Beyond Browser Online Shopping: Experience Attitude Towards Online 3D Shopping with Conversational Agents

Pedro Valente[✉], Tiago Fornelos, Rafael Ferreira, Diogo Silva, Diogo Tavares,
Nuno Correia, João Magalhães, and Rui Nóbrega

NOVA LINCS, NOVA School of Science and Technology, Caparica, Portugal
{pm.valente,rah.ferreira,dmgc.silva,dc.tavares}@campus.fct.unl.pt,
{nmc,jmag,rui.nobrega}@fct.unl.pt

Abstract. With the growing interest in recreating live and realistic outside experiences within the confines of our homes, the online shopping industry has also been impacted. However, traditional modes of interaction with online storefronts have remained mainly unchanged. This paper studies the factors influencing user experience and interaction in 3D virtual stores. We created a prototype that uses a 3D virtual environment for users to navigate, purchase items, and communicate with a conversational agent. The designed interface was tested by studying a set of variables, including the user's interaction medium with the conversational agent, the movement method within the store, the user's perception of the conversational agent, the conversational agent's usability, and the store items' visual representation. Through the user study, we gained insights into the factors that guide the user's experience in 3D virtual stores. We concluded that 80% of users preferred less intrusive conversational agents and 75% preferred agents that did not hide visual elements. Additionally, 80% of the participants favored combining 2D and 3D visualization techniques.

Keywords: 3D Interaction · ChatBot · Avatars · Online Shopping

1 Introduction

There is a growing desire to replicate outdoor experiences within the comfort of our own homes. This pursuit of enhancing the online living experience has also impacted the online shopping industry [32]. Some research has begun to explore different methods of interacting with these storefronts, from virtual reality [23] to other non-traditional approaches such as 3D virtual environments [12]. By contrast, we have well-defined guidelines for developing browser interfaces for commerce [20]. Additionally, conversational agents have become an integral part of the online shopping experience [1], providing customers with 24/7 assistance as they interact with storefronts.

Fig. 1. The studied factors: **(1)** the dialogue interface; **(2)** the locomotion system; **(3)** the agent's visual presentation; **(4)** the agent's functionality; **(5)** the product visualization;

To this end, our goal was to thoroughly examine the factors influencing user experience and interaction in 3D virtual stores with conversational agents. Furthermore, we aim to make online shopping more accessible and convenient for consumers and summarize our findings into design guidelines. Considering this, we conducted a two-part user study, starting with formative interviews that gave us insight into the most relevant features users look for in online and physical stores. These interviews enabled us to create a 3D shopping application that uses the online store catalog from Farfetch, our project's partner, and compile five variables that guide the user's interaction.

Two of the variables cover the design of the virtual store, and three, the user's interaction with the conversational agent in the virtual store environment: **(V1)** the preferred method of interaction for users with a conversational agent: textbox or voice interface; **(V2)** the most effective method for users to navigate a virtual store using keyboard and mouse controls or teleport to specific points; **(V3)** the most effective representation of the conversational agent, a humanoid avatar, or a text-based representation; **(V4)** the conversational agent's capacity to substitute parts of the visual interface, a dialogue based interface compared to a visual interface; **(V5)** the Visual representation of the items in the store, a context window or a 3D model. These variables represent the factors we identified as crucial transition points from conventional browser interfaces to 3D virtual environments (Fig. 1).

By studying these factors, we wanted to answer the following research questions: (**RQ1**) How should these factors drive the UX design of virtual stores with conversational agents? (**RQ2**) How do these variables rank by their importance regarding the design of Virtual Stores with conversational agents?

Through this research, we sought insights into which of the selected variables should be prioritized to improve the user's experience when designing 3D Virtual Stores with conversational agents and what impact these variables have on the UX design of these interfaces.

We start this article with a review of the related work (Sect. 2), followed by an examination of the formative interviews (Sect. 3). Next, we introduce the

interface we developed (Sect. 4) and discuss the results of a subsequent user study (Sect. 5). Finally, we analyze our findings (Sect. 6) and present our conclusions (Sect. 7).

2   Related Work

Online shopping has followed a steady browser-based 2D interface recipe in recent years. To change this paradigm, some companies have recently attempted to create 3D virtual environments for their online marketplaces [19,32]. This has prompted researchers to consider the most effective methods for designing virtual social environments, specifically virtual simulated stores [7].

This field has seen considerable progress since the early 1990s,s, as evidenced by works such as Burke et al.'s [5] original publication, where the authors used a simulated environment to study consumer behavior. Recent studies have demonstrated that users feel more comfortable navigating virtual stores through VR [27], indicating the potential for researching this area.

Despite the advancements in these applications' immersive and interactive features, there has been limited progress in providing task-specific assistance to users. However, conversational agents can offer users additional support in completing tasks such as purchasing products [30]. Additionally, conversational agents have proven valuable in providing systems with intelligence and natural language capabilities [15]. These tools can process natural language inputs and give innate responses, enabling a conversation with the user [9]. Furthermore, this technology can automate interactions between a company and its customers, creating the illusion of conversation with a human being [6].

The traditional chatbox is often the first consideration when discussing conversational agent user interfaces [24]. However, alternative forms of interaction may be more beneficial in some cases. One example is an interface developed by Quarteroni et al. [26]. This interface enlarged the chat window into two sections: a text box on the left and a panel on the right to present additional information about the conversation context, such as links to web pages or more informative answers to user questions.

Vaddadi et al.'s [31] conducted a similar research project. They developed a wrapper for an online shopping assistant on mobile devices that incorporates buttons, cards, and text messages. The researchers found that buttons helped select product sizes, as it is more convenient for the conversational agent to display the available sizes as buttons for the user to choose, rather than requiring the user to type in the size. The cards show images or videos of requested products, links, and text.

Likewise, Pricilla et al. [25] also researched this field and developed a mobile chatbot interface for online fashion shopping. This team took a user-centered approach to the conversational agent's development and proposed a swiping list of messages containing various products presented by the agent. Each item includes the product image, information about the product, and a link to the web page or a more informative view.

