DOI: 10.17469/O2106AISV000025

GIUSY GAUDINO, MARIA DI MARO

Conversing on Artistic Representation Topics (CART): An android audio guide with dialogic skills

The present work describes the architecture of an interactive audio guide (CART) specifically conceived for the Neapolitan Charterhouse "San Martino", as part of a project, whose aim is to collect and model multimodal data for the development of spoken technologies for museum fruitions. In addition to the cultural contents that an audio-guide can provide, this application is enriched with a spoken dialogue system, a system interacting with humans by means of spoken language commands and requests. The application was developed using Android Studio. Furthermore, the developed interface for the Museum was endowed with dialogic skills modelled in DialogFlow using its Android Software Development Kit (SDK), to which Text-to-Speech technologies were integrated to record the user query. The results are promising in that they display positive feedback for the introduction of such technologies in museums.

Key words: spoken dialogue systems, oral archive applications, audio guide, cultural heritage, android, dialogflow.

1. *Introduction*

Several authors have described the act of experiencing as a conversation: «it always involves at least two voices» and its «meaning is always changing as a consequence of what comes next» (McCarthy, 2008). The act of experiencing is indeed a dialogue of the self with the other, being it another human being or not. Consequently, museum experience is describable as a dialogic exchange with the artistic product where the conveyed messages are mainly emotional: the artwork, in fact, communicates emotions first, which are differently perceived according to the characteristics and preconception of the experimenting subject. The perceived emotions are then empowered at a deeper stage of analysis with cultural contents once the museum experience became informative and educational.

The presence of tour guides in historical places or in cultural sites traces back to the Greek Empire, when the Periegeta (gr. περιηγέω [*periegeo*] 'I show around') used to show the visitors places new to them. In this scenario, the dialogic aspect of the experience becomes physical, since the influence of the interlocutor on the main speaker can be explicitly communicated via spoken requests or multimodal feedback of understanding (i.e. nodding, smiling, vocalizations, etc.). Besides tour guides, at the end of the XVIII century, the art dealer Tommaso Puccini introduces little tags indicating the name of the author of each artwork in the Uffizi Gallery, representing the first informative 'technological' means provided to visitors (Kaplan, 2014). The

technological progress has brought, in some cases, to the substitution of tour guides with audio guides. The first audio guide was introduced in the Milan Cathedral by the electronics technician Giovanni D'Uva in 1959 and was called the "audiophone". This device enabled users to listen for the first time to informative audios without the physical presence of an expert (Kaplan, 2014). Nowadays, the use of audio guides is extremely widespread, resulting in a transformation of the museum experience, which is becoming less dialogic and more monologic, and, therefore, passive. Interestingly, some scholars pointed out how the inclusion of recorded dialogues in audio guide devices can be actually used as an engagement strategy for museum visitors (Elliston, Fitzgerald, 2012). Specifically, creating audio guides using expert-expert spontaneously generated dialogues was shown to better engage listeners with learning contents. Surprisingly, the resulting users' engagement caused visitors to increase the amount of spoken interactions among them. On the other hand, with the introduction of mobile phones in our daily life, different services started to migrate on these devices. The CIMI (Computer Interchange of Museum Information) studied the possible application of mobile computing in museums, ranging from virtual guides, to electronic maps, communication channels, and virtual diaries for collecting visitors' impressions (Gay, Spinazze & Stefanone, 2002). In this scenario, Economou and Meintani (2011) underline the power of the exploitation of mobile apps in museums, for the possibility they create to reach new audiences through personal and familiar devices. Furthermore, those tools would enable life-long learning and edutainment (i.e. game-based apps) as they could be used before, during and even after the visit. The authors' evaluation of the available systems points out that the transition from the previous audio guides to the new multimedia devices has not brought any changes in the design of the museum experience. Conversely, the aim should concern the design of a system, which is considered as an active tool in the content creation process.

For all the aforementioned reasons, we believe that users' engagement could be further increased if the audio guide would become interactive, enabling the user to personalize the visit according to their interests and curiosities. Consequently, in this work we propose to integrate a dialogue system – i.e. a conversational agent interacting with human users – in a museum guide app to enhance the visitors' experience. In the second Section, we are going to focus the attention on the description of the system architecture, specifying the developed interfaces of our app, the modules created for the different provided services, and the requested requirements. The third Section is dedicated to the interaction modelling, where we pay attention to the description of the interaction design principles and the corpus-based content extraction. In Section 4, we describe the test sessions and the collected users' feedback, which are going to be part of the future developments of our app.

