# Employing Virtual Humans for Interaction, Assistance and Information Provision in Ambient Intelligence Environments

Chryssi Birliraki[1,2](✉), Dimitris Grammenos[1](✉), and Constantine Stephanidis[1,2]

[1] Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH), N. Plastira 100, Vassilika Vouton, 70013 Heraklion, Crete, Greece
{birlirak, gramenos, cs}@ics.forth.gr
[2] Computer Science Department, University of Crete, Heraklion, Crete, Greece

**Abstract.** This paper reports on the design, development and evaluation of a framework which implements virtual humans for information provision. The framework can be used to create interactive multimedia information visualizations (e.g., images, text, audio, videos, 3D models) and provides a dynamic data modeling mechanism for storage and retrieval and implements communication through multimodal interaction techniques. The interaction may involve human-to-agent, agent-to-environment or agent-to-agent communication. The framework supports alternative roles for the virtual agents who may act as assistants for existing systems, standalone "applications" or even as integral parts of emerging smart environments. Finally, an evaluation study was conducted with the participation of 10 people to study the developed system in terms of usability and effectiveness, when it is employed as an assisting mechanism for another application. The evaluation results were highly positive and promising, confirming the system's usability and encouraging further research in this area.

**Keywords:** Virtual humans · Virtual agents · Virtual assistants · Embodied agents · Multimodal interaction · User-agent interaction · Usability evaluation

## 1 Introduction

Virtual humans are embodied agents that exist in virtual environments that look and act like humans and users can interact with them. They are employed as user interfaces and can serve various needs for human-computer interaction, including guidance, assistance, information provision and user training. Virtual humans exhibit human-like qualities and can communicate with humans, or even with each other, using natural human modalities and are capable of real-time perception, cognition and action.

Ambient Intelligence (AmI) environments are characterized by the ubiquitous and unobtrusive presence of electronic equipment in the users' environment that allows sensing the users' actions and react accordingly in order to help them achieve their


N. Streitz and P. Markopoulos (Eds.): DAPI 2015, LNCS 9189, pp. 249–261, 2015.
DOI: 10.1007/978-3-319-20804-6_23

goals. In AmI environments the interface is indistinguishable from the physical setting, as the real world assumes the role of the interface.

The incorporation of virtual humans in AmI environments can enhance the social aspects of interaction offering human-like anthropocentric communication. This work presents the design, development and evaluation of a framework which employs virtual humans able to provide multimodal interaction, assistance and information provision in AmI environments. In general, the framework:

- Is dynamic and flexible so as to fit diverse needs of other systems and the environment.
- Allows virtual humans to act as assistants to other systems offering real-time help, tutorials and user training on interaction techniques.
- Offers the ability to use virtual humans as standalone “applications”.
- Uses the virtual human’s body gestures and speech so as to provide information.
- Is able to visualize information in different forms, such as images, videos, 3D models, text and audio.
- Uses a data model for information storage and retrieval.
- Supports natural multimodal interaction using a variety of means, including verbal interaction and gestures.
- Integrates techniques regarding user interaction in three dimensions and mobile devices.

## 2 Related Work

### 2.1 Ambient Intelligence Environments and Interaction Techniques

Ambient intelligence (AmI) is an emerging discipline that brings intelligence to our everyday environments and makes those environments sensitive to us. The features that are expected in Ami environments and technologies: sensitivity, adaptation, transparency, ubiquity and intelligence. According to Weiser [1], “the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it”.

In Ambient Intelligence environments, the interaction is different from the typical interaction techniques due to the fact that the environments should be “transparent” from the technological point of view and their interaction techniques are human-centric. Multimodality is the “key” for AmI to provide systems that can be easily manipulated from different groups of people who operate in many different ways individually. Multimodality is the mixture of textual, audio, and visual modes in combination with media and materiality to create meaning [2].

In computer science, speech recognition is the translation of spoken words into text. Some systems use speaker independent speech recognition, while others use a training procedure, where an individual speaker reads sections of text into the system, like Hindi or Kaldi Speech Recognition System both using HTK [3–5]. The feature of invisibility in speech is vital because in AmI environments the transparency is important. This makes the technique challenging to voice-based interfaces in the design and use, but simultaneously helps them to provide an alternative way of interaction with impaired users or in environments where other techniques are difficult to be used or cannot be used at all.

