Use or Not: A Qualitative Study on the User Adoption and Abandonment of Voice Shopping with Smart Speakers

Jiahe Lan
Xidian University
jhlan16@stu.xidian.edu.cn

Yu Chen
San Jose State University
yu.chen@sjsu.edu

Zheng Yan
Xidian University
zyan@xidian.edu.cn

Abstract

The recent advances in smart speakers impel the emergence and prevalence of voice shopping – placing orders on voice assistants. Previous work has studied user acceptance of voice shopping and the factors influencing users' experience of voice shopping. However, despite the growing interest in the use of voice shopping, little is known about the limited usage or abandonment of voice shopping. In this paper, we address this research gap through a qualitative study of 43 users of Tmall Genie, a smart speaker popular in China. We found that participants are willing to make low-involvement purchases via voice shopping. However, after a period of use, participants tend to limit or abandon voice shopping due to time-consuming interaction and mistrust of voice shopping. Based on our findings, we discuss how our study could advance the understanding of voice shopping and present implications for researchers and practitioners on technical robustness, adaptive conversational product presentation, and cross-platform product recommendations for the future design of voice shopping systems.

Keywords: Voice shopping, intelligent voice assistants, smart speakers

1. Introduction

Smart speakers with built-in voice assistants, such as Amazon Alexa, Apple Siri, Alibaba Tmall Genie, and Xiaomi Xiao AI, are reported to be among the most promising smart home devices (Oksien, 2022). Voice assistants provide a new accessible way – the voice interface – and bring voice-based interaction to the mainstream. Voice assistants allow users to seek information (e.g., weather and news) and execute tasks on their behalf (e.g., playing music and sending messages) via conversational interactions (Bentley et al., 2018), which are considered more natural interactions than traditional graphical interactions. According to Harvard Business Review, voice assistants are the fastest-growing consumer technology since smartphones and are predicted to revolutionize the way we live, work, and play (Simms, 2019).

With the affordance of natural conversational interaction and accessibility, voice assistants also provide comprehensive features to users for voice shopping. Voice shopping is referred to as the act of placing orders through voice assistants. Although voice shopping is relatively new to e-commerce, it is growing rapidly. Market research has estimated that the voice shopping market will grow to USD 164 billion in 2025 (Sudlow-Poole, 2022). Apart from the huge market, the accessibility and natural conversational interaction of voice shopping also make it be considered a frictionless e-commerce phenomenon. The huge potential of voice shopping is attracting attention and investigation from both the industry and academia.

There has been a growing number of research studies on user acceptance of voice shopping (Ahn et al., 2019) and the factors influencing users' shopping experience, such as trust relationships (Bawack et al., 2021) and anthropomorphism of voice assistants (Son Nguyen et al., 2021). However, Newman (2020) reported that consumers are adopting voice shopping slower than expected. Although the number of smart speaker owners is growing exponentially, their main usage remains on simpler functions such as listening to music or seeking information. Additionally, we have observed in our pilot study that some users tend to abandon or limit the use of voice shopping after a novelty period. Although studying the adoption and motivation of technology

provides valuable insights, it is also important to investigate technology non-use (Baumer et al., 2013). Thus, the observations in the trend and pilot study motivate us to study the non-use of voice shopping, which is understudied. We asked the following research questions:

1. Why do users choose to adopt voice shopping?

2. Why do users limit or even abandon voice shopping?

In this paper, we investigate these questions by studying Tmall Genie, a voice assistant-embedded smart speaker developed by the Chinese e-commerce company Alibaba Group. Like other smart speakers, Tmall Genie offers various voice-driven services. In addition, like Amazon Echo, Tmall Genie is also well known for its voice shopping features. In 2019, more than one million orders were placed and paid via Tmall Genie during a 24-hour window of the biggest shopping festival in China (Li, 2019). Therefore, we started our research by focusing on Tmall Genie due to its large user base in China.

We conducted a qualitative study with 43 participants who had used voice shopping with Tmall Genie to understand why users adopt and abandon voice shopping. We found that after novelty effects have waned, participants tended to limit their usage or abandon voice shopping. They struggled with time-consuming interaction and mistrust of voice shopping. Based on the findings, we offer design implications of voice shopping for designers, practitioners, and manufacturers.

This paper contributes to the literature by extending existing knowledge on the factors that influence the use and non-use of voice shopping through an in-situ study among voice shopping users. We also provide suggestions on technical robustness, adaptive conversational product presentation, and cross-platform product recommendations for researchers and practitioners to design more frictionless future voice shopping systems.