Another critical question surrounding the presence of conversational agents in virtual spaces is how we present this type of interaction in 3D environments. The most common way is using an embodied virtual agent (EVA) [14]. EVAs are an interface where an Avatar[1] physically represents the agent in the virtual space. This avatar is usually presented as a human to create a more empathetic experience.

There have been multiple attempts to implement EVAs before, with one of the first attempts by Nijholt et al. [21], where the authors experimented with blemishing traditional dialog with a virtual environment populated with the avatar of the agent. The authors observed the possible potential for these interfaces to be used in helping people with disabilities. Another study by Martin Holzwarth et al. [11] showed that using an avatar in web-based information increased the customer's satisfaction with the retailer, attitude toward the product, and purchase intention.

Some recent research has focused on whether these interfaces can provide a better experience than regular dialog interfaces. For example, Jessica et al. [28] focused on questioning parents about how the agent's interface presentation could affect the parent's perception of a specific agent and whether the interface was a toy. They did this by questioning parents about their attitude toward multiple interfaces, including toys with chatbot functionalities. Further research has been done on the usefulness of this type of interface. Yet, in Li et al.'s [13] research, the authors conclude that the physical embodiment alone does not provide a better social presence when interacting with chatbots.

A major problem with these interfaces is that many of the used avatars fall into the uncanny valley [18]. In Nowak et al.'s [22] work, the authors observed that when EVAs try to have a more anthropomorphic design, they fall short of being realistic because they create higher expectations, making them more challenging to meet without complex technological features. Similar results can be seen in Groom et al.'s [10] research. Furthermore, in Ben Mimoun et al.'s [2] work, the authors identified another problem: many EVAs fail to meet the user's expectations when providing a realistic interaction, leading to a more frustrating interaction.

Another critical question is how we should show shopping items in the context of a 3D virtual world. In most cases, in online stores, items are shown in a 2D view with no additional 3D information, so most catalogs only contain information about the 2D representation of the items. A common technique is to have the 2D images of the items mapped onto a 3D model. This was what Aymen Mir et al. [17] did in *Pix2Surf*. Their open-source algorithm was implemented to handle input images of t-shirts, shorts, and pants, being able to render 3D models of mannequins with different poses.

---

[1] Avatar in this context is used to refer to the virtual representation of the interactable agent in the virtual world.
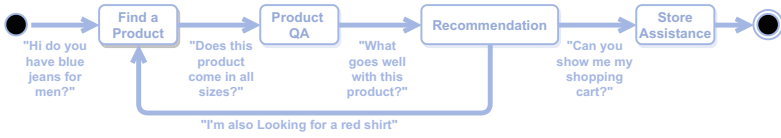
Fig. 2. Dialog graph for a common interaction with the conversational agent, where we can see all the tasks the agent has to complete

3   On the Design of Virtual Stores with Conversational Agents

Our focus was on the fashion domain. In this context, creating a conversational agent primarily consists of creating a dialog interaction that can assist the user in finding and buying items in the store. The conversational agent should be able to perform tasks grouped in the following categories:**(1)** store assistance, meaning assisting with tasks related to the main interface, **(2)** product recommendations, **(3)** product question-answering (QA), this is, answering questions about the characteristics of a particular product, and **(4)** finding products in the store. Figure 2 shows an example of a dialog graph from a conversation.

Furthermore, we designed the interface in such a way that users could simulate the purchase of items, navigate the store, and interact with the conversational agent. A major part of designing this interface is understanding the user's expectations. To achieve this, **we conducted formative interviews with six participants**. We deliberately sought participants with previous experience buying clothes online.

All the study participants were female and bought clothes online at least four times per year, with one of the users buying 24 items per year. Furthermore, all the participants had had previous experience interacting with a conversational agent. The participants also varied in age. Three participants were between 21 and 27, one was less than 21, and the other two were above 27.

In the interview, we showed users three distinct scenarios. (1st) The first scenario was focused on buying clothes in a browser store, (2nd) the second was buying clothes in a physical store, (3rd) and the third was purchasing clothes with the assistance of a voice agent. In each scenario, users were asked what their main buying habits were when shopping for clothes and what information they expected to be available in the described scenario. Furthermore, the interviewees were also asked what advantages they could identify in buying clothes online and in physical stores.

Some noteworthy findings were the following. When asked about their online practices in the first scenario, a common answer was looking first at sales and discounted items. When asked what information the users found relevant, two participants answered that shipping information was the most important. Two others said they wished that stores had better recommendation systems. For instance, a user said that they valued "(...) *showing me relevant items that have a similar style or are similar to the ones I've been searching* (...)".

In the second scenario, when describing their practices, four users said they usually go around the store looking for interesting items. Regarding what information they found relevant, three users said they do not seek additional information when buying clothes in physical stores. One said they usually avoid interacting with store assistants. For instance, a user said, "*I don't usually ask for anything from the retail worker besides when I want a clothing item in a different size, and cannot find it. (...)*".

When shown the third scenario and asked what information they expected from the conversational agent, three participants said they would ask for specific details regarding the product they were trying to buy, either shipping information or specific features. Two participants also said they would ask for recommendations or items that go well with what they previously saw or bought, "*I'd like to ask for possible suggestions based on the things that I've previously seen, or the articles of clothing we've talked about. (...)*".

When asked about the benefits of buying clothes in physical stores, all users answered unanimously that the only benefit is that they can try the items immediately without waiting for them, for instance, "*Definitely seeing how the clothes fit me. That's the only downside of buying them online. Sometimes an item looks really good on the model but doesn't fit properly on my body. (...)*".

When asked about the benefits of buying clothes online, four participants answered that a major advantage is avoiding interacting with other people, the assistants or other people in the store, "*I like the convenience of being able to shop from home, not having to deal with queues and other people. (...)*".

The interviews were a valuable tool in formulating our research questions. Through the interviews, we identified some critical factors, which later informed the design of our ranking tasks in the user study. Furthermore, we also saw that users avoid interacting with store assistants in the real world.

