, we also found no effect of the interface modality on perceived care ($F(1, 606) = 0.086$, $p = 0.77$, $\eta^2 = 0.00$; $M_{\text{VoiceInterface}} = 3.35$, $M_{\text{TextInterface}} = 3.31$; also no interaction between both factors, $F(1, 606) = 2.014$, $p = 0.16$).

Discussion

The results of Study 3 demonstrate that a greater number of conversational turns can reduce interface flow when using voice-based as opposed to text-based interfaces. While one might expect the main effect of conversational turns across the board (such that enhancing the number of conversational turns might negatively impact both voice-based and text-based interfaces alike), the high synchronicity and faster pace of voice-based interactions makes consumers more sensitive to the negative effect of an enhanced number of conversational turns. The current study also rules out that these findings can be explained merely by a general positive affect account or by enhancing attributions of perceived care.

Study 4

The objectives of Study 4 were twofold. First, the current study was designed to examine whether the current effects are specific to consumers' experience of interface flow. Flow-like experiences have been shown to be driven by creating "optimal levels of challenge", i.e., experiences that are neither considered as boredom nor as excessively challenging. Thus, for flow to emerge, there needs to be some level of challenge (see optimal stimulation theory; Holbrook & Gardner, 1993) consistent with the faster-paced, more synchronous, and more flexible cue and channel characteristics of voice-based interactions (see Table 2). The current study tests this flow-specific mechanism. Second, the current study is also designed to further examine the impact of voice-based interfaces on a range of behavioral outcomes.

Design and procedure

All methods and statistical analyses for this study were pre-registered (https://aspredicted.org/6nh8a.pdf). A total of 811 participants ($M_{\text{Age}} = 35.15$, $SD_{\text{Age}} = 12.93$, 50.92% males) were recruited from Prolific Academic and randomly assigned to either a voice-based or a text-based interface. The interface conditions used the same insurance context as in all preceding studies. The first part of the task was identical to all three preceding studies (filing a claim).

We created a MailChimp landing page for our fictitious insurance brand TER Insurance and developed and designed an advertisement that provided the user with an option to sign up as a future user, as a consequential, behavioral measure of user conversion. The advertisement was presented at the end of the study to each participant with a promotion to become a beta user to test future insurance services (see Web Appendix A4. for the advertising stimuli). In both conditions the advertisement was identical, except that in the voice condition the advertisement asked the user to sign-up to "become a test user for our new insurance services based on our *voice* interface" while in the text-based interface condition, the advertisement asked the user to sign up to "become a test user for our new insurance services based on our *text* interface". The main dependent variable was the conversion rate of the advertisemen and measured as the number of participants subscribing divided by the number of unique page visits.

Measurement The same items as in the preceding studies were used to assess consumers' experience of interface flow ($\alpha_{Flow} = 0.91$), experiential service delivery ($\alpha_{ExpService} = 0.92$), and their perception of the service firm ($\alpha_{FirmEval} = 0.94$). To assess the flow-specific mechanism for consumers' experienced challenge, we used a 6-item scale (adapted by Novak et al. (2000), $\alpha_{Challenge} = 0.89$; see Web Appendix A3.8). To assess consumer sentiment, we used a rating task inspired by Hildebrand and Bergner (2020) using an open-response technique about consumers' positive and negative thoughts and feelings during the service task. We used the tidytext package in R to process the text data and used the sentimentr package to extract sentiment scores (using both positively and negatively valenced words) (Nielsen 2011). Finally, we measured the subscriber conversion following McDowell et al. (2016) (subscribers divided by the number of page visits).

Results

Interface flow & service and firm perception In line with our theorizing and corroborating the results of our previous studies, participants who used the voice-based interface experienced significantly higher levels of interface flow compared to participants using the text-based interface ($M_{VoiceInterface} = 4.02$, $M_{TextInterface} = 3.25$, $t = 6.96$, $p < 0.001$; Cohen's $d = 0.49$). Mirroring the results of the preceding studies, we also found that the voice-based interface led to a significantly more experiential service delivery than the text-based interface ($M_{VoiceInterface} = 4.23$, $M_{TextInterface} = 3.01$, $t = 11.25$, $p < 0.001$; Cohen's $d = 0.79$) and a more positive evaluation of the service firm ($M_{VoiceInterface} = 4.75$, $M_{TextInterface} = 4.50$, $t = 2.26$, $p < 0.05$; Cohen's $d = 0.16$).

