



Voice Assistant vs. Chatbot – Examining the Fit Between Conversational Agents’ Interaction Modalities and Information Search Tasks

Christine Rzepka¹ · Benedikt Berger² · Thomas Hess³

Accepted: 18 November 2021 / Published online: 10 December 2021
© The Author(s) 2021

Abstract

Owing to technological advancements in artificial intelligence, voice assistants (VAs) offer speech as a new interaction modality. Compared to text-based interaction, speech is natural and intuitive, which is why companies use VAs in customer service. However, we do not yet know for which kinds of tasks speech is beneficial. Drawing on task-technology fit theory, we present a research model to examine the applicability of VAs to different tasks. To test this model, we conducted a laboratory experiment with 116 participants who had to complete an information search task with a VA or a chatbot. The results show that speech exhibits higher perceived efficiency, lower cognitive effort, higher enjoyment, and higher service satisfaction than text-based interaction. We also find that these effects depend on the task’s goal-directedness. These findings extend task-technology fit theory to customers’ choice of interaction modalities and inform practitioners about the use of VAs for information search tasks.

Keywords voice assistant · conversational agent · speech interaction · cognitive fit · customer service

1 Introduction

With rapid advances in artificial intelligence, specifically natural language processing, speech holds promise as an interaction modality between humans and computers (Enholm et al., 2021; Hirschberg & Manning, 2015). Voice assistants (VAs) enable speech interaction with existing devices (e.g., smartphones) and serve as the main interface of new devices (e.g., smart speakers). Similar to chatbots,

which interact in written form, VAs engage the user in conversations that are set up to assist in various tasks, such as the answering of questions and smart home control (Hoy, 2018). Both VAs and chatbots can be referred to as conversational agents (CAs) (Diederich et al., 2022; Gnewuch et al., 2017). Companies increasingly adopt CAs in banking, healthcare, and e-commerce, with customer service being a key use case (Behera et al., 2021; Woodford, 2020). A customer service encounter describes “any customer-company interaction that results from a service system that is comprised of interrelated technologies, human actors (employees and customers), physical/digital environments and company/customer processes” (Larivière et al., 2017, p. 2). CAs enable cost savings while still providing human-like interaction, thereby bridging the existing service provision gap between frontline employees and self-service technologies (Adam et al., 2021; Larivière et al., 2017). However, the availability of speech- and text-based conversational interfaces to communicate with companies raises questions about the impact of the interaction modalities on customer service encounters (De Keyser et al., 2019).

Extant literature has examined costumers’ satisfaction with service encounters to assess whether offering a CA is beneficial (Diederich et al., 2019; Diederich et al., 2020).

✉ Christine Rzepka
ch.rzepka@lmu.de

Benedikt Berger
benedikt.berger@ercis.uni-muenster.de

Thomas Hess
thess@lmu.de

¹ Institute for Digital Management and New Media, Ludwig-Maximilians-Universität München (LMU Munich), Ludwigstr. 28, 80539 Munich, Germany

² Department of Information Systems, University of Münster, Leonardo-Campus 11, 48149 Münster, Germany

³ Institute for Digital Management and New Media, Ludwig-Maximilians-Universität München (LMU Munich), Ludwigstr. 28, 80539 Munich, Germany

In the service literature, customers' satisfaction with the service is used as a measure for the company's success in addressing their expectations and needs (Dai & Salam, 2020; McKinney et al., 2002). To satisfy customers, service encounters must provide them with both utilitarian and hedonic value (Childers et al., 2001; Chiu et al., 2014; Jones et al., 2006). Because satisfied customers tend to continue using a service (Bhattacharjee, 2001) or to buy repeatedly (Bartl et al., 2013), satisfaction is an appropriate construct to evaluate the effect of CAs' distinct interaction modalities on service encounters. Indeed, existing research comparing speech- and text-based interactions shows that speaking generates positive attitudes toward the CA (Cho et al., 2019; Novielli et al., 2010; Schroeder & Schroeder, 2018). These attitudes are driven by both utilitarian and hedonic values: Studies on users' perceived efficiency find that speaking is faster than typing, which increases users' productivity (Le Bigot et al., 2007; Ruan et al., 2017). Speech is also more intuitive and natural (Kock, 2004), increasing its ease of use and decreasing cognitive effort. Users also report simply enjoying the interaction, which stresses VAs' hedonic benefits (Pal et al., 2020; Yang & Lee, 2019). However, research explicitly comparing speech- and text-based interactions reports inconclusive results. While speaking is faster and easier (Ruan et al., 2017), listening to the CA's answers reduces the interaction speed, thereby increasing the relative efficiency of text-based conversations (Le Bigot et al., 2004).

Whether speaking or texting is more beneficial during a service encounter depends on the type of task the customer wants to solve. Extant service literature distinguishes between (goal-directed) searching and (experiential) browsing tasks (Hoffman & Novak, 1996; Hong et al., 2004). Depending on the customers' goal when interacting with, e.g., a chatbot, the CA's conversational style should be more or less friendly (Chen et al., 2021). Initial research on speech- and text-based CAs found that users' preference for one interaction modality over another can change depending on the task and context of the interaction. Krämer et al. (2009) show that users' preference for text interaction decreased after service failures, while speech interaction preferences increased. Cho et al. (2019) find that attributing a human likeness mediates users' attitudes in utilitarian tasks but not in hedonic tasks. Additionally, VA users frequently adopt VAs for simple routine tasks, but remain reluctant to use them for more complex tasks (Ma & Liu, 2020). Consequently, no interaction modality is superior to the other. While previous research has examined the fit between specified tasks and technologies to support these tasks (Goodhue & Thompson, 1995), the literature has not yet considered differences between speech- and text-based interaction for information search tasks. Against this background, we aim to answer the following research question:

How do CAs' interaction modalities fit different information search tasks in the service encounter?

To address this research question, we investigate the applicability of speech- and text-based conversational interaction to different information search task types. Because speech- and text-based interactions differ in the way humans produce, transmit, and process information, we use cognitive load theory and cognitive fit theory (Vessey & Galletta, 1991) as underlying theories to derive our hypotheses. Based on this theoretical foundation, task-technology fit theory (Goodhue & Thompson, 1995) guides the inclusion of the task type in the research model. We conduct a 2 (interaction modality: speech vs. text) \times 2 (search task type: goal-directed vs. experiential) between-subject laboratory experiment to test our research model. The results show that the search task's perceived goal-directedness moderates the interaction modality's effect on users' perceptions of the interaction. Thereby, our study extends cognitive fit theory to conversational information requests and presentations and task-technology fit theory to users' choice of interaction modalities. Our findings further inform practitioners about the benefits of speech-based interactions over text-based interactions, and of their effects on customers' satisfaction. Considering that visual and auditory modalities suit distinct types of information and tasks in different ways, it is essential that service providers understand how different tasks and interaction modalities jointly affect customers' satisfaction in the service encounter (De Keyser et al., 2019; Lee et al., 2001). In this way, depending on the task's goal-directedness, system designers can learn when and how the offering of speech as an interaction modality is beneficial.

This paper is structured as follows: In section 2, as the conceptual background of our study, we provide an overview of extant research on CAs and the relevant theories. Section 3 introduces the research model and derives the underlying hypotheses. We describe the experiment including its procedure, measurements, and data collection process in section 4 and report our analyses' results in section 5. Section 6 presents the main findings and indicates their implications in the light of extant literature. The paper briefly concludes in section 7.

2 Conceptual Background

2.1 Voice Assistants, Conversational Agents, and the Service Encounter

Although first efforts to build speech-based systems date back to 1950, the diffusion of VAs was only recently enabled by advancements in natural language processing and the availability of large volumes of linguistic data (Hinton et al., 2012; Hirschberg & Manning, 2015). VAs minimally

consist of an automatic speech recognizer that interprets human speech, a dialogue manager that determines and performs the requested action, and text-to-speech synthesis that responds in a spoken form (Hirschberg & Manning, 2015). While speech- and text-based CAs face the same challenges in interpreting language, advances in artificial intelligence enabled VAs to successfully map sound signals to sequences of words, thereby overcoming what is arguably the biggest difficulty in attaching meaning to speech (Hinton et al., 2012). Various VAs followed Apple's launch of Siri in 2011, each offering a new interaction modality using speech as input and output to represent information (Hoy, 2018).

Existing information systems (IS) research on VA use distinguishes between utilitarian and hedonic drivers and inhibitors. As speech is a natural and intuitive way of interacting (Kock, 2004), various studies stress its utilitarian benefits of convenience, ease of use, and time saving through hands-free possibilities (Balakrishnan et al., 2021; McLean & Osei-Frimpong, 2019). In addition, users enjoy talking to their VAs, thus deriving hedonic value from their interactions (Pal et al., 2020; Rzepka, 2019; Yang & Lee, 2019). Although users' enjoyment could be attributed to the newness of the technology (McLean & Osei-Frimpong, 2019), users may also enjoy the social aspect of engaging with a VA. Extant research shows that VAs trigger users' perceptions of human likeness (Cho et al., 2019) and social presence (McLean & Osei-Frimpong, 2019), which also encourages more personal dialogues (Novielli et al., 2010). Users' perceptions of human likeness are driven by VAs' capabilities to communicate in natural and interactive ways, e.g., through synthesized speech, and to answer in a responsive manner as they are able to process human speech (Li, 2015). Extant research shows that these social cues elicit social responses from users who perceive the IT system as a social actor (Moon, 2000; Nass & Moon, 2000). Social cues not only promote use intentions (McLean & Osei-Frimpong, 2019), but also attenuate privacy concerns that present a major adoption barrier (Benlian et al., 2020). These cues are subject to various IS studies on text-based CAs (i.e., chatbots) (e.g., Adam et al., 2021; Diederich et al., 2019; Gnewuch et al., 2017). However, direct comparisons suggest that speech differs from text in its impact on users' attitudes (Cho et al., 2019; Novielli et al., 2010) and task performance (Le Bigot et al., 2004; Le Bigot et al., 2007). This raises the need for further research on the hedonic and utilitarian differences between chatbot and VA use.