2. Background and Related Work

2.1. Voice Assistants and Smart Speakers

The rapid development of artificial intelligence has facilitated the rise and use of voice assistants, such as Amazon Alexa, Apple Siri, Alibaba Tmall Genie, and Xiaomi Xiao AI. These voice assistants have brought voice-based interaction into the mainstream and provide a wide range of functions such as playing music and news or showing search results online. Voice assistants can be embedded in different carriers, such as smart speakers, smartphones, or laptops, which bring

different characteristics. For example, smartphones and laptops are portable, while smart speakers are usually stationary and used in fixed places such as homes. The degree of mobility offers affordability to different services such as voice-based shopping at home using Alexa versus voice-based shopping on a phone while driving. Additionally, smartphones and laptops offer both voice-based and graphical user interfaces, which allow users to view information beyond the audio-only channel on smart speakers. Furthermore, interacting with a smart speaker at home supports hand-free interaction, which lowers the barriers to use compared to the requirement of activating via a smartphone or laptop. Due to the differences in the nature of interaction and functionalities, we focus on voice shopping on smart speakers in this study, rather than voice shopping across all types of devices.

Prior studies have explored different applications of smart speakers, such as speech coaches (Wang et al., 2020), and different populations of users, such as the elderly (Trajkova & Martin-Hammond, 2020) and families at home (Voit et al., 2020). In addition to the wide variety of applications, researchers also investigated the human factors of voice assistants (Son Nguyen et al., 2021). A body of research has also studied the security and privacy issues of smart speakers with their always-on microphones (Abdi et al., 2019).

Beyond the commonly used application scenarios, both the industry and academia have started to explore the commercial potential of smart speakers in voice shopping (Ahn et al., 2019; Bawack et al., 2021; Hu et al., 2022), the act of placing orders on voice assistants. Voice shopping inherits the accessibility and natural conversational interaction of smart speakers, which makes it to be considered the future of e-commerce. In this paper, we focus on voice shopping amongst the various promising applications of smart speakers.

2.2. Voice Shopping Usage Adoption and Barriers

Previous work has investigated user acceptance of voice shopping (Ahn et al., 2019; Sorensen & Jorgensen, 2021), as well as the factors that influence the user shopping experience (Bawack et al., 2021; Rhee & Choi, 2020). Ahn et al. (2019) explored the factors influencing the consumers' intention to accept voice shopping in South Korea and found that performance expectation, effort expectation, and amusement expectation have positive effects on user acceptance. Noticing millennials use voice-activated technology more frequently than people of other age groups, Sorensen and Jorgensen (2021) explored millennials' acceptance of voice shopping. Their findings show that millennials have a

positive attitude towards voice shopping. Researchers have also investigated the impact of user personality (Bawack et al., 2021), trust relationships (Bawack et al., 2021), and anthropomorphism of voice assistants (Son Nguyen et al., 2021) on the user shopping experience. The above research has provided a good understanding of the key factors that impact the use and adoption of voice shopping.

Investigating voice shopping use and adoption also revealed barriers and breakdowns during voice shopping usage. Abdi et al. (2019) and Son Nguyen et al. (2021) found that the invisibility of products is a major obstacle that prevents users from adopting voice shopping. Tuzovic and Paluch (2018) identified technical malfunctions during voice shopping as another major barrier and suggested that in order to avoid deterring users, smart speakers could be offered as an additional transaction channel as a backup, rather than the main channel. In addition, researchers also found that users' concerns about privacy risks associated with voice shopping negatively impacted user adoption (Abdi et al., 2019). Studying barriers to the use and adoption of voice shopping helps researchers to better understand the challenges faced by voice shopping and helps to improve it.

2.3. Technology Abandonment

While previous work has provided important insights into the use and barriers to voice shopping, they have been from the perspective of understanding use rather than the non-use: limited usage or abandonment.

Previous literature has underscored the importance of studying the limited usage and abandonment of technology in a variety of contexts. Zou et al. (2020) examined the adoption and abandonment of identity theft protection practices. They found that abandonment tends to occur when users perceived such practices as low-value, inconvenient, or when users overrode them with subjective judgment. These findings lead to recommendations on designing security and privacy practices that better align with user needs. Studies that investigated the abandonment of Facebook (Baumer et al., 2013) and wearable self-tracking devices (Clawson et al., 2015) have also contributed to their design and enhancement.

