In: Maybury, M., Stock, O., Wahlster, W. (eds.): Intelligent Technologies for Interactive Entertainment.
Proceedings of the First International Conference, INTETAIN 2005, Madonna di Campiglio, Italy,
November/December 2005. Series: Lecture Notes in Computer Science, Subseries: Lecture Notes in Artificial Intelligence,
Vol. 3814, pp. 104-113, Springer, Berlin, Heidelberg, 2005.

Ambient Intelligence in Edutainment: Tangible Interaction with Life-Like Exhibit Guides

Alassane Ndiaye, Patrick Gebhard, Michael Kipp, Martin Klesen,
Michael Schneider, Wolfgang Wahlster

German Research Center for Artificial Intelligence
DFKI GmbH, Stuhlsatzenhausweg 3, D-66123 Saarbrücken
<first name>.<last name>@dfki.de

Abstract. We present COHIBIT, an edutainment exhibit for theme parks in an ambient intelligence environment. It combines ultimate robustness and simplicity with creativity and fun. The visitors can use instrumented 3D puzzle pieces to assemble a car. The key idea of our edutainment framework is that all actions of a visitor are tracked and commented by two life-like guides. Visitors get the feeling that the anthropomorphic characters observe, follow and understand their actions and provide guidance and motivation for them. Our mixed-reality installation provides a *tangible,* (via the graspable car pieces), *multimodal,* (via the coordinated speech, gestures and body language of the virtual character team) and *immersive* (via the large-size projection of the life-like characters) experience for a single visitor or a group of visitors. The paper describes the context-aware behavior of the virtual guides, the domain modeling and context classification as well as the event recognition in the instrumented environment.

1 Introduction

Ambient Intelligence (AmI) refers to instrumented environments that are sensitive and responsive to the presence of people. Edutainment installations in theme parks that are visited by millions of people with diverse backgrounds, interests, and skills must be easy to use, simple to experience, and robust to handle. We created **COHIBIT** (**CO**nversational **H**elpers in an **I**mmersive exhi**BI**t with a **T**angible interface), an AmI environment as an edutainment exhibit for theme parks that combines ultimate robustness and simplicity with creativity and fun. The visitors find a set of instrumented 3D puzzle pieces serving as affordances. An affordance is a cue to act and since these pieces are easily identified as car parts, visitors are being motivated to assemble a car from the parts found in the exhibit space. The key idea of our edutainment framework is that all actions of a visitor, who is trying to assemble a car, are tracked and two life-like characters comment on the visitor's activities. Visitors get the feeling that the anthropomorphic characters observe, follow and understand their actions and provide guidance and motivation for them.

We have augmented the 3D puzzle pieces invisibly with passive RFID tags linking these items to digital representations of the same (cf. [9]). Since we infer particular assembly actions of the visitors indirectly from the change of location of the instru-

mented car parts, the realtime action tracking is extremely simplified and robust compared with vision-based approaches observing the behavior of the visitors (see also [10]). The COHIBIT installation provides a *tangible,* (via the graspable car pieces), *multimodal,* (via the coordinated speech, gestures and body language of the virtual character team) and *immersive* (via the large-size projection of the life-like characters) experience for a single visitor or a group of visitors. Our installation offers ultimate robustness, since visitors can always complete their physical assembly task, even if the RFID-based tracking of their actions or the virtual characters are stalled. As Cohen and McGee put it, such hybrid systems "support more robust operation, since physical objects and computer systems have different failure modes" (cf. [1], p. 44). Thus, the ambient intelligence provided by the situation-aware virtual guides can be seen as an exciting augmentation of a traditional hands-on exhibit, in which the visitors do not see any computer equipment or input devices.

The fact that visitors in our AmI environment are not confronted with any computing device contrasts the work described here with previous work. Like our system, PEACH (cf. [8]) and Magic Land (cf. [6]) both use mixed reality approaches for edutainment in museums, but the PEACH user must carry a PDA and the Magic Land user must wear a HMD to experience the mixed reality installation.

The remainder of the paper is organized as follows. First, we present an overview of the instrumented edutainment installation. Then we focus on the distinguishing features of our approach: Section 3 concentrates on the context-aware behavior of the virtual guides, while Section 4 focuses on the domain modeling and context classification and Section 5 deals with the event recognition in the instrumented environment. Finally, Section 6 concludes the paper.

