

Why Another Customer Channel? Consumers' Perceived Benefits and Costs of Voice Commerce

Christine Rzepka
LMU Munich
rzepka@bwl.lmu.de

Benedikt Berger
LMU Munich
benedikt.berger@bwl.lmu.de

Thomas Hess
LMU Munich
thess@bwl.lmu.de

Abstract

Owing to rapidly increasing adoption rates of voice assistants (VAs), integrating voice commerce as a new customer channel is among the top objectives of businesses' current voice initiatives. However, customers are reluctant to use their VAs for shopping; a tendency not explained by extant literature. Therefore, this research aims to understand consumers' perceived benefits and costs when using voice commerce, based on a theoretical framework derived from prior literature and the theory of reasoned action. We evaluated and extended this framework by analyzing 30 semi-structured interviews with smart speaker users. According to our results voice commerce consumers perceive benefits in terms of efficiency, convenience, and enjoyment, and criticize the perceived costs of limited transparency, lack of trust, lack of control, and low technical maturity. The resulting model sheds light on the promoters and inhibitors of voice commerce and provides guidelines that enable practitioners to design and improve voice commerce applications.

1. Introduction

With recent technological advances in natural language processing and speech recognition, voice assistants (VAs) opened up a new customer channel [1]. Outlets such as Starbucks, Dunkin' Donuts, and Domino's already offer VA skills to place orders via voice. Urgent.ly enables stranded motorists to call and pay for roadside assistance through Amazon's Alexa. Voice commerce, that is enabling customer purchases through the application and integration of VAs, is currently among the top objectives of businesses' voice initiatives [2]. The growing adoption rate of VAs, whether in smartphones or smart speakers [3], presents companies the opportunity to reach an increasing number of consumers by way of this new channel.

However, it is reported that only 15 percent of smart speaker users in the U.S. make regular purchases by voice [4]. In order to maximize the potential of voice commerce and to leverage voice as a new customer channel, e-commerce providers need to offer services that optimize VAs' advantages and provide superior experiences compared to existing customer interfaces. Therefore, we aim to examine the benefits and costs that VA users expect and obtain from voice commerce.

Extant research on both VAs and e-commerce insufficiently explains voice commerce's benefits and costs. Literature on the use of VAs suggests that speech interaction has characteristics that encourage a positive user experience. A system's spoken output increases consumers' perceived anthropomorphism and generates a positive attitude toward the system [5]. Simultaneously, the spoken input of consumers reduces perceived mental workloads [6] and stimulates more intuitive behavior [7]. Owing to these characteristics, speech interaction based on VAs is often regarded as convenient, efficient, and enjoyable [8, 9]. In contrast, VA interaction also incurs costs for users. The major shortcomings of VAs are speech recognition errors and privacy issues [8, 10]. However, most studies of VA focus on simple tasks, thus neglecting the effect of higher task complexity. Because the buying process involves multiple alternative solutions and financial risks, it is characterized by high complexity. We therefore lack an understanding of customers' benefits and costs for voice commerce.

In addition, we know from e-commerce literature that customers' buying decisions indeed differ according to varying interaction modalities [11]. For example, verbal preference expressions activate impulsive behaviors and erode self-restraint [12]. Similarly, voice input reduces search costs and increases convenience, leading to more impulsive buys and less self-control when purchasing and consuming digital content [13]. However, voice commerce research still lacks a holistic and theoretically sound evaluation of relevant factors regarding consumers' adoption intention. To bridge this research gap, we

provide a comprehensive overview of users' expected benefits and costs when using VAs during the purchasing process. By integrating existing literature on both VAs and e-commerce, we extend initial explorative work [14] and address recent research calls to investigate customer preferences for VAs [15]. Therefore, we pose the following research question:

Which benefits and costs do consumers evaluate when deciding to use a VA to purchase a product?

To answer this research question, we conducted 30 semi-structured interviews with VA users and analyzed their responses by means of qualitative data coding techniques [16]. This approach, involving a continuous review of the resulting codes in accordance with extant literature, ensures that we capture all relevant determinants of VA users' perceptions of voice commerce. The results could be useful to researchers who intend developing appropriate quantitative models for voice commerce adoption and use, as well as to practitioners who improve existing and design new applications that satisfy customers' needs.

This paper is structured as follows: First, in order to develop a theoretical framework for our interview procedure and analysis, we review relevant literature on the adoption and use of VAs and e-commerce. Second, we present the interview findings and indicate the benefits and costs that consumers evaluate when deciding to use voice commerce. Finally, we discuss our key results and their implications, and propose new directions for future research.

2. Theoretical background

2.1. Voice assistants

While first attempts to build systems capable of communicating in a natural way date back to the 1950s [17], recent technological advancements enabled the diffusion of VAs in private and organizational contexts. Due to better computing power, data availability, and machine learning methods, which significantly improved the performance of speech recognition and natural language processing [18], VAs can now understand spoken commands and respond via synthesized voices in order to fulfill certain tasks [19]. Therefore, as a minimum, VAs consist of a speech recognizer, a dialogue manager, and a text-to-speech synthesizer [20]. The speech recognizer records spoken words and converts them into text. Then, a dialogue manager interprets the requested action and conducts the requested task based on cloud architecture. Subsequently, the system converts its answer into speech by text-to-speech synthesis. Apple launched its first successful VA, Siri, in 2011. This was followed by several VAs such as Amazon's Alexa and Google's

Assistant. Currently, VAs are integrated into more and more devices. Not only did existing devices (e.g. smartphones) receive an additional voice input channel, and thus have multimodal input and output capabilities, but new system types also emerged that exclusively build on spoken interaction (e.g. smart speakers). While most smartphone users are therefore able to use multimodal speech input on their phones, the rapidly increasing adoption rate of smart speakers affirms consumer interest in solely spoken interactions. Smart speakers enable the user to carry out various tasks, ranging from information retrieval, through smart home control, to voice commerce.