Flow & optimal challenge To test the flow-specific mechanism, we estimated a mediation model (5,000 bootstrap samples) with the interface modality condition as the independent variable, consumers' experienced challenge as the mediator, and interface flow as the dependent variable. As expected, the voice-based interface led to a significant increase in consumers' experienced challenge ($\beta_{Interface} = 0.32$, SE = 0.09, $t = 3.47$, $p < 0.001$). In turn, greater levels of experienced challenge led to a higher flow perception ($\beta_{Challenge} = 0.51$, SE = 0.03, $t = 13.03$, $p < 0.001$), and a significant indirect effect with a confidence interval excluding zero ($\beta_{Indirect} = 0.16$, 95% CI: [0.07, 0.25]), indicating full mediation.

To further assess whether more excessive levels of challenge negatively impact interface flow (as a test of the "optimal level of challenge" notion of flow), we estimated a polynomial regression in which we predicted interface flow using the interface modality condition and the second-degree polynomial of experienced challenge. The model was significant ($F(3, 807) = 78.14$, $p < 0.001$), confirming that the voice-based interface led to a significant increase of flow ($\beta_{Interface} = 0.60$, SE = 0.10, $t = 5.96$, $p < 0.001$), a significant positive main effect of experienced challenge ($\beta_{Challenge} = 18.78$, SE = 1.44, $t = 13.07$, $p < 0.001$), and also a significant negative squared effect of challenge ($\beta_{Challenge}{}^{2} = -3.18$, SE = 1.43, $t = -2.23$, $p < 0.05$), conforming the expected "optimal level of challenge" necessary for flow-like user experiences (i.e., reduced flow at extreme levels of challenge).

Behavioral outcomes We observed a systematic increase in consumers' behavioral responses across our key dependent variables (i.e., consumer sentiment, conversation rate). First, analyzing consumer sentiment scores revealed a significant effect of interface type on sentiment valence ($F(1, 809) = 23.73$, $p < 0.001$). Participants who interacted with the voice-based interface perceived the service experience significantly more positive than those who interacted with the text-based interface ($M_{VoiceInterface} = 0.17$, SD = 0.41; $M_{TextInterface} = 0.03$, SD = 0.42). For example, while participants in the voice-based interface condition mentioned that the experience was "exciting", "fun", and "like speaking to a real customer service representative", while participants in the text-based interface only mentioned how "efficient", "simple", and "practical" the claim interface was. Next, a path model using the lavaan package in R with robust standard errors demonstrates that the increase in interface flow evoked by the voice-based interface ($\beta_{VoiceInterface} = 0.77$, $z = 6.92$, $p < 0.001$) led to a significant increase in experiential service delivery ($\beta_{Flow} = 0.76$, $z = 36.81$, $p < 0.001$) and an overall more positive firm evaluation ($\beta_{Flow} = 0.66$, $z = 25.69$, $p < 0.001$), which in turn significantly enhanced consumer sentiment ($\beta_{ExpService} = 0.03$, $z = 2.52$, $p < 0.05$;

$\beta_{FirmEval} = 0.09$, $z = 8.76$, $p < 0.001$). Finally, a Chi-Square test assessing the subscription rate conditional on the type of interface further confirmed that participants in the voice-based interface condition were also significantly more likely to subscribe to become a future user of TER insurance compared to participants in the text-based interface condition ($M_{VoiceInterface} = 68.37\%$, $M_{TextInterface} = 50.05\%$; χ^2 (1, N = 201) = 5.93, $p = 0.01$; a path model analysis was not possible at the individual level due to the inability to track participants between systems).

Discussion

Study 4 demonstrates that consumers are more likely to engage in positive firm advocacy (i.e., enhanced consumer sentiment about the firm and conversion rates) when using voice-based compared to text-based interfaces. The current study provides also a more nuanced look at the underlying mechanism of flow, demonstrating that the increase in interface flow is driven by an increase in consumers' experienced level of challenge. We also provide initial evidence consistent with an "optimal challenge" account of flow-like experiences such that the effect of interface flow is weakened when consumers experience the service interaction as extremely challenging (weakened but not turned-off).