2 The Instrumented Edutainment Installation

The COHIBIT installation creates a tangible exploration experience guided by a team of embodied conversational agents (cf. [5]). While the visitor can fully engage in the manipulation of tangible, real world objects (car pieces), the agents must remain in the background, commenting on visitor activities, assisting where necessary, explaining interesting issues of car technology at suitable moments and motivating the visitor if s/he pauses for too long. Since turn-taking signals (speech, gesture, posture) from the visitor cannot be sensed by our AmI environment, the agents must infer from the sensor data (car pieces movement) and from the context (current car configuration constellation, current state of the dialog) when it is a good time to take the turn/initiative and when to yield a turn, probably interrupting themselves.

The exhibit requires interaction modalities that allow us to design a natural, unobtrusive and robust interaction with our intelligent virtual agents. We use RFID devices for this purpose. This technology allows us to determine the position and the orientation of the car pieces wirelessly in the instrumented environment. The RFID-tagged car pieces bridge the gap between the real and the virtual world. By using tangible objects for the car assembly task, visitors can influence the behavior of the two virtual characters without necessarily being aware of it. The RFID technology does not restrict the interaction with the exhibit to a single visitor. Many visitors can move car

pieces simultaneously making the car assembly task a group experience. The technical set-up of the AmI environment (see Figure 1) consists of:

- Ten tangible objects that are instrumented with passive RFID tags and represent car-model pieces on the scale 1:5. There are four categories of pieces: (a) two front ends; (b) one driver's cab, including the cockpit as well as the driver and passenger seats; (c) two middle parts providing extra cabin space, e.g. for a stretch limousine; (d) five rear ends for the following body types: convertible, sedan/coupé, compact car, van and SUV.
- A workbench and shelves: The workbench has five adjacent areas where car pieces can be placed. Each area can hold exactly one element and the elements can be placed in either direction, i.e. the visitor can build the car with the front on the left and the rear on the right hand side or the other way around.
- A high-quality screen onto which the virtual characters are projected in life-size. (We use state-of-the-art characters and CharaVirld^{TM}, the 3D-Player of Charamel.) We project also a virtual screen which is used to display technical background information in the form of graphics, images, and short video sequences.
- Three cameras facing to visitors and mounted on the ceiling for realtime detection of visitor presence.
- A sound system for audio output in the form of synthesized speech feedback. (The audio output is provided by a unit selection based Text-To-Speech system.)

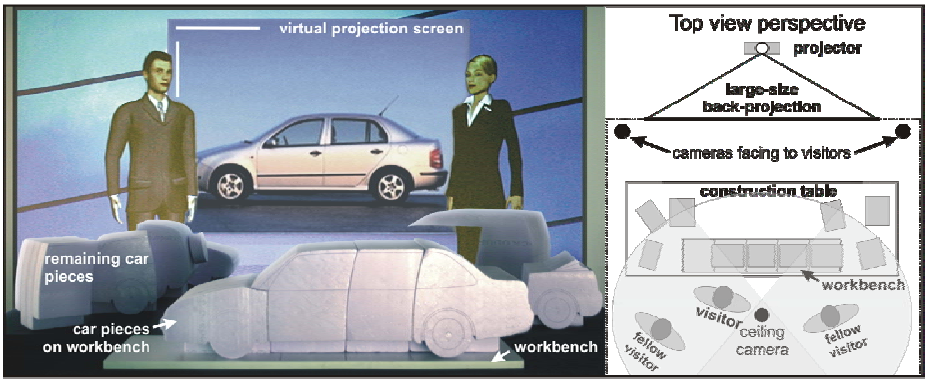

Fig. 1. Overview of the interactive installation prototype.

The installation runs in two modes. The OFF mode is assumed when nobody is near the construction table. A visitor approaching the table is being detected by the ceiling camera and lets the system switch to ON mode. The idea behind these two modes is based on our experiences with CrossTalk – a self-explaining virtual character exhibition for public spaces [7]. In the OFF mode the virtual characters try to capture the attention of potential visitors. They talk to each other about their jobs and hobbies while making occasional references to the situational context (i.e., the current time of day, the outside weather, upcoming events). Making references to current real world conditions creates an illusion of life and establishes common bonds with the visitor's world.