Research on intelligent voice assistants is highly relevant to our work. Voit et al. (2020) studied the phenomenon of smart speaker deactivation in the household context and stressed the importance of understanding the social impact of smart speakers. Trajkova and Martin-Hammond (2020) investigated the reasons why older adults abandoned Alexa as a health-tracking technology and identified difficulty

finding valuable uses, beliefs associated with ability and voice assistant use, or challenges of use in shared spaces as crucial factors in abandonment, sparking deep insights into how intelligent voice assistants can be better designed as an assistive technology to support aging and independent living in the future. Along the same line, Werner et al. (2022) conducted a focus group study on the use and non-use of smart speakers and identified concerns related to privacy and trust, lack of accuracy and reliability, and lack of knowledge regarding particular functions/features or the system as a whole.

In summary, while studying the adoption and usage of voice shopping is valuable, investigating the non-use – limited usage and abandonment – not only refines users' perspectives about real-world use but also provides new perspectives on design implications. However, to the best of our knowledge, the literature still lacks an in-depth understanding on the non-use of voice shopping. Thus, we bridge this gap by researching the adoption and abandonment of voice shopping via Tmall Genie.

3. Methodology

In order to investigate users' experience of voice shopping on smart speakers, we conducted a qualitative study by interviewing 43 users who have voice shopping experience using Tmall Genie smart speaker.

3.1. Tmall Genie

Tmall Genie is a smart speaker developed by the Chinese e-commerce company Alibaba Group, first released in July 2017. Similar to other smart speakers, Tmall Genie offers various voice-driven services, such as controlling smart home gadgets and playing music. Tmall Genie is among the three most widely used smart speakers in China, together with DuSmart Speaker and Xiaomi AI. Among them, we choose to focus only on Tmall Genie due to its large user base in China.

Various models of Tmall Genie have been released, including audio-only models such as Tmall Genie Sugar Cube 2, and audio and video models such as Tmall Genie CC. Among these models, Sugar Cube 2 has been the most widely adopted model because it is more affordable than other models: around 100 RMB (around 15 US dollars). Figure 1 shows an example of the audio-only model Sugar Cube 2, which includes a microphone and four buttons, but does not support any visual display.
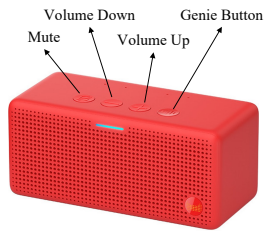
Figure 1. Tmall Genie Sugar Cube 2 (Genie Button refers to personalized function).

3.2. Study Design

We conducted semi-structured interviews with forty-three participants with voice shopping experiences using the Tmall Genie Sugar Cube 2. We first contacted qualified participants about the purpose of the study and recommended participants be aware of their interaction with Tmall Genie, and then conducted the interview two weeks after the initial contact with each participant.

During the interviews, which lasted between twenty minutes and fifty minutes, we first asked for participants' demographic information. We then asked about their online shopping habits, such as commonly used shopping applications and types of products typically purchased online. We continued to invite participants to share their experience using their Tmall Genie devices, e.g., frequency, duration of usage (for how long), and location of using Tmall Genie. Next, we asked the participants to share their experiences of voice shopping using Tmall Genie, including what, when, and how they purchased using the device. We encouraged them to tell stories and show examples when sharing their experiences. After that, we also asked participants to reflect on what they liked and disliked about using voice shopping, or any pleasant or unpleasant experiences using it. We audio-recorded and transcribed all interviews for further analysis.

3.3. Participants

We recruited participants who owned and used Tmall Genie and voice shopping via three approaches in China: direct contact, snowball sampling, and recruitment advertisement through university forums. Eventually, forty-three users, who had used voice shopping more than ten times and had used Tmall Genie for over one year, participated in our survey. Among the forty-three participants, seven were our direct contacts, eleven were selected by snowball sampling, and twenty-five were recruited through the university forums.

The participants include twenty-three males and twenty females aged between seventeen and forty-eight. Thirty-seven of the participants were between the ages

of seventeen and twenty-seven. Their educational backgrounds ranged from high school to doctorate. Their occupations included undergraduate students (25), graduate students (6), office clerks (4), Ph.D. students (3), government officials (3), and teachers (2). Except for nine participants who placed Tmall Genie in a home environment, such as bedrooms and kitchens, all other participants lived and used Tmall Genie in a shared room.

3.4. Data Analysis

We transcribed the audio recordings verbatim. We applied the Grounded Theory (Birks & Mills, 2015) approach to analyze the transcribed interview data. Two researchers who did not participate in the interviews first generated initial coding themes in the open coding stage. The collection of codes from the two researchers formed our initial codes. In the systematic axial coding stage, to find the relationship between the initial codes, the first two authors presented all the initial codes in a tree diagram to obtain secondary codes, which are more general than the initial codes. After that, the two authors wrote down the emergent themes, compared their results, and then identified and categorized the themes, as presented in the findings (Section 4). The quotes included in this paper were translated from Simplified Chinese to English.