Gestures can be defined as a form of non-verbal communication in which visible body actions communicate particular messages. The recognition of gestures is an issue on which much work is found in literature [6, 7]. There is a diversity of gesture types, including hand gestures [8] and body gestures [9]. Gestures can also be used for manipulating "smart humans" [10], like robots, and controlling their movements.

Context aware systems generally track users and through that offer a large range of information on demand. Context awareness may be used as a mean to provide navigation in 3D space. Another form in which context awareness appears in literature is the progressive information visualization. When the user is far away from the system then the information displayed is more like an overview and when the user come closer the system displays more details of information [11].

Nowadays, many people have smartphones and it is reasonable to expect that they carry them even when they are playing games. Researchers have proposed the use of motion gestures for a variety of input tasks, e.g., to navigate maps or images, to input text, to control a cursor and to verify user identity [12]. Users generally enjoy gestures as a way to use their mobile phones to intuitively interact with other devices [13].

### 2.2 Existing Approaches for Virtual Humans

Virtual environments can become useful training tools if used properly and for the appropriate training application. Training systems of the future need to simulate all aspects of a virtual world, from the physics of the scene objects to realistic human behavior. Therefore, projects, like Virtual Humans [14, 15] are concentrating on high fidelity embodied agents that are integrated into these environments.

Embodied conversational agents (ECA) [16] are a form of intelligent user interface. Graphically embodied agents aim to unite gesture, facial expression and speech to enable face-to-face communication with users, providing a powerful means of human-computer interaction. A virtual human can itself serve as an exhibit of technology. Consolidating this idea InterFaces project has created virtual museum guides that are in use at the Museum of Science in Boston [17].

Virtual humans can be powerful tools in a wide range of areas. The ICT Virtual Human Toolkit [14] was designed to support researchers with the creation of embodied conversational agents. It offers a collection of modules, tools, libraries, a framework and an open architecture that integrates these components. Apart from virtual agents, the Toolkit services full coverage of subareas such as speech recognition, audio-visual sensing, natural language processing, dialogue management, nonverbal behavior generation and realization, text to speech and rendering. Another powerful animation engine is Maxine [18], which allows management of scenes and virtual characters and focuses on multimodal and emotional interaction in order to establish more effective communication with the user.

Multimodal interaction seems straightforward in everyday life, but a closer look exposes that such interaction is indeed complex and contains a variety of coordination levels. For a better understanding of how multimodal behaviors are coordinated, Zhang at [19] proposes a real-time multimodal human-agent interaction system in which human contributors interact with a virtual agent in a virtual environment.

## 3 Motivation and Rationale

Various systems already exist that offer diverse features related to virtual humans including conversational abilities, user training, adaptive behavior and virtual human creation, but they constitute limited, “monolithic” attempts that do not provide a comprehensive solution that can cater for broader needs, such as the ones imposed by AmI environments. In this context, the aim of the work described in this paper is the design, development and evaluation of a generic and adaptable framework that combines all the aforementioned features and employs virtual humans in order to provide interaction, assistance and information provision in ambient intelligence environments. Towards this end, the envisioned framework needs to be highly dynamic and flexible so as to be independent of its context of use. Since the information flow among the three fundamental components of AmI environments – i.e., the users, the individual systems and the integrated “smart” environment - is omnidirectional, the system should also be able to communicate with all of them.

There are three distinct roles that virtual humans must be able to support by acting as: (a) assistants for third-party systems; (b) standalone applications and (c) integral parts of emerging smart environments. When acting as an assistant, the framework should provide the tools needed for presenting information in the form of tutorials and real-time help. Furthermore, the virtual humans should support the users’ training on interaction techniques applied for the assisted system. As a standalone application, the framework should provide categorized information visualization and interactive help about the interaction techniques and information presented. Finally, when the virtual humans are embedded in smart environments, the framework should be able to support all the aforementioned functionalities individually or even in combination.

