Conversing on Artistic Representation Topics (CART): an Android Audio Guide with Dialogic Skills.

With the present work we intend to describe the architecture of an interactive audio guide (CART)[1] 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.

The starting point of the development was collecting materials for textual and spoken contents to be added to the application. To fulfil this requirement, the audio-visual corpus of guided tours collected for the project was used (Self citation 2018). Precisely, the corpus contains audio-visual material collected from guided tours at San Martino Charterhouse in Naples. For this purpose, three art history experts were recruited to give one-hour-long tours to four groups of four persons. From each tour, spoken and paralinguistic materials were recorded. Furthermore, each visitor was provided with a questionnaire with Likert scale evaluations and open questions (Self citation 2018). The spoken data and the suggestions coming from the questionnaires were used to model the contents to be inserted in the application.

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. As a matter of fact, the content can be further explored by asking for information. Differently from other dialogue systems, which make use of different separated modules to be trained – specifically Automatic Speech Recognition, Natural Language Understanding, Dialogue Manager, Natural Language Generation and Speech Synthesis – the presented system can be classified as an end-to-end system. This kind of systems does not consider the separation into different components. Conversely, it indeed considers them as a single system to be trained as a whole (Vinyals et al. 2015, Serban et al. 2018). In particular, *DialogFlow*[2], which uses Machine Learning algorithms to enable the system to communicate with users based on specific modelled *intents*, was here used for training. Each *intent* is made up of training sentences given to the system to learn to recognize the intention of the user, and textual answers to be generated. 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*, *artist*, *sculptor*, *architect*, *author*, *creator*, *designer*, *inventor*, *master*. The use of annotation is advantageous to correctly classify an intent uttered with variable terms, and to improve disambiguation processes.

The dialogue was structured on the basis of the information added to the audio-guide, making use of different strategies. On 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, some details were left pending with rhetorical or provocative questions, and highlighting pieces 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 exclamation points represent the missing details that can be asked to the system, and specifically *Which other frescos are situated in the Parlor? 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).

[1] The acronym refers to the Chartusian logo.

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

(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?

The application was developed using Android Studio[3]. The developed interface for the Museum was then enriched with the dialogue modelled in DialogFlow using Android SDK for DialogFlow[4], to which Text-to-Speech technologies were integrated to give answers to user queries.

Besides the audio and vocal channels, the multimodal nature of the application also made use of the visual one, as comparisons with other similar environments, work of arts, or Charterhouses were visually shown, when requested or implied in the offered content. The future application of this system is intended to be further developed to comprise users' position knowledge. This will enable the system to use spatial references and deitics (i.e. spatial, temporal and personal pronouns) focusing the listener's attention to a specific object of the specific environment, thus enriching the linguistic potential of the system itself.

The application was tested on 10 users, who made use of the interactive audio guide in the modelled areas of the Charterhouse, and precisely the Pronaos, the Great Cloister, the Parlor, the Chapter Hall, the Choir, and the Treasure Hall. After the visit, users were asked to fill a questionnaire 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 et al. 2010). With the additional use of open questions we were able to collect opinions and impressions to get the usefulness of such systems and how to improve them. As a matter of fact, the generically positive results were paired up with useful suggestions, such as returning vivid and significant contents which are not too long-winded, as the attention tends to decrease when listening at someone talking without having that someone in the same room.

References

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

Self citation (2018).

Serban I. V., Lowe R., Henderson P., Charlin L., and Pineau J. (2018). "A Survey of Available Corpora For Building Data-Driven Dialogue Systems: The Journal Version". In *Dialogue & Discourse*, 9(1), 1-49.

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

Vinyals O. and Le. Q (2015). *A neural conversational model*. arXiv preprint arXiv:1506.05869.

[3] Android Studio: https://developer.android.com/studio/

[4] Android SDK for DialogFlow: https://github.com/dialogflow/dialogflow-android-client