The OLOS® Way to Cultural Heritage: User Interface with Anthropomorphic Characteristics

Daniele Baldacci, Remo Pareschi

***Abstract*—**Augmented Reality and Augmented Intelligence are radically changing information technology. The path that starts from the keyboard and then, passing through milestones such as Siri, Alexa and other vocal avatars, reaches a more fluid and natural communication with computers, thus converting the dichotomy between man and machine into a harmonious interaction, now heads unequivocally towards a new IT paradigm, where holographic computing will play a key role. The OLOS® platform contributes substantially to this trend in that it infuses computers with human features, by transferring the gestures and expressions of persons of flesh and bones to anthropomorphic holographic interfaces which in turn will use them to interact with real-life humans. In fact, we could say, boldly but with a solid technological background to back the statement, that OLOS® gives reality to an altogether new entity, placed at the exact boundary between nature and technology, namely the holographic human being. Holographic humans qualify as the perfect carriers for the virtual reincarnation of characters handed down from history and tradition. Thus, they provide for an innovative and highly immersive way of experiencing our cultural heritage as something alive and pulsating in the present.

***Keywords*—**Human-computer interfaces, holographic simulation, digital cinematography, interactive museum exhibits.

I. Introduction

HUMANIZING user interfaces has for quite a while been around as a goal of HCI design, nonetheless for a long time such interfaces were both a challenge and an option. They were a challenge because engineering an interface with a human face is not a simple task, even with the support of advanced computer graphics. They were an option because it is far from obvious that a human face endowed with speech capabilities is the best way of interacting with the screen of a device like a desktop or a laptop computer, which has for a long time provided the natural context for user interfaces; keyboards, mouse and icons could well be up to the task, if not better. However, in the past twenty years the world has changed, and has undergone a new revolution after the one that has brought computers into everyday life during the second half of the past century. The coming of age of mobile networks has created mobile telephony, and this in turn has spurred the possibility of putting computing and Internet communication capabilities within mobile phones, thus making them “smart”. With a smartphone people carry around a full-fledged operating system such as IOS or Android, which makes it by all means into a computer, but with characteristics and context of use very different even from last generation laptops. With smartphones users are immersed in the environment, rather than solely focused on the device. Keyboards are still available, mainly through touchscreens, but in this case they provide a far from optimal interaction, particularly when the user is in situations of effective mobility within a bustling environment. In fact, the execution of the action of typing, that was conceived for the quiet and sheltered environment of the home and of the office, diverts users from a condition of alertness and awareness of the surrounding environment. Consequences range from suboptimal user experiences to safety threats, as witnessed by the car accidents related to the notorious behavior of “texting” while driving.

Smartphones have in turn spawned more mobile devices, such as tablets and smartwatches, all of which can be carried and even worn in person. Thus, mobile and wearable computing is now part of the everyday life of people in the world. Internet of Things (IOT) has completed the picture by making the physical environment itself part of the overall information infrastructure, with the potential of creating a single information ecosystem encompassing humans and the things that surround them — in other words, of effectively ushering in the era of ubiquitous and pervasive computing. Clearly, an evolution of this kind calls for radical innovation in the design of user interfaces, and human semblance as well as human-like communication appear as natural candidates for interaction in the present and in the near future since they are general and universal, and thus suit a user who is constantly in touch with the communication infrastructure through a variety of devices. Therefore, what was previously a still shaky and untested option, has now acquired full concreteness, as consequent from the stability and maturity achieved in the meantime by the enabling technologies for anthropomorphic interfaces, starting from natural language understanding and speech recognition. To proof, there are hugely popular software such as Amazon's Alexa, Microsoft's Cortana and Apple's Siri, which come in the form of voice assistants and interact with humans by interpreting natural language and integrating into speakers and other smart devices. In [1] it is shown how IoT does indeed open a scenario where things not only perceive but also communicate, thus creating the conditions for an ecosystem where communication surrounds us universally, and can happen at any time and in any place. However, talking to a disembodied human voice is not like talking to a human figure. Aside the fuller user experience that it provides, human figures can interact with users whenever there is need of visual demonstration, as in training, trying on

Daniele Baldacci is with Blue Cinema Tv, 00153 Rome, Italy (e-mail: d.baldacci@bluecinematv.com).
Remo Pareschi is with University of Molise, 86090 Pesche (IS), Italy (corresponding author; phone +39 0874 404100; fax: +39 0874 404123; e-mail: remo.pareschi@unimol.it).

clothes, running a make-up session. Furthermore, they are suited for one to many user interactions, in as diverse situations as museum touring, information kiosks, live entertainment. That's where our OLOS® technology [2], providing audiovisual interfaces based on the human figures, picks up. An OLOS® interface typically deploys an architecture based on components that include the hardware for visual display, event management capabilities, IoT communication infrastructure, speech and natural language understanding, user profiling, plus content provided by real humans performing in the role and the actions that can be executed by the interface; thus, it is hybrid between 21st century engineering and the ancient creativity of the tradition of performing arts. It is obviously aimed at full size displays rather than wearable smart devices, with which it can however be effectively integrated. In this way it contributes substantially to fulfil the potential of holographic displays in cultural heritage settings, as proposed also in [3]-[6].