4. Findings

In this section, we first present participants' usage scenarios of voice shopping. We then elaborate on the reasons for the limitation or abandonment of voice shopping.

4.1. Voice Shopping Usage Scenarios

We first aim to understand what users use voice shopping for and why. In general, participants mainly reported utilizing voice shopping for low-involvement purchases, referring to scenarios when consumers can make a purchase decision without much effort. In particular, three types of low-involvement purchases emerged from our data: routine payment, repurchase, and low-risk purchase.

4.1.1. Routine Payment Routine payment refers to the payment of various bills in daily life. As a type of routine payment, paying phone bills through voice command is one of the most frequently mentioned features of Tmall Genie in our study. More than half of the participants have used Tmall Genie to pay their phone bills. Instead of a recurring monthly bill or automatic payment, it is common for mobile phone

users in China to top up and refill their prepaid accounts.

"Since I have Tmall Genie, the task of paying phone bills has been delegated to it. I can pay phone bills through Tmall Genie while watching TV. Tmall frees my hands." [P21]

It is convenient for P21 to use Tmall Genie to pay phone bills without involving his hands, therefore freeing up space for multitasking. Besides convenience, accessibility of payment is one of the important reasons why P21 uses Tmall Genie to pay phone bills.

"My grandparents do not know how to pay phone bills online. In the past, they went to a business hall far away to pay. After I brought a Tmall Genie for them, they can pay phone bills by themselves at home via talking." [P21]

In the above scenario, voice shopping could facilitate purchases and daily living for elderly users, who might not be familiar with smartphones, since talking makes payment much more accessible for them to operate.

4.1.2. Repurchase
Repurchase – buying products that have been bought before – has also been reported frequently by our participants. In the interviews, most participants said that they hardly used voice shopping to buy products they were not familiar with; however, they used voice shopping to buy the products they used. This may be because users tend to be more cautious when buying new products in general, especially on a comparatively new platform. Note that previous purchases were not necessarily purchased on Tmall Genie.

"I do not like to use voice shopping to buy products I have not bought. It is too unreliable. I do not even know what it looks like. I prefer to buy products that I already have bought via voice." [P29]

For P29, since he has already purchased certain items multiple times, which he had trust in, the purchase decision is relatively easy and straightforward without much need for an in-depth investigation on product details or comparison. Therefore, P29 felt more confident and comfortable completing the purchase through voice shopping.

In addition to products, participants also used voice shopping for repurchasing services, such as food delivery services. For example, P23 would order take-out food from the restaurant he frequently visited.

"I usually order takeout food while playing games. For that, I can simply say 'Tmall Genie, order the same takeout I ordered yesterday.', and then complete the payment. It only takes a few words without even needing to move my fingers." [P23]

For P23, when focuses on playing games, ordering food via voice only takes him a few words, which not only saves time and effort in choosing, frees up his hands, supports multitasking, but also has little chance for mistakes.

4.1.3. Low-risk Purchase
Low-risk purchase, the third type of low-involvement purchase, refers to scenarios when the products purchased pose a low risk to the buyer if he or she makes a mistake in the purchase or encounters an unsatisfactory purchase (Zaichkowsky, 1985). One type of low-risk purchase is the low-price purchase. Due to the low price, users would not care too much for a less-than-desirable purchase, and therefore they tend to pay less attention to the details of a product as long as it has the basic features they expect.

"After I noticed my shoes were dirty, I bought a shoe cleaner via Tmall Genie. And I bought the first one it recommended to me because the price is reasonable and does its work." [P33]

Seeing dirty shoes, P33 bought the first shoe cleaner recommended by Tmall Genie without spending much time gathering information about the product or evaluating other similar products. Low price and low risk are reasons for her to make a quick purchase decision through voice shopping.

Similar to P33, P35 shared her thoughts about factoring in the price of purchased product.

"I prefer to use (Tmall) Genie to buy miscellaneous groceries, such as snacks and detergent because they are cheap and stable in price. However, for other goods, such as smartphones and clothes, I prefer to buy them offline." [P35]

P35 was happy buying some groceries by voice shopping since the price is typically low and does not fluctuate, but she was not comfortable with buying anything expensive.

To summarize, users tend to use voice shopping for routine payments, repurchases, and low-risk purchases. In such scenarios, voice shopping makes the transaction more convenient and accessible, and supports multitasking.

4.2. Reasons for Limited Usage or Abandonment of Voice Shopping

Despite the usage scenarios and advantages of voice shopping, participants also reported scenarios of limited usage or abandonment. Overall, participants felt that the limitation or abandonment of voice shopping stemmed from the time-consuming interactions and mistrust of voice shopping.