2. *Experimental Work*

2.1 CART App

CART is an interactive audio guide that allows visitors to interact with the audio guide itself with the aim of asking questions about artworks and history of a cultural site. The integration of a conversational agent is the true innovation of this product that adds to the static information provided from traditional audio guides the possibility for the users to deepen some topics of interest, just as they would do with a human guide. For the implementation of the contents of the audio guide, the transcriptions of experimental guided tours within the San Martino Charterhouse in Naples were used. The guided tours were recorded for the data collection projected for the Italian national project CHROME (Cultural Heritage Resources Orienting Multimodal Experiences), whose goal is to define a methodology to collect, to analyse, and to model multimodal data in designing virtual agents serving in museums[1] (Cutugno, Dell'Orletta, Poggi, Savy & Sorgente, 2018). The cultural sites chosen for the CHROME project are the Charterhouses in Campania. A trivia on the name of our application is indeed that it takes back the official logo of the *Chartusian order* that had four crossed letters, a C, an A, an R and a T (from the initials of Carthusia – Latin for Chartreuse, the place near Grenoble where the first Charterhouse was born). Therefore, the app's logo is inspired to the Chartusian one.

The front-end architecture of the application is divided into four different interfaces:

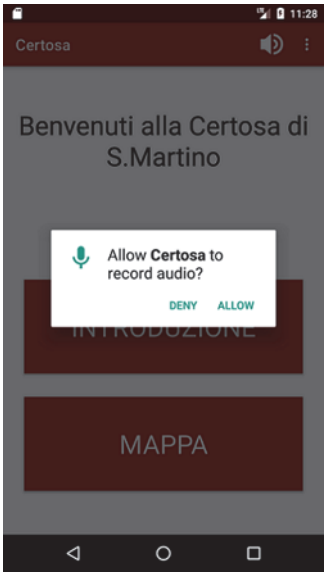
a. Introduction;
b. Map;
c. Description;
d. Dialogue interface.

Furthermore, the application uses a module for operating the Text-to-Speech service and another control module (cfr. Section 2.7) to manage data required by the application.

The architectural choice was designed to maximize the application's usability and the software modularity at the same time. The reduced number of interfaces has the effects to allow the user to have a good learning curve and to facilitate browsing the application. Furthermore, the ordered display of the interfaces is designed to guide the user to a progressive visit of the tour stages without, however, restricting its freedom and autonomy to explore the place as they prefer. The application was designed for newbies, for this reason it follows learning and memorability principles (Polillo, 2010).

[1] http://www.chrome.unina.it.

Figure 1 - *Introduction screen asking for "record audio"*



2.2 Introduction Interface

The introduction interface (Figure 1) is the main interface of the application. Deliberately plain, it drives the user to discover the app and the place of the visit. The two provided buttons (Introduzione, and Mappa) lead respectively to the first descriptive screen (first step of the tour) and to the map screen where an overall view of the tour is provided. In Figure 2 is shown a pop-up asking for "recording audio" permission (cfr. Section 2.8), later required to interact with the vocal assistant. In view of future developments, this screen could provide options for selecting the language to be used for visitors of different nationalities.

2.3 Map Interface

The map screen (Figure 2) shows to the user a general distribution of the points of interest (POIs) described within the app. This is useful to the visitors to get oriented and to verify their position according to the description being provided. From this screen, the descriptive screen of all tour's stages can be accessed.

Figure 2 - *Map screen*

2.4 Description Interface

The description interface (Figure 3) takes up the classic functionality of the audio guide, giving information on the single step of the visit. There is an image that can help the visitor to recognize the place in which they should be while watching a specific description interface, or rather the described place. The description is both textual and vocally synthesized. Everywhere in the application the audio option can be enabled and disabled and the preamble interface, the map or the other steps of the tour can be reached. Through this interface, one can easily pass from the description of the current stage to the next or previous one with the side arrows (respectively) ">" and "<". There is a rating-bar on the screen, with a score from 1 to 5 stars that allows users to leave a feedback about the quality of the description related to that place. This function can be used to collect the user satisfaction within the app fruition, in order to improve the different sections. On the bottom right, the icon of a microphone, is strikethrough if the user doesn't grant microphone access permission, enabled otherwise. This icon, if clicked, give access to the dialogue interface, where users can ask questions to the virtual assistant about the current section they are seeing.