In addition to the above, virtual humans created using the framework should be expressive in order to be as realistic as possible and enhance the human-oriented character of the system. Therefore, the information provided by the virtual human should be accompanied by body animations and synthesized speech. Additional information may be offered in various forms, including text, images, audio, videos and 3D models. These means of information visualization should be dynamic, adaptable and interactive so that different users can manipulate them in different ways.

Finally, a fundamental characteristic of systems embedded in AmI environments is the adoption of natural user interaction techniques. As interaction with virtual humans in ambient intelligence environments presented in related literature mainly focuses on conversational interaction and user tracking in space, the suggested framework should support but also enhance and extend these techniques by including verbal communication, kinesthetic interaction and mobile devices.

# 4 System Overview

## 4.1 Basic Components and System Visualization

**The Virtual Human (Bryan).** Bryan (below) is an animated human-like virtual assistant aiming to provide assistance to every application using it as well as to every user handling the application. In the framework, applications can modulate the assistance they wish to be given. The assistance obtained from Bryan comes through hand and body gestures and speech. The system allows the description to be dynamic and adaptive at any point. Furthermore, Bryan responds to the users' wishes and obeys to real time system commands that may come and performs them at once.


**Fig. 1.** (Left) Bryan, (middle) T-pose of Bryan, (right) hand cursor - projection screen

**Projection Screen.** The projection screen (Fig. 1) is another important design component, as a large proportion of the information visualization is displayed on it, e.g., images, videos and text. As the entire system needs to give the impression of an informative, helpful and training environment, the information is displayed on a projection screen as a typical presentation.

**Virtual Hand Cursor.** The user employs a virtual hand cursor (Fig. 1) to communicate with the system and select items. This hand cursor moves along the user's hand and follows precisely the movements of the user's hand. The goal is to give the user the impression that they actually position their real hand on the scene and that they are in control of the interaction in every way.

**Button Components.** A number of ready to use interactive buttons are available that contain animations on their idle, hover and click states and provide actions that affect the entire environment, such as the mute, show subtitles, previous stage, next/previous, play/pause and others.

**Environment.** The environment (Fig. 2) that encloses all the visual components is represented as an ordinary room with walls and pictures on them. Lights surround the scene to give a more realistic graphical environment with shadows on the objects. The user's viewpoint and field of view can dynamically change depending on the current information provision needs.

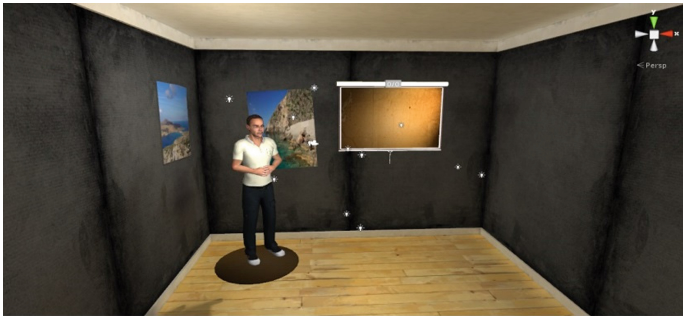


Fig. 2. The environment

**Information Categorization.** This subsystem allows the provision of information through distinct categories. Each category is visualized using a title and a representative item (a 3D model or a 2D image). Each category item is dynamic, interactive and has its own animation states (idle, hover, click). The available information presentation media include image, video, text, 3D object and audio. Upon selection of an item the category selection menu hides and the camera moves in space so that the projection screen resides on the one side of the screen and the virtual human on the other (Fig. 3).