General discussion

Across four studies, we shed light on how the affordances of voice-based interfaces impact consumers' service experiences. We demonstrate that voice-based compared to text-based interfaces cause more absorptive, flow-like service experiences. In addition, we provide evidence for two novel boundary conditions (semantic fluency and conversational turns) that can reduce flow-like experiences and in turn dampen the observed positive effects of voice-based interfaces. Finally, we show how an increase in interface flow impacts both firm and service perceptions (how consumers perceive the overall service experience and evaluate the firm) as well as behavioral outcomes (contract renewal, consumer sentiment, conversion rate; see Table 3 for a summary of results and key findings across studies).

Theoretical implications

The findings of this research make four novel contributions. First, to the best of our knowledge, the current research is one of the first to contrast the effects of voice-based versus text-based self-service technologies on perceptual and behavioral service outcomes. We provide a novel theory of voice-mediated service interactions, hypothesizing and empirically demonstrating that the sensory-rich experience of voice-based interfaces leads to more positive service perceptions and an increase in interface flow. Specifically, we provide a novel look at prior work on flow (e.g., Berger et al., 2018; Hoffman & Novak, 1996; Zanjani et al., 2016), demonstrating that the voice modality is a key driver of flow and in turn enhances both perceptual (consumers' perception of the firm and service delivery) as well as behavioral service outcomes (from contract renewal to conversion rates).

Second, the current findings offer novel insights into consumer-technology relationships (Hoffman & Novak, 2018; Novak & Hoffman, 2018) and how the natural interaction through a voice-based interface creates a new "assemblage" of consumer experiences. Specifically, this research provides evidence that even though speaking to a voice-based interface represents merely a change in communication modality, consumers tend to form experiences that impact not only the assemblage itself (i.e., the relationship between the consumer and the interface) but also the entity providing the interface experience (i.e., the firm) by both attributing more positively-valenced perceptions toward the firm and to extend the relationship with that firm in the future.

Third, the current work introduces and empirically tests two novel boundary conditions that can either enhance or dampen the positive effects of voice-based user interfaces. First, we demonstrate that the use of easy-to-process phrases (i.e., high semantic fluency) is critical to boost greater levels of interface flow and in turn positively impact both service and firm perceptions, as well as behavioral outcomes. Second, we also show that consumers' experience of flow decreases with an increasing number of conversational turns throughout the conversation due to the more disjointed flow of the conversation. These findings provide a new look at the conditions under which voice-based interfaces can either enhance or reduce flow-like service experiences and the critical importance of the "conversational design" of voice-based interfaces.

Finally, the current findings also contribute to prior work on experiential service delivery. Experiential service delivery has been often considered to depend upon a multisensory service design (de Oliveira Santini et al., 2020), one that often requires a systematic orchestration of different service and context variables (Zomerdijk & Voss, 2010). However, the current findings demonstrate that a more experiential service perception can be induced by merely altering the interface modality without changing any other aspect of the service experience. The results of the current work suggest that voice-based interactions can transform even mundane tasks such as filing an insurance claim into an absorptive and ultimately more satisfactory service experience for consumers.

Table 3 Summary of studies

Study	Condition	Interface flow		Exp. Serv. delivery		Firm evaluation		Contract renewal	Consumer sentiment		Conversion rate	Sample size	Key findings
		M	SD	M	SD	M	SD	%	M	SD	%		
1	Text	2.31	1.67	2.90	1.92							184	Voice-based (as opposed to text-based) interfaces lead to enhanced levels of interface flow. Enhanced levels of interface flow in turn increase experiential service perceptions
	Voice	3.09	1.84	3.78	1.73								
2	Text – Low Fluency	2.86	1.57	2.65	1.57	4.21	1.72	64.24%				612	Semantic fluency moderates the effect of interface modality on interface flow such that the positive effect of voice-based interfaces on interface flow is enhanced (reduced) under conditions of high (low) fluency. Voice-based (as opposed to text-based) interfaces enhance attributions of experiential service experience and firm evaluation
	Text – High Fluency	2.59	1.47	2.39	1.38	4.67	1.45	69.18%					
	Voice – Low Fluency	3.38	1.63	3.40	1.67	4.58	1.64	63.92%					
	Voice – High Fluency	3.77	1.61	3.76	1.46	5.16	1.47	64.34%					
3	Text – Low Turns	2.85	1.63	2.79	1.76	4.98	1.43					610	The extent of conversational turns moderates the effect of interface modality on interface flow such that the positive effect of voice-based interfaces is reduced with a greater number of conversational turns
	Text – High Turns	3.20	1.69	2.98	1.78	5.10	1.48						
	Voice – Low Turns	3.65	1.85	3.97	1.80	4.95	1.65						
	Voice – High Turns	3.43	1.58	3.61	1.74	5.20	1.42						
4	Text	3.25	1.52	3.01	1.51	4.50	1.54		.03	.42	50.05%	811	Voice-based (as opposed to text-based) interfaces enhance positive firm-related behaviors such as more positive consumer sentiment and an increase in conversion rates
	Voice	4.02	1.62	4.23	1.58	4.75	1.61		.17	.41	68.37%		