Therefore, when studying conversational agents within a virtual store environment, we aimed to test various levels of interactivity and the use of different representations, each with varying levels of presence and multimodality. Three of our research variables explored the extent to which the conversational agent's interaction should be hidden or revealed. The study also included a task that evaluates the store's usability and the effectiveness of product visualization, two other concerns raised during the interviewing process.

4   System Description of the 3D Shopping Experience

The conceived interface is a 3D virtual store where the user navigates in the first person. We created a 3D store environment (Sect. 4.1) and implemented multiple methods of locomotion (Sect. 4.2), different dialog interfaces (Sect. 4.3), and multiple visualization techniques (Sect. 4.4). In the sections below, we will cover every element of the developed interface.
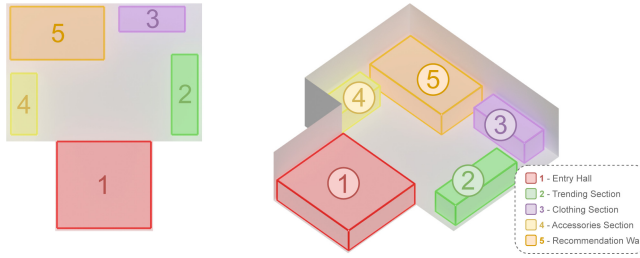
Fig. 3. The store's overall layout, with a top-down view on the **left** an isometric view on the **right**

4.1 Virtual Store Environment

The virtual environment can be divided into multiple sections. A section is an area of the store. Each section can contain a variable number of display screens, including none, that show a preview of the available items. These areas are organized based on the type of items they contain and what activities can be performed in that section. The store has five sections (Fig. 3):

– **Entry Hall** (Red zone - 1 in Fig. 3): This section corresponds to the store's starting area. From here, they can see every other section of the store. It is also the only section that does not contain any article of clothing;
– **Trending Section** (Green zone - 2 in Fig. 3): In this section, users can visualize a set of premade outfits that correspond to the trending outfits (Fig. 4a);
– **Clothing Section** (Purple zone - 3 in Fig. 3): This section of the store corresponds to the place where users can visualize multiple clothing items, with every article category mixed in the same display window (Fig. 4b);
– **Accessories Section** (Yellow zone - 4 in Fig. 3): Here, users can find items that do not fit in the clothing item category, such as bags and watches (Fig. 4c);
– **Recommendation Wall** (Orange zone - 5 in Fig. 3): In this section, users can use a set of three mannequins to preview outfits with a three-dimensional presentation.

4.2 Virtual Store Navigation

To navigate the store, the user can use a mouse or touchscreen. To facilitate navigation, we created a point-of-interest (PoI) system. Every section of the store has its point of interest. To navigate to a specific PoI, the user must select one of the 3D arrows in the interface by pressing it with their finger or the mouse cursor (Fig. 5a).

Each PoI also defines a focus point, so the camera rotates to shift the user's attention toward a specific position when traveling to a PoI. The camera is

(a) Trending Section   (b) Clothing Section   (c) Accessories Section

Fig. 4. Some of the multiple sections of the 3D store

controlled by clicking and dragging the mouse. To smooth the navigation around the store's geometry, we used a pathfinding algorithm to find the shortest path between two points of interest. Then we smoothed the navigation along the track with a bezier curve (Fig. 5b).

We have incorporated an alternative locomotion system within our study, namely a conventional first-person control scheme utilizing a keyboard and mouse. In this system, the camera turns using a mouse, while navigation uses the arrow keys on the keyboard. This solution is an ideal benchmark due to its extensive adoption in video games over the course of several decades. As a result, users who are familiar with this scheme may have developed ingrained motor skills or muscle memory and perform better [16, 29].
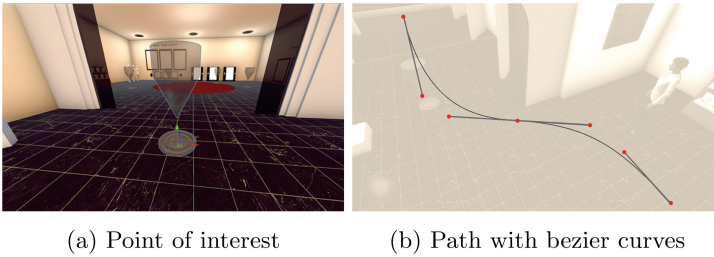

(a) Point of interest   (b) Path with bezier curves

Fig. 5. The various components that make the point of interest system

4.3   The Conversational Agent

The conversational agent was designed to effectively understand and respond to the user's intent using Automated Speech Recognizers and multiple Natural Language Processing (NLP) algorithms [7]. To interact with the conversational agent, we implemented a chatbox that contains the history of the conversation between the user and the agent located in the bottom right corner of the screen.

When designing the dialog interface, we had to present the agent's responses to the user. These responses are a mixture of text, actions to be performed in the interface, and product recommendations. Therefore, we implemented three

interfaces (Fig. 6). The first one uses the chatbox interface. Here the text is presented as a dialog bubble in the chat window that sometimes contains a preview of specific products (Fig. 6a).

We also implemented a speech interface using the Cortana voice API[2]. Users can activate this interface by pressing the microphone icon in the screen's bottom right corner, which will bring up a window displaying the system's detected voice input. This interface aims to provide a more multimodal interaction while removing the necessity for an on-screen chatbox. This speech interface uses a visual representation we called the *subtitle interface*, where text is presented at the bottom of the screen, similar to a movie's subtitles, and recommendations are shown in a context window above the text (Fig. 6b).