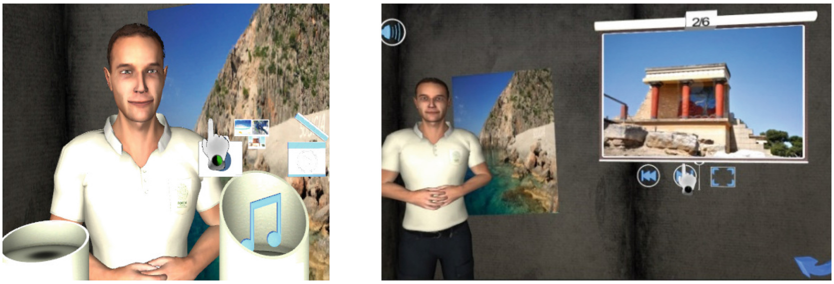



Fig. 3. (Left) selection of presentation type, (right) image gallery presentation

**User Interaction Training Tool.** This tool employs the virtual human in order to teach to the user the interaction techniques supported by the respective application. In this case, the scene is divided in two areas: (a) the assistant area and (b) the interaction techniques area. In the assistant area, the virtual human is presented in full body view and can perform any interaction technique employing hand/full body gestures or speech commands. In the training technique area, a list comprising all the available interaction techniques to be taught is presented. Users can select a technique to practice and the assistant gives the respective instructions.

**Structured Tutorial.** This component can provide structured instructions regarding the target application, its functions and how one can interact with it. The presentation of the tutorial is divided into three areas. The assistant stands on the left side of the scene and presents information through speech (in the user's preferred language) accompanied by body, hand and face animations. In the central part of the scene visual information is rendered on a projection screen through images, videos and written text. Appropriate control buttons are provided for each distinct medium. Finally, a list of the chapters in which information is divided is displayed on the right side of the scene.

**Real Time Assistance.** The virtual human subsystem allows applications to send explicit commands directly to the assistant in order to help the users when needed. These commands vary in complexity and can range from overly simple, such as "speak X", to very complex ones, as for example "speak X and show Y image/video while you animate Z motion for E time". To this end, a communication system has been developed in order for the application and the assistant to collaborate.

## 4.2 Existing Installations

Up to now, the presented system has been used in various environments supporting diverse applications. For instance, in the context of a smart hotel room, it was employed to provide assistance to the users regarding the room's functionality. In this case, a virtual assistant is displayed in the room's television. In another setup, the virtual human is projected at the entrance door of a building and is used to welcome visitors and provide access to it via voice recognition and speaker identity verification. Finally, in a third case three distinct agents are used to provide assistance in different rooms of a building. In this case, agents can also communicate with each other, while users can also act as intermediaries, passing messages of one agent to another, as a means of personalization (e.g., a phrase like "Bryan sent me" can be said to another virtual agent as a means of user identification evoking all the related context and adaptation that Bryan has already instantiated for the specific user).

## 4.3 Implementation

The system is developed in C# programming language on the Unity Platform[1]. The system is represented by Unity Game Objects which contain two main parts that manipulate different aspects of an element: the component part, which contains the necessary information for each element and implements all the needed functions in order to handle it at a higher level. Each component may be part of a hierarchy, but is able to render itself and contains values for: transformation, rendering and physics and the script part, which defines the behavior of each element and controls every component associated with. The scripts are attached as components to other components or Game Objects in Unity. Scripting architecture is adopted aiming at keeping each element's distinct implementation from its behavior. This way, code reusability and the system extendibility are supported.

[1] Unity, http://unity3d.com/.

# 5 Evaluation

## 5.1 Set-up and Participants

The evaluation session took place at FORTH's Ambient Intelligence Facility. The system was set up in a room with a 55" display. A total of 10 users participated in the evaluation, 5 females and 5 males. The age of the participants varied from 20 to 35 years old. Six of the users had intermediate or high computer expertise whereas the other participants had limited expertise. Even though the majority of the users were familiar with computers and touch screen systems, they did not have familiarity with hand gesturing or speech recognition as a mode of interaction.

## 5.2 The Evaluation Scenario

The scenario of the evaluation involves a virtual human, Bryan, being used as an assistive system to an interactive exhibit in a museum. The interactive exhibit that was assisted was TimeViewer [9, 20] a system with complex interaction involving hand and leg gesturing and therefore requiring user guidance. Therefore, assistance and gesture training was imperative for the users to be able to interact with the system, regardless of how natural the gestures were.

## 5.3 The Evaluation Process

The evaluation process started with the users being informed about the goals of the evaluation process. A series of tasks were assigned to the participants in order to measure the usability of the system. The tasks covered all the primary functionalities of the system in order to assess both the system's design and interaction techniques.