Although different research disciplines investigate phenomena related to human speech, human-machine communication, and the use of VAs, research in the information systems (IS) domain is still at an early stage. Initial explorative studies focused on users' adoption intentions, related expectations, and actual use experiences (e.g. [8, 9, 21]). The major benefits that users expect to derive from the use of VAs are gains in efficiency, convenience, and enjoyment [8, 9]. Established models that explain technology acceptance determinants, for example TAM [22] and UTAUT2 [23], are still able to account for the intention to adopt this natural way of interacting with machines. However, extant research also suggests that additional factors should be considered when evaluating the use behavior of consumers, for example their privacy concerns [10] and trust beliefs [21]. For example, Easwara Moorthy and Vu [10] conclude that users are reluctant to share private information via voice, especially in public locations, compared with keyboard entry methods. Nasirian et al. [21] empirically show that VA interaction quality affects user trust, which in turn positively impacts on the intention to use a VA.

Differences in the relevance of established constructs and theoretical models can be explained by the human-like behavioral characteristics of VAs. Through their ability to understand spoken input and to answer in a spoken manner, VAs can engage in human-like conversations with their users, thereby establishing a sense of anthropomorphism, that is the attribution of human-like characteristics to a VA by the user [24]. In respect of VAs, these human-like characteristics may even attenuate the negative effects of privacy invasions in the smart home context [25]. While the system's spoken output and resulting anthropomorphism change consumers' perceptions, the users' ability to interact in a spoken way also affects their overall interaction experience. Compared to text input, speech is more intuitive [7] and requires less mental workload [6]. As a result, spoken interactions are more personal, they foster warmer user attitudes [26], and they encourage users to trust machines with

more personal information [27]. In contrast, especially for high complexity tasks that may require user confirmation and control, text is preferred to voice [8].

These findings suggest that VAs, and spoken interaction as a differentiating characteristic in particular, differ from conventional technologies and interaction modes. However, thus far research on voice interaction focused on simple information search tasks or basic assistant functions. There still is a lack of research on the use of VAs for complex tasks. The purchasing process presents such a complex task as it involves multiple alternatives to achieve the goal [28]. As our current knowledge may not apply to this kind of task, we seek to identify the benefits and costs that VA users ascribe to voice commerce and to shed light on the motivation of users to purchase products via voice.

2.2. Electronic and voice commerce

E-commerce is broadly defined as the transaction of information and products from vendors to customers via the internet [29]. While consumers and vendors can interact online at various stages of the customer journey, the purchase of a product is the most studied consumer behavior in this context [30]. Therefore, this research project focuses on the customer’s decision making process that concludes with the actual purchase. Since the rise of e-commerce, technological advancements have changed conventional e-commerce. These developments included the emergence of new interaction modes and touch points via the internet; advancements that also changed the customer’s experience. For example, mobile commerce added the benefits of personalization, flexibility, and localization through wireless devices [31]. More recently, conversational commerce extended the functionality of these devices by adding new interaction modes such as messenger apps, chatbots, and VAs.

Table 1. Prior research on VAs in e-commerce

Ref.	Main result
[32]	Attractive use cases of voice commerce include “request delivery status” and “find product”; however, most customers remain indifferent.
[33]	Interactivity through VAs increases the effectiveness of advertisements.
[12]	Speaking encourages more indulgent choices, compared to manual preference expression.
[34]	Convenience has a larger impact on satisfaction in voice commerce than in e-commerce.
[35]	Information search via speech is less efficient but yields less mental workload compared to text.
[13]	VA use increases hedonic consumption but decreases content completion.

As a part of conversational commerce, voice commerce refers to the transaction of services between consumers and vendors via VAs [32]. Hence, VAs enable consumers to interact in a spoken manner with the online vendor when purchasing a product. In respect of voice commerce, there is a scarcity of empirical research on consumer behavior (see Table 1).

While the sources in Table 1 focus on narrow aspects and the implications of voice commerce, a holistic view that integrates both theory and empirical data is absent. Tuzovic and Paluch [14] provide a starting point by exploratively investigating consumers’ perceptions associated with conversational commerce. Building on this research, we aim to deepen our understanding of the impact of VAs on the customer’s purchase experience. This is of particular importance since extant research shows that new interaction modes are likely to affect consumer behavior and experience [11]. For example, Brasel and Gips [11] conclude that touch-based devices can enhance product valuations compared to mouse-driven desktop computers. They point out that the interface changes consumers’ perceptions of products and marketing activities made online, and stress that research on interfaces could be as important as research on the content itself. In a similar vein, Shen et al. [36] identify a “direct-touch” effect, namely the preference of consumers for an affect-laden alternative over a cognitively superior one, which originates in the enhanced mental simulation of interacting with the more affective choice alternative on touch interfaces.