Lastly, we experimented with having a fully embodied conversational agent represented by a hologram (Fig. 6c). Each section of the store has its point. At runtime, the system deciphers the closest visible point to the user's camera and instantiates the avatar. When the system receives or sends a message, the assistant does an animation to give the user feedback. The conversational agent's text is shown as a speech bubble floating over the avatar in the screen space. The recommendations are offered inside the bubble.
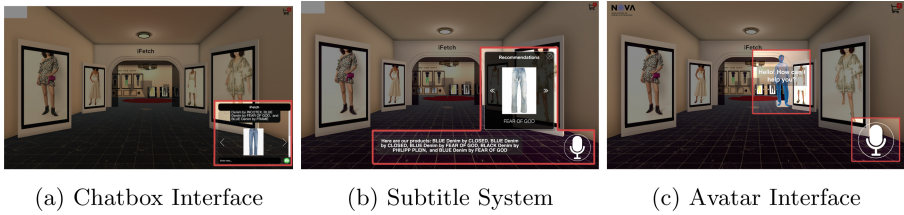

(a) Chatbox Interface   (b) Subtitle System   (c) Avatar Interface

Fig. 6. The multiple dialog interfaces

4.4 Product Visualization

A problem with migrating from a traditional 2D viewport to 3D is how we should display the products available around the store and what product information should be presented to the user. The items around the store combine items from an online fashion store catalog with manually selected items. As a result, we can see a representative mix of each type of clothing.

Multiple display screens around the store show the available items, as seen in Fig. 7a. Each section of the store has its own set of displays. Objects are displayed in frames with a 2D image of the product. Clicking on one of these frames opens a context window containing information about the selected item. Here we display the product's brand, price, available sizes, and a short description.

---

[2] Cortana Speech detection, Unity API, https://learn.microsoft.com/en-us/windows/mixed-reality/develop/unity/voice-input, Last Access 2023.

An alternative approach to presenting these items is to show them in a 3D viewport. To achieve this, we mapped the images of our 2D catalog to a 3D mannequin. We did this using Pix2Surf[3] [17]. To work with this model, we had to restrict our catalog further, as it only works with short-sleeved t-shirts, trousers, and shorts.

To see an item in a mannequin, one has to select the item they want to preview and mark it as "*Interested.*" This will add that item to the recommendation tab. After that, in the Recommendation Section, one can drag and drop an item from one frame to another, updating the mannequin's clothes. Furthermore, depending on the interface, recommendations are shown as a special message with arrows and cards, where every card has the item's preview and name. Alternatively, recommendations can be displayed in a context window with arrows and information about the articles (Fig. 7b).
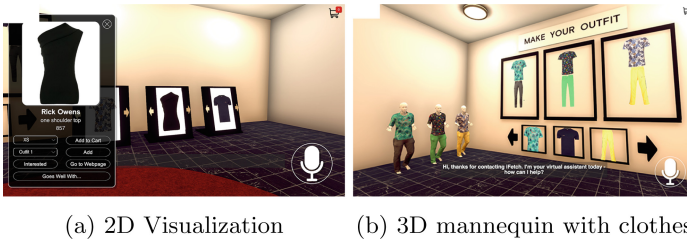

(a) 2D Visualization   (b) 3D mannequin with clothes

Fig. 7. The multiple visualization interfaces

5   Evaluation

Considering the described interface, we tested the variables stated in Sect. 1. To do so, we conducted a user study with multiple interfaces, two for each variable, interfaces A and B. The variables can be seen in Table 1.

Our study focused on whether these variables could affect the user's experience while interacting with the 3D virtual store and how they stack against each other to improve their experience.

5.1   Protocol

When designing the questionnaire for our user study, we based many of our questions on existing literature [3,4] and the interviews that we previously conducted (Sect. 3). The data collected from the users was anonymized, and users were informed that they could leave at any point during the test.

The experience was composed of five tasks (T1, T2, T3, T4, and T5) with a climatization task (T0). T1 through 5 were meant to evaluate each of the

---

[3] Pix2Surf, repository, https://github.com/aymenmir1/pix2surf, Last Access: 2023.

Table 1. The variables being studied and their respective interfaces

Variable	Interface A	Interface B
V1: The medium that the user employs to communicate with the conversational agent	**1A**: Speech Interface	**1B**: Chatbox
V2: The most effective method of locomotion in the store	**2A**: Mouse and Keyboard	**2B**: Points of Interest
V3: The way the user visualizes the agent	**3A**: Humanoid Avatar	**3B**: Text-based Representation
V4: The conversational agent's capacity to substitute parts of the visual interface	**4A**: Dialogue Based Interface	**4B**: Visual Interface
V5: The Visual representation of the items in the store	**5A**: Context Window	**5B**: 3D Model

corresponding variables. For each of the five main tasks, users had to test two interfaces, A and B. The order was alternated in a Latin Square order to reduce learning bias. For the context of every task, A differed from B, and every task was independent of the other.

After every task, the user answered ten questions, some regarding Interface A or B and some about both interfaces. Questions comprised Likert scale evaluations (1 to 5) and ranking questions. At the end of the questionnaire, users would evaluate both interfaces using a Likert scale (1 to 5) and are asked what was their favorite. For the fifth task, users had to rank the features of both interfaces. At the end of the test, users responded to some questions about a complete version of the interface, including the System Usability Scale (SUS) [4].

For every task, we annotated whether the user could finish the task (if they finished the task in less than 4 min) and if they asked for guidance. For T1 and T2, we recorded the time it took for participants to finish.

The setup for the experience was comprised of a computer where the user would test the multiple interfaces. Every user also used a microphone to communicate with the conversational agent. Users were also given paper instructions containing all the tasks they had to perform and a map of the store with every section labeled. Users could consult this map at any time during testing.