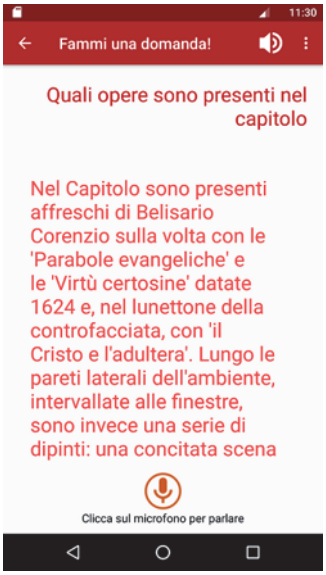
Figure 3 - *Description interface*

2.5 Dialogue Interface

The dialogue interface (Figure 4) handles the true innovation of this digital audio guide that supplies the service offered from traditional audio guides with querying services, peculiar of human guides. The screen recalls the model of the classic vocal assistants commonly integrated in smartphones, such as Google Assistant. By clicking on the microphone icon an audio feedback is provided to let users know that the application is starting to listen, and their input can therefore be processed. A different feedback audio informs the user, that their question has been received. The online service DialogFlow[2] (cfr. Section 3.3) is consequently questioned, and the question-answer pair appears on the screen well distinguished by different colours and different alignments (respectively the question on the right and the answer on the left). If the user doesn't disable the audio service, the answer will be acoustically synthetized. If an Internet connection is not available, the query cannot be submitted to the DialogFlow service. This information is given to the user through a pop-up message.

[2] https://dialogflow.com/.

Figure 4 - *Dialogue interface*



2.6 Text-to-Speech Manager

This module is responsible for correct loading and setting of the voice synthesizer used in the application. The vocal synthesis is useful for two reasons: i) to reproduce the POI's description (as for the traditional audio guide) ii) to vocally synthesize the answer of the assistant (Hoy, 2018). The acoustic description of the POI would for sure sound more natural with a recorded file audio rather than with a speech synthesiser. Recognizing this, the choice fell on speech synthesis because of the scalability and the generalization of the application's content. In fact, despite the changes concerning the application's contents, vocal synthesis grants an acoustic reproduction of information of interest, making the service much more versatile.

2.7 Information Module

"Information Module" is the focal point for the modularity of the application. It lets the application to scale in size and it is the point where all the information showed to the user are collected. It is the only section where the developer can edit the theme of the application. By editing the information contained and saving the aesthetic design of the application, the same design can be fitted to any other museum, creating an interactive audio guide ready to be used. For each step of the tour, the stored information comprises: i) chronological order of the point of interest (POI) one is visiting (as part of the general tour); ii) name and description; iii) an image of the place one is visiting; iv) a link to the vocal assistant trained on this particular argument (step of the tour).

2.8 Required Permissions

An access permission to the microphone is required in order to interact with the vocal assistant. The request is carried out at the first access to the application and, if denied, it is requested again when the user tries to access to the dialogue interface, as this screen is not available when the permission is denied. Furthermore, an internet access is required in order to interact with the vocal assistants integrated in the app and for the Speech-to-Text service whose results are used to query the vocal assistant.

2.9 Used Services

The app uses DialogFlow service that allows the building of conversational interfaces. To include the DialogFlow service into the Android application the API dialogflow-android-client "api.ai"[3] has been used. The API includes Speech-to-Text services that turn the user's vocal input into textual information and submit the information achieved in the DialogFlow service through an HTTP request. The answer obtained is shown to the user in a textual way and worked out from a Text-to-Speech service, specifically the one offered by Google and integrated for Android.

