IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Artworks' Features Discovery Through Engaging Conversations for Children

To cite this article: Juan Carlos Jaramillo *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **364** 012097

IOP Conf. Series: Materials Science and Engineering **364** (2018) 012097 doi:10.1088/1757-899X/364/1/012097

Artworks' Features Discovery Through Engaging Conversations for Children

Juan Carlos Jaramillo, Daniele Occhiuto, Franca Garzotto

i3lab – Politecnico di Milano – Via Golgi 39 (Milano)

juancarlos.jaramillo@mail.polimi.it, daniele.occhiuto@polimi.it, franca.garzotto@polimi.it

Abstract. Dialog-based interfaces have seen a renewed interest in the last years. Smart assistants like Siri, Google Assistant or Cortana are pervasive nowadays. Messaging platforms like Facebook Messenger, Telegram or Skype are integrating automatic responders commonly known as bots or chatbots. Industries started incorporating smart assistants as web-services to help, retain and profile their customers. We see a considerable potential in using conversational interfaces to enhance guided tours in cultural institutions. Museums user experiences may vary depending on the exhibition. Nonetheless, hearing is the common channel museums use to inform visitors, in particular by exploiting audio-guides. The existing guides are very well structured and provide indeed a complete guided experience but are generally intended for adults and neuro-typical users. As a consequence, museums may have difficulties in targeting all kinds of visitors, e.g. children or users with cognitive disability. They may require assistance to pass over their heritage to all audiences. Our research purpose is to design an infrastructure to help cultural institutions in broadcasting their heritage to all guests. We propose interactivity over information: the user takes an active part to the visit and is engaged though the whole experience. Using conversational interfaces, we switch from the perspective of a passive listener to the point of view of a dynamic "player": the user can ask for information. The engagement is a necessary aspect to maintain a high interest in children. We explore how bi-directional conversations can provide accessible pieces of information to users with special needs. We design a plug-and-play pervasive solution for a seamless adoption for the interested institutions. We allow holders of the system to fine-tune its characteristics based on the artwork, the exhibition and the target users.

1. Introduction

Conversational interfaces are a new interaction paradigm that is gaining popularity as the underneath technology becomes more and more robust. In the last decade, technological advancements have guaranteed new forms of interaction, some of them are re-shaping the user experience for a particular domain (e.g. smartphones impact in communications: SMS have almost completely been replaced by messaging apps). Still, we believe that shaping the user experience is not the outcome of technological breakthroughs alone, but it is strongly linked to the way tech advancements are put into place. In particular, our interests in human computer interaction (HCI) leads to the study of new interaction paradigms for empowering all users, with special attention in children and users with disability.

In this work we describe the design and development of MUSU a conversational application to trigger the user engagement in museums' exhibitions. MUSU is an app able to converse with the user and proactively ask questions about a recognized piece of artwork. The app purpose is not to replace classical audio-guides that excel in their task for adults, but to develop new forms of interest triggers in children or empower users with disability through conversations as in [1]. In fact, the role of museums is moving from the main conservation and exhibition responsibility to the enhancement of artworks in order to achieve engagement, entertainment and learning [2]. Some studies revealed that technology use during the visit may isolate the visitors from the environment [3]. Isolation during the visit is counterproductive in case of museums since a great effort is put in the environment itself. For example, the disposition of artworks is determined with care and the environmental lightning and sounds are the output of an attentive analysis. To avoid isolation, we decided to adopt a dominant interaction modality based on auditory cues, following the existing *de-facto* standard for museums tours that are audio-guides. However, one issue with audio-guides is the loss of interactivity. To cope with interactivity losses, we devised a system with customizable responses under the form of a conversational "agent". *Agents* are pieces of software integrated on different devices (smartphones, computers, boards) to assist users in general purpose tasks, e.g., event reminders, navigation, information retrieval. Of course, our guide cannot and does not aspire at replacing a human guide, while other works try to mimic a human interaction with robots [4][5], we try to trigger a comprehensive guided tour through conversations without displaying any representation of the agent. We decided to avoid humanoids to allow MUSU to be portable, we think of our system as a plug and play solution to be deployed on pervasive devices (smartphones or other museum proprietary solutions). Moreover, we bypass the agent graphical representation when possible to insist on the auditory channel only: visual stimuli shall come exclusively from the environment set on purpose for the visit.