The question regarding differences between speech and text specifically becomes relevant in domains where humans can choose between interaction modalities, e.g., service encounters (De Keyser et al., 2019; Larivière et al., 2017). The application of CAs in this context is promising as they offer the customer efficient, yet still personalized service, thus addressing the key challenges of customer service:

service efficiency and service quality (Adam et al., 2021; De Keyser et al., 2019). As indicated, CAs provide both utilitarian and hedonic benefits, which drive customers' satisfaction with the service and related outcomes (Childers et al., 2001; Chiu et al., 2014; Jones et al., 2006). While Mero (2018) shows that a text-based CA's interactivity has positive effects on customers' satisfaction, prior research on haptic interfaces emphasizes that the interaction modality can be as important as the content itself (Brasel & Gips, 2014). For instance, extant studies comparing speech and manual expression modalities show that speech reduces customers' self-control, leading to different product choices (Klesse et al., 2015) and consumption behaviors (Son & Oh, 2018).

CAs are implemented in multiple steps and tasks along the service process, starting with the information search and continuing through purchasing products to filing a complaint. In this context, extant research identified two major search task types that customers conduct online: goal-directed (searching) tasks and experiential (browsing) tasks (Hong et al., 2004). Searching is characterized by a clearly defined goal and is extrinsically motivated. In contrast, browsing is exploratory and non-directed, being intrinsically motivated (Hoffman & Novak, 1996). These two task types describe two instantiations of the specificity of customers' search objectives, which translate into a continuum of the customers' perceived goal-directedness. Although customers may have already successfully adopted CAs for tasks they perceive as highly goal-directed, they show a reluctance when perceiving tasks as less goal-directed, especially when using VAs (Ma & Liu, 2020). However, until now, service literature has mainly focused on visual search behaviors and distinguished different display characteristics and information formats such as lists and matrices (e.g., Hong et al., 2004; Nadkarni & Gupta, 2007). More recently, Chen et al. (2021) distinguished between these tasks to examine the fit of a chatbot's distinct conversational styles to the customers' goal. Overall, little is known on the differences between spoken and written interactions and the interplay between customers' information requests and the system's information presentation in a two-way conversation. In order to effectively design voice-based experiential search, it is necessary to understand the differences between speech- and text-based searches for different information search tasks (Ma & Liu, 2020) and to extend initial research efforts in this field (Chen et al., 2021).

2.2 Cognitive Fit Theory and Task-Technology Fit Theory

Research in human-computer interaction and linguistics confirms that speech and text differ in their modes of production, transmission, and reception. Speaking and listening develop naturally and intuitively, while writing and reading

develop only through formal learning (Akinnsaso, 1982). Being more natural than writing, speaking requires less cognitive effort (Kock, 2004; Le Bigot et al., 2007). Moreover, physical effort is reduced through hands-free use (Akinnsaso, 1982). In addition, speech is of a temporal nature, requiring sequential information presentation, whereas written text can be processed in parallel as the reader can skip parts of the text (Rubin et al., 2000). Hence, depending on the information and the task, these distinct properties of speech and text can affect customers' preferences for one or the other modality. To understand how the interplay between the interaction modality and the task type affects customers' cognitive resources, we draw on cognitive fit theory and task-technology fit (TTF) theory.

Cognitive fit theory offers an explanation for the match or mismatch between interaction modalities and task types (Vessey & Galletta, 1991). Depending on the match between the form of information representation and the task, users' performance within a given system could be enhanced by reducing their cognitive load when carrying out the task. However, if there is a mismatch and the users' cognitive ability for information processing and their attention span is exceeded, they experience overload and lose information. The term "cognitive load" describes the human's working memory capacity used to accomplish the task, depending on the task's characteristics and the users' cognitive resources (Paas et al., 2004). Using this concept of "fit," Goodhue and Thompson (1995) established a generic TTF model to measure the interplay between different characteristics of the task, technology, and human through the users' perceived TTF. The higher the discrepancy between task and technology, the lower the perceived TTF and the lower the value users get from using a technology. High TTF, by contrast, positively affects users' utilization of the technology and performance in achieving "improved efficiency, improved effectiveness, and/or higher quality" (Goodhue & Thompson, 1995, p. 218). The TTF theory allows us to assume that either text or speech has a better fit with certain tasks, which eventually affects users' performance. Initial findings support this assumption, showing that speech interaction is better evaluated than text-based interactions for utilitarian tasks, but not for hedonic tasks (Cho et al., 2019).

Hence, we adapt TTF and cognitive fit theory to investigate task-modality fit (Lee et al., 2001) between speech- and text-based CAs for two information search tasks. Because the interaction modality describes a distinct technical property of CAs, which is shown to affect users' perceptions and behavior, task-modality fit describes a TTF problem. Cognitive fit theory and TTF theory are frequently used to study the adoption of diverse IT systems such as social networking sites (Lu & Yang, 2014) or green IS (Yang et al., 2018). In the service context, TTF theory has already been adapted to the whole customer journey (Wells et al., 2003; You et al., 2020) and to specific tasks such

as information search (Dang et al., 2020; Hong et al., 2004). Additionally, Chen et al. (2021) apply cognitive fit theory to investigate the matching of a chatbot's interaction style with goal-directed and experiential tasks. Based on this theoretical framework, the following section derives our research model on the applicability of CAs' interaction modalities for distinct information search tasks.

3 Research Model and Hypotheses

The use of CAs promises benefits in terms of efficiency, specifically in the service encounter (De Keyser et al., 2019; Kraus et al., 2019). However, we do not know to what extent these benefits depend on the interaction modality. Le Bigot et al. (2004), for instance, find that text interaction is faster than speech. With current technical advancements, however, VAs offer more natural conversational interaction possibilities and faster processing speeds (Dubiel et al., 2018). Consequently, comparisons in prior literature may only hold to a certain extent (Schmitt et al., 2021). Recent qualitative investigations on VA use suggest that users particularly value the efficiency of speech interaction (Moussawi, 2018; Rzepka, 2019). From a users' perspective, speech production is faster than writing because it does not involve the physical process of writing (Akinnsaso, 1982). Hence, we assume that customers achieve and perceive more efficiency in speech interaction compared to text. At the same time, extant research shows that reading a system's prompts requires less time than listening to them because humans can selectively retrieve relevant information from text (Rubin et al., 2000). In contrast, listening to the VA's information in a sequential manner takes time and reduces users' efficiency (Dubiel et al., 2020; Schmitt et al., 2021). The amount of information that users must listen to depends on the tasks they conduct. Because customers listen to less information during tasks they perceive to be more goal-directed (Hong et al., 2004), the advantages of speech interaction should be more pronounced if the level of goal-directedness is high. However, if customers want to compare and evaluate different alternatives in experiential tasks, they need to listen to more information sequentially in speech-based interactions, which reduces their efficiency compared to text-based interactions (Dubiel et al., 2020). Therefore, we hypothesize:

H1a: Customers achieve higher levels of efficiency with speech interaction than with text interaction.

H1b: Goal-directedness positively moderates the effect of interaction modality on efficiency.

Besides efficiency, cognitive effort is a suitable variable to account for differences between interaction modalities and tasks (Hong et al., 2004). While Le Bigot et al. (2004,

2007) consistently show that cognitive effort is higher for speech than for textual interaction, Ruan et al. (2017) report the opposite. Recent technological advancements and media naturalness theory offer a potential explanation for this phenomenon (Kock, 2004; Schmitt et al., 2021). Humans perceive speech interaction as more natural and intuitive, which reduces cognitive effort (Kock, 2004). However, any decrease in naturalness through speech recognition errors or synthetic speech output weakens this effect. As VAs have recently advanced technologically, we refer to recent empirical results (Dubiel et al., 2018; Foley et al., 2020; Ruan et al., 2017) and suggest that speech interaction requires less cognitive effort than text interaction. This is in line with the basic assumption of media naturalness theory and is based on the fact that speech does not require the formal written structuring of information (Akinaso, 1982). Producing text involves a process of consciously applying intentionally learnt rules, while humans intuitively build and process speech. In particular, we expect this effect to be strong if customers have a high goal-directed perception of a task. For low goal-directedness, customers are likely to use more interaction turns to explore the CA's information, analogous to the increasing number of clicks on a website (Hong et al., 2004). These interactions increase cognitive effort because customers need to remember the agent's information, which becomes more difficult considering the increasing information load of multiple turns (Paas et al., 2004). Consequently, when evaluating different alternatives in experiential search, the sequential information representation in speech-based interactions applies cognitive pressure on customers as they need to retain different options in their memory (Dubiel et al., 2020). Customers can, however, selectively retrieve relevant information from the system's written answers, which reduces their cognitive effort compared to listening to and memorizing the information given in spoken interactions (Rubin et al., 2000). Therefore, we assume speech requires more cognitive effort in experiential tasks:

H2a: Customers perceive less cognitive effort when they interact via speech compared to interacting via text.

H2b: Goal-directedness positively moderates the effect of interaction modality on cognitive effort.