3. *Dialogue Management*

3.1 Conversation Design

In this section, particular attention is drawn to the development of the interaction module integrated in the application. Before adding contents to the app, design principles were defined in order to build an interaction, which could be easily carried out by human users, with no need to teach them how. The first rule chosen for the conversation design is *Identify your user*. This rule is essential when designing interfaces of any kind, as it can be used to circumscribe what the interface is actually going to do, considering technological constraints as well. Knowing what kind of users is going to make use of the interface can help the designer to predict which contents have to be included in it. For the definition of the user and its needs we adopted the questionnaire results collected during the collection phase of the CHROME corpus (cfr. Section 3.2). For the CHROME project, audio-visual recordings of gatekeepers during tour guides were collected. Immediately after the recorded visits, the recruited visitors were asked to compile a questionnaire composed of 23 items including both Likert scale evaluations and open answer questions. The items were designed to collect anagraphic data, self-evaluations of artistic competence, evaluations of personal satisfaction, and evaluations of the gatekeeper's performance (Origlia, Savy, Poggi, Cutugno, Alfano, D'Errico, Vincze & Cataldo, 2018). Therefore, the questionnaire was used to get an idea of the possible visitors' profiles and, since open questions

[3] https://github.com/dialogflow/dialogflow-android-client (last access 03/10/2019). The v1 version was used, since v2 version does not support Android.

about what they liked the most or what they would like to hear from a guide were included, even topic preferences could be modelled.

The second rule that was taken into account is *Give a reason why*. With this principle, we wanted the agent to guide the interaction by stimulating the interlocutor to ask questions. In order to trigger questions in the users we followed two different strategies. On the one hand, the strategy of the unsaid (Todd, 2013) was exploited to stir curiosity in the users: instead of giving all the information to them in the different description screens (cfr. Section 2.2.4), some details were left pending with rhetorical or provocative questions, and highlighting pieces were introduced without giving explanations, as in (1).

(1) In uno degli affreschi (?) del Parlatorio è ritratto lo stesso Sant'Ugo (?) mentre dorme su un letto a baldacchino. Su di lui aleggiano sette stelle (?).

In one of the frescos of the Parlor (?), St. Ugo is represented while sleeping on a canopy bed (?). Seven stars hover on him (?).

In (1), the interrogation points represent the missing details that can be asked to the system, and specifically Which other frescos are situated in the Parlor? Who is St. Ugo? What's the name of the fresco representing St. Ugo? What does it mean? What do the seven start represent? On the other hand, curiosity was guided through direct questions, options, and class of topics to further be investigated (2).

(2) Vuoi saperne di più sul Parlatorio? Vuoi conoscere il significato del ritratto di Sant'Ugo? Vuoi passare al Coro?

Do you want to learn more about the Parlor? Do you want to know the meaning of the painting representing St. Ugo? Do you want to move on to the Choir?

There is another designed rule connected to the second strategy, named *Provide help*. This means that, when the user does not know what to ask, they can rely on options given by the agent.

The fourth regulation principle for the conversation design is *Be adaptable*. With this rule, we refer to the necessity to adapt to different sentence structures that can be uttered by the users. Therefore, the first important step, after the definition of the semantic category to develop, was to define the possible questions a user could pose to ask for information for each specific category (cfr. Section 3.3).

As far as the interaction style was concerned, other important rules were adopted. The first one was *Be relevant*, that is do not give more information than the one needed. When a question could be answered with more details, an in-depth option was used (cfr. Section 3.3). On the other hand, it was also found to be important to *Be conversational*. In fact, in order to sound more natural, no formal words were used, and second-person references were adopted. Furthermore, chitchat intents were included to build this conversational perception (cfr. Section 3.3). Especially as far as the chitchat intents were concerned, we also followed the *Use variations* rule, randomizing answers when appropriate. Moreover, with the aim of not only entertaining but also educating the user, the *Be clear* principle was adopted, as am-

biguity or obscure words were avoided. In fact, one important goal to reach when designing *infotainment* devices is to be clear and be sure that the contents are well received by the users.

Particularly important was also to react promptly to scenarios where the agent could not understand the user input. *Error handling* is indeed another principle adopted. With the use of DialogFlow (cfr. Section 3.3), this aspect was automatically taken into account, as rephrasing requests were activated by the agent (i.e. *I'm not sure I understand your question. Could you repeat?*), therefore avoiding assigning a wrong category to the misunderstood user input.

In the next two sections we are going to explain in detail the contents of the CHROME Corpus used as a reference corpus for the application development and how the contents were modelled using DialogFlow. The rules here displayed are then going to be enriched with the feedback collected from the testing phase (cfr. Section 4).

3.2 CHROME Corpus

In order to achieve the main aim of the CHROME project[4] (cfr. Section 2.1), data need to be collected, analysed and annotated. In fact, the first step taken was to record audio-visual material comprising real gatekeepers giving tours to four groups of four visitors in the Neapolitan Charterhouse "San Martino". The visit is divided into six points of interest (POIs), selected as the most relevant parts of the Charterhouse from an architectural and artistic point of view, specifically *Pronaos*, *Great Cloister*, *Parlor*, *Chapter Hall*, *Wooden Choir*, *Treasure Hall*, corresponding to the description screens selected for the app development (*Pronaos* was left out). The collected data are going to be analysed and annotated multimodally by using the ELAN software (Wittenburg, Brugman, Russel, Klassmann & Sloetjes, 2006). In fact, gestures and speech are going to be annotated: on the one hand, gestures are analysed as far as their textual or interactional goals are concerned; on the other, linguistics features are pointed out, as far as orthographic, phonetic, syllabic, disfluency, macro-syntactic, syntactic, and intra-syntactic levels (Origlia et al., 2018). For the development of the app, these multimodal annotation levels were not taken into account, as only contents and user preferences were exploited (cfr. Section 3.1).

3.3 DialogFlow Web Services

Following the principles explained in Section 3.1, the interaction was built within the DialogFlow framework. DialogFlow is a Google platform that enables designer to exploit Google's Machine Learning algorithms and Natural Language Processing techniques to easily create conversational agents. DialogFlow provides a powerful natural language understanding (NLU) engine to process and understand natural language inputs. "DialogFlow allows to create *agents* that can understand the vast and varied nuances of human language and translate them to standard

[4] http://www.chrome.unina.it.

and structured meaning that a service can understand". As different users can express the same meaning differently, as far as syntax and vocabulary are concerned, DialogFlow parse the requests in a way that each information is mapped to the pertinent data in order to find the appropriate response[5].

The semantic and conceptual category corresponding to the classes of questions that the agent is capable of answering to are named *intents*. Each intent is made up of training sentences given to the system to learn to recognize the intention of the user which can be differently expressed (cfr. *Be adaptable* in Section 3.1), and of textual, visual or multimodal answers (i.e. text, cards, links, pictures). Training sentences are annotated with *entities*, representing the semantic classes to which words belong. For instance, given the entity *artist*, different lexical items can be used to express it, such as *painter*, *sculptor*, *architect*, *author*, *creator*, *designer*, *inventor*, *master*. The use of such annotations is advantageous to improve disambiguation processes and to correctly classify an intent that can be uttered with different terms according to the user. Follow-up intents can also be added to ask for details concerning the previous intent or for yes/no confirmations.

The conversation was designed for the five of the six POIs reported in Section 3.2. For each of the selected POIs specific intents (41 in total) were developed, as reported in Appendix A.

4. *Testing and Results*

CART functionality was tested by 10 users. The test phases are divided in two different steps. At first, users were provided with a task accomplishment assignment, for which they were asked to compile information requested in a questionnaire. In details, the questionnaire was made up of four tasks: i) collect information about the fresco in the Treasure Hall Cupola (i.e. name of the fresco, name of the artist, characters represented, meaning of the fresco), ii) collect missing information about the Carthusian in the Great Cloister and in the Parlor screens, iii) collect information about Wooden Choir frescos, iv) converse freely with the agent.

The second part of the test was a usability questionnaire given to users to measure the Quality of Service (QoS) and the Quality of Experience (QoE), where QoS is used to evaluate the effectiveness of the system in giving the right information and QoE to capture the user perception of interaction quality (Fiedler, Hossfeld & Tran-Gia, 2010). The evaluation was divided in 5 sections: i) utility, ii) ease of use, iii) ease of learning, iv) satisfaction, v) comments. The evaluation was computed through a five value Likert Scale (Never, Rarely, Sometimes, Often, Always). The results, as shown in the figures 5-8, are generally positive in all users.

[5] https://dialogflow.com/docs/intro.