Managerial implications

The findings of this research have immediate implications for practitioners. In what follows, we provide an overview of the implications and design guidelines for industry practice.

Simple language and conversational flows The semantic and structural characteristics of voice-based interfaces greatly impact the user experience and firm perception. Practitioners should carefully test and ensure that the conversational design of their service interaction is adequate given the user and task (i.e., ensuring high semantic fluency). In short, the boundary conditions we examined in this study highlight that the conversational design of a voice-based interface is sensitive to both the number of conversational turns during a service interaction and the specific words used by a voice interface during the conversation. When implementing voice-based interfaces, practitioners should be careful in limiting the number of conversational turns required to complete the service, as well as optimizing semantic fluency by using simpler and more familiar words to allow easier processing and in turn an enhanced service experience.

More inclusive user experiences Voice-based interfaces not only provide a richer, more absorptive user experience as demonstrated by this research but are also more inclusive across consumer demographics. Specifically, consumers with impaired writing abilities such as dyslexia or impaired sight (such as older consumers or consumers with a poor vision) can directly interact with a voice-based interface across service settings. Traditional interfaces often lead to detrimental outcomes and heightened frustration particularly for those otherwise disadvantaged consumer segments (Abdolrahmani et al., 2018). Thus, the well-selected use of voice technology provides an unexplored potential to deliver not only more sensory-rich but ultimately more inclusive user experiences.

Voice-based interfaces as a sales channel The findings of this research revealed that voice-based interfaces can boost downstream firm-related outcomes such as conversion or contract renewal rates compared to text-based interfaces. It is plausible that this greater persuasiveness is driven by a combination of both enhanced flow (as per our theorizing) and also mechanisms studied in earlier work (such as enhanced social presence; Qiu & Benbasat, 2009). The important implication for firms is that voice-based interfaces are an effective lever in a marketer's sales automation toolbox. However, our boundary conditions also reveal that firms are well-advised to ensure that consumers' processing capacities are not overwhelmed (i.e., by reducing the complexity of the language used and by avoiding a large number of conversational turns). Thus, firms need to balance the effectiveness of voice-based interfaces for sales automation purposes with the affordances of the task and the user.

Context-dependent channel choice Current industry practice in choosing the right channel to interact with consumers is often more driven by trial and error and technology vendors as opposed to a strategic mapping of task affordances and business objectives. We hope that the current research can provide guidelines to decide more effectively under which conditions a voice-based interface is appropriate and when not (see also Table 2 for a summary of key properties of voice-based communication). Given the high channel synchronicity of voice-based interfaces, service firms are advised to choose text-based interfaces for tasks that require more deliberation from the consumer and that are better suited for asynchronous interactions. Specifically, service tasks that require parallel processing such as comparing multiple product configurations during a shopping task are better suited for text-based interfaces as the consumer can re-read and re-type information if needed.

Monitoring channel choice Multimodal interfaces are becoming increasingly common across the service value chain and allow consumers to interact with self-service technologies via voice or text interchangeably. This begs the question of whether consumers should be able to choose their preferred interaction modality or not. Our results suggest that some consumers may feel overwhelmed by the speed and immediacy of voice-based interfaces (Study 4). Hence, firms are advised to carefully monitor drop-out rates and also transition periods when consumers switch from one modality over to the other. Allowing consumers to choose their preferred channel is a critical ingredient in enhancing consumers' agency during service interactions which we expect to positively impact both service and firm outcomes.