Moreover, CAs promise to provide personal customer encounters, similar to relationships with human personnel (Larivière et al., 2017). With their anthropomorphic characteristics and social cues, CAs can promote feelings of humanness and social presence (Diederich et al., 2019). Speech-based interactions trigger such perceptions more than text-based interactions do, which leads to more positive (Cho et al., 2019) and personal attitudes (Novielli et al., 2010). This indicates that users prefer VAs because they enjoy the interaction itself (Rzepka, 2019). Indeed, extant

research shows that perceived enjoyment has a significant positive effect on humans' intention to use VAs (Pal et al., 2020; Yang & Lee, 2019). Hence, we propose that speech interaction will lead to greater enjoyment than textual interaction. We furthermore assume that this effect will be weaker if customers perceive tasks to be highly goal-directed because their focus on the particular goal leaves less space for enjoying the actual interaction (Hoffman & Novak, 1996). In contrast, for tasks perceived as less goal-directed, the interaction serves to explore different options. In this way, the benefits of speaking in a two-sided conversation should become more pronounced as the duration of the interactions increases and because of the hedonic nature of experiential tasks (Li et al., 2012; Novak et al., 2003). So, our next hypotheses are:

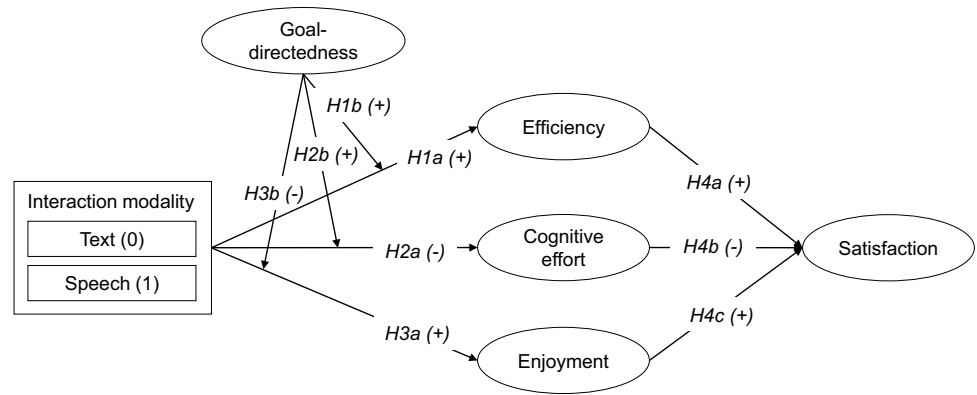
H3a: Customers perceive more enjoyment when they interact via speech, compared to interacting via text.

H3b: Goal-directedness negatively moderates the effect of interaction modality on enjoyment.

Finally, we assume that cognitive effort, efficiency, and enjoyment affect customers' satisfaction with the service outcome, i.e., the search result. Satisfaction is mostly conceptualized as a state in which an interaction meets or exceeds a customer's expectation (McKinney et al., 2002), and, therefore, is also dependent on the customers' expectations of the service encounter. Technology is used in the service encounter to provide efficient service which, in turn, should enhance satisfaction with the encounter (Larivière et al., 2017). Because customers using CAs in the service encounter have similar expectations and request fast service (Kraus et al., 2019), meeting these expectations should positively affect their satisfaction. Furthermore, initial results on customers' satisfaction determinants of voice commerce show high convenience expectations (Kraus et al., 2019), which indicates that keeping customers' cognitive effort low promotes their satisfaction regarding service encounters via CAs. This relationship has already been examined for online shopping, showing that reducing cognitive effort through electronic decision aids has a positive impact on customers' satisfaction (Bechwati & Xia, 2003). Hence, we assume that high cognitive effort negatively impacts customers' satisfaction with the search outcome. Furthermore, extant research has shown that customers' satisfaction are not only driven by extrinsic motivations, but also by intrinsic expectations (Lowry et al., 2015). Since CAs offer an enjoyable experience through their interactive conversational capabilities (Cho et al., 2019; Novielli et al., 2010), we expect that enjoyment also drives customers' satisfaction with the service encounter:

H4a: Perceived efficiency has a positive impact on customers' satisfaction with the service encounter.

Figure 1 Research model



H4b: Cognitive effort has a negative impact on customers' satisfaction with the service encounter.

H4c: Enjoyment has a positive impact on customers' satisfaction with the service encounter.

Figure 1 illustrates the hypotheses on the fit between interaction modality and task goal-directedness as regards their impact on customers' satisfaction. Following our research question and the extant literature, we hypothesize that speech differs from textual interaction in its perceived efficiency, cognitive effort, and enjoyment, thus ultimately affecting customers' satisfaction with the service encounter. However, this relationship can change depending on the search task's perceived goal-directedness.

4 Experimental Design

4.1 Experimental Procedure

To test our research model, we designed a 2 (interaction modality: speech vs. text) × 2 (search task type: goal-directed vs. experiential) between-subject laboratory experiment. Originally, we designed and conducted the experiment as a mixed design experiment with interaction modality serving as a between-subject factor and task type

as a within-subject factor. In this experiment, we randomly assigned participants to a group who would interact either via speech or via text with a specifically created CA to solve – for experimental purposes – a single goal-directed and a single experiential search task in a randomized order. We used the randomization algorithm integrated in the survey software Qualtrics, which we also used for the instructions and questionnaires of the experiment. Unfortunately, we found confounding effects caused by the experience of the first task and, thus, had to omit the second measurement from further analyses. Because we randomized the order of the tasks, we were able to report the remaining between-subject laboratory experiment based on the data of the first task only.

All interactions took place in single-person laboratory sessions to ensure controlled and private conditions for each participant. Participants selected an appointment via an online registration form. Upon arrival, a researcher provided participants with information about the general procedure of the study and the technical devices (Laptop and CA) used in the experiment. Participants could briefly familiarize themselves with the setup and ask open questions. Thereafter, the researcher left the room but remained accessible at all times should problems occur. At this stage, participants started the experiment on the laptop and were guided through the experimental process depicted in Figure 2.

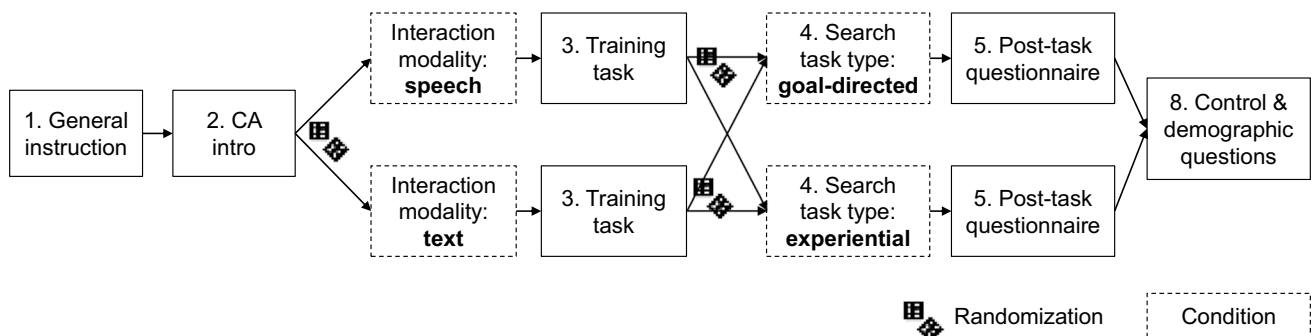


Figure 2 Experimental procedure

After giving general instructions on the experiment, we briefly defined CAs and asked participants to answer questions on their prior experience with CAs. Next, we randomly assigned participants to either the text or the speech condition and provided them with specific instructions on how to interact with the CA. The participants conducted an exemplary task to familiarize themselves with the system. When participants felt confident using the system, they went ahead with the actual task which, by chance, was either goal-directed or experiential. After the completion of the task, we administered a post-task questionnaire with manipulation and attention checks, and posed questions on subjective measures (perceived efficiency, cognitive effort, enjoyment, and satisfaction with the service encounter). The experiment concluded with demographic questions.

4.2 Experimental Treatments

To manipulate the task’s perceived goal-directedness, we adapted a restaurant search task that had already been used successfully to represent goal-directed and experiential search tasks and to compare speech and text-based interactions (Le Bigot et al., 2004; Liu et al., 2016). The goal-directed search task asked participants to search for a restaurant that, for them to have dinner with a friend, had to fulfil specific criteria (cuisine, location, price). In the experiential search task, we asked participants to freely explore restaurants according to their own preferences without indicating specific objectives or prescriptive criteria. To prevent any scenario-dependent effects, we altered the scenario slightly, assigning either a restaurant or a café search. Participants could end the interaction at any time. Table 1 indicates the conditions.

We developed a CA called Restaurant Finder based on Google’s cloud platform DialogFlow to enable both speech- and text-based interactions using the same dialogue structure and eliciting identical responses. DialogFlow is a conversational platform that uses machine learning algorithms for natural language processing and has already been successfully used in similar experiments (Diederich et al., 2019; Diederich et al., 2020). We based the implemented dialog tree on Le Bigot et al.’s (2004) structure. The CA provided

a recommendation for a restaurant or a café based on the participants’ preferences of cuisine, location, and price. Depending on the condition, either a smart speaker or a smartphone was placed next to the laptop. In the speech condition, participants interacted with the CA via the Google Nest Mini smart speaker by saying “ok google, start restaurant finder.” We implemented the DialogFlow agent using a demo integration for Google’s assistant application on the smart speaker. In the text condition, participants started the interaction through a messenger chat window on a Google Pixel smartphone by typing “ok google, start restaurant finder.” In this condition, we implemented the DialogFlow agent as a demo in a browser window on the smartphone. Both implementations were based on the same DialogFlow agent and, therefore, engaged the user in the exact same conversational flow. Table 2 presents an exemplary dialogue for each task.