Figure 5 - *Utility Results*

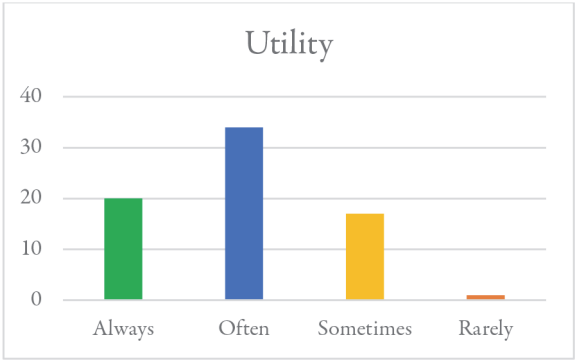


Figure 6 - *Ease of use Results*

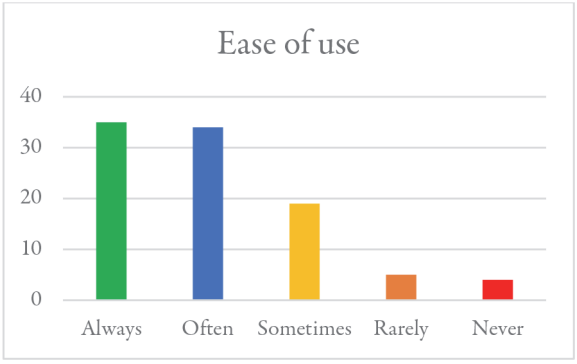


Figure 7 - *Ease of learning Results*

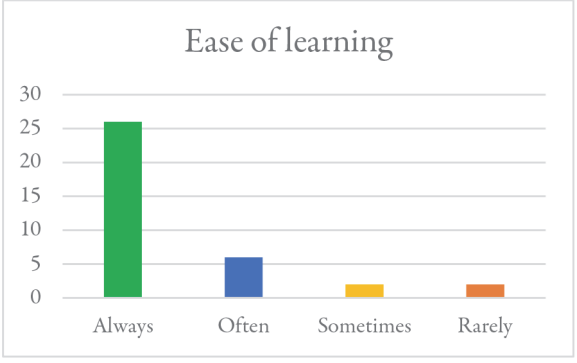


Figure 8 - *Satisfaction Result*

Nevertheless, the application needs further development as suggested in the comments collected from the open question of the evaluation questionnaire. In fact, user's feedbacks can help to improve the definition of conversational design principles that may be used for next implementations of CART. One important need highlighted by the users may be referred to one of the Grice's Maxims (Grice, Cole & Morgan, 1975) that was not entirely followed in the conversation design of CART, namely the maxim of quantity, or in other words *Be efficient*. Users signalled the necessity to get vivid and significant contents, which are not too long-winded. Indeed, the attention of users tends to decrease when listening at someone talking without having that someone in the same room (as for a conversational agent which is not embodied). The necessity to provide relevant and detailed information to users who are to be educated on a particular subject collides with the necessity of expressing a concept using few and simple words. This contradiction can be solved by using shorter descriptions for which a *read/hear more* option can be added. Among the other feedback we mention the need to model more intents, as some information asked were missing. This means that the maxims of *Identify your user* and *Be adaptable* should be further improved. This necessity is going to be easily fulfilled also thanks to the conversation which were collected during the testing phase. In fact, DialogFlow permits to save the user inputs, enabling the improvement of the intents design. Moreover, it has interestingly been noticed that users find multimodality as an important expected need in such interfaces. As a matter of fact, they suggested that links to events organised in the descripted site, or links to further details, or more images or even videos should be added within the interaction interface. This leads to another important principle that is *Conversation is multimodal*. In fact, no matter with whom or what one is interacting, because the demand for the multimodal conveyance and/or reception of meanings will be always on the spotlight.

5. *Conclusions and Future Work*

The present paper reported on the development of an Android interactive audio guide thought for the museum fruition. The interactivity of the application is concerned with the possibility for users to ask further information besides the descriptions provided within different sections of the guide. For the development of the interaction module DialogFlow was used. The dialogue was modelled following specific design principles in order to ensure the naturalness and the success of the interaction. The testing phase pointed out that further principles need to be considered for the improvement of the audio guide, specifically as far as efficiency, adaptability, and multimodality are concerned. In details, we aim at enriching the intents collection of our agent, starting from user inputs collected during the testing phase. Furthermore, multimodal functionality, such as pictures, links, buttons, and videos, need to be added in our framework. In fact, some information needs visual support to be better received by users. Furthermore, in order to give users a more fluent interaction with the app and allow them to pause and play the audio description of the place (vocally synthesized), like a traditional audio file, the description text could be chopped into more sentences which could be sequentially synthesized. Indexing the sentences and tracking the last played one, the impossibility of "pause" a vocally synthesized text could be therefore overcome and the interaction could increase in naturalness.