Smartphones are by design interactive through the visual channel, since we target auditory interaction to let the user focus on the environment, we need to find a way to trigger the experience by limiting as much as possible the use of the graphical interface. The solution we came up with is the positioning of Bluetooth Low Energy (BLE) beacons in the environment. Bluetooth is a wireless technology that works in proximity, BLE radio signal range can arrive up to 100m but the signal is strongly reduced in presence of obstacles. Beacons are tags, under the form of cards, stickers, or badges, that act as continuous transmitters of a predefined message. In summary, beacons can be thought as lighthouses that broadcast messages to devices in range by means of radio signals, just as lighthouses broadcast light signals from a fixed position. In this way, when the visitor approaches an artwork, the device will automatically "wake-up", receive an identifier of the artwork in front of the user and start a relevant conversation with the visitor.

In the next sections we review exiting interactive solutions employed in museums nowadays, we explain the design and development of MUSU, we clarify its implementation process, we report the results of an exploratory study investigating what information users would like to know about artworks, and finally we discuss our findings and future works.

2. Related Work

Museums bear the responsibility to transfer their heritage to visitors, their task is complicated by a generation gap where every interaction is becoming digital while most artworks belong to an analogical world. Is there a way to "sample" artworks and make them accessible to users from the digital age? We would like to insist on the fact that technology is not a comprehensive answer, but we hope it may help in bridging the two worlds together. Should the digital and analogical world fail to communicate, there would be a considerable loss in heritage transfer.

IOP Conf. Series: Materials Science and Engineering **364** (2018) 012097 doi:10.1088/1757-899X/364/1/012097

Children born in the digital age are more engaged when dealing with interactive interfaces: in a research considering virtual assistants, children interacted much more than adults with the assistant (47% of gathered data from speakers belongs to children, compared to the 13% from adult males and 8% from adult females, while 31% did not respond to the virtual assistant) [6]. The same research showed an increase in children engagement and interests for the exhibit. Others demonstrated that experiences embedded in the context of the child's world, e.g. play and story, provide greater meaning to the exhibit [7]. Incorporating virtual guides in museums is not a new practice, in the past virtual guides have been extensively explored: Max is a screen-projected virtual guide enabling the interaction through a keyboard [8]. Tinker is another virtual guide situated in the Museum of Science in Boston, its particularity is the ability to recognize visitors over time, creating a bond with the visitors and engaging them using human relationship-building behaviors [9]. Anew, in Boston Museum of Science, Ada and Grace are twins virtual agents for guided tours with an educational goal for children [10]. MUSU exclusive channel for interaction will be auditory, hence, we shall put a great effort in making the experience pleasant, both by providing meaningful conversations and analyzing the user's sentiment in her/his responses.

Virtual agents are only one instance of museums guides; other studies leverage interactivity through physical objects instead. For example, sometimes physical robots are used as tour guides in presence of users, while in other examples tabletops and touchscreens are the means of interaction. Robots have the advantage of being mobile, they are a tangible product and do not need to be advertised to visitors. Indeed, as opposite to virtual agents that may go un-noticed if the users do not see the screen or do not start an eventual app, robots are self-explanatory in the sense they are explicitly apparent to visitors. Robots have been deployed in museums and fairs as guides in several studies [4][11]. One main motivation for robots' deployment is the possibility for multimodality: they capitalize on different interaction paradigms to simulate the natural multimodal interaction that takes place between humans. In fact, when people interact, they use gestures, body language and facial expressions alongside speech. Acquiring the features of natural human interaction is a challenging task for a machine; for now, humans can easily recognize robots, even if they are designed to look like and behave as real persons. As a consequence, the robot guide may not be seen as a peer, humans are not attracted to robots if they have a too much photo-realistic human-like appearance, this effect is known as "the uncanny valley" [12]. The author explains the uncanny valley as a function that is not strictly monotonic, providing the example of a climber arriving at the top of a mountain only to discover a valley on the other side with a bigger summit across it. Imagine the disappointment of the climber when finding that (s)he is not on the highest summit. Similarly, when humans interact with robots there is a point where they feel familiar enough with it to interact in a natural way. However, such seamless interaction is brittle, a simple mismatch in the expected communication would reveal that the interlocutor is not human, with the destructive consequence of breaking the illusion of interacting with a peer.