4.3 Measures

In the post-task questionnaire, we measured all latent variables with scales adapted from extant literature to ensure content validity while modifying the items to fit the target context. We used 7-point Likert-type scales ranging from ‘strongly disagree’ (1) to ‘strongly agree’ (7) to measure cognitive effort (adapted from Hong et al. (2004) and Pereira (2000)), efficiency (Cho, 2004), and enjoyment (Davis et al., 1992; Venkatesh, 2000). Satisfaction was assessed on a 7-point semantic differential scale based on Bhattacharjee (2001). The construct for the task’s goal-directedness was adapted from Stawski et al. (2007) on a 7-point Likert-type scale. Furthermore, we collected information on participants’ individual background (age, gender, profession, field of study, and income), personality traits (personal innovativeness (Agarwal & Prasad, 1998), web skills (Novak et al., 2000), and extraversion (John & Srivastava, 1999)), and on prior interaction frequencies with CAs. Moreover, we assessed attention checks, control questions on the CA’s technical performance, and questions on the scenario’s realism. As attention checks, we asked participants to select the name of the CA out of four alternatives and to indicate whether they had interacted via speech or text. If

Table 1 Experimental conditions

		Interaction modality	
		Speech (Google Nest Mini)	Text (Google Pixel 3a)
Search task type	Goal-directed	Search for the cheapest Indian restaurant near the University	Search for the cheapest Indian restaurant near the University
	Experiential	Search for any restaurant without predefined criteria	Search for any restaurant without predefined criteria

Table 2 Exemplary dialogue of each search task type

Goal-directed dialogue	Experiential dialogue
<p>CA: Welcome to Restaurant Finder! I am happy to give you a restaurant recommendation. What would you like to eat?</p> <p>User: I am looking for an Indian restaurant near the University.</p> <p>CA: Ok, I am searching for Indian food near the University. How much money would you like to spend?</p> <p>User: It should be the cheapest restaurant.</p> <p>CA: Alright, I have found a restaurant matching your criteria. The cheapest Indian restaurant is Sankar's Food Corner. You can now ask for further information or close the application.</p>	<p>CA: Welcome to Restaurant Finder! I am happy to give you a restaurant recommendation. What would you like to eat?</p> <p>User: I don't know yet. What would you recommend?</p> <p>CA: I am happy to give you a recommendation! Please choose a cuisine, for example, Asian, Italian, German, Greek, or Mexican food.</p> <p>User: Hm, what about African options?</p> <p>CA: Ok, I am searching for African food. Where would you like to eat?</p> <p>User: In the city center.</p> <p>CA: Ok, I'll search in the city center. How much would you like to spend?</p> <p>User: Less than 20 euros.</p> <p>CA: Alright, I have found three restaurants matching your criteria. Makamba offers various African food specialties. You can now ask for further information or close the application.</p> <p>User: What is the other recommendation?</p> <p>CA: The second recommendation is Savanna, which offers African tapas. You can ask for further information or close the application.</p> <p>User: What are the reviews? ...</p>

a participant failed these attention checks, we excluded the data set from further analysis. As a control question, participants were asked to indicate whether they had been able to solve the task successfully. We used this question to assess the CA's performance. The question on the scenario's realism assessed how well participants were able to acquaint themselves with the described situation and task (Paschall et al., 2005). Finally, we asked for free form feedback on the interaction with the CA. The participants' comments did not reveal any serious problems or misunderstandings.

4.4 Data Collection and Sample

We conducted two pre-tests to train and evaluate our CA with objective and subjective measures (Shawar & Atwell, 2007). During the first pre-test, 16 participants interacted with the CA and noted everything that came to their mind. Based on this feedback, we revised the CA and adjusted the conversational structure for a second objective evaluation. To reduce system errors in the main laboratory experiment, the aim of the second pre-test was to train the CA for variations of user requests. We therefore distributed the Dialog-Flow demo link to researchers and students at our university to collect real conversation data from them. Based on this conversation data, we manually assigned unmatched user queries from 64 conversations. As our CA aims to provide a specific service, we further assessed whether participants were able to solve the task. This yielded satisfactory results as 96.6% of the participants were able to find a restaurant (Shawar & Atwell, 2007).

The main data collection took place in Q4 2020. To recruit students, we advertised the laboratory experiment via student mailing lists and on site at a large public university, as is frequently done in related studies (e.g., Polites and Karahanna (2012)). To ensure completion, participation was incentivized with a five-euro reward. In succession, a total of 119 participants completed the experiment. We discarded three participants from further analysis as their attention checks and interaction logs showed that they did not follow the instructions. Of the remaining 116 participants, 60.3% were female and the average age was 24 ($M=23.97$; $SD=4.10$). Most participants were students in the field of business (63.8%), followed by medicine (6.9%) and social sciences (6.9%). 80.9% of the participants that disclosed their income earned less than €1,500 per month (5.2% chose not to disclose). Nearly all participants had previously made use of speech- or text-based CAs. More than 80% of the participants used text-based CAs at least weekly; only 4.3% indicated that they had never interacted with a chatbot. Speech-based CAs were used weekly by 74% of our participants while 7.8% had never used them. Detailed information about the sample distribution is presented in Table 7 of the appendix.

5 Results

We tested our research model using partial least squares (PLS) structural equation modelling carried out with Smart-PLS3 (Ringle et al., 2015). Therefore, we determined the

significance of path coefficients by running the bootstrapping resampling approach with 5,000 subsamples (Chin, 1998). For all other analyses, we relied on IBM SPSS Statistics 26.

5.1 Control Variables and Manipulation Check

We assessed the participants’ successful randomization to the experimental groups using Fisher’s exact tests for the categorical variables and analyses of variance (ANOVAs) for the metric variables. The results showed no significant differences between the participants’ gender, field of study, income, and prior experience with both CAs ($p > .1$). Hence, the participants were distributed homogeneously across the treatment groups. We further conducted one-way ANOVAs for age, personal innovativeness, web skills, and extraversion, revealing no significant differences ($p > .1$). Thus, we can assume that group differences in demographic or control variables did not confound our treatment effects. To ensure that our experimental design did not affect the results, we further conducted ANOVAs to assess differences between the café and restaurant search scenario. Our analyses reveal no significant differences between the café and restaurant search for the mediating and dependent variables ($p > .1$). Moreover, across all groups, participants indicated that they perceived the experiment as realistic ($M = 6.17$; $SD = .97$). We further conducted a one-way ANOVA for the manipulation check, which revealed significant group differences for perceived goal-directedness of the task ($F = 17.79$, $p < .001$). Thus, we conclude that the treatment was successful. In the following analyses of the research model, we use participants’ perceptions of the task’s goal-directedness – rather than our binary treatment variable – to account for humans’ diverse individual perceptions (Lowry et al., 2013).

5.2 Measurement Validation

To ensure construct validity, we evaluated our measurement model following established validation procedures (Chin, 2010). Because our latent constructs were adapted from extant literature, we conducted a confirmatory factor analysis to validate our scales. We determined internal consistency reliability by calculating each construct’s Cronbach’s alpha (CA) and composite reliability (CR). Our constructs showed satisfactory CA and CR values above the threshold of 0.7 (see Table 3). Factor loadings should also lie above the threshold of 0.7 (Hair et al., 2011). The indicators fulfilled this requirement, as the lowest observed factor loading in our data was .805. Hence, all constructs were reliable. Each item and the factor loadings are shown in Table 5 of the appendix. Furthermore, we evaluated the validity of our constructs. Convergent validity was evaluated by assessing the average variance extracted (AVE) using the critical threshold of 0.5 (Chin, 2010), which was exceeded by all constructs. Discriminant validity was determined by assessing indicator’s cross loadings, the Fornell-Larcker criterion, and the heterotrait-monotrait (HTMT) ratio of correlations. All factor loadings of indicators exceeded cross loadings. The Fornell-Larcker criterion was met, as each construct’s square root of the AVE was greater than the interconstruct correlations (Fornell & Larcker, 1981). As all HTMT ratio values were below the most conservative threshold of 0.85 (Henseler et al., 2015), our constructs exhibited sufficient discriminant validity. To summarize (see Table 3), all constructs showed satisfactory psychometric properties.

5.3 Hypotheses Testing

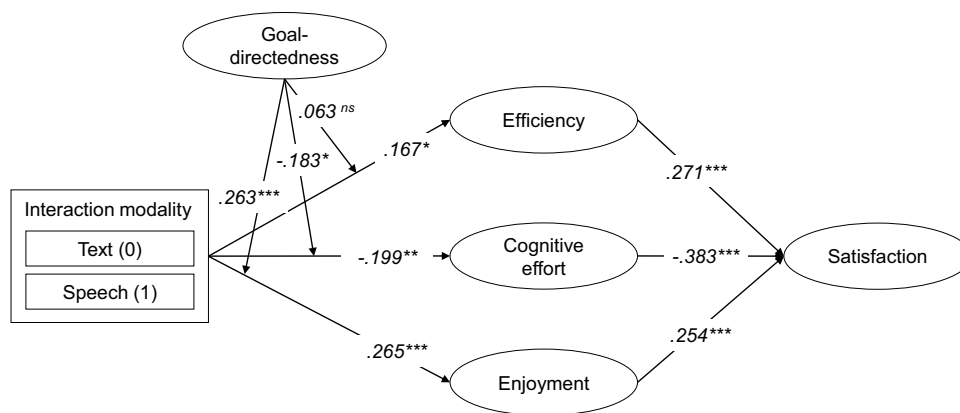
We examined our structural model for collinearity by computing inner VIF values. With a maximum inner VIF of 1.77,

Table 3 Factor loadings, internal consistency criteria, AVE, and correlation matrix

Construct	Loadings	CA	CR	AVE	(1)	(2)	(3)	(4)	(5)
(1) Efficiency	.845 - .921	.858	.914	.780	.883				
(2) Cognitive effort	.810 - .903	.888	.922	.747	-.572	.864			
(3) Enjoyment	.870 - .949	.930	.950	.827	.551	-.450	.909		
(4) Goal-directedness	.805 - .939	.917	.938	.791	-.016	-.043	.069	.889	
(5) Satisfaction	.861 - .912	.913	.939	.793	-.630	-.652	.576	.177	.890

Elements in italic represent the square root of AVE for the corresponding construct

Fig. 3 PLS estimation results (N = 116; ***p < .01, **p < .05, *p < .1)



all values were below the critical threshold of 5, confirming that collinearity is not an issue in our structural model. By assessing path coefficients and their significance, we found support for six of our nine hypotheses, as presented in Figure 3.