4.2.1. Time-consuming Interactions
When discussing why they abandoned purchasing items or discontinued using voice shopping, one common

reason participants mentioned was the time-consuming interaction limited by the audio-only interface.

Participants often mentioned the initial excitement of using voice-based interaction. They were curious about voice assistants due to novelty effects (Bentley et al., 2018). Thus, in the initial phase, participants tend to play around with voice shopping, even if they had no real desire or need to purchase.

"I was curious if it was possible to use voice shopping to purchase any item. Once, I said to Tmall Genie 'I want to buy a mainboard'. You know, a mainboard is expensive, and it is impossible for me to buy it via voice. I just wonder how it [Tmall Genie] will answer me." [P32]

However, over time they noticed that their efforts out-weighed the benefits of voice shopping. They found audio-only modality more cumbersome than shopping methods that support multiple modalities such as videos and images that they are familiar with, especially when exploring an unfamiliar product.

"Listening to one item description via Tmall Genie takes me one or two minutes and becomes challenging when I have multiple items to compare." [P21]

For P21, it is not only time-consuming to learn about a product's description, but also difficult to compare between products. In addition, some participants reported that it was difficult to absorb and retain all the information with an audio-only modality. Therefore, they need to revisit the information multiple times, costing them even extra time.

"I cannot keep up with the speed of Tmall Genie and have to listen to it over and over again, which is time-consuming and has led to me gradually giving up using voice shopping." [P22]

Because before the advent of voice shopping, most participants are already used to other online shopping methods, such as mobile shopping and web shopping, they usually find it easier to use these online shopping methods than voice shopping. Therefore, they gradually limit or even abandon the use of voice shopping.

"I prefer mobile shopping than voice shopping. I almost never leave my phone and I can access it whenever and wherever. More importantly, I can learn about a product in a few seconds by taking a look at its images. However, I need several minutes to do the same thing [learn about a product] via Tmall Genie." [P17]

The product presentation in an audio modality brings new challenges to users. Visuals have always been one of the main channels for users to access information about products in online shopping. However, voice shopping on Tmall Genie only offers the audio channel without visual support, which is not as intuitive as users' current online shopping platforms and modalities.

"In fact, when the Tmall Genie tells me about an item, I often need to imagine what it looks like and builds an image in my mind, which is more time-consuming than shopping via my phone. Voice shopping is actually not convenient for me." [P11]

As P11 reflected, appearance can be a critical specification for a product that users care about. Without the proper visual support, voice shopping takes away a key factor to help users make a purchase decision, which has been repeatedly reported to frustrate users.

Finally, a few participants complained that sometimes it could take a few rounds of conversational interaction for users' queries to be correctly understood by voice assistants.

"I felt that Tmall Genie is not intelligent enough. Sometimes, it gives an irrelevant answer to me, which is probably because it does not recognize my words correctly. Thus, it sometimes required multiple interactions to successfully deliver my commands, which make me feel really frustrated." [P7]

In this case, participants attributed the time wasted in multiple rounds of interaction to the "limited intelligence" of voice assistants.

To summarize, the time-consuming interaction with voice assistants makes users gradually abandon voice shopping. Users sometimes reported cases of the need for multiple attempts before their queries for products to be understood accurately, and the need for a few minutes to learn about an item or to compare multiple items, whereas, with other online shopping platforms, such as websites or mobile apps, users can accomplish the same goal in just a few seconds.

4.2.2. Mistrust of Voice Shopping Mistrust of voice shopping is another key reason for abandoning it. Based on the interviews, we found that users' mistrust of voice shopping stems from insufficient product information, lack of transparency in product recommendations, and occasional failures in correctly understanding users.

Insufficient product information is a frequent factor that lowered participants' confidence in making a purchase decision or using voice shopping. As mentioned earlier, because users typically obtain less information via the audio channel than textual or visual channels, manufacturers tend to display only the most essential product information. However, the preference for the types of product information in making a purchase decision is subjective. For example, when buying beverages, some users value the brand, while others care more about the ratings. The abridged or condensed product information provided by voice shopping did not seem to meet the needs of every user.

"I think the search results of products are good, but I'm missing a lot of important product information, e.g., brand, composition, especially packaging. I can hardly imagine what the products look like without their pictures." [P24]

Besides the product information, some participants need product reviews for making a purchase decision with confidence. For example, one participant expressed his frustration.

"Consumer reviews of products are one of the key factors I use to filter and select products. However, I cannot get consumer reviews in voice shopping at all." [P2]

Similar to P2, many participants reported that ratings and reviews are peripheral information to a product and yet very important for them to make a purchase decision for a product they have never used before. However, since review texts can be lengthy to display in an audio format, which tends to be linearly presented and time-consuming, voice shopping might skip product reviews, which were valued by many customers.