When the visitors enter the installation, the system recognizes their presence and switches to ON mode. The virtual characters welcome them and explain what the purpose of the exhibit is and encourage the visitors to start with the car assembly task. In the following construction phase visitors can remove pieces from the shelves and put them on the workbench. They can also modify a construction by rearranging pieces or by replacing them with other elements. In the ON mode the virtual characters play various roles, e.g., as *guides*, by giving context-sensitive hints how the visitor can get the construction completed; *commentators and expert critic*s, by accompanying the assembly process through personalized comments on the visitors' actions and on the current task state; *motivators*, by encouraging the visitors to continue playing; *tutors*, by evaluating the current state of the task and providing additional background information about car assembly. Figure 2 illustrates the visitor's experience in our AmI exhibition space by showing the three basic phases of the ON modus (welcome, construction and completion).

Fig. 2. Examples of visitor actions and corresponding character behavior.

3 Creating Context-Aware Behavior of the Virtual Guides

In the COHIBIT installation, the characters' verbal and non-verbal behavior is defined as a function of the visitors' actions and the current state of the construction. The characters should be able to react instantly (e.g., by interrupting their current talk), adapting their conversation to the current situation. At the same time, we must avoid that their comments and explanations become too fragmented and incoherent as they react and interrupt each other. Commenting every move would be boring or even irritating. Instead, the system has to pick significant events like a soccer commentator who does not comment on every move of the many players on the field but selects those events that are most significant for the audience. Our design guidelines for the mixed-initiative interactive dialogue were developed by aiming at three major goals: (1) believability, (2) assistance, and (3) edutainment.

Believability means to strengthen the illusion that the agents are alive. One means to achieve this is to let the agents live beyond the actual interaction in what we call the OFF mode. If no visitors are present, the two agents are still active, engaged in private conversation. Thus, from a distance, passers-by take them for being alive and, as an important side-effect, they might be attracted to enter the installation [4]. In ON mode, believability is created by making the agents react intelligently to visitor actions, i.e. being aware of the construction process and knowing possible next actions to complete the car assembly. The second goal is that of assistance concerning the car assembly task. Although it is not a difficult task, observation of naïve visitors showed that the agents' assistance can speed up the task by identifying relevant car pieces, giving suggestions where to place pieces and how to rebuild configuration constellations that can never result in a valid car model. The third goal is to convey information in an entertaining fashion on two domains: cars and virtual characters. Both topics arise naturally in our AmI environment. The car pieces serve as a starting point for explanations on topics such as air conditioning or safety features of cars. Virtual character technology is explained mainly in the OFF mode, during smalltalk, but also in the ON mode. For example, if the same situation has occurred so often that all pre-scripted dialog chunks have been played, the characters react by apologizing for using only a limited number of pre-scripted dialog contributions.

3.1 Scenes and Sceneflow

The installation's behavior is defined in terms of structure (*sceneflow*) and content (*scenes*) [2]. The sceneflow is modeled as a finite state machine (currently consisting of 94 nodes and 157 transitions) and determines the order of scenes at runtime using logical and temporal conditions as well as randomization. Each node represents a state, either in the real world (a certain car construction exists on the workbench) or in the system (a certain dialog has been played). Every transition represents a change of these states: e.g., a visitor places another piece on the workbench or a certain time interval has passed. Each node and each transition can have a playable scene attached to it. A scene is a pre-scripted piece of dialog that the two agents perform. Our 418 scenes are defined separately from the sceneflow definition in the form of a multimodal script which contains the agents' dialog contributions, gestures, facial expressions

and optional system commands (e.g., displaying additional information on the virtual screen). Scenes can be edited by an author with standard text processing software.