Strictly speaking, MUSU does not have any human-like aspects, in fact it does not even have a shape; it is pure software. Nonetheless, when speaking on the phone with somebody unknown, there are several cues that expose the interlocutor's human attitude. For example, the tone of voice, the duration of pauses, and the speed variation during speech, are all factors that influence how we picture and how we perceive the emotional state of the person talking with us. While there is no explicit definition of an uncanny valley for conversations we find affinities with the uncanny valley theory that are detectable in the conversational guidelines we follow. It is important to stress that in the uncanny valley theory, the recommendation when designing robots is to create artifacts that clearly do not appear as humans but share some features of human-like behaviors. For example, the user distinctly realizes that (s)he is interacting with a machine but is not bothered by it. Going back to the uncanny valley metaphor, it is like if the user aims at climbing the first summit and is perfectly satisfied with it because (s)he had no pretense to arrive at the highest top across the valley. To conclude this section, we report an overview of other systems designed to augment the museum's interactivity that are further away from MUSU's approach but share the common goal of boosting the users' engagement. In particular, virtual reality (VR) applications are well suited for the creation of immersive environment. Immersion has been proven to augment the feeling of presence in the virtual environment, with the consequence of incrementing the user's participation to the experience [13]. Virtual reality (VR) and augmented reality (AR) effects are comprehensively investigated in recent studies [14][15]. One study focused on mixed realities experiences, combining both VR and AR, leveraging the experience economy (the focus is no longer in the product but is shifting to the visitor experience) [16]. Last but not least, touchscreen tables have been explored as well, and are reported to be effective for collaborative activities in a comparative study demonstrating that ICT plays a significant role in enhancing the user experience [17].

3. Design and development

When designing conversational interfaces, there are several requirements to keep in mind [18]. Agents need an appropriate task to be successful, "small-talk" capabilities are useful to humanize the agent (yet avoiding the uncanny valley), but in the end, agents shall prove their utility. In order to be successful, a conversational agent must answer the user's needs. Like in real-life conversations, there is a context of the dialog, and users expect the agent to remember the context. Shall the context fail, the agent can adopt social smarts, for example exhibiting its interest for the previous statement. The use of narration can be a trick to avoid context losses and enhancing the conversation: it is easier to talk about something the agent is prepared about. The conversational agent can keep up with the context if the focus of the discussion is known a priori e.g. the dialog is contextualized in a story. Finally, people tend to interact with peers, and don't have the same discussion at home or at the office. We can summarize five main requirements for natural language interaction with machines as follows:

1. The agent requires an appropriate task answering the users' needs
2. The agent awareness of the context is expected from users
3. The agent shall implement a reasonable number of social smarts (small-talk)
4. The agent shall support narrative conversations for bi-directional communication
5. The agent shall consider the dialog setting's influence on the conversation itself

The last point in particular was illuminating in the design of MUSU, conversations are not specifically suited to museums settings at first sight. Firstly, the user has no knowledge of the context and it is hard to participate to the narration: not all users are experts in the art domain, so what would they talk about with an agent? Secondly, people tend to respect silence in museums, thus conversations may not be the best form of interaction. The first objection where users have no knowledge on the conversation domain, is solved by curiosity: with real guides the interaction is mostly mono-directional: the guide explains and the visitors listen. Questions, if any, are raised at the end of the explanation. With real guides and group of visitors, the setting is collaborative, so even if a question may raise the awareness in the other visitors, asking it in the first place may be hard. In fact, the visitor can be shy or altruist, and avoid slowing down the tour for the others by asking questions. The same attitude is not present with audio-guides, the visitor can take all the time needed to appreciate the artwork and has no constraints with the group of visitors. The second objection, where users do not tend to talk in museums, limits the use of MUSU. Indeed, shall everyone talk with the guides, the visits would end up in a messy murmuring and visitors would no longer enjoy the experience. However, suppose you go visiting with a friend, you can share your opinion in a one-to-one talk without bothering the other visitors. Moreover, the number of users in front of an artwork is proportional to the artwork itself: as a consequence, there is a limited number of users asking information in front of each artwork, in such setting the noise would be strongly reduced, granting a pleasant tour and allowing an enjoyable use of MUSU.