5.2   The Population

Users were selected by surveying college students. All the participants had at least a K12 education level and were fluent in English. The study was conducted with 20 users, 11 female (55%) and 9 male users (45%). Users were between the ages of 19 and 49. Many users had rarely interacted with a conversational agent before (30%) or interacted yearly (25%). The rest of the participants interacted monthly (20%) or weekly (25%). Most users played video games, with only 1 (5%) saying they rarely played. 75% said they played games daily or weekly, and the rest played monthly or yearly. We further questioned the users about how frequently they play FPS games. Although 35% users still played FPS games weekly, 25% said they didn't play FPS games.
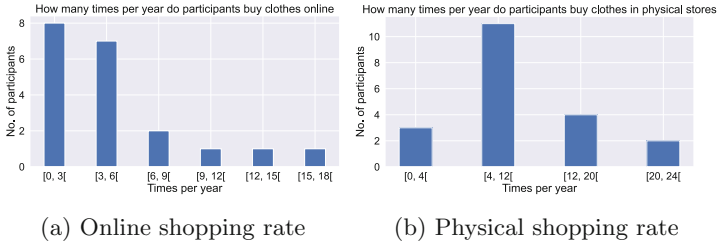
How many times per year do participants buy clothes online

How many times per year do participants buy clothes in physical stores

(a) Online shopping rate   (b) Physical shopping rate

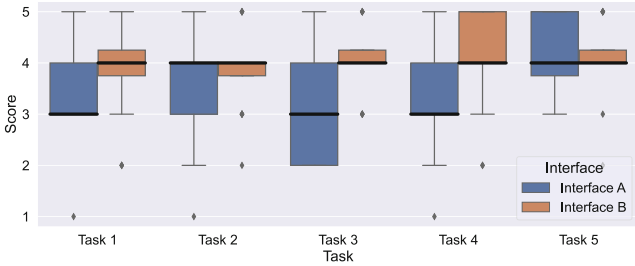**Fig. 8.** How many times do users buy items in online (a) and physical (b) stores?

Fig. 9. Interface ratings for every task (median in bold)

Furthermore, when asked if they follow the most influential trends in fashion, 85% of the users answered no. Following this trend, 5 users said they do not buy clothes in online stores. Still, the rest of the users said they buy at least one clothing item per year online, with a user even they buy around 15 clothing items per year in online stores (Fig. 8a). Moreover, when asked how frequently users bought clothes in physical stores, the most common answer (45%) was between 4 and 11 times per year (Fig. 8b).

5.3   Results

As previously mentioned, users completed five tasks while freely interacting with the virtual store. Starting with **T1**, users had to interact with the conversational agent using a voice interface (1A) and a chatbox (1B). When observing the overall scores of both interfaces, 55% of users said they preferred interface 1B. We can see this reflected in the general scores of the two interfaces, where we could see that users rated 1B higher but without having a notable enough statistical significance ($p=0.176$) (Fig. 9). Furthermore, on average, users took more time to complete the task using 1B. Yet, this difference was not major at a 5% significance level. See Table 2.

An observation where we saw a major difference was the number of times the users had to repeat commands. In 1A, 90% of users had to repeat utterances, while in 1B, only 35% had to repeat. Repeated commands happened either

when the agent didn't understand the user's intent or when the voice detection algorithm didn't correctly pick up the user's utterance.

Regarding **T2**, users had to perform the task using traditional FPS controls (2A) and the PoI system (2B). Although a larger number of users preferred 2B (55%), it was not a large difference. This was reflected in the data where 2B had marginally better results than 2A (Fig. 9). Furthermore, observing the times in Table 2 we can see that 2A and 2B had similar times. When analyzing this data, we must remember that many users are familiar with this interface type, as seen in Sect. 5.2.

On task **T3**, we tested the presence of the agent's avatar in the store, where one interface had the avatar (3A) while the other didn't (3B). 80% of the users said they preferred 3B to 3A. This is observed in the rest of the collected data (Fig. 9). One such example is seen in the scores of each interface, where users rated 3B much higher than 3A (p=0.007). The preference for 3B is further verified by the users' responses to questions about readability and uncanniness. See the first two rows in Table 3.

When looking at these values, we can infer that users felt more comfortable interacting with 3B than with 3A, yet they didn't feel as if the dialog was disconnected from the store. During testing, users even commented on the presence of the avatar in the store being weird or uncomfortable. When we observe the boxplots for Q6 (I liked the presence of the avatar in the store) and Q7 (The avatar contributed to the experience of interacting with the chatbot) (Fig. 10), we can see that users did not enjoy the presence of the avatar in the store.

In **T4**, users were asked to complete the task with the assistance of the conversational agent (4A) and without (4B). When users were asked what their favorite interface was, most said they preferred 4B to 4A (75%). This answer is well represented in the rating given by the users, where we verified a significant difference between the scores of both interfaces (p=0.006). 4A had a median score of three, while 4B had a median score of four (Table 4).

Although we saw this significant difference in the ratings, this did not extend to the answers users gave in questions about frequency of use and cumbersomeness of the interface (see the first two rows of Table 4). Furthermore, when the users were asked whether they agreed with "I found the interaction with the agent unnecessary," they answered with a mode of 4 and a median of 3.5. This indicates that when presented with the option of utilizing the conversational agent, the participants preferred not utilizing it.

Table 2. Average time it took for users to finish each task (T1 and T2), the standard deviation, and the t-test p-value for every interface

#	time (m)	mean	Std. Dev.	t-test
T1 : 1A	Complete T1 with 1A	**02:12,9**	00:51,6	$p = 0,099$
T1 : 1B	Complete T1 with 1B	02:55,5	00:57,9	$p > 0,05$
T2 : 2A	Complete T2 with 2A	**00:46,9**	00:23,4	$p = 0,147$
T2 : 2B	Complete T2 with 2B	01:05,5	00:31,0	$p > 0,05$

Table 3. Median, first quartile ($Q1$), third quartile ($Q3$), and chi-square test p-value (X^2) of the scores of both interfaces in questions about readability, uncanniness, frequency of use and consistency in Task 3

Interface	Question	median	Q1	Q3	X^2
3A	I could read the text well.	4	3	5	$p = 0{,}015$
3B		**5**	4	5	$p < 0{,}05$
3A	I found the experience unnatural.	3	2	4,25	$p = 0{,}014$
3B		**2**	1	3	$p < 0{,}05$
3A	I want to use the chatbot more often.	3	2	3,25	$p = 0{,}053$
3B		**3.5**	3	4	$p > 0{,}05$
3A	The dialog felt disconnected.	2	1	3	$p = 0{,}825$
3B		2	1	3	$p > 0{,}05$

Table 4. Median, first quartile ($Q1$), third quartile ($Q3$), and chi-square test p-value (X^2) of the scores of both interfaces in questions about frequency, cumbersomeness, and the ratings in Task 4

Interface	Question	median	Q1	Q3	X^2
4A	I would use this interface frequently.	3,5	3	4	$p = 0{,}372$
4B		**4**	3	5	$p > 0{,}05$
4A	I found the interface cumbersome.	2	1,75	3	$p = 0{,}470$
4B		**1,5**	1	3	$p > 0{,}05$
4A	How would you rate this interface?	3	3	4	$p = 0{,}006$
4B		**4**	4	5	$p < 0{,}05$

In **T5**, the participants were presented with two distinct interfaces for visualizing clothing items, a traditional visualization interface (5A) and a 3D item visualization interface using a mannequin (5B). When asked to indicate their preferred interface, participants had the option to select 5A, 5B, or both interfaces simultaneously. Results of the study revealed that 80% of the participants preferred utilizing both 5A and 5B simultaneously.
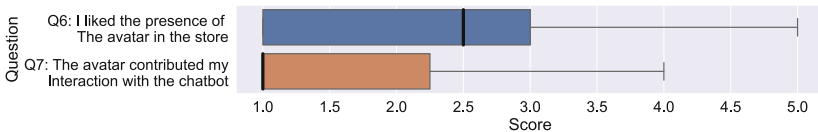

Fig. 10. Data for presence and interaction in T3 (median in bold)

Furthermore, the participants were requested to rank various features from 5A and 5B (Fig. 11). These features were product information and visualization techniques. About 5A, 70% of the participants considered the price the most crucial feature to be shown on the user interface. At the same time, the material used was considered the least important feature (35%) to be shown. The participants were also asked about which features they would include in the visualization of the product. Some examples of the mentioned features were the brand of the product and a size guide.

Concerning 5B, the participants deemed that the most salient features were the ability to map the clothes directly onto an image of themselves (25%) and a 360-degree view of the mannequin with the clothes (25%). However, unlike 5A, there was no consensus among the participants as to which feature was the most desirable, as illustrated in Fig. 11b. Additionally, features such as having a 360-degree view of the mannequin and the ability to adjust the clothes according to the users' size were not rated as the least important feature. In contrast, 35% of the participants stated that having multiple lighting options in the mannequin was the least important feature.
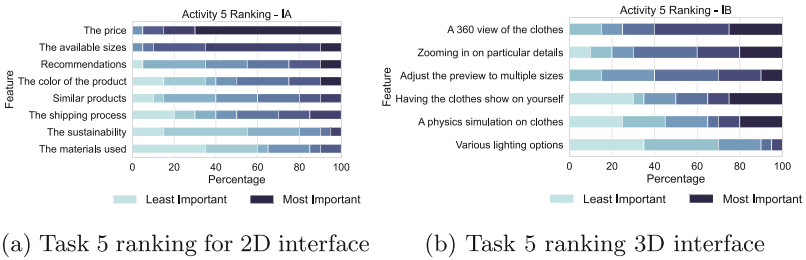

(a) Task 5 ranking for 2D interface   (b) Task 5 ranking 3D interface

Fig. 11. Task 5 rankings, for 2D and 3D product visualization

After the questionnaire, the participants were instructed to rank every task they performed during the study. The results of this ranking can be observed in Fig. 12. Upon examination of this graph, we can see that the participants prioritized the visualization of items over all other factors. Additionally, although it elicited the strongest reaction from the participants, the agent's avatar was primarily considered the least important feature, with 75% of the participants rating it as the least important.

The SUS score was calculated at the end of the test. We obtained an average SUS score of 70.625 with a standard deviation of 9,516. The lowest score we obtained was 45, and the highest was 82,5. For reference, a study by Debjyoti Ghosh et al. [8] found that Siri had a SUS mean value of 54,167.
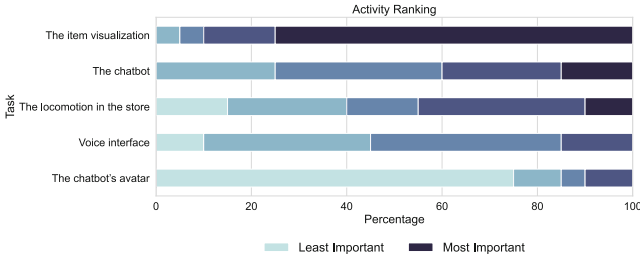
Fig. 12. Ranking of each task

6 Discussion

Our objective was to determine which factors are crucial when designing and developing 3D virtual stores and which can be ignored. By examining the data collected from the study, we will gain insights into the most effective solutions for enhancing the user experience in 3D virtual stores and how to prioritize the importance of different factors when planning such interfaces.

We observed no significant findings after examining the results from **T1**. However, we saw a trend where participants tended to prefer the chatbox interface. This may be attributed to many users repeating commands when interacting with the voice interface, as reported in Sect. 5.3. Specifically, 18 participants had to repeat their utterances in 1A, while 13 had to repeat them in 1B. This caused users to become frustrated with the system while testing 1A and react more negatively toward this interface. A common error we observed was the voice-to-text algorithm misreading the user's words, for example, interpreting "Nike shorts" as "knight shorts". Despite the conversational agent being designed to handle this type of error, when users saw their utterances misspelled, they still felt the need to repeat their command, even when the system responded correctly. This suggests that, in future designs, hiding the user's utterance from them might improve the user experience and reduce frustration.

In **T2**, participants, after answering the questionnaire, were asked a follow-up question regarding their preference of interface if 2B (the Point of Interest system) were to be on a tablet device. In response, 80% of users said they prefer 2B to 2A. This represents a major difference from the results obtained when tested on a laptop, where 55% preferred 2B over 2A.