Cost-effective integration in existing infrastructure With the rapid development of technology platforms such as Google's natural language processing platform Google Dialogue or Amazon's Polly, voice-based interfaces can be easily integrated into the corporate infrastructure. Instead of developing a native and firm-owned voice-based interface channel, companies can easily tweak existing language models (such as Google WaveNet as in the current research) and incorporate them into their current service delivery process. As highlighted in our technical documentation in the Web Appendix (see Web Appendix A1), the deployment and technical integration into existing enterprise solutions are both cost-effective (infrastructure provided, maintained, and continuously improved by firms like Google or Amazon) and customizable (such as developing a unique "voice character" that matches the personality of the brand). In short, the use of voice-based interfaces is highly modular and can be integrated flexibly across different stages of the customer journey and without requiring substantive changes to the corporate technology infrastructure.

Future research

We see three immediate directions for future research. First, future work could further explore the effective "voice design" and to which extent the voice of the assistant or interface (such as the depth, pitch, or other vocal features of the voice-based interface) can be mapped onto the "brand personality" that the voice-based interface represents. Second, one unexplored area of research is whether and to which extent the specific features of a voice-based interface can evoke specific affective responses from the user interacting with the interface. As softer and slower-paced voices tend to induce greater psychological comfort and calmness (Elkins & Derrick, 2013; Nass & Lee, 2001), future work might explore which specific features of the voice-based interface affect which specific psychological user experience. Third, the use of voice technology may not be beneficial across all settings or contexts (Cambre et al., 2020; Seaborn et al., 2021). We acknowledge that the impact of voice-based interfaces on consumers is arguably multiply determined, which is also reflected in the varying effect sizes across our studies. Future work could therefore further explore under which conditions voice technology can enhance consumers' experience versus when it becomes a nuisance or even detrimental.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11747-022-00868-5.

Funding This research received funding by the Basic Research Fund of the University of St.Gallen (2120255/2200687) and the Swiss National Science Foundation (100013_192718). Open access funding provided by University of St.Gallen.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdolrahmani, A., Kuber, R., & Branham, S. M. (2018). Siri talks at you: An empirical investigation of voice-activated personal assistant (VAPA) usage by individuals who are blind. In *ASSETS 2018 - Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 249–258). https://doi.org/10.1145/3234695.3236344

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quaterly, 24*(4), 665–694.

Berger, A., Schlager, T., Sprott, D. E., & Herrmann, A. (2018). Gamified interactions: Whether, when, and how games facilitate self–brand connections. *Journal of the Academy of Marketing Science, 46*(4), 652–673. https://doi.org/10.1007/s11747-017-0530-0

Berger, J. A., Rocklage, M., & Packard, G. M. (2021). Expression modalities: How speaking versus writing shapes what consumers say, and its impact. *Journal of Consumer Research*, (Forthcoming). https://doi.org/10.2139/ssrn.3791506

Blake, M. (2018). 4 Ways to Disrupt Insurance with Customer Technology. *Forbes*. https://www.forbes.com/sites/blakemorgan/2018/05/02/4-ways-to-disrupt-insurance-with-customer-technology/?sh=467684b1382e

Brannon Barhorst, J., McLean, G., Shah, E., & Mack, R. (2021). Blending the real world and the virtual world: Exploring the role of flow in augmented reality experiences. *Journal of Business Research, 122*, 423–436. https://doi.org/10.1016/j.jbusres.2020.08.041

Cambre, J., Reig, S., Kravitz, Q., & Kulkarni, C. (2020). "All rise for the AI director": Eliciting possible futures of voice technology through story completion. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 2051–2064). https://doi.org/10.1145/3357236.3395479

Clark, H. H., & Brennan, S. E. (1991). Grounding in Communication. In *Perspectives on Socially Shared Cognition* (pp. 222–233). https://doi.org/10.1037/10096-006

Clarke, S. G., & Haworth, J. T. (1994). 'Flow' experience in the daily lives of sixth-form college students. *British Journal of Psychology, 85*(4), 511–523. https://doi.org/10.1111/j.2044-8295.1994.tb02538.x

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: Experiencing flow in work and play*. Jossey-Bass Publishers.

Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychologyo Title. In *Optimal experience: Psychological studies of flow in consciousness* (2, pp. 15–35).