Furthermore, an additional series of tasks were given for the usage of TimeViewer so as to measure the effectiveness of the assistance provided by the virtual human. During the evaluation process, the evaluators provided assistance only when asked for, in order to examine whether users were able to manipulate the system and utterly depict the effectiveness of the virtual assistant.

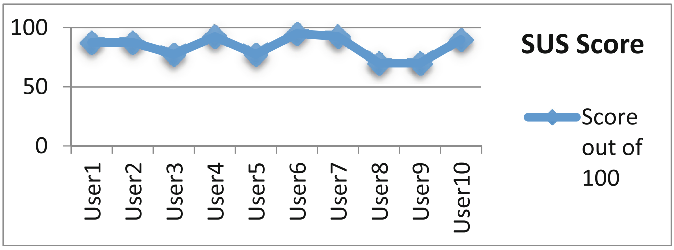
In addition to the tasks, the participants filled in questionnaires in order to assess the opinion of the users and retrieve qualitative results in a formal way (Table 1.). Furthermore, the participants were encouraged to express their thoughts throughout the evaluation process, which were written down using notes. Finally, a system usability scale (SUS) questionnaire was assigned in order measure the usability of the system in terms of design, interaction and effectiveness as an assistive tool.

## 5.4 Results

**Usability and Interaction.** The users found the virtual human very helpful and pleasurable to interact with. The information displayed by the system was clear and users did not face difficulties in perceiving any aspect of the system. Although there were a few comments regarding some design decisions, the users found the user interface self-explaining and intuitive.

The System Usability Scale [21] provides a reliable tool for measuring usability. It consists of a 10 item questionnaire with five response options for the participants; from strongly agree to strongly disagree.

A SUS score above a 68 would be considered above average and anything below 68 is below average. The results of System Usability Scale were very encouraging with a final score of 84. The graph above displays the SUS score per participant (ranging from 70 to 92.5 – Graph 1).




**Graph 1:** (Left) selection of presentation type, (right) image gallery presentation

Gestural interaction involved hand movement in any direction and the manipulation of a virtual cursor. The users performed the supported gestures without difficulty and found them intuitive for specific tasks, such as displaying the next image in a slide-show. Furthermore, they were able to control the virtual cursor in order to select items shown at the display. Four of the users (40 %) initially faced difficulties using the cursor when moving it to the lower part of the display. However, once they got familiar to it they could manipulate the system without problem.

The participants were provided with the list of all the available voice commands in order to assess the responsiveness, the exactness and the usefulness of speech recognition. Despite not being accurate at all times, users generally managed to manipulate the system using voice commands with a maximum of three tries. Despite the fact that the system did not focus on human-assistant conversation, voice recognition proved to be pleasing and successful, motivating future research in this area.

A smartphone was given to the participants of the evaluation and the available gestures were described. The users enjoyed interacting using mobile devices but preferred the other interaction techniques (hand gestures and speech recognition) which were considered more natural.

**Table 1.** The users' answers to the evaluation questionnaire regarding the three supplied means of interaction: gestures, voice commands and smartphone gestures

| Questions | Gestures | Voice commands | Smartphone interaction |
|---|---|---|---|
| **I like interacting with this method** | Agree: 10<br>Disagree: 0<br>Neutral: 0 | Agree: 9<br>Disagree: 0<br>Neutral: 1 | Agree: 8<br>Disagree: 1<br>Neutral: 1 |
| **No special training is needed to learn** | Agree: 9<br>Disagree: 0<br>Neutral: 1 | Agree: 9<br>Disagree: 0<br>Neutral: 1 | Agree: 6<br>Disagree: 0<br>Neutral: 4 |
| **Corresponding precisely to my actions** | Agree: 6<br>Disagree: 0<br>Neutral: 4 | Agree: 6<br>Disagree: 2<br>Neutral: 2 | Agree: 6<br>Disagree: 0<br>Neutral: 4 |
| **It was awkward to use** | Agree: 3<br>Disagree: 5<br>Neutral: 2 | Agree: 6<br>Disagree: 4<br>Neutral: 0 | Agree: 0<br>Disagree: 5<br>Neutral: 5 |
| **Responded promptly to my actions** | Agree: 7<br>Disagree: 0<br>Neutral: 3 | Agree: 3<br>Disagree: 2<br>Neutral: 5 | Agree: 3<br>Disagree: 0<br>Neutral: 7 |
| **It was tiring to use** | Agree: 0<br>Disagree: 9<br>Neutral: 1 | Agree: 2<br>Disagree: 6<br>Neutral: 2 | Agree: 2<br>Disagree: 5<br>Neutral: 3 |
| **I would prefer another method of interaction** | Agree: 0<br>Disagree: 7<br>Neutral: 3 | Agree: 1<br>Disagree: 8<br>Neutral: 1 | Agree: 0<br>Disagree: 6<br>Neutral: 4 |