This effect implies that prior findings on other input modalities, such as text-based chatbots, are insufficient to understand users’ evaluations of voice commerce. As we know, speech differs from text in the mental production, transmission, and reception processes of consumers [7]. The main values that users evaluate when deciding whether they should purchase a product online or in a conventional store are their expected maximization of convenience and minimization of time, effort, and costs associated with the purchase experience [29]. We still need to determine whether and how these values can be achieved for voice commerce. Indeed, issues regarding trust, privacy, and anthropomorphism known from extant literature on VAs are not new to e-commerce (see e.g. [5]). A better understanding of their role in and impact on voice commerce could improve VA applications and adequately address customers’ concerns.

2.3. Theoretical framework

Drawing on the theory of reasoned action (TRA), we assume that consumers’ decisions to purchase a product via voice are determined by their intention to

purchase the product via voice. This theory postulates that the decisions of individuals on a certain behavior depend on their intention to engage in this behavior. This intention, in turn, is determined by the individuals' attitudes (as a result of salient beliefs) and subjective norms (i.e. normative beliefs) associated with this behavior [37, 38]. As this theory provides the foundation for various technology acceptance models (e.g. TAM [22], UTAUT2 [23]), we believe that it provides a suitable theoretical basis for users' cost-benefit evaluations regarding voice commerce. We further argue, in line with Benlian and Hess [39], that consumers' perceived benefits and costs are one way to represent the salient beliefs that determine their attitudes, intentions, and actions. Hence, the positive beliefs of consumers about voice commerce enhance their perceived benefits, while negative beliefs translate into perceived costs.

From extant research we know that the positive beliefs and perceived benefits related to VA use are efficiency, convenience, and enjoyment [8, 9]. Research on VA characteristics further suggests that consumers may anthropomorphize VAs and form positive beliefs about their human-likeness [5, 24]. To further account for the positive impact of anthropomorphism on consumers' trust beliefs, we add trust to the positive beliefs toward voice commerce in Table 2.

Table 2. Benefits and costs in extant literature

Benefits	Description
Efficiency	Voice interaction is faster than other input modes, due to hands free use [8].
Convenience	VAs are valued for their ease of use [8], also in voice commerce [34].
Enjoyment	Users have fun when speaking to the VA as the interaction is more personal [9].
Anthropomorphism	Voice output may induce the attribution of human-like characteristics to VAs [5].
Trust	Users' willingness to accept vulnerability toward a VA denotes a positive belief in VAs [21].
Costs	Description
Privacy concerns	Users are reluctant to share personal information via VAs as they are concerned about their data [10].
Low technical maturity	Speech recognition errors reduce users' perceived technical maturity, leading to negative user experiences [8].

In contrast, potential costs and negative beliefs include privacy and security issues. These are found to affect individuals' adoption intentions of VAs [10] and may therefore also apply to consumers' evaluations of voice commerce. Furthermore, we add technical maturity as a cost since extant research shows that

speech recognition errors have a negative effect on users' experience with the system [8]. Table 2 summarizes the proposed benefits and costs that we associate with voice commerce, as seen from the perspective of the customer. Overall, the extant literature mainly presents positive beliefs that could be expected from voice commerce.

3. Methodology

We chose a qualitative research approach to address our research question and to extend the knowledge derived from extant literature. Since voice commerce is a recent phenomenon and only partially understood in respect of narrow aspects, a qualitative approach is appropriate to explore consumers' beliefs regarding purchasing products via VAs in more depth and breadth [40]. Therefore, we conducted semi-structured interviews with VA users to validate the extant literature's findings and to explore additional aspects. In order to ensure a high comparability with related research, we based our interview guideline on the theoretical constructs in Table 2. To analyze our data, as well as to identify, analyze, and report themes in the data and evaluate them against extant literature, we followed the guidelines on qualitative data coding proposed by Miles et al. [16].

3.1. Data collection

For our interviews, we used purposeful sampling [40] to recruit regular users of Amazon's Alexa. As prior research had found that inexperienced consumers were mostly indifferent to commercial VA applications [32], our approach would capture the future potential and requirements of more innovative users who are familiar with the capabilities of VAs. We limited our sample to users of Amazon's Alexa as it is the most popular smart speaker and also has the highest market share of all voice shoppers worldwide [3]. Between November 2018 and February 2019, we conducted semi-structured one-to-one interviews with 30 VA users, either in person or via telephone. Participants were recruited at a large university and through social media. The resulting sample included 12 female and 18 male participants, aged between 21 and 82 years. Only five participants claimed to be adopters (i.e. they had tried voice commerce at least once); two continued and three discontinued the use of voice commerce.

We divided the interview process into three parts. First, participants reported on their general online purchasing behavior and the use of their smart speaker. After these introductory questions, we asked the participants about their prior experience with voice commerce. Depending on their level of experience,

additional questions were concerned with participants' purchasing behavior via voice, such as their purchase frequency and preferred product types. In order to familiarize non-adopters with the voice purchasing process and to prevent potential biases because of their lack of experience, we demonstrated the order process via Alexa in a self-made video. Second, we questioned the participants on their general attitude toward voice commerce and on their perceived advantages and disadvantages of purchasing products via voice. Following this, the participants were asked about their perception of the theoretical constructs listed in Table 2. Accordingly, we formulated questions based on the construct definitions. For example, the question "do you think purchasing with your smart speaker is easy and convenient?" refers to convenience. Third, participants could suggest desired improvements and provided information about their age, gender, and profession, depicted in Table 3.

Table 3. Information on participants

Age	Frequency (Percentage)
20-29	26 (86.7%)
> 30	4 (13.3%)
Gender	Frequency (Percentage)
Female	12 (40.0 %)
Male	18 (60.0 %)
Profession	Frequency (Percentage)
Student	14 (46.7 %)
Employee	15 (50.0 %)
Other	1 (3.3 %)

3.2. Data analysis

Two researchers independently coded all literal transcripts of the interviews, based on the applicability of what the interviewees said to either the benefits or the costs associated with voice commerce. Therefore, following qualitative data coding recommendations [16], the coding process was divided into first-cycle and second-cycle coding. The first-cycle coding started with deductive coding based on the constructs of our theoretical framework (see Table 2). We then inductively coded data that could not be assigned to the existing codes, for example 'no visual representation'. After discussing mismatches and reaching consensus on the naming of concepts, the coders derived 18 final codes from the data. In the subsequent second-cycle coding process, we grouped these codes to build thematic categories and overarching themes [16]. In this way 'no visual representation' and 'no comparison function' were grouped together to produce the thematic category 'limited transparency'.

4. Findings

Of the five benefits of voice commerce derived from the literature, and based on our interview data, we confirmed three positive beliefs: *convenience*, *efficiency*, and *enjoyment*. However, the perceived costs of voice commerce differed from our anticipated findings, with the negative beliefs toward voice commerce being *limited transparency*, *low technical maturity*, *limited control*, and *lack of trust*. Table 4 summarizes the codes derived from the interviews.

Table 4. Benefits and costs in the interviews

Benefits	Codes
Efficiency	Hands and eyes free use
Convenience	Less mental effort
	Ease of use
Enjoyment	Usage enjoyment
Anthropomorphism	Personal shopping experience
Costs	Codes
Limited transparency	No visual representation
	No comparison function
	Limited product information
	No independent reviews
Lack of trust	Vendor's competence
	Vendor's benevolence
	Technology reliability
Limited control	Potential misuse by strangers
	No manual input modality
	Risk of misunderstanding
Low technical maturity	Limited interactivity
	Speech recognition errors
Anthropomorphism	Feelings of uneasiness

Contrary to our expectations, privacy concerns are not negatively related to voice commerce. As regular users of VAs, our interviewees do not indicate any additional privacy risks that affect their intention to adopt voice commerce: "I know that I am continuously monitored. But if it really bothers me, I can switch it off. I do not have any concerns regarding voice commerce [...] not any more than I would have for traditional online shopping. It does not make a difference" [P9]. This quote illustrates that although consumers may indeed have privacy concerns regarding online shopping or VAs in general, these concerns are not higher for voice commerce although they may still exist. Only one participant states that, after talking about a certain topic, he would feel particularly uneasy about the advertisements on his phone. Since this issue does not affect his general use intention, we nevertheless exclude privacy concerns from the resulting conceptual model.

In addition, there is no evident trend in users' perceptions of the VA's anthropomorphism and its impact on voice commerce. Although most of the participants do not attribute any human-likeness to their VA, they feel differently about it. Some participants desire more human-likeness and a personal shopping experience: "I am used to [the VA], I am used to her voice. She is more personable" [15]. "If she would be like a real shopping advisor in a store, I would like to consult her" [P1]. In contrast, other participants state that they would rather not want their VA to resemble a human being as this makes them feel uneasy: "I prefer that I talk to a computer that does not sound like a human, [...] that makes it less creepy" [P16]. "When I speak to Alexa, I feel like I would tell another person to buy something for me. I find this very insensitive" [P6]. Hence, we cannot categorize anthropomorphism as either a benefit or a cost, as it can have a positive and/or a negative connotation for the participants. From the interviews, it seems as if consumers prefer a human-like experience in terms of high responsiveness and competence, but refrain from human imitations, for example through a human voice.

4.1. Perceived benefits

Efficiency: VA users expect time saving from voice commerce when performing their purchase activity. The participants expect efficiency gains when ordering routine products, similar to the functionality of the Amazon Dash Buttons. Since they do not have to open the app and deliberately select a product, the buying process is expected to be much quicker: "I don't lose time for the order. [...] I can do other things at the same time, i.e. multitasking" [P8]. "It is much faster [...] You don't have to search for it by typing and to scroll through a thousand items" [P4].

Convenience: Interviewees anticipate convenience when they use VAs to purchase products. Voice commerce is regarded as being easier than conventional e-commerce. The purchase process is less complicated as consumers do not have to browse through the products online and type on the keyboard. In addition, by requiring less demanding mental efforts, the decision process itself is easy: "I can also simplify my decision choice by entrusting Alexa to make the choice for me" [P8].

Enjoyment: VA users indicate that they experience general enjoyment when using their smart speaker. However, they expect that their usage enjoyment will also transpire in voice commerce. "I have fun using Alexa and speaking to her, making my life easier. I would imagine that this would also apply to voice commerce" [P9].

4.2. Perceived costs

Limited transparency: Perceived transparency refers to the user's understanding of the inner workings of a system, its underlying motives and the characteristics that drive its behavior [41]. As a negative belief, participants mostly fear that when they order products by simply saying "Alexa, buy garbage bags", the VA would choose a product by default without any comparison shopping. They would rather like to know how the VA makes this decision: "I find it random how it makes a decision. It suggests a product and I do not know how it decided on that one" [P8]. "I don't know on which basis she chooses a product. Does she take the best one, how does she decide?" [P14]. The participants also miss visual product representations: it is exhausting when, simultaneously, they have to remember the information of different products. In addition, the participants state that they do not want to rely on the vendor's product choices as they would rather prefer independent customer reviews, test results, or "reports about whether the [product] had any malfunctions or what people found to be pros and cons" [P5].