Given the increasing impact of tablet interfaces on e-commerce, as noted in previous studies [33], this large difference in user preference is noteworthy and merits further investigation. We posit that the improved reception of 2B as a tablet interface may be due to its reduced degrees of freedom. When using touchscreen devices, users are limited to controlling the camera's orientation with virtual inputs. Additional degrees of freedom for user locomotion would require additional clutter in the user interface. This explanation may also be applied to the voice interface tested in **T1**, as the inclusion of a chatbox would imply the presence of a virtual keyboard on the screen.

A notable finding in our study was that participants in **T3** did not appreciate the avatar's presence in the store, as outlined in Sect. 5.3. We attribute this adverse reaction to two factors. First, the avatar used to represent the conversational agent in the store employed a semi-realistic, anthropomorphic model that attempted to mimic a hologram. This model made participants uneasy, as they felt the chosen representation was unnatural, which is consistent with the findings of Nowak et al.'s [22] work on the uncanny valley applied to avatars.

In addition, the avatar's non-interactive nature and inability to create empathy with users contributed to its negative reception. Looking at Sect. 5.3, we obtained a negative response when participants were asked if the avatar had a positive effect on their interaction (Fig. 10). Furthermore, when considering this result in conjunction with the participants rating this aspect of the interface as the least important (Fig. 12), we can infer that users found the avatar unnatural and unnecessary. With this in mind, we can conclude that when designing this type of interface, this aspect should not be the development focus if we cannot ensure a realistic and meaningful interaction.

Another noteworthy finding was in **T4**, where users expressed a preference for the interaction where they didn't have to use the conversational agent, in contrast to the one in which they did (Table 4). Users performed a recommendation task, they either asked the agent for clothing items that would complement a selected product or clicked a button on the visual UI. We posit that this outcome resulted from users perceiving the interaction as unnecessarily complex for a task that could be accomplished by simply pressing a button. Although some studies [6] have shown the benefit of using chatbox interfaces to aid users, they should not be seen as alternatives to traditional interfaces.

In **T5**, participants still considered the visualization of the product the most important feature (Fig. 12). Additionally, users demonstrated a high receptivity to using a three-dimensional representation of the item they were seeking to purchase, indicating that this type of visualization may offer a superior solution to traditional visualization methods.

Information was gathered during the data collection process to divide the study population into sub-groups. However, upon analysis of the data, we observed no statistically meaningful differences among the sub-groups based on variables such as age, gender, frequency of interaction with games and chatbots, and frequency of usage of online stores.

7 Conclusion

With the valuable insights we gained from our research on creating 3D virtual stores with conversational agents, we identified several domains that require further inquiry. Primarily, while our study encompassed a broad range of variables, other factors may require investigation in this field, for example, the capacity of the conversational agent to interrupt the user's interaction. Furthermore, we acknowledge that delving deeper into 3D visualization techniques can reveal the complete advantages of utilizing this interface.

We studied the impact of several variables on the user's experience when interacting with a 3D virtual store with conversational agents in the fashion domain. The study found that the interface type, either a chatbox or a speech interface, impacted the user experience. Participants preferred the chatbox interface, possibly due to the repetition of commands in the voice interface. The study also revealed that the point-of-interest system was helpful for users (Fig. 9). The study also found that intrusive agents negatively impacted the user's experience (Fig. 10). The study also suggested that conversational agents should be unobtrusive in their visual representation and not hide any features of the visual interface (Fig. 9) (**RQ1**).

Our research also revealed that 3D visualization techniques in a virtual store environment significantly impact the user's shopping experience (Fig. 12). This feature is perceived as crucial by participants when shopping for clothes online and should be prioritized in designing a 3D virtual store. Furthermore, our study suggests that the point-of-interest system benefits users (Fig. 9). In addition, we observed that users generally prefer the chatbox interface over the speech interface (Fig. 9), and it was considered one of the least important features (Fig. 12). Beyond this, our study showed that the agent's presentation should not be prioritized as it could harm the user's experience. Also, hiding the visual elements of the interface can lead to a more frustrating interaction (Fig. 9). However, users still value using the conversational agent as an alternative to the main interface (Fig. 12) (**RQ2**). We can summarize our findings into the following **guidelines**:

1. We recommend using a chatbox instead of a speech interface for user interaction, as the latter may elicit a higher frequency of utterance repetition and subsequent user frustration.
2. Implement a point-of-interest system for navigating the virtual store. Users often prefer this system, and it's more suitable for touchscreens.
3. It is crucial to refrain from using intrusive agents, as users strongly rejected them and found them irrelevant to their interaction.
4. Conversational agents must not obscure visual interface features through dialogue. It will adversely affect the user's experience.
5. Emphasize 3D visualization techniques, such as mapping clothes to 3D models that allow you to rotate and zoom in on particular details, allowing for meticulous examination of specific details.

In conclusion, we highlight the preference for a chatbox interface over a voice interface, the importance of a point-of-interest system, the negative effect of intrusive agents, the need to avoid obscuring visual interface features, and the significance of emphasizing 3D visualization techniques.