We designed our system by prioritizing speech, nevertheless we kept in mind that in some situations speech is not the best channel. Since MUSU aims at being accessible, we diversified the interaction starting from the auditory channel, which is the default interaction modality of our system. MUSU received two subsequent enhancements, on the one hand to be adaptable to a museum setting, and on the other hand to be flexible for users who prefer chats over conversations or users who have speaking or hearing difficulties. MUSU interaction modalities are:

- *Sound vs Sound*: the interaction is on the auditory channel; the user speaks with MUSU through natural language like in a human dialog.
- *Text vs Sound*: the user inputs her/his request as text message, and the app answers with audio messages.
- *Text vs Text*: the interaction with MUSU takes place through textual messages like in chatting applications.

4. Prototype

The first version of MUSU was developed for iOS systems but we plan to develop a version for Android as well. The app presents an initial screen asking the user's permission to access the device location, this process is done only once after installation (Figure 1). Then it is possible to choose the preferred interaction modality to communicate with MUSU (Figure 2); the interaction modality can be changed at any time from the main menu (Figure 3).

Figure 1. MUSU home screen.

Figure 2. MUSU selection of interaction modality.

Figure 3. MUSU main menu.

At any time, the user can access MUSU's main functions from the bar menu at the bottom of the screen visible in Figure 3. The bottom bar holds a description of the museum, some suggestions about the use of MUSU and a button to open the main menu. The artwork dialogue is triggered by the user proximity to a beacon. To better understand the app's functions and its integration with the Bluetooth Low Energy (BLE) beacons, we report the general architecture of MUSU in Figure 4. The smartphone acts as end-point to communicate with the user, via chat messages, text-to-speech or speech-to-text, depending on MUSU's selected interaction modality. Through the smartphone's Bluetooth capabilities, MUSU receives the broadcasted identifier of a beacon, as shown by the line connecting MUSU app and the beacon tag. The received identifier corresponds to a precise target artwork, the identifier acts as key to retrieve the artwork information that will be provided during the conversation. We displayed dotted lines between the museum institution and the artwork since there is no digital communication among them. The same happens among MUSU app and the museum institution. The artwork information is stored in the museum catalogue which is matched in the cloud services used by MUSU, and is retrieved for the conversation when needed. In particular, we used IBM Watson conversations to implement the dialog with the users.

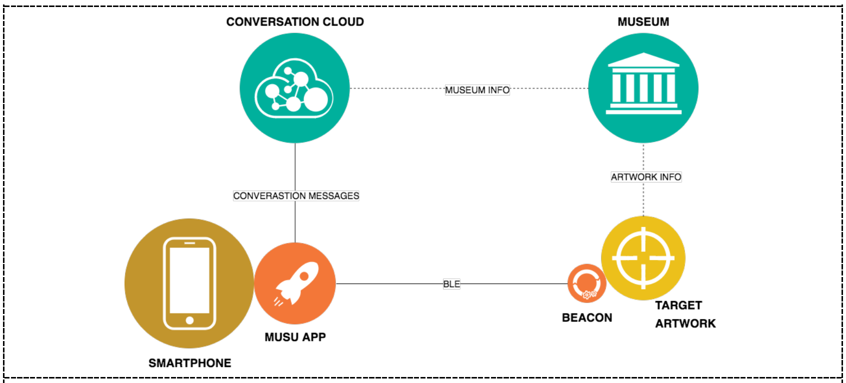

Figure 4. MUSU architecture overview.

5. Exploratory investigation

In this section we report an exploratory investigation about the contents that may be helpful in MUSU's conversations. Past research considered the information that users requested about a single artwork, in particular the painting of a military man, and classified the dialog statements based on who initiated the statement and on the type of the requested information (e.g. biographical questions, personal preferences, general purpose knowledge) [19]. One interesting finding was that 96,7% of user-initiated requests were about the portrayed individual biographical information.