The lack of transparency in the product recommendation mechanism exacerbated their mistrust of voice shopping. The recommendation mechanism is a crucial part of online shopping. Accurate recommendations can enhance users' shopping willingness and increase their trust in the shopping system. Conversely, poor recommendations can negatively impact user attitudes toward the shopping system. Some participants said that the items recommended to them by voice shopping made them doubt the effectiveness of the recommendation mechanism.

"Once, I wanted to buy napkins. The napkins recommended by Tmall Genie were neither a brand I often buy nor a popular one, but an unknown brand. I was worried about the quality of this unknown brand. I strongly suspected it was a sponsored product on Tmall Genie." [P2]

For P2 and other participants, when customers cannot rationalize the mechanisms behind the recommendations from an online shopping platform, they lose trust not only in the recommended items but the shopping platform overall.

Furthermore, occasional failures in understanding users become another reason why users mistrust voice shopping. In addition to current limitations and constantly improving capabilities in natural language understanding, language differences add complexity to the scenario. Different from many other languages (such as English), Mandarin Chinese has intonation making, and therefore different intonations of the same spelling can become distinct characters and words and thus represent different meanings. In this scenario, the words with the same spelling but different intonations are called homophones. P32 told us his story.

"I told Tmall Genie that I wanted to buy a mainboard (pronounced as Zhǔbǎn in Mandarin), but it was recognized as a bamboo board (also pronounced as Zhúbǎn in Mandarin) by Tmall Genie." [P32]

In this cases, even though Tmall Genie successfully recognized the spelling components of the words, but failed to capture users' actual intention due to the subtle differences in homophones in Mandarin. Therefore, participants were not confident about voice shopping's ability to understand users' requests in product searches.

To summarize, insufficient product information, lack of transparency in product recommendations, and failure in understanding users' requests can reduce users' trust in voice shopping, which may lead to limitation and abandonment of voice shopping.

5. Discussion

Our findings reveal the common usage scenarios for voice shopping and the challenges or breakdowns users typically encountered, including the time-consuming interaction and mistrust of voice shopping. These challenges led participants to limit or even abandon voice shopping. Even though there has been an increasing number of studies on voice shopping and more studies on voice assistants, to the best of our knowledge, little work has studied non-use or abandonment of voice shopping. Therefore, we discuss our findings in relation to prior work on the use and non-use of voice shopping, highlight how our study can advance voice shopping experience design and provide practical implications for designers and manufacturers.

5.1. Technical Robustness as the Foundation

Based on our interviews, we found that the current voice-based interface and interaction present a barrier to using voice shopping, particularly interaction fluency. Our findings confirm with prior work that the lack of technical robustness is a primary factor of users abandoning technology (Clawson et al., 2015). Our study adds to this line of research by confirming the lack of technical robustness contributes to user abandonment of voice shopping, especially the perceived intelligence.

Perceived intelligence mainly refers to participants' perception of natural conversational interaction fluency in voice shopping. Voice shopping using natural language introduces new challenges related to customer queries, like handling mispronounced, mis-expressed, and misunderstood queries, any one of which can affect conversational interaction fluency. Our work took a

step back from prior research in a lab setting that investigates the factors that influence users' acceptance of voice shopping assuming a fluent interaction between users and voice assistants in voice shopping. For example, prior studies manually generated customized scripts for each user, meaning that the user was actually interacting with a human rather than a voice assistant (Rhee & Choi, 2020). Admittedly, the factors highlighted by prior work such as the social relationship between users and voice assistants are important for a satisfying user shopping experience (Rhee & Choi, 2020). However, our participants highlighted the frustration with interaction fluency mostly limited by the natural language understanding and processing capabilities of current conversational agents. We believe that it is important to situate research on user experience and acceptance of voice shopping without isolation it from the actual technical capabilities. In this respect, our findings confirm with Son Nguyen et al. (2021) who recommended that while human attributes of voice agents such as anthropomorphism play an important role in changing the consumer's attitude and behavioral intentions, the impact could easily backfire with technical constraints such as interaction fluency and lack of visual displays. With the advances of natural language processing technologies, it is hopeful that users may embrace smoother fluency and more natural interaction in future voice shopping through voice assistants with more advanced technical capabilities.

Design implication #1: Designers and manufacturers should invest efforts in improving the technical capabilities involved in voice shopping to reduce interaction friction, especially conversational interaction fluency.

5.2. Adaptive Conversational Product Presentation

Our findings show that participants held different opinions towards product presentation – how products are presented or displayed to users – depending on the types of products they browse. Overall, participants encountered few issues in low-involvement purchases. By contrast, for high-involvement purchases, our participants reported frustration either because of too much information read to them all at once, or too little information to trust a product. In a word, due to the audio-only modality, participants demand essential supplementary information about products to be presented in a concise way. Unfortunately, participants' experience showed that current product presentation is mainly inherited from mobile or web commerce, which is a naturally evolving process and understandable. However, it cannot be applied "as is" in

voice-based product presentations (Penha et al., 2022), as the voice-based interface and interaction modality are different from the graphical user interface on mobile and web commerce. Powered by natural language processing and artificial intelligence capabilities, voice shopping is intended to produce a natural conversational interaction with consumers, rather than presenting all information to consumers at once. Along this line, it might be more intuitive to present the product through a conversation-based and consumers' most valued information. To this end, we propose adaptive conversational product presentation.

Adaptive conversational product presentation introduces product information in a conversational style by asking users the type of information that they care about. It might increase user involvement to reduce boredom when listening to a long product description and enhance their perceived control power in the decision-making process by proactively looking for the information they need. Hu et al. (2022) showed that perceived power over AI assistants can reduce risk perception regarding voice shopping. This may alleviate users' mistrust, which is one of the biggest barriers to using voice shopping. Under this premise, the future conversational design could further integrate verbal cues that include more anthropomorphic elements, which was suggested by Son Nguyen et al. (2021). This adaptive conversational product presentation also relies on the advances of natural language processing and technical robustness mentioned earlier.

Design implication #2: Consider integrating conversational product information presentation that adapts to consumer needs to promote user involvement and trust while reducing time in product assessment.

5.3. Cross-platform Product Recommendation

Our findings show that the voice-based product recommendation mechanism is another barrier to using voice shopping. The occasional irrelevant recommendations or lack of transparency in recommendation mechanisms lead to mistrust and discontinuation in voice shopping. Personalized product recommendations are a possible way to mitigate these issues. However, voice shopping poses new challenges for providing personalized recommendations. First, voice shopping is still in its infancy and therefore lacks sufficient consumer behavior data to start with, also known as the cold start problem caused by data sparsity. Second, even with a robust recommendation mechanism, the voice modality makes it difficult for users to compare multiple recommendations, echoing the findings of Penha et al. (2022). Given these, we further propose cross-platform personalized product

recommendation.

Cross-platform personalized product recommendation leverages users' mobile or web shopping history to provide personalized recommendations for voice shopping. It is most feasible for companies that support multi-platform shopping. This is the case for our study since Tmall Genie is under the same parent company as its web- and mobile-based counterparts Taobao, which could be a similar scenario for voice shopping on Alexa and web-/mobile-shopping on Amazon. In addition, although we found that users behave differently in voice shopping versus mobile or web shopping, cross-platform product recommendations could still benefit each other. On the one hand, users do acknowledge the convenience of low-involvement purchases brought by voice shopping, which tend to be routine and repeated. It might be worthwhile transferring existing web-based recommendations to support new lightweight, low-involvement product discovery to introduce some novelty to routine products or services. On the other hand, as for high-involvement purchases, our findings show that users found it more convenient to indicate their purchase interest through product search on voice assistants in a hand-free way. Such scenarios usually happen in specific locations, such as dorms and kitchens, which makes it easier for users to remember and create a shopping list. This initial search provides valuable data on user intent for web or mobile shopping and could support users to complete their purchases on other platforms afterward.

Design implication #3: Consider leveraging cross-platform user purchasing and interaction data to mutually enhance product recommendations on web-, mobile-, and voice-based shopping.

5.4. Limitations

Our study has limitations. First, we focused on user experience on one type of voice assistant – Tmall Genie – and in one culture, similar to prior work which mainly focuses on Alexa and Amazon in voice shopping studies (Ahn et al., 2019; Penha et al., 2022). In the future, we will extend our research to multiple devices among a more diverse culture and user population. In addition, we only talked about voice shopping on standalone voice assistants. We haven't investigated the scenario of mobile voice assistants such as Siri, which might reveal different barriers and opportunities. Another limitation of our study is that most of our participants are aged between 17 and 27, i.e., Generation Z users. Generation Z users are considered the first generation born into a digital world. They are highly educated, technologically savvy, innovative, and creative (Priporas

et al., 2017). As such, their views may differ from other users. Furthermore, most participants placed their voice assistants in a shared space, which makes their usage scenarios likely to be different from a private space. Research on the environmental settings of voice shopping could be future work.