References

1. Bavaresco, R., et al.: Conversational agents in business: a systematic literature review and future research directions. Comput. Sci. Rev. **36**, 100239 (2020)
2. Ben Mimoun, M.S., Poncin, I., Garnier, M.: Case study-embodied virtual agents: an analysis on reasons for failure. J. Retailing Consum. Serv. **19**(6), 605–612 (2012). https://doi.org/10.1016/j.jretconser.2012.07.006
3. Borsci, S., et al.: The chatbot usability scale: the design and pilot of a usability scale for interaction with AI-based conversational agents. Personal Ubiquitous Comput. **26**(1), 95–119 (2022)
4. Brooke, J., et al.: SUS-a quick and dirty usability scale. Usability Eval. Indust. **189**(194), 4–7 (1996)
5. Burke, R.R., Harlam, B.A., Kahn, B.E., Lodish, L.M.: Comparing Dynamic Consumer Choice in Real and Computer-simulated Environments. J. Consum. Res. **19**(1), 71–82 (1992). https://doi.org/10.1086/209287
6. van Eeuwen, M.: Mobile conversational commerce: messenger chatbots as the next interface between businesses and consumers (2017). www.essay.utwente.nl/71706/
7. Fornelos, T., et al.: A conversational shopping assistant for online virtual stores. Association for Computing Machinery. https://doi.org/10.1145/3503161.3547738
8. Ghosh, D., Foong, P.S., Zhang, S., Zhao, S.: Assessing the utility of the system usability scale for evaluating voice-based user interfaces. In: Proceedings of the Sixth International Symposium of Chinese CHI (2018)
9. Griol, D., Carbó, J., Molina, J.M.: An automatic dialog simulation technique to develop and evaluate interactive conversational agents. Appl. Artif. Intell. **27**(9), 759–780 (2013). https://doi.org/10.1080/08839514.2013.835230
10. Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J.K., Robles, E.: Evaluating the effects of behavioral realism in embodied agents. Int. J. Human-Comput. Stud. **67**(10), 842–849 (2009)
11. Holzwarth, M., Janiszewski, C., Neumann, M.M.: The influence of avatars on online consumer shopping behavior. J. Market. **70**(4), 19–36 (2006). https://doi.org/10.1509/jmkg.70.4.019
12. Kang, H.J., Shin, J.h., Ponto, K.: How 3D virtual reality stores can shape consumer purchase decisions: the roles of informativeness and playfulness. J. Interact. Market. **49**, 70–85 (2020)
13. Li, J.: The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. Int. J. Human-Comput. Stud. **77**, 23–37 (2015)
14. Lugrin, B., Pelachaud, C., Traum, D.: The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics, volume 2: Interactivity, platforms, application (2022)
15. Luo, B., Lau, R.Y., Li, C., Si, Y.W.: A critical review of state-of-the-art chatbot designs and applications. Wiley Interdiscip. Rev. Data Mining Knowl. Disc. **12**(1), e1434 (2022)
16. McClymont, J., Shuralyov, D., Stuerzlinger, W.: Comparison of 3d navigation interfaces. In: 2011 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems Proceedings, pp. 1–6 (2011). https://doi.org/10.1109/VECIMS.2011.6053842
17. Mir, A., Alldieck, T., Pons-Moll, G.: Learning to transfer texture from clothing images to 3D humans (2020). https://doi.org/10.48550/ARXIV.2003.02050

18. Mori, M., MacDorman, K.F., Kageki, N.: The uncanny valley [from the field]. IEEE Robot. Autom. Mag. **19**(2), 98–100 (2012)
19. Mystakidis, S.: Metaverse. Encyclopedia **2**(1), 486–497 (2022). https://doi.org/10.3390/encyclopedia2010031
20. Nah, F.F.H., Davis, S.: HCI research issues in e-commerce. J. Electron. Comm. Res. **3**(3), 98–113 (2002)
21. Nijholt, A., Heylen, D., Vertegaal, R.: Inhabited interfaces: attentive conversational agents that help. In: Proceedings 3rd International Conference on Disability, Virtual Reality and Associated Technologies (2000)
22. Nowak, K.L., Rauh, C.: Choose your "buddy icon" carefully: the influence of avatar androgyny, anthropomorphism and credibility in online interactions. Comput. Human Behav. **24**(4), 1473–1493 (2008)
23. Peukert, C., Pfeiffer, J., Meißner, M., Pfeiffer, T., Weinhardt, C.: Shopping in virtual reality stores: the influence of immersion on system adoption. J. Manage. Inform. Syst. **36**(3), 755–788 (2019)
24. Piro, L., Desolda, G., Matera, M., Lanzilotti, R., Mosca, S., Pucci, E.: An interactive paradigm for the end-user development of chatbots for data exploration. In: Ardito, C., et al. (eds.) Human-Computer Interaction - INTERACT 2021–18th IFIP TC 13 International Conference, Bari, Italy, 30 August–3 September 2021, Proceedings, Part IV. Lecture Notes in Computer Science, vol. 12935, pp. 177–186 (2021). https://doi.org/10.1007/978-3-030-85610-6_11
25. Pricilla, C., Lestari, D.P., Dharma, D.: Designing interaction for chatbot-based conversational commerce with user-centered design. In: 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), pp. 244–249 (2018). https://doi.org/10.1109/ICAICTA.2018.8541320
26. Quarteroni, S., Manandhar, S.: A chatbot-based interactive question answering system. Decalog **2007**, 83 (2007)
27. Schnack, A., Wright, M.J., Holdershaw, J.L.: Immersive virtual reality technology in a three-dimensional virtual simulated store: Investigating telepresence and usability. Food Res. Int. **117**, 40–49 (2019), special issue on "Virtual reality and food: Applications in sensory and consumer science". https://doi.org/10.1016/j.foodres.2018.01.028
28. Szczuka, J.M., Güzelbey, H.S., Krämer, N.C.: Someone or something to play with? an empirical study on how parents evaluate the social appropriateness of interactions between children and differently embodied artificial interaction partners. In: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, pp. 191–194. IVA 2021. Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3472306.3478349
29. Tan, D.S., Robertson, G.G., Czerwinski, M.: Exploring 3D navigation: combining speed-coupled flying with orbiting, pp. 418–425. CHI 2001. Association for Computing Machinery, New York (2001). https://doi.org/10.1145/365024.365307
30. Tan, S.M., Liew, T.W.: Designing embodied virtual agents as product specialists in a multi-product category e-commerce: the roles of source credibility and social presence. Int. J. Hum.-Comput. Interact. **36**(12), 1136–1149 (2020). https://doi.org/10.1080/10447318.2020.1722399
31. Vaddadi, S., Asri, S., Ghemi, Y., Aytha, R.: Developing chatbot wrapper for online shopping: a case study of using generic mobile messaging system (2020)
32. Xi, N., Hamari, J.: Shopping in virtual reality: a literature review and future agenda. J. Bus. Res. **134**, 37–58 (2021)
33. Xu, K., Chan, J., Ghose, A., Han, S.P.: Battle of the channels: the impact of tablets on digital commerce. Manage. Sci. **63**(5), 1469–1492 (2017)