Our interviews were divided in two questions with seven options each: the questions refer to the artwork in front of the user and aim at uncovering what users would like to learn about the artwork:

- *Question A*: What would you ask about this artwork?
- *Question B*: What is calling your attention?

Table 1. Participants' answers to Question A: What would you ask about this artwork?

Answer	*Question A*	*Children (%)*	*Adults (%)*
A1	What does it mean?	85%	67%
A2	Who is the portrayed person?	33%	55%
A3	Who painted it?	69%	45%
A4	Where is the painter from?	38%	42%
A5	What year is the painting from?	25%	27%
A6	How much does it cost?	4%	33%
A7	Other	46%	30%

Table 2. Participants' answers to Question B: What is calling your attention?

Answer	*Question B*	*Children (%)*	*Adults (%)*
A1	The colors in the painting	98%	67%
A2	The style of painting	44%	70%
A3	The shapes	60%	45%
A4	The realism	35%	58%
A5	A specific element (tree, eyes, dog)	25%	15%
A6	Nothing	4%	24%
A7	Other	33%	21%

Participants were assigned to two groups based on their age: children (age range 4-7 years) and adults (age range 24-55 years). The procedure followed a small interview were participants were asked to answer the two questions with a total of fourteen options. We recruited 33 adults (16 women and 17 men) from a university campus in Milan, and 48 children (21 girls and 27 boys) from a kindergarten in Milan (22 children) as well as in a day care center in Ecuador (26 children). As in [19] we chose one

IOP Conf. Series: Materials Science and Engineering **364** (2018) 012097 doi:10.1088/1757-899X/364/1/012097

representative painting from a local museum (Museo del Novecento) portraying an old woman: "La signora Virginia" from Umberto Boccioni. We report in Table 1 the options for *Question A* and the percentage of participants who selected each possible answer (A1 to A7). We did the same for *Question B*, and report the results in Table 2. Figure 5 and Figure 6 present the answers (A1 to A7) for *Question A* and *Question B* in a visual form.

Figure 5. *Question A* – Children's percentages (inner circle) versus adults' percentages (outer circle) for each answer (A1 to A7).

Figure 6. *Question B* – Children's percentages (inner circle) versus adults' percentages (outer circle) for each answer (A1 to A7).

6. Discussion and future work

We can see that children would ask more naturally an explanation about the meaning of the painting (answer A1 is selected by 85% of children against the 67% for adults), children were also interested in knowing who is the painter, on the contrary adults seemed to be less interested in the painter and stated they would ask biographical questions about the person being portrayed, in accordance with the findings in [19]. The biggest difference in the answers from children and adults for *Question A* can be seen for answer A6: asking about the price of the painting (33% of adults would ask the price - A6, against 4% for children). This could help MUSU in filtering what information it shall provide to its users, for adults some price information could be interesting, while it would be totally irrelevant for children. Concerning *Question B* instead, 98% of children (against 67% of adults) are attracted by the painting's colors, while adults seem to be predominantly attracted by the style of the painting (70% of adults against 44% of children). Adults accord more attention to the realism of the painting (58% pay attention to it, against 35% for children). As a result of *Question B,* we can design MUSU to talk more about colors and shapes when interacting with children, and explain the style of painting and the techniques used to create realistic portraits when conversing with adults.

In conclusion, we presented MUSU, a conversational application that acts as an audio guide designed to boost the visitors' engagement in interactive conversations. We leveraged BLE beacons as triggers to start the conversation seamlessly when the visitor is in front of the identified artwork. We did a small exploratory investigation to better understand what users would expect from a dialogue with MUSU and direct the context of conversation towards the visitors' needs. MUSU is at an early stage and it is a preliminary prototype investigating the role of the auditory channel in guided tours. We are aware of its limitations and hope we can overcome them in future versions of our app. Next steps include analyzing the user interaction more in depth, gathering further museums' needs and re-engineering the technology underneath the app. In the end, we hope MUSU could become an effective support tool to bridge the analogical and digital worlds and guarantee the heritage transfer to posterity.