6. Conclusion

In this paper, we conducted a semi-structured interview study with 43 users of voice shopping on Tmall Genie. We first investigated the adoption and abandonment of voice shopping. We found that after a period of use, participants tend to limit or abandon voice hopping due to time-consuming interaction and mistrust of voice shopping. Based on our findings, we discussed how our study could advance understanding of voice shopping and provided design suggestions on technical robustness, adaptive conversational product presentation, and cross-platform product recommendations.

Acknowledgements

References

Abdi, N., Ramokapane, K. M., & Such, J. M. (2019). More than smart speakers: Security and privacy perceptions of smart home personal assistants. *SOUPS@ USENIX Security Symposium*.

Ahn, S., Jo, W., & Chung, D. (2019). Factors affecting users to adopt voice shopping: Empirical evidence from the utaut model. *Journal of Technology Innovation*, *27*(4), 111–144.

Baumer, E. P., Adams, P., Khovanskaya, V. D., Liao, T. C., Smith, M. E., Schwanda Sosik, V., & Williams, K. (2013). Limiting, leaving, and (re) lapsing: An exploration of facebook non-use practices and experiences. *Proceedings of the SIGCHI conference on human factors in computing systems*, 3257–3266.

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021). Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping: Evidence from sem and fuzzy set qualitative comparative

analysis. *International Journal of Information Management*, *58*, 102309.

Bentley, F., Luvogt, C., Silverman, M., Wirasinghe, R., White, B., & Lottridge, D. (2018). Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), 1–24.

Birks, M., & Mills, J. (2015). *Grounded theory: A practical guide*. Sage.

Clawson, J., Pater, J. A., Miller, A. D., Mynatt, E. D., & Mamykina, L. (2015). No longer wearing: Investigating the abandonment of personal health-tracking technologies on craigslist. *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 647–658.

Hu, P., Lu, Y., & Wang, B. (2022). Experiencing power over ai: The fit effect of perceived power and desire for power on consumers' choice for voice shopping. *Computers in Human Behavior*, *128*, 107091.

Li. (2019). Millions in china tapped 'voice shopping' during 11.11 [November 13, 2019].

Newman, P. (2020). The smart speaker report: Smart speakers could be the fastest-growing digital platform ever - here's how to engage with customers through the devices [February 12, 2020].

Oksien, J. (2022). Best smart home devices 2022: Automate your home with these gadgets [August 18, 2022].

Penha, G., Krikon, E., Murdock, V., & Avula, S. (2022). Helping voice shoppers make purchase decisions. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–8.

Priporas, C.-V., Stylos, N., & Fotiadis, A. K. (2017). Generation z consumers' expectations of interactions in smart retailing: A future agenda. *Computers in Human Behavior*, *77*, 374–381.

Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior*, *109*, 106359.

Simms, K. (2019). How voice assistants could change the way we shop [May 15, 2019].

Son Nguyen, H., Mladenow, A., Strauss, C., & Auer-Srnka, K. (2021). Voice commerce: Anthropomorphism using voice assistants. *The*

23rd International Conference on Information Integration and Web Intelligence, 434–442.

Sorensen, K., & Jorgensen, J. J. (2021). "hey alexa, let's shop": Millennials' acceptance of voice-activated shopping. *International Journal of E-Services and Mobile Applications (IJESMA)*, *13*(1), 1–14.

Sudlow-Poole, E. (2022). Conversational commerce: Business model analysis, value-added services market forecasts 2022-2027 [November 14,2022].

Trajkova, M., & Martin-Hammond, A. (2020). "alexa is a toy": Exploring older adults' reasons for using, limiting, and abandoning echo. *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–13.

Tuzovic, S., & Paluch, S. (2018). Alexa-what's on my shopping list? investigating consumer perceptions of voice-controlled devices. *SERVSIG Conference Proceedings 2018 Paris: Opportunities for services in a challenging world*, 107–110.

Voit, A., Niess, J., Eckerth, C., Ernst, M., Weingärtner, H., & Woźniak, P. W. (2020). 'it's not a romantic relationship': Stories of adoption and abandonment of smart speakers at home. *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia*, 71–82.

Wang, J., Yang, H., Shao, R., Abdullah, S., & Sundar, S. S. (2020). Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety. *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–13.

Werner, L., Huang, G., & Pitts, B. J. (2022). Smart speech systems: A focus group study on older adult user and non-user perceptions of speech interfaces. *International Journal of Human–Computer Interaction*, 1–13.

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of consumer research*, *12*(3), 341–352.

Zou, Y., Roundy, K., Tamersoy, A., Shintre, S., Roturier, J., & Schaub, F. (2020). Examining the adoption and abandonment of security, privacy, and identity theft protection practices. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15.