First, we examined the hypothesized relationships between the interaction modality, the task's perceived goal-directedness, and efficiency. We find that the interaction modality speech has a significant positive effect on efficiency ($\beta=.167, p<.1$). However, goal-directedness does not moderate the relationship between the interaction modality and efficiency ($\beta=.063, p>.1$). Thus, we find support for the corresponding hypotheses H1a, but not for H1b. Second, our results show that the interaction modality speech has a significant negative effect on participants' cognitive effort ($\beta=-.199, p<.05$), hence reducing participants' cognitive effort compared to text-based interactions. Additionally, we find a significant negative moderation effect between interaction modality and goal-directedness on cognitive effort ($\beta=-.183, p<.1$). Thus, we find support for the corresponding hypothesis H2a,

but reject H2b since we hypothesized a moderation effect in the opposite direction. Third, the interaction modality speech has a significant positive effect on enjoyment ($\beta=.265, p<.01$). This effect is strengthened by the task's goal-directedness ($\beta=.263, p<.01$), which contradicts the proposed direction in our hypothesis. Therefore, we confirm hypothesis H3a, but reject H3b. Lastly, we find empirical evidence that efficiency ($\beta=.271, p<.01$) and enjoyment ($\beta=.254, p<.01$) have a significant positive effect on participants' satisfaction with the service encounter, while cognitive effort has a significant negative effect on the latter ($\beta=-.383, p<.01$). Hence, we confirm the corresponding hypotheses H4a, H4b, and H4c. Additionally, we also find that the total effect of speech on participants' service satisfaction is significant ($\beta=.189, p<.01$). Overall, perceived efficiency, perceived cognitive effort, and perceived enjoyment explain 56,7 % of the variance in participants' satisfaction with the service encounter. The means and standard deviations of all investigated variables are shown in Table 6 of the appendix. Table 4 summarizes the hypotheses and results.

Table 4 Summary of hypotheses and results

Hypotheses		Results			
H1a	Speech modality	→	Efficiency ↑	$\beta = .167, p = .071$	Supported
H1b	Speech modality x Task goal-directedness	→	Efficiency ↑	$\beta = .063, p = .476$	Rejected
H2a	Speech modality	→	Cognitive effort ↓	$\beta = -.199, p = .021$	Supported
H2b	Speech modality x Task goal-directedness	→	Cognitive effort ↓	$\beta = -.183, p = .056$	Rejected (other direction)
H3a	Speech modality	→	Enjoyment ↑	$\beta = .265, p = .002$	Supported
H3b	Speech modality x Task goal-directedness	→	Enjoyment ↓	$\beta = .263, p = .007$	Rejected (other direction)
H4a	Efficiency	→	Satisfaction ↑	$\beta = .271, p = .007$	Supported
H4b	Cognitive effort	→	Satisfaction ↓	$\beta = -.383, p = .000$	Supported
H4c	Enjoyment	→	Satisfaction ↑	$\beta = .254, p = .002$	Supported

6 Discussion

6.1 Main Findings

Our paper's aim was to assess the impact of interacting with a CA, either via speech or via text, on customers' satisfaction with two search tasks that differ in respect of their perceived goal-directedness. For this purpose, we developed a CA that could answer, in an identical manner, either speech- or text-based queries. We evaluated our research model in a controlled laboratory experiment with 116 participants.

The results provide interesting insights into the moderating effect of the customers' perceived goal-directedness of the task on the relationship between the interaction modality and their perceptions. Regarding H2b, we assumed that the higher the goal-directedness, the stronger the effect of interacting via speech on cognitive effort. In contrast, our data provide evidence in the opposite direction: higher goal-directedness weakens the effect of speech on customers' cognitive effort. Indeed, our data shows that customers' cognitive effort from interacting via speech stays the same for both the goal-directed and experiential tasks, while their cognitive effort from interacting via text is lower for the goal-directed task. This contradicts our assumption that listening to and memorizing information become more effortful for lower goal-directedness. However, we may explain the effect with reference to the effort associated with typing text input (Le Bigot et al., 2007). Over time, this effort increases along with the number of interaction turns. Another possible explanation lies in customers' satisficing behavior (Dubiel et al., 2020). Dubiel et al. (2020) argue that customers would approve a minimally acceptable option as soon as the interaction becomes too costly in terms of cognitive load. Following this argumentation, customers will not experience higher cognitive loads in speech-based interactions if they are allowed to end the conversation at any time. However, more research is needed to better understand the joint impact and weighting of the interaction modality's input and output in a two-way conversation.

Furthermore, H3b hypothesized that the more goal-directed the task is, the weaker is the effect of interacting via speech on enjoyment. We assumed that, during the conversation, customers would enjoy interacting via speech more with increasing interaction turns. Instead, customers' enjoyment of text-based interaction increases for experiential tasks, but their enjoyment of speech-based interaction does not depend on the task. A possible explanation for this finding could be the "novelty effect" (Fryer et al., 2017; Kanda et al., 2004). As interacting via

speech is still a new phenomenon, customers may enjoy the interaction – regardless of any task characteristics – because of its mere newness. In contrast, text-based service encounters are more established and hence more dependent on the task's characteristics. These findings further imply that more research is needed on the long-term effects of VA usage as the "novelty effect" can decrease over repeated interactions, thereby increasing the importance of utilitarian benefits for continuous usage. Nevertheless, our results remain important for the use of CAs in customer service contexts, especially in contexts where customers do not interact on a regular basis (compared to using VAs as part of users' smart homes) but have a single contact point, allowing them to benefit from this "novelty effect."

Lastly, our results also support existing research and show that speech-based interaction modalities indeed mean higher perceived efficiency and enjoyment and less cognitive effort (H1a, H2a, and H3a). These findings are in line with qualitative research on the determinants of voice assistant use (Luger & Sellen, 2016; Moussawi, 2018; Rzepka, 2019). The positive effects of interacting via speech on users' perceived efficiency and cognitive effort also support our assumption that speech recognition technology has improved in recent years, enabling users to realize the benefits of natural speech – it being faster and requiring less cognitive effort than text-based interactions. The results further add to a better understanding of how speech affects customers' utilitarian and hedonic evaluations of CA-based service encounters. Extant research does not fully agree in this regard. Whereas McLean and Osei-Frimpong (2019) highlight the influence of utilitarian benefits and show that hedonic benefits are weaker predictors of VA usage, other studies stress the relevance of hedonic benefits (Pal et al., 2020; Yang & Lee, 2019). In this study, we find that speech indeed offers both utilitarian and hedonic benefits if compared to text-based interactions. Additionally, all three mediators (efficiency, cognitive effort, and enjoyment) similarly predict customers' satisfaction with the service encounter (H4a-c). Hence, we conclude that in a service context, speech offers the benefits of customer efficiency, cognitive effort, and enjoyment, which, in turn, significantly predict customers' satisfaction with the service encounter.

6.2 Theoretical Contributions and Practical Implications

Our results contribute to extant research on the applicability of CAs' conversational interaction modalities for differing tasks. First, in general, we add to existing

research on CAs and human-computer interaction and, more specifically, to VA adoption by investigating, on a comparative basis, how and why speech and text interactions differ. Specifically, we show that customers perceive speech-based interactions to be more beneficial in terms of efficiency, cognitive effort, and enjoyment, and, subsequently, in terms of service satisfaction. In this way, for both speech- and text-based interactions, we add to the existing research stream on CAs (Diederich et al., 2022; Zierau et al., 2020). We extend this stream by highlighting the distinct characteristics of both interaction modalities, which need further in-depth investigations in line with Schmitt et al. (2021) and Zierau et al. (2020). Furthermore, we provide an initial understanding of differences in speech- and text-based interactions in information searches, addressing research calls by Ma and Liu (2020) and Vakulenko et al. (2020). Our results show that speech, when compared to text-based interactions, has significant benefits that are utilitarian and hedonic in nature. So far, VAs' hedonic and utilitarian benefits have only been investigated for VA adoption in general (McLean & Osei-Frimpong, 2019; Pal et al., 2020; Zimmermann et al., 2021). We show that these benefits positively affect customers' service satisfaction. Previously, this relationship had only been investigated for text-based interactions (Diederich et al., 2019; Diederich et al., 2020), thus receiving less attention in extant VA research.

Second, the interaction effects between interaction modality and task type extend cognitive fit theory and TTF theory to the applicability of different interaction modalities. Specifically, we show that speech and text are not similarly suited to various kinds of tasks and that the tasks' perceived goal-directedness influences their effect. In this way, we extend cognitive fit theory to the study of conversational human-computer interactions, specifically speech- and text-based interactions, and their interplay in two-way communications. These communications not only include information representation by the system, but also information requests from the user. So far, extant research using cognitive fit theory has been concerned with the examination of visual information representation formats, e.g., in lists and matrices. Until now, only Chen et al. (2021) had investigated cognitive fit for text-based CAs, showing that a friendlier conversational style better fits a goal-directed task. We extend this research stream and the TTF theory by distinguishing between different interaction modalities based on their inherent, distinct technologies (VAs, compared to text-based CAs, additionally include a speech recognizer and speech-to-text synthesis). While our results do not show that text-based interaction matches any information search task better than does speech-based interaction, we believe that our results open new avenues

for future research on the moderating effects of task characteristics and on the applicability of each interaction modality.

Third, our results also inform research on the use of CAs in the service encounter. So far, prior literature has mainly focused on the application of text-based CAs (Adam et al., 2021; Schuetzler et al., 2021). By addressing De Keyser et al.'s (2019) research calls, our study sheds light on the impact that different interaction modalities have on customer perceptions and, therefore, it has implications for the development of suitable speech- or text-based CAs in the service journey. Specifically, we show that extant research on the use of text-based CAs in the customer encounter may not be equally generalizable to speech-based interactions. Rather, speech-based interactions have considerable benefits that need further exploration in their customer interaction usage. In this context, our study also enters the existing research stream investigating customers' goal-directed and experiential search behavior (Hong et al., 2004; Nadkarni & Gupta, 2007). More specifically, we not only show that the task's perceived goal-directedness affects how different website designs influence customers' behavior, but that it also influences the applicability of speech- and text-based interaction modalities.

Furthermore, we provide insights for practitioners assessing different types of CAs offered to customers for various kinds of task. This is most relevant because companies increasingly apply CAs in service encounters to reduce costs while also providing personal interactions at all times. We show that speech-based interaction modalities can be beneficial as they increase customers' satisfaction with service encounters. Hence, providers should not only offer text-based touchpoints to their customers, but also consider the provision of speech-based services. As a consequence, questions relevant to the deployment of chatbots (Schuetzler et al., 2021) need to be re-evaluated for VAs. For example, when providing product information and recommendations, providers should consider offering VA applications for their customers' smart speakers. They should also implement speech-based interaction touchpoints on their websites to increase customers' efficiency and enjoyment. We further suggest that practitioners should not refrain from using VAs for more complex tasks as our results show that the benefits of speech are present for both search tasks. Finally, our results inform service providers and offer developers intending to design VA dialogues for more complex tasks an initial understanding of the differences in speech- and text-based interactions in information searches. Specifically, system providers should not only focus on designing efficient speech-based interactions but also ensure that customers enjoy them.

6.3 Limitations and Future Research

This study is not without limitations; limitations that also open future research avenues. First, our analyses build on a comparatively small sample size because participants had to take part in the experiment one by one. While the laboratory setting offered benefits in terms of a high internal validity, future research could re-examine our research model using larger sample sizes in an online experiment. An online experiment also offers the possibility of recruiting participants globally and of considering diverse cultural backgrounds in the analyses. Apart from increasing the sample size and including user characteristics in the research model, the comparison of speech- and text-based interactions offers a wide range of new research opportunities. Most research on the use of CAs has focused on text-based interactions, which – based on our results – differ from speech-based interactions. However, we know little about the right design and the value of speech-based interactions. While social cues are well researched for text-based interactions, more research is needed – for instance – on types of cues that are beneficial for speech-based interactions, e.g., loudness, pitch, or pauses (Schmitt et al., 2021).

Second, it would be worthwhile to investigate the structure of the conversations in more detail, depending on the nature of the task (e.g., depending on whether the user is a sender or a receiver of information). In this vein, it could be interesting to examine the exact way in which the number of interaction turns affects users' interaction outcomes. Would users prefer very brief and efficient interactions that convey a large amount of information in each turn – information that they would need to convey or memorize – or would they rather interact in multiple turns, each comprising a single piece of information? Similarly, it would be interesting to investigate this relationship with different numbers of required user-confirmation steps, especially when this results in actual product purchases as outcomes of the information search behavior. Would the user prefer longer interactions if they included more user confirmations?

Third, we only compared purely speech-based interactions with purely text-based interactions. Although we deliberately opted for this comparison, as these interaction modalities are similar but constitute different CA configurations, further research projects should investigate combinations or extensions of these conditions. For example, future studies could examine

how different combinations of speech input and text output, or vice versa, affect users' satisfaction with the interactions. Future research could also compare speech interactions to website information searches or product purchasing, extending initial research by Kraus et al. (2019). Another promising research area is the comparison of speech-based, text-based, and human interactions, in particular the impact of disclosing the CA's non-human identity. While this topic has already received initial attention for text-based CA interactions (Cheng et al., 2021; Mozafari et al., 2021a, 2021b), it becomes progressively important given the increasing naturalness of synthetic speech output.

Fourth, our study focuses on a single task characteristic in a particular context. Further research should also examine the impact of the task's goal-directedness for other contexts, e.g., in organizational settings. Moreover, other task characteristics such as associated risks or the task's importance pose interesting research questions. Is there a certain degree of task importance or risk that makes text-based interactions more preferable than speech-based interactions? And would customers prefer human interactions for an even higher degree of risk or do they already prefer humans to text-based interactions? In this regard, it is essential to investigate trust in particular, being an important determinant of humans' usage of artificial intelligence-enabled systems in general and of VAs in particular (Mari & Algesheimer, 2021).

7 Conclusion

Overall, our study sheds first light on the impact of task characteristics on the benefits of speech- based interaction modalities for information search tasks. We specifically show that, in terms of perceived efficiency, cognitive effort, enjoyment, and satisfaction, speech-based interaction modalities are perceived as being superior to text-based interactions. We further show that the task's goal-directedness influences these effects. In this way, we provide initial evidence that opens new avenues for future research; research that can build on these results and that can inform practitioners about the applicability of both speech- and text-based CAs. With continuous technical advancements, companies increasingly need to decide whether a human, a VA, or a chatbot should answer customers' service queries. We hope to inspire and motivate future research in this area.

Appendix

Table 5 Constructs, items, factor loadings, and scales

Construct	Items	Loading	Scale
Efficiency (EFF)	Interacting with the [CA]...	.844	7-point Likert-type scale (Cho 2004)
	EFF1 ... takes less time for making decisions.	.921	
	EFF2 ... takes less time for investigating different options.	.882	
Cognitive effort (COG)	COG1 To complete the task, using this [CA] was very frustrating.	.896	7-point Likert-type scale (Hong et al. 2004; Pereira 2000)
	COG2 To complete the task, using this [CA] took too much time.	.845	
	COG3 To complete the task, using this [CA] required too much effort.	.903	
	COG4 To complete the task, using this [CA] was too complex.	.810	
Enjoyment (ENJ)	ENJ1 I find using the [CA] to be enjoyable.	.949	7-point Likert-type scale (Davis et al. 1992; Venkatesh 2000)
	ENJ2 The actual process of using the [CA] is pleasant.	.893	
	ENJ3 I have fun using the [CA].	.923	
	ENJ4 Using the [CA] is exciting.	.870	
Goal-directedness (GD)	GD1 I knew clearly what I wanted to do.	.906	7-point Likert-type scale (Stawski et al. 2007)
	GD2 I had a strong sense of what I wanted to do.	.939	
	GD3 I know what I wanted to achieve.	.902	
	GD4 My goals were clearly defined.	.805	
Satisfaction (SAT)	How do you feel about your overall experience of [CA] use?	.912	7-point semantic differential scale (Bhattacharjee 2001)
	SAT1 Very dissatisfied/Very pleased.	.876	
	SAT2 Very displeased/Very pleased.	.911	
	SAT3 Very frustrated/Very contented.	.862	
	SAT4 Absolutely terrible/Absolutely delighted.		

Table 6 Means (and standard deviations)

	Speech & Goal-directed	Speech & Experiential	Text & Goal-directed	Text & Experiential	All
Age	24.57 (5.57)	24.33 (3.26)	24.27 (3.99)	22.84 (3.19)	23.97 (4.10)
Personal innovativeness	4.71 (1.39)	4.64 (1.37)	4.98 (1.54)	4.96 (1.53)	4.83 (1.45)
Web skills	5.68 (1.08)	5.49 (1.15)	5.59 (1.08)	5.75 (1.22)	5.63 (1.13)
Extraversion	4.64 (1.23)	5.06 (1.16)	5.01 (1.28)	4.70 (1.25)	4.85 (1.24)
Efficiency	5.07 (1.27)	4.77 (1.64)	4.26 (1.85)	4.46 (1.82)	4.63 (1.68)
Cognitive effort	1.82 (1.01)	1.80 (1.05)	2.29 (1.43)	2.42 (1.68)	2.10 (1.35)
Enjoyment	5.12 (1.34)	5.08 (1.19)	4.03 (1.53)	4.52 (1.68)	4.67 (1.51)
Satisfaction	5.54 (.87)	5.08 (1.03)	5.04 (1.32)	4.82 (1.24)	5.11 (1.15)
Goal-directedness	6.42 (.90)	5.84 (1.21)	6.48 (.65)	5.44 (1.31)	6.04 (1.13)
N	28	27	30	31	116

Table 7 Sample distribution

Variable		Total
Gender	female	70
	male	46
	other	0
Field of study	Business and Economics	74
	Cultural Studies, Linguistics, Arts, Music, Design	2
	Engineering	5
	Math, Natural Sciences	7
	Medicine, Healthcare	8
	Social Sciences	8
	other	12
Income	<500€	31
	500-999€	43
	1,000-1,499€	17
	1,500-1,999€	4
	2,000-2,499€	9
	2,500-2,999€	4
	3,000-3,499€	1
	3,500-3,999€	1
	no disclosure	6
Use frequency speech-based CA	daily	30
	weekly	56
	monthly	14
	less than monthly	7
Use frequency text-based CA	never	9
	daily	22
	weekly	74
	monthly	9
	less than monthly	6
	never	5

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of Interest The authors have no conflicting interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Adam, M., Wessel, M., & Benlian, A. (2021). AI-Based Chatbots in Customer Service and Their Effects on User Compliance. *Electronic Markets*, 31, 427–445. <https://doi.org/10.1007/s12525-020-00414-7>

Agarwal, R., & Prasad, J. (1998). A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. *Information Systems Research*, 9(2), 204–215. <https://doi.org/10.1287/isre.9.2.204>

Akinnaso, F. N. (1982). On the Differences between Spoken and Written Language. *Language and Speech*, 25(2), 97–125. <https://doi.org/10.1177/002383098202500201>

Balakrishnan, J., Dwivedi, Y. K., Hughes, L., & Boy, F. (2021). Enablers and Inhibitors of AI-Powered Voice Assistants: A Dual-Factor Approach by Integrating the Status Quo Bias and Technology Acceptance Model. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10203-y>

Bartl, C., Gouthier, M. H., & Lenker, M. (2013). Delighting Consumers Click by Click: Antecedents and Effects of Delight Online. *Journal of Service Research*, 16(3), 386–399. <https://doi.org/10.1177/1094670513479168>

Bechwati, N. N., & Xia, L. (2003). Do Computers Sweat? The Impact of Perceived Effort of Online Decision Aids on Consumers' Satisfaction with the Decision Process. *Journal of Consumer Psychology*, 13(1-2), 139–148. https://doi.org/10.1207/S15327663JCP13-1&2_12

Behera, R. K., Bala, P. K., & Ray, A. (2021). Cognitive Chatbot for Personalised Contextual Customer Service: Behind the Scene and Beyond the Hype. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10168-y>

Benlian, A., Klumpe, J., & Hinz, O. (2020). Mitigating the Intrusive Effects of Smart Home Assistants by Using Anthropomorphic Design Features: A Multi-Method Investigation. *Information Systems Journal*, 30(6), 1010–1042. <https://doi.org/10.1111/isj.12243>

Bhattacharjee, A. (2001). Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Quarterly*, 25(3), 351–370. <https://doi.org/10.2307/3250921>

Brasel, S. A., & Gips, J. (2014). Tablets, Touchscreens, and Touchpads: How Varying Touch Interfaces Trigger Psychological Ownership. *Journal of Consumer Psychology*, 24(2), 226–233. <https://doi.org/10.1016/j.jcps.2013.10.003>

Chen, J., Le, H., & Tran, S. (2021). Understanding Automated Conversational Agent as a Decision Aid: Matching Agent's Conversation with Customer's Shopping Task. *Internet Research, ahead-of-print*. <https://doi.org/10.1108/INTR-11-2019-0447>

Cheng, X., Bao, Y., Zarifis, A., Gong, W., & Mou, J. (2021). Exploring Consumers' Response to Text-Based Chatbots in E-Commerce: The Moderating Role of Task Complexity and Chatbot Disclosure. *Internet Research, ahead-of-print*. <https://doi.org/10.1108/INTR-08-2020-0460>

Childers, T. L., Carr, C. L., Peck, J., & Carson, S. (2001). Hedonic and Utilitarian Motivations for Online Retail Shopping Behavior. *Journal of Retailing*, 77(4), 511–535. [https://doi.org/10.1016/S0022-4359\(01\)00056-2](https://doi.org/10.1016/S0022-4359(01)00056-2)

Chin, W. W. (1998). The Partial Least Squares Approach to Structural Equation Modeling. *Modern Methods for Business Research*, 295(2), 295–336.

Chin, W. W. (2010). How to Write up and Report PLS Analyses. In *Handbook of Partial Least Squares* (pp. 655–690). Springer. https://doi.org/10.1007/978-3-540-32827-8_29

Chiu, C. M., Wang, E. T., Fang, Y. H., & Huang, H. Y. (2014). Understanding Customers' Repeat Purchase Intentions in B2c E-Commerce: The Roles of Utilitarian Value, Hedonic Value and Perceived Risk. *Information Systems Journal*, 24(1), 85–114.

- Cho, E., Molina, M. D., & Wang, J. (2019). The Effects of Modality, Device, and Task Differences on Perceived Human Likeness of Voice-Activated Virtual Assistants. *Cyberpsychology, Behavior, and Social Networking*, 22(8), 515–520. <https://doi.org/10.1089/cyber.2018.0571>
- Cho, J. (2004). Likelihood to Abort an Online Transaction: Influences from Cognitive Evaluations, Attitudes, and Behavioral Variables. *Information & Management*, 41(7), 827–838. <https://doi.org/10.1016/j.im.2003.08.013>
- Dai, H., & Salam, A. F. (2020). An Empirical Assessment of Service Quality, Service Consumption Experience and Relational Exchange in Electronic Mediated Environment (EME). *Information Systems Frontiers*, 22(4), 843–862. <https://doi.org/10.1007/s10796-019-09894-1>
- Dang, Y., Zhang, Y., Brown, S. A., & Chen, H. (2020). Examining the Impacts of Mental Workload and Task-Technology Fit on User Acceptance of the Social Media Search System. *Information Systems Frontiers*, 22(3), 697–718. <https://doi.org/10.1007/s10796-018-9879-y>
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. *Journal of Applied Social Psychology*, 22(14), 1111–1132. <https://doi.org/10.1111/j.1559-1816.1992.tb00945.x>
- De Keyser, A., Köcher, S., Alkire, L., Verbeeck, C., & Kandampully, J. (2019). Frontline Service Technology Infusion: Conceptual Archetypes and Future Research Directions. *Journal of Service Management*, 30(1), 156–183. <https://doi.org/10.1108/JOSM-03-2018-0082>
- 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.
- Diederich, S., Janssen-Müller, M., Brendel, A. B., & Morana, S. (2019). Emulating Empathetic Behavior in Online Service Encounters with Sentiment-Adaptive Responses. Proceedings of the 40th International Conference on Information Systems, Munich, Germany.
- Diederich, S., Lembcke, T.-B., Brendel, A. B., & Kolbe, L. M. (2020). Not Human after All: Exploring the Impact of Response Failure on User Perception of Anthropomorphic Conversational Service Agents. Proceedings of the 28th European Conference on Information Systems, A Virtual AIS Conference.
- Dubiel, M., Halvey, M., Azzopardi, L., Anderson, D., & Daronnat, S. (2020). Conversational Strategies: Impact on Search Performance in a Goal-Oriented Task. Proceedings of the 3rd International Workshop on Conversational Approaches to Information Retrieval, Vancouver, Canada.
- Dubiel, M., Halvey, M., Azzopardi, L., & Daronnat, S. (2018). Investigating How Conversational Search Agents Affect User's Behaviour, Performance and Search Experience. Proceedings of the 2nd International Workshop on Conversational Approaches to Information Retrieval, Ann Arbor, USA.
- Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial Intelligence and Business Value: A Literature Review. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10186-w>
- Foley, M., Casiez, G., & Vogel, D. (2020). Comparing Smartphone Speech Recognition and Touchscreen Typing for Composition and Transcription. Proceedings of the CHI Conference on Human Factors in Computing Systems, New York, USA.
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.1177/002224378101800104>
- Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and Sustaining Interest in a Language Course: An Experimental Comparison of Chatbot and Human Task Partners. *Computers in Human Behavior*, 75, 461–468. <https://doi.org/10.1016/j.chb.2017.05.045>
- Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service. Proceedings of the 38th International Conference on Information Systems, Seoul, South Korea.
- Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly*, 19(2), 213–236. <https://doi.org/10.2307/249689>
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-Sem: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, 19(2), 139–152. <https://doi.org/10.2753/MTP1069-6679190202>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., & Sainath, T. N. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, 29(6), 82–97. <https://doi.org/10.1109/MSP.2012.2205597>
- Hirschberg, J., & Manning, C. D. (2015). Advances in Natural Language Processing. *Science*, 349(6245), 261–266. <https://doi.org/10.1126/science.aaa8685>
- 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/10.1177/002242999606000304>
- Hong, W., Thong, J. Y., & Tam, K. Y. (2004). The Effects of Information Format and Shopping Task on Consumers' Online Shopping Behavior: A Cognitive Fit Perspective. *Journal of Management Information Systems*, 21(3), 149–184. <https://doi.org/10.1080/07421222.2004.11045812>
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37(1), 81–88. <https://doi.org/10.1080/02763869.2018.1404391>
- John, O. P., & Srivastava, S. (1999). *The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives* (Vol. 2). University of California, Berkeley, USA.
- Jones, M. A., Reynolds, K. E., & Arnold, M. J. (2006). Hedonic and Utilitarian Shopping Value: Investigating Differential Effects on Retail Outcomes. *Journal of Business Research*, 59(9), 974–981. <https://doi.org/10.1016/j.jbusres.2006.03.006>
- Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human-Computer Interaction*, 19(1-2), 61–84. https://doi.org/10.1207/s15327051hci1901&2_4
- 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>
- Kock, N. (2004). The Psychobiological Model: Towards a New Theory of Computer-Mediated Communication Based on Darwinian Evolution. *Organization Science*, 15(3), 327–348. <https://doi.org/10.1287/orsc.1040.0071>
- Krämer, N. C., Bente, G., Eschenburg, F., & Troitzsch, H. (2009). Embodied Conversational Agents: Research Prospects for Social Psychology. *Social Psychology*, 40(1), 26–36. <https://doi.org/10.1027/1864-9335.40.1.26>
- Kraus, D., Reibenspiess, V., & Eckhardt, A. (2019). How Voice Can Change Customer Satisfaction: A Comparative Analysis between E-Commerce and Voice Commerce. Proceedings of the 14th International Conference on Wirtschaftsinformatik, Siegen, Germany.

- Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., Wunderlich, N. V., & De Keyser, A. (2017). “Service Encounter 2.0”: An Investigation into the Roles of Technology, Employees and Customers. *Journal of Business Research*, 79, 238–246. <https://doi.org/10.1016/j.jbusres.2017.03.008>
- Le Bigot, L., Jamet, E., & Rouet, J.-F. (2004). Searching Information with a Natural Language Dialogue System: A Comparison of Spoken Vs. *Written Modalities*. *Applied Ergonomics*, 35(6), 557–564. <https://doi.org/10.1016/j.apergo.2004.06.001>
- Le Bigot, L., Rouet, J.-F., & Jamet, E. (2007). Effects of Speech-and Text-Based Interaction Modes in Natural Language Human-Computer Dialogue. *Human Factors*, 49(6), 1045–1053. <https://doi.org/10.1518/001872007X249901>
- Lee, H.-K., Suh, K.-S., & Benbasat, I. (2001). Effects of Task-Modality Fit on User Performance. *Decision Support Systems*, 32(1), 27–40. [https://doi.org/10.1016/S0167-9236\(01\)00098-7](https://doi.org/10.1016/S0167-9236(01)00098-7)
- Li, J. (2015). The Benefit of Being Physically Present: A Survey of Experimental Works Comparing Copresent Robots, Telepresent Robots and Virtual Agents. *International Journal of Human-Computer Studies*, 77, 23–37. <https://doi.org/10.1016/j.ijhcs.2015.01.001>
- Li, M., Dong, Z. Y., & Chen, X. (2012). Factors Influencing Consumption Experience of Mobile Commerce. *Internet Research*, 22(2), 120–141. <https://doi.org/10.1108/10662241211214539>
- Liu, F., Xiao, B., Lim, E. T., & Tan, C.-W. (2016). Is My Effort Worth It? Investigating the Dual Effects of Search Cost on Search Utility. Proceedings of the 20th Pacific Asia Conference on Information Systems, Chiayi, Taiwan.
- Lowry, P. B., Gaskin, J., & Moody, G. D. (2015). Proposing the Multi-Motive Information Systems Continuance Model (MISC) to Better Explain End-User System Evaluations and Continuance Intentions. *Journal of the Association for Information Systems*, 16(7), 515–579.
- Lowry, P. B., Gaskin, J., Twyman, N., Hammer, B., & Roberts, T. (2013). Taking ‘Fun and Games’ Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM). *Journal of the Association for Information Systems*, 14(11), 617–671.
- Lu, H.-P., & Yang, Y.-W. (2014). Toward an Understanding of the Behavioral Intention to Use a Social Networking Site: An Extension of Task-Technology Fit to Social-Technology Fit. *Computers in Human Behavior*, 34, 323–332. <https://doi.org/10.1016/j.chb.2013.10.020>
- Luger, E., & Sellen, A. (2016). “Like Having a Really Bad PA” the Gulf between User Expectation and Experience of Conversational Agents. Proceedings of the CHI Conference on Human Factors in Computing Systems, San Jose, USA.
- Ma, X., & Liu, A. (2020). Challenges in Supporting Exploratory Search through Voice Assistants. Proceedings of the CHI Conference on Human Factors in Computing Systems, Hawaii, USA.
- Mari, A., & Algesheimer, R. (2021). The Role of Trusting Beliefs in Voice Assistants During Voice Shopping. Proceedings of the 54th Hawaii International Conference on System Sciences, Hawaii, USA.
- McKinney, V., Yoon, K., & Zahedi, F. M. (2002). The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach. *Information Systems Research*, 13(3), 296–315. <https://doi.org/10.1287/isre.13.3.296.76>
- McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa... Examine the Variables Influencing the Use of Artificial Intelligent in-Home Voice Assistants. *Computers in Human Behavior*, 99, 28–37. <https://doi.org/10.1016/j.chb.2019.05.009>
- Mero, J. (2018). The Effects of Two-Way Communication and Chat Service Usage on Consumer Attitudes in the E-Commerce Retailing Sector. *Electronic Markets*, 28(2), 205–217. <https://doi.org/10.1007/s12525-017-0281-2>
- Moon, Y. (2000). Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers. *Journal of Consumer Research*, 26(4), 323–339. <https://doi.org/10.1086/209566>
- Moussawi, S. (2018). User Experiences with Personal Intelligent Agents: A Sensory, Physical, Functional and Cognitive Affordances View. Proceedings of the ACM SIGMIS Conference on Computers and People Research, New York, USA.
- Mozafari, N., Weiger, W., & Hammerschmidt, M. (2021a). Resolving the Chatbot Disclosure Dilemma: Leveraging Selective Self-Presentation to Mitigate the Negative Effect of Chatbot Disclosure. Proceedings of the 54th Hawaii International Conference on System Sciences, Hawaii, USA.
- Mozafari, N., Weiger, W., & Hammerschmidt, M. (2021b). Trust Me, I’m a Bot - Repercussions of Chatbot Disclosure in Different Service Frontline Settings. *Journal of Service Management*, 33, 1–33. <https://doi.org/10.1108/JOSM-10-2020-0380>
- Nadkarni, S., & Gupta, R. (2007). A Task-Based Model of Perceived Website Complexity. *MIS Quarterly*, 31(3), 501–524. <https://doi.org/10.2307/25148805>
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Novak, T. P., Hoffman, D. L., & Duhachek, A. (2003). The Influence of Goal-Directed and Experiential Activities on Online Flow Experiences. *Journal of Consumer Psychology*, 13(1-2), 3–16. https://doi.org/10.1207/S15327663JCP13-1&2_13
- Novak, T. P., Hoffman, D. L., & Yung, Y.-F. (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>
- Novielli, N., de Rosis, F., & Mazzotta, I. (2010). User Attitude Towards an Embodied Conversational Agent: Effects of the Interaction Mode. *Journal of Pragmatics*, 42(9), 2385–2397. <https://doi.org/10.1016/j.pragma.2009.12.016>
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, 32(1/2), 1–8. <https://doi.org/10.1023/B:TRUC.0000021806.17516.d0>
- Pal, D., Arpanikanondt, C., Funilkul, S., & Chutimaskul, W. (2020). The Adoption Analysis of Voice-Based Smart Iot Products. *IEEE Internet of Things Journal*, 7(11), 10852–10867. <https://doi.org/10.1109/JIOT.2020.2991791>
- Paschall, M. J., Fishbein, D. H., Hubal, R. C., & Eldreth, D. (2005). Psychometric Properties of Virtual Reality Vignette Performance Measures: A Novel Approach for Assessing Adolescents’ Social Competency Skills. *Health Education Research*, 20(1), 61–70. <https://doi.org/10.1093/her/cyg103>
- Pereira, R. E. (2000). Optimizing Human-Computer Interaction for the Electronic Commerce Environment. *Journal of Electronic Commerce Research*, 1(1), 23–44.
- Polites, G. L., & Karahanna, E. (2012). Shackled to the Status Quo: The Inhibiting Effects of Incumbent System Habit, Switching Costs, and Inertia on New System Acceptance. *MIS Quarterly*, 36(1), 21–42. <https://doi.org/10.2307/41410404>
- Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. Boenningstedt. *SmartPLS GmbH*.
- Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. A. (2017). Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4), 1–23. <https://doi.org/10.1145/3161187>
- 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. (2019). Examining the Use of Voice Assistants: A Value-Focused Thinking Approach. Proceedings of the 25th Americas Conference on Information Systems, Cancun, Mexico.
- Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2021). Voice as a Contemporary Frontier of Interaction Design. Proceedings of the 29th European Conference on Information Systems, A Virtual AIS Conference.
- Schroeder, J., & Schroeder, M. (2018). Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines. Proceedings of the 51st Hawaii International Conference on System Sciences, Hawaii, USA.
- Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Rosser, H. K. (2021). Deciding Whether and How to Deploy Chatbots. *MIS Quarterly. Executive*, 20(1), 1–15. <https://doi.org/10.17705/2msqe.00039>
- Shawar, B. A., & Atwell, E. (2007). Different Measurement Metrics to Evaluate a Chatbot System. Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies, Rochester, USA.
- Son, Y., & Oh, W. (2018). Alexa, Buy Me a Movie!: How AI Speakers Reshape Digital Content Consumption and Preference. Proceedings of the 39th International Conference on Information Systems, San Francisco, USA.
- Stawski, R. S., Hershey, D. A., & Jacobs-Lawson, J. M. (2007). Goal Clarity and Financial Planning Activities as Determinants of Retirement Savings Contributions. *The International Journal of Aging and Human Development*, 64(1), 13–32. <https://doi.org/10.2190/13GK-5H72-H324-16P2>
- Vakulenko, S., Savenkov, V., & de Rijke, M. (2020). Conversational Browsing. *arXiv preprint arXiv:2012.03704*.
- Venkatesh, V. (2000). Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation and Emotion into the Technology Acceptance Model. *Information Systems Research*, 11(4), 342–365. <https://doi.org/10.1287/isre.11.4.342.11872>
- Vessey, I., & Galletta, D. (1991). Cognitive Fit: An Empirical Study of Information Acquisition. *Information Systems Research*, 2(1), 63–84. <https://doi.org/10.1287/isre.2.1.63>
- Wells, J. D., Sarker, S., Urbaczewski, A., & Sarker, S. (2003). Studying Customer Evaluations of Electronic Commerce Applications: A Review and Adaptation of the Task-Technology Fit Perspective. Proceedings of the 36th Hawaii International Conference on System Sciences, Hawaii, USA.
- Woodford, S. (2020). *Why Chatbots Are Essential to Retail*. <https://www.juniperresearch.com/document-library/white-papers/why-chatbots-are-essential-to-retail>
- Yang, H., & Lee, H. (2019). Understanding User Behavior of Virtual Personal Assistant Devices. *Information Systems and e-Business Management*, 17(1), 65–87. <https://doi.org/10.1007/s10257-018-0375-1>
- Yang, Z., Sun, J., Zhang, Y., & Wang, Y. (2018). Peas and Carrots Just Because They Are Green? Operational Fit between Green Supply Chain Management and Green Information System. *Information Systems Frontiers*, 20(3), 627–645. <https://doi.org/10.1007/s10796-016-9698-y>
- You, J.-J., Jong, D., & Wiangin, U. (2020). Consumers' Purchase Intention of Organic Food Via Social Media: The Perspectives of Task-Technology Fit and Post-Acceptance Model. *Frontiers in Psychology*, 11(579274). <https://doi.org/10.3389/fpsyg.2020.579274>
- Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020). A Review of the Empirical Literature on Conversational Agents and Future Research Directions. Proceedings of the 41st International Conference on Information Systems, A Virtual AIS Conference.
- Zimmermann, S. K., Wagner, H.-T., Rössler, P., Gewalt, H., & Krcmar, H. (2021). The Role of Utilitarian Vs. Hedonic Factors for the Adoption of AI-Based Smart Speakers. Proceedings of the 27th Americas Conference on Information Systems, Montreal, Canada.

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

Christine Rzepka is a Ph.D. candidate and research assistant at the Institute for Digital Management and New Media at Ludwig-Maximilians-Universität (LMU), Munich, Germany. Her doctoral thesis addresses benefits and costs of voice interaction and its value for human-computer interaction. She received an M.Sc. in business administration from LMU. Her research has been published in the proceedings of the International Conference on Information Systems and other international conference proceedings in the field of information systems.

Benedikt Berger is an assistant professor of Digital Transformation and Society at the Department of Information Systems at the University of Münster, Germany. He holds a Ph.D. in management from Ludwig-Maximilians-Universität (LMU), Munich, Germany and an M.Sc. in management from the University of Mannheim, Germany. He was a visiting scholar at the Department of Information Technologies at HEC Montréal, Canada. His current research focuses on digital products and services as well as on AI-based information systems. His work has appeared in *Journal of Management Information Systems*, *Electronic Markets*, *Business and Information Systems Engineering*, and in various international conference proceedings.

Thomas Hess is a professor of Information Systems and Management at Ludwig-Maximilians-Universität (LMU), Munich, Germany, where he also serves as director of the Institute for Digital Management and New Media. He furthermore is director of the Bavarian Research Institute for Digital Transformation and member of the board of the Internet Business Cluster Munich. Thomas holds a Ph.D. in management from the University of St. Gallen, Switzerland. His research focuses on the digital transformation of companies and the management of media, internet, and IT companies. His work has appeared in journals such as *Journal of Management Information Systems*, *European Journal of Information Systems*, *Electronic Markets*, *Business and Information Systems Engineering*, *Communications of the ACM*, *Decision Support Systems*, *Information Systems Journal*, *International*, and others.