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology, 56*(5), 815–822. https://doi.org/10.1037/0022-3514.56.5.815

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

de Oliveira Santini, F., Ladeira, W. J., Pinto, D. C., Herter, M. M., Sampaio, C. H., & Babin, B. J. (2020). Customer engagement in social media: A framework and meta-analysis. *Journal of the Academy of Marketing Science, 48*(6), 1211–1228. https://doi.org/10.1007/s11747-020-00731-5

Dennis, A. R., & Kinney, S. T. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research, 9*(3), 256–274. https://doi.org/10.1287/isre.9.3.256

Diederich, S., Brendel, A. B., Morana, S., & 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*, Forthcoming.

DuBay, W. H. (2004). *The principles of readability: A brief introduction to readability research*. Impact Information.

Elkins, A. C., & Derrick, D. C. (2013). The sound of trust: Voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation, 22*(5), 897–913. https://doi.org/10.1007/s10726-012-9339-x

Fernández-Sabiote, E., & López-López, I. (2020). Discovering call interaction fluency: A way to improve experiences with call centres. *Service Science, 12*(1), 26–42. https://doi.org/10.1287/SERV.2019.0251

Fuscaldo, D. (2019). Bank Of America's Virtual Assistant Now Has More Than 10 Million Users. *Forbes*. https://www.forbes.com/sites/donnafuscaldo/2019/12/11/bank-of-americas-virtual-assistant-now-has-more-than-10-million-users/?sh=1ceeb18ef69b

Ghani, J. A., Supnick, R., & Rooney, P. (1991). The experience of flow in computer-mediated and in face-to-face groups. In *ICIS 1991 Proceedings*.

Globerson, T. (1983). Mental capacity, mental effort, and cognitive style. *Developmental Review, 3*(3), 292–302.

Guan, Z., Hou, F., Li, B., Phang, C. W., & Chong, A. Y. L. (2021). What influences the purchase of virtual gifts in live streaming in China? A cultural context-sensitive model. *Information Systems Journal*, 1–37. https://doi.org/10.1111/isj.12367

Haeckel, S. H. (1998). Abou th nature and future of interactive marketing. *Journal of Interactive Marketing, 12*(1), 63–71.

Harmeling, C. M., Moffett, J. W., Arnold, M. J., & Carlson, B. D. (2017). Toward a theory of customer engagement marketing. *Journal of the Academy of Marketing Science, 45*(3), 312–335. https://doi.org/10.1007/s11747-016-0509-2

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.

Hayes, P., & Wagner, J. (2018). Prepare for the voice revolution. https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html

Hess, T., Fuller, M., & Campbell, D. (2009). Designing interfaces with social presence: Using vividness and extraversion to create social recommendation agents. *Journal of the Association for Information Systems, 10*(12), 889–919. https://doi.org/10.17705/1jais.00216

Hildebrand, C., & Bergner, A. (2020). Conversational robo advisors as surrogates of trust: Onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science, 49*, 659–676. https://doi.org/10.1007/s11747-020-00753-z

Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W. H., Hoffman, D. L., & Novak, T. P. (2020). Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications. *Journal of Business Research, 121*, 364–374. https://doi.org/10.1016/j.jbusres.2020.09.020

Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing, 60*(3), 50–68. https://doi.org/10.2307/1251841

Hoffman, D. L., & Novak, T. P. (2009). Flow online: Lessons learned and future prospects. *Journal of Interactive Marketing, 23*(1), 23–34. https://doi.org/10.1016/j.intmar.2008.10.003

Hoffman, D. L., & Novak, T. P. (2018). Consumer and object experience in the internet of things: An assemblage theory approach. *Journal of Consumer Research, 44*(6), 1178–1204. https://doi.org/10.1093/jcr/ucx105

Holbrook, M. B., & Gardner, M. P. (1993). An approach to investigating the emotional determinants of consumption durations: Why do people consume what they consume for as long as they consume it? *Journal of Consumer Psychology, 2*(2), 123–142. https://doi.org/10.1016/S1057-7408(08)80021-6

Hollebeek, L. D., Sprott, D. E., & Brady, M. K. (2021). Rise of the machines? Customer engagement in automated service interactions. *Journal of Service Research, 24*(1), 3–8. https://doi.org/10.1177/1094670520975110

Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155–172. https://doi.org/10.1177/1094670517752459

Huang, M. H., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research, 24*(1), 30–41. https://doi.org/10.1177/1094670520902266

Johnson, J. L., Lee, R. P. W., Saini, A., & Grohmann, B. (2003). Market-focused strategic flexibility: Conceptual advances and an integrative model. *Journal of the Academy of Marketing Science, 31*(1), 74–89. https://doi.org/10.1177/0092070302238603

Khawaja, M. A., Chen, F., & Marcus, N. (2010). Using Language Complexity to Measure Cognitive Load for Adaptive Interaction Design. In *International Conference on Intelligent User Interfaces, Proceedings IUI* (pp. 333–336). https://doi.org/10.1145/1719970.1720024

Kim, D., & Ko, Y. J. (2019). The impact of virtual reality (VR) technology on sport spectators' flow experience and satisfaction. *Computers in Human Behavior, 93*, 346–356. https://doi.org/10.1016/j.chb.2018.12.040

King, D., Auschaitrakul, S., & Lin, C. W. J. (2021). Search modality effects: merely changing product search modality alters purchase intentions. *Journal of the Academy of Marketing Science*, (0123456789). https://doi.org/10.1007/s11747-021-00820-z

Klesse, A. K., Levav, J., & Goukens, C. (2015). The effect of preference expression modality on self-control. *Journal of Consumer Research, 42*(4), 535–550. https://doi.org/10.1093/jcr/ucv043

Lee, G. G., & Lin, H. F. (2005). Customer perceptions of e-service quality in online shopping. *International Journal of Retail and Distribution Management, 33*(2), 161–176. https://doi.org/10.1108/09590550510581485

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing, 80*(6), 69–96. https://doi.org/10.1509/jm.15.0420

Levinson, S. C. (2016). Turn-taking in human communication - origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Mathwick, C., & Rigdon, E. (2004). Play, flow, and the online search experience. *Journal of Consumer Research, 31*(2), 324–332. https://doi.org/10.1086/422111

McDowell, W. C., Wilson, R. C., & Kile, C. O. (2016). An examination of retail website design and conversion rate. *Journal of Business Research, 69*(11), 4837–4842. https://doi.org/10.1016/j.jbusres.2016.04.040

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research, 56*(4), 535–556. https://doi.org/10.1177/0022243718822827

Moffett, J. W., Folse, J. A. G., & Palmatier, R. W. (2021). A theory of multiformat communication: Mechanisms, dynamics, and strategies. *Journal of the Academy of Marketing Science, 49*, 441–461. https://doi.org/10.1007/s11747-020-00750-2

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied, 7*(3), 171–181. https://doi.org/10.1037/1076-898X.7.3.171

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proceedings, 718,* 93–98.

Novak, T. P., & Hoffman, D. L. (2018). 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, 47*(2), 216–237. https://doi.org/10.1007/s11747-018-0608-3

Novak, T. P., Hoffman, D. L., Yung, Y. F., Novak, T. P., & Hoffman, D. L. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science, 19*(1), 22–42. https://doi.org/10.1287/mksc.19.1.22.15184

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., et al. (2016). WaveNet: A Generative Model for Raw Audio, 1–15. http://arxiv.org/abs/1609.03499

Pagani, M., Racat, M., & Hofacker, C. F. (2019). Adding voice to the omnichannel and how that affects brand trust. *Journal of Interactive Marketing, 48*, 89–105. https://doi.org/10.1016/j.intmar.2019.05.002

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peer, E., Rothschild, D., Evernden, Z., Gordon, A., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research data quality of platforms and panels for online behavioral research. *Social Science Research Network*, 1–46.

Perez, S. (2017). Starbucks unveils a virtual assistant that takes your order via messaging or voice. *TechCrunch*. https://techcrunch.com/2017/01/30/starbucks-unveils-a-virtual-assistant-that-takes-your-order-via-messaging-or-voice/

Pogacar, R., Shrum, L. J., & Lowrey, T. M. (2018). The effects of linguistic devices on consumer information processing and persuasion: A language complexity × processing mode framework. *Journal of Consumer Psychology, 28*(4), 689–711. https://doi.org/10.1002/jcpy.1052

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems, 25*(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Qiu, L., & Benbesat, I. (2005). Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction, 19*(1), 75–94.