Acknowledgment

This work is funded by the Italian ongoing PRIN project CHROME - *Cultural Heritage Resources Orienting Multimodal Experience* (Cutugno et al. 2018) #B52F15000450001. We also thank Valentina Lipari who is working on the improvement of CART.

Bibliography

Cutugno, F., Dell'Orletta, F., Poggi, I., Savy, R. & Sorgente, A. (2018). The CHROME Manifesto: integrating multimodal data into Cultural Heritage Resources. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLIC-IT 2018)*, Turin, 10-12 December 2018.

Economou, M., Meintani, E. (2011). Promising beginning? Evaluating museum mobile phone apps. Rethinking Technology. Paper presented at *Museums 2011: Emerging experiences*, University of Limerick, 26-27 May 2011.

Elliston, B., FitzGerald, E. (2012). Encouraging museum visitor engagement using spontaneous talk-in-interaction audio guides. Paper presented at the *4th International Conference on Computer Supported Education (CSEDU 2012)*, Porto, 16-18 April 2012.

Fiedler, M., Hossfeld, T. & Tran-Gia, P. (2010). A generic quantitative relationship between quality of experience and quality of service. In *IEEE Network*, 24, 2, 36-41.

Gay, G., Spinazze, A. & Stefanone, M. (2002), Handscape: Exploring potential use scenarios for mobile computing in museums. In *Cultivate Interactive*, 8, 15.

Grice, H.P., Cole, P. & Morgan, J.L. (1975). *Logic and conversation*. In Cole, P., Morgan, J.L. (Eds.), *Syntax and Semantics vol. 3. Speech Acts*. New York: Academic Press, 41-58.

Hoy, M.B. (2018). Alexa, siri, cortana, and more: An introduction to voice assistants. In *Medical reference services quarterly*, 37, 1, 81-88.

Kaplan, F. (2014). Fantasmagories au musée. In *Alliage*, 73, 132-140.

McCarthy, J., Ciolfi, L. (2008). Place as Dialogue: Understanding and Supporting the Museum Experience. In *International Journal of Heritage Studies*, 14, 3, 247-267.

Origlia, A., Savy, R., Poggi, I., Cutugno, F., Alfano, I., D'Errico, F., Vincze, L. & Cataldo, V. (2018). An Audiovisual Corpus of Guided Tours in Cultural Sites: Data Collection protocols in the CHROME Project. In *Proceedings of the AVI-CH 2018 Workshop on Advanced Visual Interfaces for Cultural Heritage*, Castiglione della Pescaia, 29 May 2018.

Polillo, R. (2010). *Facile da usare. Una moderna introduzione all'ingegneria della usabilità.* Milan: Apogeo.

Todd, S. (2013). *Learning desire: Perspectives on pedagogy, culture, and the unsaid.* London: Routledge.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Genoa, 24-26 May 2006, 1556-1559.

Appendix A

In this appendix, the list of intents developed for each POI is displayed.

a. *Great Cloister*: Cemetery, Fanzago_info, Cloister_function, Carthusian_day, Carthusian_meal, Carthusian_crest, Small_talk.
b. *Parlor*: Bruno_Calabria, Bruno_Italia, Bruno_born_dead, Bruno_studies, StUgo_Painting_meaning, StUgo_Painting_name, Parlor_artwoks, Parlor_info, Silence, Heremitage, Small_talk.
c. *Chapter Hall*: Chapter_info, Chapter_artworks, Chapter_Name_meaning, Lunette_artworks, LayBrother, Prior, Preator, Small_talk.
d. *Wooden Choir*: Choir_info, Choir_artworks, Carthusian_day, Carthusian_crest, Women_ban, Small_talk.
e. *Treasure Hall*: Ribera_painting, Ribera_info, Fapresto_meaning, Giordano_info, Giuditta_info, Fresco_name, Fresco_meaning, Treasure_info, Small_talk.