**Effectiveness in Manipulating TimeViewer.** The users were able to start interacting with TimeViewer immediately after they were informed by Bryan. They were aware of the supported gestures and knew what to expect to see. Especially after training, users were confident and able to perform the gestures right ahead.

Indicatively, 30 % of the users stated (without being asked) that the training was very helpful and considered themselves experienced users after experimenting using the virtual assistant as they had "tried it in action before". The confidence was so apparent that one of the users stated his belief that "gesture tracking is more precise in TimeViewer", although the same software and hardware was used in both cases.

Supplementary to the assigned tasks, the users were given extra questions in order to extract additional information and urge them to express their opinion regarding the efficiency and the helpfulness of the virtual assistant. The users believed that TimeViewer would be difficult to use without interactive guidance, as other methods such as leaflets would be insufficient. Moreover, Bryan as an assistant was described as sufficient, comprehensive and did not display redundant information.

## 6 Conclusion and Future Work

This paper reports on the design, development and evaluation of a framework which implements virtual humans supporting body gestures and speech synthesis that can be used for information provision. The framework can be used to create interactive multimedia information visualizations (e.g., images, text, audio, videos, 3D models) and provides a dynamic data modeling mechanism for storage and retrieval and implements communication through multimodal interaction techniques. The interaction may involve human-to-agent, agent-to-environment or agent-to-agent communication. The generated virtual agents can have diverse roles, as they can be used as assistive tools for existing systems, standalone "applications" or even as vital parts of smart environments.

When acting as an assistant, the framework provides the tools needed for presenting information in the form of tutorials and real-time help. Furthermore, the virtual humans support the users' training on interaction techniques applied for the assisted system. As standalone application, the framework provides categorized information visualization and interactive help about the interaction techniques and information presented. Finally, when the virtual humans are embedded in other environments the framework is be able to create hybrid mode that supports all the aforementioned functionalities individually or even in combination.

The developed framework supports multiple multimodal techniques in order to fit to various ambient intelligence environments and offer natural interaction, such as gestural, verbal and tangible interaction, in a wide range of setups.

A preliminary evaluation has already been conducted in order to assess the framework in terms of usability, effectiveness and likeability. Additional, more extensive assessments, with a larger number of participants covering a broader age range, are planned in order to further validate and improve the current implementation.

Based on the encouraging evaluation results, further research work is already underway in regarding the use of virtual humans in ambient intelligence environments. Improvements in the system's design are also foreseen based on observation results and specific user comments. Since speech recognition proved to be very popular among the users, the conversational abilities of the supported virtual humans will also be extended and improved.

A key drawback of the framework when used in practice was that lack of an interactive means for providing all the required digital content to the system. This will be addressed by the development of a visual tool for content insertion and editing. Finally, since the framework is built using an engine that supports interoperability with mobile devices future work will seek to port and adapt the current implementation to address the needs imposed by user mobility, as well, as small and multi-touch displays and sensor-enabled interaction.

**Acknowledgements.** The work reported in this paper has been conducted in the context of the AmI Programme of the Institute of Computer Science (ICS) of the Foundation for Research and Technology-Hellas (FORTH). The authors would like to express their gratitude to Anthony Katzourakis for the artistic work and 3D modelling and to the Signal Processing Laboratory (SPL) of ICS-FORTH and especially Elena Karamichali for creating the speech recognition system.