The major challenge when using pre-scripted dialog segments is variation. For the sake of believability the characters must not repeat themselves. Therefore, the multimodal script allows for context memory access (identifier of the relocated car piece, session time, current date and weather situation) and making scenes conditional (play a scene only on a Wednesday or only between 9:00 and 11:00 am). The contextual information is verbalized at runtime. In addition, we increase variation by *blacklisting*: already played scenes are blocked for a certain period of time (e.g., 5 minutes), and, variations of these scenes are selected instead. For each scene, we have 2-9 variations that make up a so-called scene group. Scene groups that need a high degree of variation are identified by analyzing automatic transcripts of real-life test runs for high density of scene groups. In case that all scenes of a scene group are played, a generic diversion scene is triggered that can be played in many contexts. Various sets of such scenes exist where the agents comment on the current car piece, the number of today's visitors, the current weather etc. As a third strategy for variation, long scenes are decomposed into smaller sections that are assembled at runtime based on the current conditions and blacklisting constraints.

3.2 Design Optimization Based on User Studies

In an iterative development approach, the system was tested repeatedly with 15 naïve visitors at two different stages of development to identify necessary changes. In these informal tests, an evaluation panel, including research staff and members of the theme park management, observed the visitors (including groups of up to three) who interacted with the system to check on system performance and robustness under realistic conditions. A number of visitors were interviewed by the panel after the test. Critical discussion of our observations and the interviews led to essential changes in the prototype's interaction design:

- Visitors are grateful for instructions and assistance. Some visitors seem to be afraid of doing "wrong" in presence of fellow visitors → Focus on the assembly task first (concise comments), assist where necessary, tell about more complex issues later.
- Visitors move multiple pieces at once. Multiple visitors even more so. The system is bombarded with many simultaneous events → Introduce lazy event processing as a form of wait-and-see strategy (see Section 5).
- Visitors memorize only few facts presented as background information being too busy with the assembly task. → Give in-depth information at points of "natural rest" (car is completed), and visualize the information using multimedia.
- Visitors may be more interested in the agents than in the car. They try to find out how the agents react to various situations like "wrong" car piece combinations. → Include "meta-dialogs" about virtual character technology to satisfy these interests and cover a large range of "error cases" to give the visitor space for exploration.

The user studies led to strategies for deciding when the agents should start speaking and for how long, based only on the knowledge about the current visitor action and the state of the dialog. A piece being placed on the workbench is a possible time to

start talking because the visitor (1) may expect the agents to talk (many visitors placed a piece on the table and looked at the agents expectantly) or (2) may need assistance on how to continue. However, starting to speak each time a piece is placed could irritate visitors focused on the assembly task. Consequently, (1) depending on the visitor's assembly speed (number of actions per time unit) we give a lengthy, short or no comment and (2) we let the agents interrupt themselves as new events arrive. To make the interruption "smooth" we insert transitory scenes, which depend on the interrupted speaker. For example, if agent A is being interrupted, agent B would say: "Hold on for a second." Alternatively, agent A could say: "But I can talk about that later." An even more important time to initiate talking by the agent is when a car is completed. Depending on the context, (a) the completed car is described as a "reward" for correct assembly or (b) a certain aspect of the car is explained in-depth or (c) a recommendation for another exhibit is given. As our system is based on probabilistic state transitions, the density of information given at the various stages of the interaction can be easily adapted according to user studies.

4 Domain Modeling and Context Classification

As mentioned in Section 2, there are ten instrumented pieces that the visitors can use to build a car. Elements can be placed on each of the five positions on the workbench in either direction. This leads to a large number of possible configurations, totaling 802,370 different combinations! The RFID readers provide the AmI environment with a high-dimensional context space. It is obvious that we cannot address each configuration individually when planning the character behavior. On the other hand we need to be careful not to over-generalize. If the virtual exhibit guides would just point out "This configuration is invalid." without being able to explain why or without giving a hint how this could be rectified, their believability as domain experts would be diminished, if not destroyed. We use a classification scheme for complete and partial constructions that consists of the following five categories:

1. *Car completed*: a car has been completed (30 valid solutions).
2. *Valid construction*: the construction can be completed by adding elements.
3. *Invalid configuration*: an invalid combination of elements (e.g., driver's cab behind rear element).
4. *Completion impossible*: the construction cannot be completed without backtracking (e.g., driver's cab is placed on an outermost workbench position where there is no possibility to add the rear)
5. *Wrong direction*: the last piece was placed in the opposite direction with respect to the remaining elements.

An important concept in our classification scheme is the orientation of the current construction (car pointing to left or right). The orientation is determined by the majority of elements pointing in the same direction. If, for example, two elements point to the left but only one element points to the right then the current overall orientation is "left". The current orientation can change during construction. Workbench configurations that only differ in orientation are considered equivalent. Using the concept of

orientation the agents can assist if pieces point in the "wrong" direction by saying: *"Sorry, but we recommend flipping this new piece so it points in the same direction as the other pieces."*.

In many cases only the category of an element needs to be considered and not the instance. A category is denoted by F for front element, C for cockpit, M for middle element, and R for rear element. Each configuration context is represented by a construction code that has a length between one and five. Empty positions on the left and on the right hand side of the placed pieces are omitted and empty positions between pieces are marked with a hash symbol. The construction code F#M, for example, describes a state in which a front element and a middle element are placed somewhere on the workbench with an empty position between them. If the user remains inactive in this situation the agents will take initiative: *"May we propose to put the cockpit in the gap between the front end and the middle piece."*. The construction code is orientation-independent. Configurations on the workbench are treated as equivalent if they have the same construction code. The car type is defined by the construction code and the rear element used. A roadster, for example, has the construction code FCR and the rear of a convertible. Using the available pieces, visitors can build 30 different cars.

If the visitor produces erroneous configurations we only evaluate the *local* context of the configuration, which comprises the neighboring positions of the currently placed piece. The error code FR#, for example, describes a situation in which a rear element has been placed behind a front element. This configuration is invalid since a front element must be followed by a driver's cab. Here, the agents will help: *"We are sorry, but you have to put first the cabin behind the front, before placing a rear piece."*. The evaluation of the local context can also result in an error situation, where a completion is impossible.

The three concepts current orientation, construction code, and local context enable us to reduce the combinatorial complexity by classifying each configuration context into five distinct categories. The construction code and the error code can be seen as succinct descriptions of the current situation on the workbench. They provide enough context information to react intelligently to the current state of the construction.

5 Recognizing Events in the Instrumented Environment

The system's sensors consist of cameras and RFID readers. Sensory data is interpreted in terms of visitor actions (RFID readers) and visitor arrival/departure (cameras). Visitor actions consist of placing and removing car pieces to/from the workbench. Since these actions eventually control the way the characters behave, the major challenge is how to process the raw sensory data, especially if many actions happen at the same time (multiple visitors moving multiple pieces).

The processing of the sensory data has 2 phases: (1) mapping of the visitors' actions onto internal transition events, and (2) updating the representation of the current car construction. In terms of transition events, we distinguish four types that are ranked in a specific priority scheme:

1. *Visitor appeared*: visitor has entered the installation.
   Visitor disappeared: visitor has left the installation.

2. *Car completed*: visitor has completed the car assembly.
   Car disassembled: visitor has disassembled a car by removing a piece.
3. *Piece taken*: visitor has removed a piece from the workbench.
   Piece placed: visitor has placed a piece on the workbench.
4. *Piece upheld*: visitor is holding a piece above the workbench.

The events *visitor appeared* and *visitor disappeared* have the highest priority since they trigger ON and OFF mode. The events *car completed* and *car disassembled* have the second highest priority since they trigger the transitions between the construction phase and the completion phase in the sceneflow. The other three events are used in the construction phase to trigger scenes in which the characters comment on the current state of the construction, give in-depth information about the used pieces and hints on how to continue. We use this priority scheme to decide which transition events are considered for further processing, e.g., if two visitors move pieces simultaneously. If the first visitor completes the construction of a car and the second visitor places another piece on the workbench shortly afterwards, the characters start commenting on the completed car instead of the single unused piece.

The generation of transition events is followed by the classification of the current state of the construction using the five categories in our domain model (cf. Section 4). In addition, context information like the category, instance, and orientation of the currently placed piece, the number of pieces on the workbench, the overall creation time, the number of errors so far is updated. This context information along with the transition event is stored in an *action frame*. In a last step, the action frame that contains the transition event with the highest priority is selected for further processing. Transition events are used to branch in the sceneflow graph and context information is used in the selected scenes. The decision when to handle the next event is an integral part of our dialog/interaction model and explicitly modeled in the sceneflow (cf. [3] for more details). This enables us to keep the comments and explanations of the characters consistent and to balance reactivity and continuity in their interactive behavior.