Low technical maturity: As expected, the interviewees poorly rate the technical maturity of their VA. They miss the ability to interact in a responsive way and demand that the VA should understand the context of a conversation: "It should not be complicated. I do not want that she reads out loud product lists, but that it is an interactive dialogue in which she can tell me the differences between products based on their main features" [P1]. "I can't imagine how it should work. [...] I tried to add a [product], but it only reads out loud the first result, the product name. [...] I think it would have to be much more intelligent, it needs to know what information I need" [P26].

Limited control: The consumers report a perceived loss of control. Perceived control refers to peoples' perception that outcomes are the results of their own behavior [42]. The participants fear that other people could misuse their smart speaker to order products without their consent: "It really annoys me. When friends come to visit, they can make fun of it and purchase any product." [P2]. Instead, the interviewees prefer a manual input modality that enables them to consciously choose a product or swipe through different options. In addition, VA users fear the *risk of misunderstandings*, leading to wrong product purchases: "I would fear that I order the wrong thing and that I could not stop it anymore. That there are things in my shopping cart that I did not want" [P14].

Lack of trust: In contrast to an anticipated positive belief, the participants state that they do not trust the smart speaker with their product purchases and

therefore form negative beliefs. Hence, in accordance with the definition of trust, consumers want to avoid being vulnerable to a VA for voice commerce. The interviews reveal different dimensions: First, VA users do not trust the provider to be sufficiently competent to know their preferences and to be benevolent when making the best choice on their behalf: “I don’t know if she would be able to do this. And if Amazon would want that. Because Amazon wants to sell products that bring the highest profits – thus the most expensive ones. I would think that they would rather sell a [product] for thousand euro than for fifty” [P2]. Second, users do not regard technology as sufficiently reliable to be entrusted with their purchases: “Imagine you would buy something at the DriveIn counter at McDonalds. It would be a shame if the Chicken McNuggets were missing. And if Alexa would forget my washing powder, it would be the same” [P3].

4.3. Summary

Overall, the evaluation of users’ cost-benefit beliefs regarding voice commerce points to a trade-off that may not easily resolved. Our research results imply that consumers, in order to feel confident about making the right choice, demand control through manual input modalities and transparency through visual representations, comparison functions, reviews, and product details. While these factors are embedded in the e-commerce literature, since they provide positive outcomes on customers’ evaluations, their applicability to the context of voice commerce remains a challenge. When providing visual output, manual interaction modalities, and transparent product comparisons and reviews, the main benefits of VAs – being efficient, effortless, and enjoyable (through hands and eyes free use and the opportunity to multitask) – cannot simultaneously be assured. This point is also illustrated in our interviews: “Of course, it is convenient to

interact via voice in general. However, it is inconvenient when it takes your decision choice. And then it is rather inconvenient to access your decision alternatives again. Consequently, the first encounter is convenient, but every following query is becoming increasingly inconvenient” [P7]. To date and based on the interviews, it seems as if existing voice commerce applications are unable to balance customers’ cost-benefit evaluations in a way that guarantees the achievement of their expected benefits without incurring higher perceived costs. The resulting negative attitude to the use of a VA for purchasing products could further affect consumers’ intentions and actual behavior, as the TRA postulates [37, 38], thus providing possible explanations for consumers’ non-adoption behavior in respect of voice commerce. Figure 1 depicts this relationship, which also provides a starting point for quantitative evaluations.

5. Discussion

This study was motivated by the technological advancements in speech recognition and natural language processing that encourage businesses to increasingly market and monetize their services via a new customer channel, namely VAs. However, as consumers remain reluctant to make product purchases via voice, our aim is to understand the cost-benefit evaluations of consumers when they decide to use voice commerce. Building on extant literature on both VAs and e-commerce, we derived a list of relevant factors that relates to users’ (non-)adoption behavior of voice commerce. We evaluated these factors through 30 semi-structured interviews with VA users.

While our results confirm three of the anticipated benefits that consumers would expect from voice commerce, namely *efficiency*, *convenience*, and *enjoyment*, and one cost factor, namely *low technical*

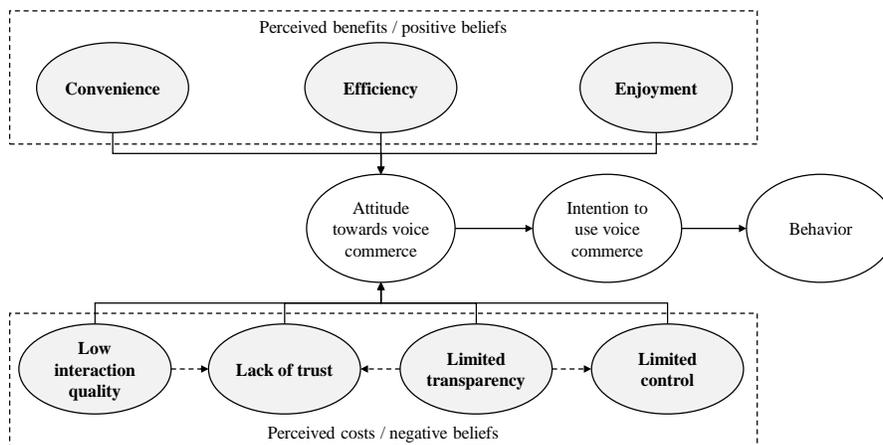


Figure 1. Conceptual model of the benefits and costs of voice commerce adoptions.