## References


1. Weiser, M.: The Computer for the twenty-first century. Sci. Am. **265**, 94–104 (1991)
2. Murray, J.: Composing multimodality. In: Lutkewitte, C. (ed.) Multimodal Composition: A Critical Sourcebook. Bedford/St. Martin's, Boston (2013)
3. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The Kaldi speech recognition toolkit. In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. IEEE Signal processing Society (2011)
4. Kumar, K.: Hindi speech recognition system using HTK. Int. J. Comput. Bus. Res. **2**, 2229–6166 (2011)
5. Hidden Markov Model Toolkit (HTK) speech Recognition. http://htk.eng.cam.ac.uk/develop/atk.shtml
6. Suarez, J., Murphy, R.: Hand gesture recognition with depth images: a review. In: 2012 IEEE RO-MAN, pp. 411–417. IEEE (2012)
7. Biswas, K.K., Basu, S.K.: Gesture recognition using Microsoft Kinect. In: 2011 5th International Conference on Automation, Robotics and Applications (ICARA), pp. 100–103 (2011)
8. Drossis, G., Grammenos, D., Birliraki, C., Stephanidis, C.: MAGIC: developing a multimedia gallery supporting mid-air gesture-based interaction and control. In: Stephanidis, C. (ed.) HCI International 2013 - Posters' Extended Abstracts. Springer, Heidelberg (2013)
9. Drossis, G., Grammenos, D., Adami, I., Stephanidis, C.: 3D visualization and multimodal interaction with temporal information using timelines. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013, Part III. LNCS, vol. 8119, pp. 214–231. Springer, Heidelberg (2013)
10. Grzejszczak, T., Mikulski, M., Szkodny, T., Jędrasiak, K.: Gesture based robot control. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 407–413. Springer, Heidelberg (2012)
11. Grammenos, D., Zabulis, X., Michel, D., Sarmis, T., Georgalis, G., Tzevanidis, K., Argyros, A., Stephanidis, C.: Design and development of four prototype interactive edutainment exhibits for museums. In: Stephanidis, C. (ed.) Universal Access in HCI, Part III, HCII 2011. LNCS, vol. 6767, pp. 173–182. Springer, Heidelberg (2011)
12. Ruiz, J., Li, Y., Lank, E.: User-defined motion gestures for mobile interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 197–206. ACM, 2011
13. Kray C., Nesbitt D., Dawson J., Rohs M.: User-defined gestures for connecting mobile phones, public displays, and tabletops. In: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 239–248. ACM (2010)
14. Hartholt, A., Traum, D., Marsella, S.C., Shapiro, A., Stratou, G., Leuski, A., Morency, L.-P., Gratch, J.: All together now, introducing the virtual human toolkit. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS, vol. 8108, pp. 368–381. Springer, Heidelberg (2013)
15. Virtual Human Toolkit. https://vhtoolkit.ict.usc.edu
16. Cassell, J., Prevost, S., Sullivan, J., Churchill, E.: Embodied Conversational Agents. MIT Press, Cambridge (2000)

17. Swartout, W., et al.: Ada and Grace: toward realistic and engaging virtual museum guides. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6536, pp. 286–300. Springer, Heidelberg (2010)
18. Baldassarri, S., Cerezo, E., Seron, F.J.: Maxine: a platform for embodied animated agents. Comput. Graph. **32**(4), 430–437 (2008)
19. Zhang, H., Fricker, D., Smith, T.G., Yu, C.: Real-time adaptive behaviors in multimodal human-avatar interactions. In: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, p. 4. ACM (2010)
20. Drossis, G., Grammenos, D., Bouhli, M., Adami, I., Stephanidis, C.: Comparative evaluation among diverse interaction techniques in three dimensional environments. In: Streitz, N., Stephanidis, C. (eds.) DAPI 2013. LNCS, vol. 8028, pp. 3–12. Springer, Heidelberg (2013)
21. Brooke, J.: SUS-a quick and dirty usability scale. Usability Eval. Ind. **189**, 4–7 (1996)