The rest of this article is structured as follows. Section II describes the visual components of the OLOS® platform in terms of the holographic simulations through which the actions of the interface are visualized and of the event management process that drives their execution, with focus on cultural heritage and in particular on the possibility of reviving historical, artistic and archaeological contexts by animating them with holographic humans who re-incarnate characters from the past. Section III identifies the components that can be connected in order to evolve the intelligence of the interface, such as the ability to understand natural language, Internet access, user profiling. Section IV concludes the document.

II. OLOS® Visual Components

We can characterize the three fundamental aspects of an OLOS® audiovisual interface in terms of the following features:

a. The interface is given by a human figure taken from a human original;
b. The interface is visualized at ultra-high definition (UHD) resolution levels;
c. An event management system supports the execution of changes in the state of the interface, in response to its interaction with the user.

Creating the interface from human originals obviously provides maximal anthropomorphic features. In addition, human subjects that give their image to the interface can actually "act" their roles, thus bringing in all the characteristics of expressiveness typical of acting for theater and film. UHD resolution in 3D (at least 4K) makes very attractive and effective the effect of holographic illusion in the perception by the user of the interface.

In a typical deployment, as exemplified in the installation at EXPO 2015 in Milan (Fig. 1; video at [7]), where the historical characters of Teodolinda and Vergil provided visitors with information about the Italian region of Lombardy, the images are reproduced on a sheet of transparent polycarbonate 1 mm thick (a thickness that guarantees the minimization of the refraction of the two surfaces of the sheet itself, thus avoiding their overlapping). The sheet is placed at 45 ° with respect to the matrix. The final images perceived by the observer are on a vertical plane behind the polycarbonate sheet, completely detached from the backdrop, with which they form a considerable effect of parallax.

Fig. 1 An example of OLOS® interactive characters

A recent evolution of the OLOS® display technology offers additional flexibility and expressiveness, by letting holographic humans into outdoor environments. HI®Human Interface [8] is a patented method for virtual transparency through flat audiovisual devices of customizable dimensions, with the possibility of going from Indoor to Outdoor mode and back. It enables a perfect visual perception outdoors even in broad daylight (in a vandal-proof and waterproof environment).

The workflow for the development of a HI®Human Interface consists of three phases:

a. prospective study of the user's point of view and related real-time image acquisition;
b. acquisition and storage of images of the subject to be reproduced;
c. real-time composition of the images obtained during the previous phases.

Phase (a) involves modeling the point of view of the potential user with respect to the audiovisual system housed in the support equipped with a camera with variable focal lens capable of reproducing in real time, through a standard television closed circuit connection, the exact portion of space behind the audiovisual system, as in Fig. 2. Thus, the exact perspective of the user with respect to his standpoint in the environment is identified, so as to reproduce images in real time that, in the audiovisual system, take the same view over such a portion of space, hence obtaining for the user a visual continuity with respect to the environment. The result is to extract the image from the environment actually seen and perceived by the user, as in Fig. 3.

Phase (b) involves generating stored images of the holographic human that will animate the landscape through reproduction by a computer video player housed in the structure.

Finally, during phase (c) the images generated during the previous phases are processed in real time through a channel composition technology housed in the structure, resulting in a new image obtained by overlap. This new image projects the holographic humans over the landscape with proportions suitably adapted to the identified perspective (Fig. 4).

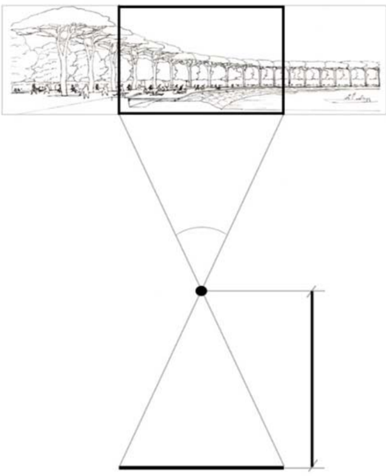

Fig. 2 Optical study of user point of view

A typical application is to let the user observe the landscape and then make holographic humans enter the scene, for instance with the role of narrators. The effect is of a journey through time, where virtual background and interactive narration bring landscapes and characters from the past back to life. For example (Fig. 5), an archaeological area can be animated by going back in time through the reconstruction of its 3D architectural history in an engaging and unique way. In fact, starting from the virtual transparency of the place as it is today, the system then performs a perfect perspective overlap on current reality with a 3D reconstruction in virtual background. This experience can be enriched with the interactive narrating by a multilingual historical figure who acts as Genius Loci, to whom questions can be asked via simple devices such as smartphones, touchscreens and audio guides.