6 Conclusions

We have presented COHIBIT, an AmI edutainment installation that guides and motivates visitors, comment on their actions and provides additional background information while assembling a car from instrumented 3D puzzle pieces. The fact that visitors in our AmI environment are not confronted with any computing devices contrasts the work described here with previous work in which visitors have to deal with additional hardware to experience the mixed reality installation.

The system is currently fully implemented and will be deployed in a theme park at the beginning of 2006, using professionally modeled car pieces and exhibit design. Although we have already conducted various informal user studies during the development of the various prototypes, the actual deployment will give use the opportunity for a large-scale empirical evaluation.

In future work we intend to exploit the full potential of the vision module by recognizing visitors approaching or in the near of the exhibit to invite and encourage them to visit the installation. The computer vision module could also be used to

determine whether the visitors are looking at the characters or are engaged in the car assembly task, in order not to interrupt. A further issue we plan to investigate is the appropriate dealing with groups of individuals interacting with the system. A prerequisite for this project is the ability to track many visitors during the whole time they are on the exhibit.

Acknowledgments: We are indebted to our colleagues Gernot Gebhard and Thomas Schleiff for their contributions to the system. We thank our partners of Charamel (www.charamel.de) for providing us with the 3D-Player and the virtual characters and our colleagues of the department "Multimedia Concepts and their Applications" at the University of Augsburg for the realtime video-based presence detection system.

Parts of the reported research work have been developed within VirtualHuman, a project funded by the German Ministry of Education and Research (BMBF) under grant 01 IMB 01.

References

[1] Cohen, P. R., McGee D. R.: Tangible multimodal interfaces for safety-critical applications. In: Communications of the ACM 47(1), 2004, pp. 41-46.

[2] Gebhard, P., Kipp, M., Klesen, M., Rist, T.: Authoring scenes for adaptive, interactive performances. In: Proc. of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems, ACM Press, New York, 2003, pp. 725-732.

[3] Gebhard, P., Klesen, M.: Using Real Objects to Communicate with Virtual Characters. In: Proc. of the 5th International Working Conference on Intelligent Virtual Agents (IVA'05), Kos, Greece, 2005, pp. 48-56.

[4] Klesen, M., Kipp, M., Gebhard, P., Rist, T.: Staging exhibitions: methods and tools for modelling narrative structure to produce interactive performances with virtual actors. Virtual Reality, Vol. 7(1), Springer, 2003, pp. 17-29.

[5] Prendinger, H. and Ishizuka, M. (eds.) Life-like Characters: Tools, Affective Functions and Applications, Springer, 2004.

[6] Nguyen, T., Qui, T., Cheok, A., Teo, S., Xu, K., Zhou, Z., Mallawaarachchi, A., Lee, S. Liu, W., Teo, H., Thang, L., Li, Y., Kato, H.: Real-Time 3D Human Capture System for Mixed-Reality Art and Entertainment. In: IEEE Transactions on Visualization and Computer Graphics, vol. 11 (6), November/December 2005, pp. 706-721.

[7] Rist, T., Baldes, S., Gebhard, P., Kipp, M., Klesen, M., Rist, P., Schmitt, M.: Crosstalk: An interactive installation with animated presentation agents. In: Proc. of the Second Conference on Computational Semiotics for Games and New Media, Augsburg, September 2-4, 2002, pp. 61-67.

[8] Rocchi, C., Stock, O., Zancanaro, M., Kruppa, M., Krüger, A.: The museum visit: generating seamless personalized presentations on multiple devices. In: Nunes, N. J., Rich, Ch. (ed.): International Conference on Intelligent User Interfaces 2004. January 13-16, 2004, Funchal, Madeira, Portugal. pp. 316-318.

[9] Ullmer, B. and Ishii, H.: Emerging Frameworks for Tangible User Interfaces. In: Carroll, J.M., (ed.): "Human-Computer Interaction in the New Millennium", Addison-Wesley, 2001, pp. 579-601.

[10] Wasinger, R., Wahlster, W.: The Anthropomorphized Product Shelf: Symmetric Multimodal Interaction with Instrumented Environments. To appear in: Aarts, E., Encarnação, J. (eds.): True Visions: The Emergence of Ambient Intelligence, Springer, 2005.