maturity, the remaining factors were not supported. The proposition that trust, which is often associated with voice, would also encourage positive beliefs in voice commerce, was refuted by the interviews. Users state that they *trust* neither the vendor's competence or benevolence, nor the technology's reliability. Although trust is not a new phenomenon in either VA research or e-commerce, our results confirm the relevance and negative effects of a lack of trust for voice commerce. Especially regarding the purchase decision, research needs to consider the duality of the customers' trust beliefs in both technology and the vendor, as is the case in respect of mobile commerce [43].

In addition, the interviewees do not attribute any human-like characteristics to their smart speaker. While some prefer more human-likeness, others feel uneasy about it. This was a rather surprising finding as extant research suggests that voice would encourage *anthropomorphism* and positive beliefs on the user's side. We present two possible explanations for this phenomenon: First, anthropomorphism may happen unconsciously without the users' control or direct attention. As the literature shows that individuals speak intuitively [7] and that spoken interactions are more personal [26], it is likely that individuals do not deliberately anthropomorphize a VA, but still behave differently during the interaction. Second, the extant literature also shows that there are various ways to infer anthropomorphism [5]; voice output, without any visual component, may not be sufficient to provide the impression of a human being. Regarding the result that users' unease increases the more human-like the VA becomes, a possible explanation is provided by the Uncanny Valley hypothesis. The latter states that a user's familiarity with a system decreases as soon as the system becomes human-like without having a life-like appearance [44]. Overall, our findings emphasize the relevance of recent research calls to determine the optimal level of anthropomorphism [15].

Finally, we identified two negative beliefs from the data, beliefs not derived from the literature: limited control and limited transparency. Although we did not account for them in our literature review, as they were not among the major topics in both research streams, these negative beliefs describe two fundamental costs that users assign to voice commerce. Perceived *control* refers to individuals' beliefs that outcomes are determined by their own behavior [42]. Although investigated in mobile commerce, there is no research on the effect of limited control in the VA context. For example, perceived control is shown to positively affect users' perceived value of mobile services and their intention to use them [45], and is positively associated with transaction efficiency and trust in the provider [46]. Since users cannot directly determine

what the smart speaker ultimately does, the negative belief of limited control seems reasonable. Based on the results of the literature on mobile commerce, we therefore observe the opposite effect of perceived control on efficiency and trust in the VA provider.

Lastly, users' perceived *transparency* refers to their understanding of the inner workings of a system [41]. Although this theoretical construct has to our knowledge not yet been investigated for VAs, we draw on prior literature in the e-commerce context to gain a better understanding. For recommender agents, improvements of their transparency by providing explanations are shown to positively affect all three trusting beliefs (competence, integrity, and benevolence) [47]. Also, the use of trade-off displays between products to increase transparency has positive effects on users' perceived enjoyment and product diagnosticity, that is the extent to which a consumer believes that a system is helpful for the full evaluation of a product [48]. Our results therefore imply that without comparison functions and the provision of product details, higher levels of product diagnosticity cannot be achieved for voice commerce.

6. Conclusion

Based on a literature review and 30 semi-structured interviews, we identified benefits and costs that VA users evaluate when making the decision to use voice commerce. Our results show that users evaluate benefits in *efficiency*, *convenience*, and *enjoyment*, against the costs of limited *transparency*, limited *control*, limited *technical maturity*, and lack of *trust*. The trade-offs between these cost-benefit evaluations point to a possible explanation of the low adoption rates of voice commerce thus far: for example, the provision of visual output and manual input reduces the expected efficiency gains of consumers.

However, this research is not without limitations. First, the interview sample shows weaknesses in the representation of the population. Most of the participants were below 30 years of age and without voice commerce experience. Although we compensated for this lack of knowledge by showing all participants a video of the purchasing process via voice, interviews with more regular and older voice commerce adopters could provide deeper insights into actual use experiences. Furthermore, we used purposeful sampling and only interviewed users of Amazon's Alexa. Although our results specifically show why VA users do not adopt voice commerce, they do not indicate why users do not adopt VAs at all. Consequently, we did not detect the voice commerce costs incurred by VA use in general (i.e. privacy concerns). Second, we did not differentiate between the

applicability of different product types for voice commerce. Although many participants state that they prefer using voice for commodity products and repeated purchases, our results did not include any product distinction. Future research could therefore examine our propositions for different product categories. Third, we did not depict the application of VAs throughout the whole customer journey, but only for the purchase of a product, thus neglecting the post-purchase stage of customer service. As customer service is a suitable application for VAs, future research should also investigate this stage.

Our results enhance existing theory in two ways: First, based on a sound theoretical model grounded in the TRA [37, 38], we provide an explanation for users' non-adoption behavior in respect of voice commerce. Second, the identified constructs describe relevant factors explaining consumers' intentions and actual behavior in regard to voice commerce, which can serve as basis for quantitative models. Therefore, future research should examine the proposed relationships that we derived from our data. For example, our results imply that increasing transparency attenuates the negative effects of perceived control losses by providing explanations for the VA's behavior. In addition, as a topic of future experimental studies, transparency could serve as a way to increase consumer trust in both the vendor and technology.

Our research results also have practical implications. To address consumer preferences, voice commerce providers and vendors must carefully balance consumers' cost-benefit evaluations in their product configurations. For example, to further increase consumers' trust and perceived control, providers can improve the transparency of VAs by adding visual displays, providing relevant and independent product information, and explaining the VA's choice based on comparable products and prices. In addition to new VA generations that already address these shortcomings, it is possible to enhance existing screenless devices through television or smartphone connections. These connections can provide additional information to the customer and decrease their effort of memorizing the system's spoken output. Another established way to increase transparency is to provide explanations. Providers can easily add them to the dialogue configurations of the VA. Furthermore, providers should differentiate between dialogue interactivity and the human-likeness of speech configurations. While users prefer a fluent dialogue and an understanding of the conversational context, they are uncomfortable with the impression that is created of a human being. Providers must consider this when designing new applications and can consider using a synthesized voice instead of a human one.

7. Acknowledgements

We thank Verena Thürmel, Maria Reschka, and Sergei Titov for their support in collecting the interview data.

8. References

- [1] Buvat, J., Jacobs, K., Khadikar, A., Sengupta, A., and Taylor, M., "Conversational Commerce", https://www.capgemini.com/wp-content/uploads/2018/01/Conversational-Commerce-Report_Digital.pdf, accessed 31.05.2019.
- [2] Besik, H., "91% of Brands are Investing in Voice: How to Make it Work", <https://theblog.adobe.com/91-of-brands-are-investing-in-voice/>, accessed 31.05.2019.
- [3] Kinsella, B., and Mutchler, A., "Smart Speaker Consumer Adoption Report", https://voicebot.ai/wp-content/uploads/2018/03/smart_speaker_consumer_adoption_report_2018.pdf, accessed 12.01.2019.
- [4] Kinsella, B., "Voice Purchasing Rose in 2018 Among US Smart Speaker Owners", <https://voicebot.ai/2019/03/20/voice-purchasing-rose-in-2018-among-us-smart-speaker-owners/>, accessed 31.05.2019.
- [5] Qiu, L., and Benbasat, I., "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems", *Journal of Management Information Systems*, 25(4), 2009, pp. 145-182.
- [6] D'Mello, S.K., Graesser, A., and King, B., "Toward Spoken Human-Computer Tutorial Dialogues", *Human-Computer Interaction*, 25(4), 2010, pp. 289-323.
- [7] Akinlasi, F.N., "On the Differences Between Spoken and Written Language", *Language and Speech*, 25(2), 1982, pp. 97-125.
- [8] Luger, E., and Sellen, A., "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 Syst.*, 2016, pp. 5286-5297.
- [9] Moussawi, S., "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*, 2018, pp. 86-92.
- [10] Easwara Moorthy, A., and Vu, K.-P.L., "Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space", *International Journal of Human-Computer Interaction*, 31(4), 2015, pp. 307-335.
- [11] Brasel, S.A., and Gips, J., "Tablets, Touchscreens, and Touchpads: How Varying Touch Interfaces Trigger Psychological Ownership and Endowment", *Journal of Consumer Psychology*, 24(2), 2014, pp. 226-233.
- [12] Klesse, A.-K., Levav, J., and Goukens, C., "The Effect of Preference Expression Modality on Self-Control", *Journal of Consumer Research*, 42(4), 2015, pp. 535-550.
- [13] Son, Y., and Oh, W., "Alexa, Buy Me a Movie!: How AI Speakers Reshape Digital Content Consumption and Preference", *Proceedings of the International Conference on Information Systems*, 2018.
- [14] Tuzovic, S., and Paluch, S., "Conversational Commerce—A New Era for Service Business Development?", in (Bruhn, M., and Hadwich, K., eds.): *Service Business Development*, Springer, 2018, pp. 81-100.

- [15] De Keyser, A., Köcher, S., Alkire, L., Verbeeck, C., and Kandampully, J., "Frontline Service Technology Infusion: Conceptual Archetypes and Future Research Directions", *Journal of Service Management*, 30(1), 2019, pp. 156-183.
- [16] Miles, M., Huberman, A., and Saldana, J., *Qualitative Data Analysis: A Methods Sourcebook*, Thousand Oaks, CA: Sage Publications Inc, California, 2013.
- [17] Juang, B.-H., and Rabiner, L.R., "Automatic Speech Recognition—A Brief History of the Technology Development", in (Brown, K., ed.): *Encyclopedia of Language and Linguistics*, Elsevier, 2005, pp. 67.
- [18] Hirschberg, J., and Manning, C.D., "Advances in Natural Language Processing", *Science*, 349(6245), 2015, pp. 261-266.
- [19] Hoy, M.B., "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants", *Medical Reference Services Quarterly*, 37(1), 2018, pp. 81-88.
- [20] Glass, J., "Challenges for spoken dialogue systems", *Proceedings of the 1999 IEEE ASRU Workshop*, 1999
- [21] Nasirian, F., Ahmadian, M., and Lee, O.-K., "AI-Based Voice Assistant Systems: Evaluating from the Interaction and Trust Perspectives", *Proceedings of the Americas Conference on Information Systems (AMCIS)*, 2017
- [22] Davis, F.D., Bagozzi, R.P., and Warshaw, P.R., "User Acceptance of Computer Technology", *Management Science*, 35(8), 1989, pp. 982-1003.
- [23] Venkatesh, V., Thong, J.Y., and Xu, X., "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology", *MIS Quarterly*, 36(1), 2012, pp. 157-178.
- [24] Schroeder, J., and Epley, N., "Mistaking Minds and Machines: How Speech Affects Dehumanization and Anthropomorphism", *Journal of Experimental Psychology: General*, 145(11), 2016, pp. 1427.
- [25] Benlian, A., Klumpe, J., and Hinz, O., "Mitigating the Intrusive Effects of Smart Home Assistants by using Anthropomorphic Design Features: A Multi-Method Investigation", *Information Systems Journal*, forthcoming.
- [26] Novielli, N., De Rosis, F., and Mazzotta, I., "User Attitude Towards an Embodied Conversational Agent: Effects of the Interaction Mode", *Journal of Pragmatics*, 42(9), 2010, pp. 2385-2397.
- [27] Schroeder, J., and Schroeder, M., "Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines", *Proceedings of the Hawaii Internat. Conference on System Sciences*, 2018.
- [28] Campbell, D.J., "Task Complexity: A Review and Analysis", *Academy of Management Review*, 13(1), 1988, pp. 40-52.
- [29] Keeney, R.L., "The Value of Internet Commerce to the Customer", *Management Science*, 45(4), 1999, pp. 533-542.
- [30] Pavlou, P.A., and Fygenson, M., "Understanding and Predicting Electronic Commerce Adoption", *MIS Quarterly*, 30(1), 2006, pp. 115-143.
- [31] Anckar, B., and D'inau, D., "Value Creation in Mobile Commerce: Findings from a Consumer Survey", *Journal of Information Technology Theory and Application*, 4(1), 2002.
- [32] Baier, D., Rese, A., and Röglinger, M., "Conversational User Interfaces for Online Shops? A Categorization of Use Cases", *Proceedings of the International Conference on Information Systems (ICIS)*, 2018.
- [33] Kim, D., Park, K., Park, Y., Ju, J., and Ahn, J.-H., "Alexa, Tell Me More: The Effect of Advertisements on Memory Accuracy from Smart Speakers", *Proceedings of the Pacific Asia Conference on Information Systems*, 2018.
- [34] Kraus, D., Reibenspiess, V., and Eckhardt, A., "How Voice Can Change Customer Satisfaction: A Comparative Analysis between E-Commerce and Voice Commerce", *Proceedings of the International Conference on Wirtschaftsinformatik*, 2019.
- [35] Le Bigot, L., Jamet, E., and Rouet, J.-F., "Searching Information with a Natural Language Dialogue System: a Comparison of Spoken vs. Written Modalities", *Applied Ergonomics*, 35(6), 2004, pp. 557-564.
- [36] Shen, H., Zhang, M., and Krishna, A., "Computer Interfaces and the "Direct-Touch" Effect: Can iPads Increase the Choice of Hedonic Food?", *Journal of Marketing Research*, 53(5), 2016, pp. 745-758.
- [37] Ajzen, I., and Fishbein, M., "Understanding Attitudes and Predicting Social Behavior": *Advances in Experimental Social Psychology*, Prentice Hall, Englewood Cliffs, 1980.
- [38] Fishbein, M., and Ajzen, I., "Belief, Attitude, Intention and Behavior: an Introduction to Theory and Research", Addison-Wesley, 1975.
- [39] Benlian, A., and Hess, T., "Opportunities and Risks of Software-as-a-Service", *Decision Support Systems*, 52(1), 2011, pp. 232-246.
- [40] Patton, M.Q., *Qualitative Evaluation and Research Methods*, Sage, Thousand Oaks, CA, 1990.
- [41] Sinha, R., and Swearingen, K., "The Role of Transparency in Recommender Systems", *CHI Extended Abstracts on Human Factors in Computing Systems*, 2002, pp. 830-831.
- [42] Ajzen, I., "The Theory of Planned Behavior", *Organizational Behavior and Human Decision Processes*, 50(2), 1991, pp. 179-211.
- [43] Siau, K., Sheng, H., Nah, F., and Davis, S., "A Qualitative Investigation on Consumer Trust in Mobile Commerce", *International Journal of Electronic Business*, 2(3), 2004, pp. 283-300.
- [44] Mori, M., "The Uncanny Valley", *Energy*, 7(4), 1970, pp. 33-35.
- [45] Kleijnen, M., De Ruyter, K., and Wetzels, M., "An Assessment of Value Creation in Mobile Service Delivery and the Moderating Role of Time Consciousness", *Journal of Retailing*, 83(1), 2007, pp. 33-46.
- [46] Collier, J.E., and Sherrell, D.L., "Examining the Influence of Control and Convenience in a Self-Service Setting", *Journal of the Academy of Marketing Science*, 38(4), 2010, pp. 490-509.
- [47] Wang, W., and Benbasat, I., "Empirical Assessment of Alternative Designs for Enhancing Different Types of Trusting Beliefs in Online Recommendation Agents", *Journal of Management Information Sys.*, 33(3), 2016, pp. 744-775.
- [48] Xu, J.D., Benbasat, I., and Cenfetelli, R.T., "The Nature and Consequences of Trade-Off Transparency in the Context of Recommendation Agents", *MIS Quarterly*, 38(2), 2014, pp. 379-406.