Redeker, G. (1984). On differences between spoken and written language. *Discourse Processes, 7*(1), 43–55. https://doi.org/10.1080/01638538409544580

Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using artificial intelligence to create value in insurance. *International Journal of Bank Marketing, 36*(6), 1145–1168. https://doi.org/10.1108/IJBM-01-2017-0015

Rubin, D. L., Hafer, T., & Arata, K. (2000). Reading and listening to oral-based versus literate-based discourse. *Communication Education, 49*(2), 121–133. https://doi.org/10.1080/03634520009379200

Rzepka, C., Berger, B., & Hess, T. (2021). Voice assistant vs . Chatbot – examining the fit between conversational agents' interaction modalities and information search tasks. *Information Systems Frontiers*, (Forthcoming). https://doi.org/10.1007/s10796-021-10226-5

Schouten, J. W., McAlexander, J. H., & Koenig, H. F. (2007). Transcendent customer experience and brand community. *Journal of the Academy of Marketing Science, 35*, 357–368. https://doi.org/10.1007/s11747-007-0034-4

Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in human-agent interaction. *ACM Computing Surveys, 54*(4), 1–43. https://doi.org/10.1145/3386867

Singh, A., Ramasubramanian, K., & Shivam, S. (2019). Processes in the Banking and Insurance Industries. In *Building an Enterprise Chatbot* (1st edn.). Apress.

Singh, J., Nambisan, S., Bridge, R. G., & Brock, J. K. U. (2020). One-voice strategy for customer engagement. *Journal of Service Research, 24*(1), 42–65. https://doi.org/10.1177/1094670520910267

Smith, S. (2018). Digital voice assistants in use to triple to 8 billion by 2023, Driven by Smart Home Devices. *Juniper Research*. https://www.juniperresearch.com/press/digital-voice-assistants-in-use-to-8-million-2023

Son, Y., & Oh, W. (2018). "Alexa, buy me a movie!": How AI speakers reshape digital content consumption and preference. In *39th International Conference on Information Systems*.

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science, 19*(10), 986–988. https://doi.org/10.1111/j.1467-9280.2008.02189.x

Steenkamp, J.-B.E.M., & Baumgartner, H. (1992). The role of optimum stimulation level in exploratory consumer behavior. *Journal of Consumer Research, 19*(3), 434–448. https://doi.org/10.1086/209313

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Thompson, D. V., & Ince, E. C. (2013). When disfluency signals competence: The effect of processing difficulty on perceptions of service agents. *Journal of Marketing Research, 50*(2), 228–240. https://doi.org/10.1509/jmr.11.0340

Trevino, L. K., & Webster, J. (1992). Flow in computer-mediated communication. *Communication Research, 19*(5), 539–573.

Unger, L. S., & Kernan, J. B. (1983). On the meaning of leisure: An investigation of some determinants of the subjective experience. *Journal of Consumer Research, 9*(4), 381–392. https://doi.org/10.1086/208932

van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo Arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research, 20*(1), 43–58. https://doi.org/10.1177/1094670516679272

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457–479. https://doi.org/10.1093/applin/ams074.

Walther, J. B. (2005). Let me count the ways: The interchange of verbal and nonverbal cues in computer- mediated and face-to-face affinity. *Journal of Language and Social Psychology, 24*(1), 36–65. https://doi.org/10.1177/0261927X04273036

Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior, 9*(4), 411–426. https://doi.org/10.1016/0747-5632(93)90032-N

Wiemann, J. M., & Knapp, M. L. (1975). Turn-taking in Conversations. *Journal of Communication, 25*, 75–92. https://doi.org/10.1111/j.1460-2466.1975.tb00582.x

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management, 29*(5), 907–931. https://doi.org/10.1108/JOSM-04-2018-0119

Zanjani, S. H. A., Milne, G. R., & Miller, E. G. (2016). Procrastinators' online experience and purchase behavior. *Journal of the Academy of Marketing Science, 44*(5), 568–585. https://doi.org/10.1007/s11747-015-0458-1

Zomerdijk, L. G., & Voss, C. A. (2010). Service design for experience-centric services. *Journal of Service Research, 13*(1), 67–82. https://doi.org/10.1177/1094670509351960

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.