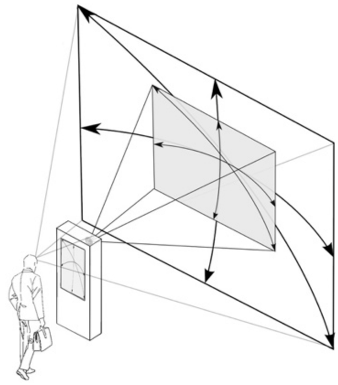

Fig. 3 Perspective image from the environment actually seen and perceived by the user

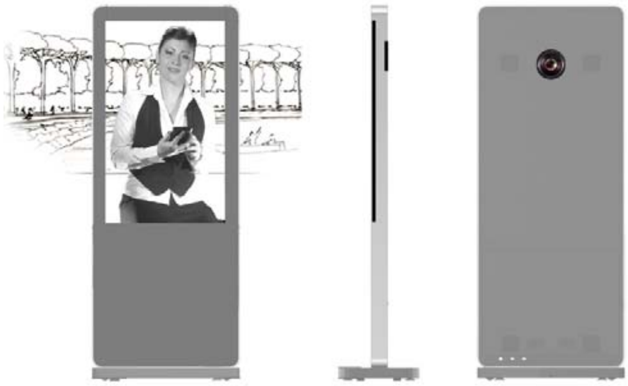

Fig. 4 Final image composition

Fig. 5 HI® Time Gate, a typical application

At the core of fully interactive OLOS® (video at [9]) there is the event manager, whose functioning is depicted in Fig. 6.

The OLOS® event manager amounts to a loop where different images and corresponding actions of the interface are triggered in relationship to actions and request of the user. For instance, a tourist arriving at the airport of Rome may ask an OLOS® located in an information kiosk, in one of the supported languages, how much it costs to travel to the city center respectively by train, by bus and by taxi, in which case the OLOS® will speak in reply so as to provide the appropriate information. The tourist may then inquire about current art exhibitions in the capital city of Italy, in which case the OLOS® may launch a search on the Web on the theme and then come out with a list. Or it may just get stuck, either because it does not know how to comply to the request or because the request was not clearly formulated, in which case it will politely ask to rephrase, eventually admitting defeat if there is no way it can comply. Clearly, logs of interactions can be maintained, and the knowledge upon which the OLOS® relies in order to provide satisfactory answers to its users can be incrementally upgraded by taking into account former failures to respond to its users.

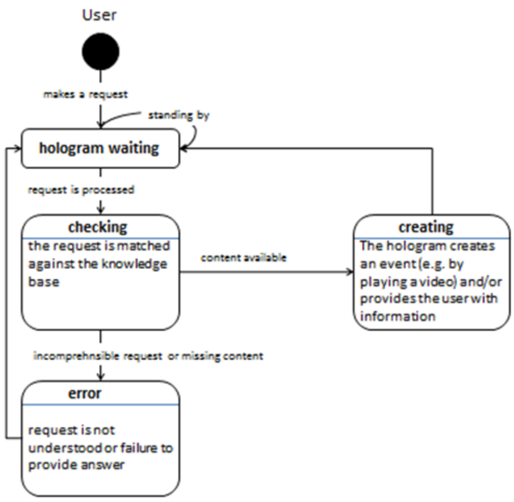

Fig. 6 The OLOS® Event Manager flow diagram

III. The OLOS® Modular Architecture

Modern morphing techniques from digital cinematography provide effective support to the shifting of the OLOS® interface into a new image in the course of user interactions. However, the capability of involving users into truly engaging interactions derives also from the easiness with which the OLOS® event manager can be integrated within a modular architecture so as to take advantage of a variety of technologies from IT and AI that provide both the form and the substance to the interactions, among which:

- Speech recognition
- Face recognition
- Natural language understanding
- Dialogue management
- Semantic content management
- IOT
- User profiling
- Multimedia DBMS

The OLOS® system architecture is composed by two main subsystems, each including modules supporting the functionalities listed above. Such subsystems are represented by:

- An IOT Application infrastructure
- The OLOS® Event Manager

The *IoT application infrastructure* manages the entry of human users into areas of interaction with OLOS®. In fact, by detecting the presence of people in the vicinity of a kiosk or an OLOS® station, it is then able to alert the OLOS® event manager to open dialogue sessions with users in the vicinity. This subsystem is, at a glance, composed by three main components, represented by:

- A *Wireless Sensor Network* (WSN), typically made of smart beacons devices, to detect human users' locations and movements, by leveraging Bluetooth low power technologies;
- An *application*, developed in the shape of an App for smart personal devices or an advanced interface for touchscreen; through this component, users are detected (by exchanging information with the WSN) and are enabled to submit question, in natural language, to the holographic figure, by exploiting the microphone of their smart device or the one provided by the touch display.
- A *Message Delivery Service* (MDS), responsible for receiving vocal inputs incoming from the App or the touchscreen and for delivering it to the OEM.

The implementation of the OLOS® Event Manager represents the core system orchestrating the whole process underlying the human-like dialogue management. It includes:

- A *Natural Language Processing Engine (NLPE),* engaged with processing vocal signals forwarded by the MDS, and translating it into textual messages. It prepares textual sentences to be processed by the OLOS® Event Manager. It is based on Voice Recognition and SpeechToText services, implemented in the current prototypal version, as a Microsoft DotNet service library, exploiting the Speech Recognition Platform API;
- A *Questions & Answers (Q&A) engine,* deputed to process the textual request against a Knowledge Base containing a set of possible answers. In the prototypal version, the engine is implemented as a set of logic and semantic rules to extract the most meaningful terms composing the request; such terms list is used as a key to find the most suitable answer. Each answer is scored against the terms list and the one with the highest evaluation is the winning one. Answers are generated through a set of grammatical rules.
- A *Video Dispatcher*, receiving the selected answer and using it as a key to query a Multimedia Database, in order to select the corresponding video segment to be played, thus effectively implementing a Text to Video process. Such a component, implemented as a software service, enables the hologram to comply with the request from the human user in a spoken form, and hence provides explanation for further actions that may follow and may be triggered by the answer itself (e.g. launching a browser).

IV. Conclusion

Summing up, the OLOS® platform can be seen as an enabling factor and a generator of holographic humans, interfaces that maintain the expressiveness and familiarity of their human originals and blend it with the skills derived from the progress made in the field of natural language understanding and with the widespread connectivity allowed by the cloud and the IOT. In this way, OLOS® marries the technical strength of advanced engineering with the aesthetic pleasure of the cinematographic tradition and appears as a well-qualified candidate to contribute significantly to an interface paradigm suitable for the era of omnipresent and pervasive information technology. The applications to cultural heritage are clear and straightforward, by letting us bring the past back into the present and experience it as something alive, pulsating and communicating, rather than as a relic to keep under the display case. OLOS® can therefore, thanks to a felicitous combination of craftsmanship and technology, make history truly speak before us.

References

[1] F. Marulli, R, Pareschi, D. Baldacci "The Internet of Speaking Things and Its Applications to Cultural Heritage" In proceedings of IoTBD 2016, pp. 107-117.
[2] OLOS® (Patent pending EPO/PCT N.14723500.6, Italy Patent MISE UIBM n. 000141612, European trademark UAMI n.011115367, European design UAMI n. 002572685-001).
[3] Isabel Pedersen, Nathan Gale, Pejman Mirza-Babaei, Samantha Reid: More than Meets the Eye: The Benefits of Augmented Reality and Holographic Displays for Digital Cultural Heritage. JOCCH 10(2): 11:1-11:15 (2017).
[4] Giuseppe Caggianese, Luigi Gallo, Pietro Neroni: Evaluation of spatial interaction techniques for virtual heritage applications: A case study of an interactive holographic projection. Future Gener. Comput. Syst. 81: 516-527 (2018).
[5] Paolo Clini, Ramona Quattrini, Emanuele Frontoni, Roberto Pierdicca, Romina Nespeca: "Real/Not Real: Pseudo-Holography and Augmented Reality Applications for Cultural Heritage" in Handbook of Research on Emerging Technologies for Digital Preservation and Information Modeling, IGI Global, (2017).
[6] Manuela Chessa, Matteo Garibotti, Valerio Rossi, Antonio Novellino, Fabio Solari: A Virtual Holographic Display Case for Museum Installations. EAI Endorsed Trans. Serious Games 2(7): e3 (2015).
[7] https://www.youtube.com/watch?v=dB6wUGG9Oys&feature=youtu.be
[8] HI® - Human Interface (Italy Patent UIBM n. 0001428984, Italy trademark UIBM no. 302019000082782 11/11/2019).
[9] https://www.youtube.com/watch?v=blT6rHx_Ktw&feature=youtu.be