References

[1] Buonamici, F., Carfagni, M., Furferi, R., Governi, L., & Volpe, Y. (2016). Are We Ready to Build a System for Assisting Blind People in Tactile Exploration of Bas-Reliefs?. *Sensors*, *16*(9), 1361.

[2] Hall, T., & Bannon, L. (2005, June). Designing ubiquitous computing to enhance children's interaction in museums. In Proceedings of the 2005 conference on Interaction design and children (pp. 62-69). ACM.

[3] van Dijk, E. M., Lingnau, A., & Kockelkorn, H. (2012, October). Measuring enjoyment of an interactive museum experience. In Proceedings of the 14th ACM international conference on Multimodal interaction (pp. 249-256). ACM

[4] Faber, F., Bennewitz, M., Eppner, C., Gorog, A., Gonsior, C., Joho, D., ... & Behnke, S. (2009, September). The humanoid museum tour guide Robotinho. In Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on (pp. 891-896). IEEE.

[5] Boboc, R. G., Horaţiu, M., & Talabă, D. (2014). An Educational Humanoid Laboratory Tour Guide Robot. Procedia-Social and Behavioral Sciences, 141, 424-430.

[6] Traum, D., Aggarwal, P., Artstein, R., Foutz, S., Gerten, J., Katsamanis, A., ... & Swartout, W. (2012). Ada and Grace: Direct interaction with museum visitors. In Intelligent Virtual Agents (pp. 245-251). Springer Berlin/Heidelberg.

[7] Anderson, D., Piscitelli, B., Weier, K., Everett, M., & Tayler, C. (2002). Children's museum experiences: Identifying powerful mediators of learning. Curator: The Museum Journal, 45(3), 213-231.

[8] Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005, September). A conversational agent as museum guide–design and evaluation of a real-world application. In International Workshop on Intelligent Virtual Agents (pp. 329-343). Springer, Berlin, Heidelberg.

[9] Bickmore, T. W., Vardoulakis, L. M. P., & Schulman, D. (2013). Tinker: a relational agent museum guide. Autonomous agents and multi-agent systems, 27(2), 254-276.

[10] Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., ... & Lane, C. (2010, September). Ada and Grace: Toward realistic and engaging virtual museum guides. In International Conference on Intelligent Virtual Agents (pp. 286-300). Springer, Berlin, Heidelberg.

[11] Burgard, W., Cremers, A. B., Fox, D., Hähnel, D., Lakemeyer, G., Schulz, D., ... & Thrun, S. (1999). Experiences with an interactive museum tour-guide robot. Artificial intelligence, 114(1-2), 3-55.

[12] Mori, M. (1970). Bukimi no tani [The uncanny valley], Energy,[Online] 7 (4), 33-35.

[13] Slater, M., Khanna, P., Mortensen, J., Yu, I.. (2009). Visual Realism Enhances Realistic Response in an Immersive Virtual Environment. IEEE Comput. Graph. Appl. 29, 76–84.

[14] Guttentag, D. A. (2010). Virtual reality: Applications and implications for tourism. Tourism Management, 31(5), 637-651.

[15] Jung, T., tom Dieck, M. C., Lee, H., & Chung, N. (2016). Effects of virtual reality and augmented reality on visitor experiences in museum. In Information and Communication Technologies in Tourism 2016 (pp. 621-635). Springer, Cham.

[16] Chung, N., Han, H., & Joun, Y. (2015). Tourists' intention to visit a destination: The role of augmented reality (AR) application for a heritage site. Computers in Human Behavior, 50, 588-599.

[17] Michael, D., Pelekanos, N., Chrysanthou, I., Zaharias, P., Hadjigavriel, L. L., & Chrysanthou, Y. (2010, November). Comparative study of interactive systems in a museum. In Euro-Mediterranean Conference (pp. 250-261). Springer, Berlin, Heidelberg.

[18] Luger, E., & Sellen, A. (2016, May). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286-5297). ACM.

[19] Robinson, S., Traum, D. R., Ittycheriah, M., & Henderer, J. (2008, May). What would you Ask a conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting.