Conversational Characters that Support Interactive Play and Learning for Children

Andrea Corradini[1], Manish Mehta[2] and Klaus Robering[1]
[1]University of Southern Denmark, IFKI
[2]Georgia Institute of Technology, College of Computing
[1]Kolding, Denmark
[2]Atlanta, GA, USA

1. Introduction

Over the last decade, digital technology has been increasingly used for educational purposes. The rationale behind this emerging trend is the belief that information technology can be utilized as a powerful means to assist learners with the acquisition of general knowledge, literacy, narrative competence, social skills like teamwork and negotiation capabilities, logical and spatial reasoning, eye-hand coordination and fine motor control, etc. In particular computer games and interactive technology have attracted a great deal of attention and much active research is currently being carried out in an effort to investigate their possible benefits in education (Squire, 2005).

In this paper, we report on the development and evaluation of an educational agent architecture that allows for instructive communication and natural human-computer interaction between humans, in particular children and adolescents, and embodied historical characters. Our e-learning software platform is built around a three-dimensional animated portrayal of the world famous writer Hans Christian Andersen. Users can interact with the graphical character through a combination of speech, (pen and/or mouse) gesture, and keyboard and, in so doing, ask the virtual character questions about Andersen's life, family, historical background, fairy tales and stories. The avatar is endowed with basic affective and emotional capabilities, and uses both synthesized speech and gestures to converse with the pupils. The intent of the system is educational while the interactive nature of the character is designed to help children learn and retain information about the subject matter to a greater extent than would be possible with traditional media. The educational content is pervasive for it is blended within the game-like system interface. In fact, there is no clear cut distinction between the system and the educational content which is always presented in different ways according to the conversational situation thus providing children with new opportunities for effective, immersive and rewarding computer-aided learning experiences.

Conceptually, our system is backed by theories of user-centered learning as well as cognition. The constructivist theory of education (Mantovani, 2001) that recognizes the importance of active experience of students with the learning material supports our approach to education. Findings in the theory of multimedia learning (Mayer, 2001) that

Source: Multiagent Systems, Book edited by: Salman Ahmed and Mohd Noh Karsiti,
ISBN 978-3-902613-51-6, pp. 426, February 2009, I-Tech, Vienna, Austria

show how multimedia presentations are more effective in terms of measured learning outcomes when auditory information is presented alongside of visually complementary images or redundant textual information furnish instead the conceptual justification for the deployment of graphical characters capable of multimodal information generation.

A series of studies that we ran with school pupils show that the implemented agent system is a promising step towards our ambitious long term goal to develop interactive game-like interfaces that enable multimodal communication between users and game characters for edutainment purposes.

The remainder of this paper is structured as follows. First, we present the motivation for this research and discuss a few relevant works that links the game community with educational science and practitioners. We then briefly outline our interactive software system and a few main issues encountered during its development. Further, we focus on a user study that we carried out to gauge our implementation and to assess the system performance. Eventually, we conclude with a discussion of lessons learned and propose a number of possible extensions and future improvements.

2. Convergence of education and game technology

2.1 Our approach: a platform for e-learning

E-learning is a domain that attempts to use information technology to improve people's learning process and skills. A variety of software systems are used within this area in order to organize and present content for didactic purpose according to a variety of design methodologies, target audience and interaction styles. One may distinguish two main forms of e-learning: asynchronous and synchronous e-learning. In asynchronous e-learning the content is created beforehand for learners to access and study the information they need at any time from anywhere. In synchronous e-learning the teacher and the student(s) interact in an actual class, usually over a broadband network. In both situations, the learning material is stored and communicated digitally as text, audio and/or video clips and can be accessed via Internet or from a storage device such as CDROMs, etc.

The development of e-learning systems relies on the assumption that software instruments could enable students to easily assimilate the learning material while simultaneously help decongesting overcrowded classes by allowing students (teachers) to attend (lecture) a class from their own location. Moreover, e-learning inherently supports the idea of life-long education and represents a valuable opportunity for some specific groups of individuals, such as children or disabled people, who can have learning materials tailored and made accessible to them according to their needs.

Real-life training of people and children in particular, is a highly challenging and demanding task. There are no standards or agreed upon guidelines for computer supported learning. Typically, special-purpose multimedia software tools that resort to experimental learning strategies and methods have been deployed in a variety of applications targeted at different audiences (Savidis et al., 2007). These approaches are backed by constructivist theories of education (Mantovani, 2001; Steffe & Gale, 1995) and by the theory of multimedia learning (Mayer, 2001). Constructivist learning theory emphasizes the importance of active experience of students with the learning material. The theory of multimedia learning is relevant in this context because of its "dual-channel" assumption based on the evidence that humans have separate processing systems for visual/pictorial versus auditory/verbal channels of information. Both processing systems have limited processing capacity.

Meaningful learning requires mental processing in both verbal and visual channels, in order to build connections between them (Wickens, 2002). Thus multimedia presentations are more effective in terms of measured learning outcomes when auditory information is presented alongside of visually complementary images or redundant textual information (Mayer & Moreno, 2003; Moreno & Mayer, 2002). The fact that multimodal redundancy serves to focus attention is buttressed by studies on infant language acquisition (Gogate et al., 2001; Bahrick et al., 2004).

The importance of emotions in learning processes is also more and more acknowledged (Anolli et al., 2006; Kort et al., 2001; LeDoux, 1998). Regarding this aspect though, multimedia presentations are only a little more powerful than conventional textbooks in conveying the rich emotional and social information that is exchanged by people when they communicate with each other and that also occur in traditional classes. While teachers can modify their teaching style and methods based on perceived learner's feedback, e-learning platforms are often too rigid in their design and rarely account and accommodate for the emotional and motivational aspects of their users.

Insights in the theory of user-centered learning, constructivist learning theory, multimedia learning along with the need to develop an e-learning software platform endowed with basic affective and emotional capabilities, have been taken into account in the development of a 3D animated conversational character that impersonates the Danish fairy tale writer Hans Christian Andersen (HCA). The target audience of our system is pupils of age 10 to 18. The intended goal is to teach them about the writer's life, historical period, works, family and the characters featured in his writings in a game-like interface. The educational content is presented within a gaming application and depends on the conversational situation. This provides children with new opportunities for a more effective, immersive and rewarding computer-aided learning experience.

2.2 Learning theory desiderata

A growing number of researchers and practitioners in education have recognized the inherent limitations of the traditional large lecture class as pedagogical instrument (Foreman, 2003). The major critic to this way of delivering information is that it cannot deal with the very specific needs of the individual participants. Every student confronted with a new learning situation has a unique knowledge level and a given set of dispositions that cannot be accounted for in a traditional lecture. Moreover, when students encounter problems they must be given immediate feedback in terms of clarification or amplification. This is not always possible in a standard class.

More and more pedagogues and educators advocate for constructive learning environments that allow students to explore learning materials, ideally multi-sensorially, and that encourage comprehension through the active discovery of new knowledge (Mantovani, 2001; Narayanan & Potamianos, 2002; Steffe & Gale, 1995) while also integrating cognitive and emotional factors (Goleman, 1995, Picard et al., 2004). If the learning process is perceived as a rewarding activity, students are motivated to learn regardless of any external imposed requirements and devote more time to the material that is to be learnt. This contributes to the creation of enduring neural structures that ensure that concepts are stored into long-term memory rather than just crammed up for the upcoming test or examination.

2.3 Interactive game technology and game-based learning

Computer games have the power to engage and maintain the attention of players over long periods of time. Most commercial computer games are built around the common conceit of game as a karmic wheel: the user plays, dies, reloads and repeats this operative loop until he gets it right. Digital game-based learning detaches from and goes beyond this common pattern. It aims at exploiting the pedagogical promise of interactive technologies and applies them to the development of games that have educational goals. This kind of learning has its roots in both interactive adventure games and simulation games.

Before successful computer games such as *Zork* and *PacMan* hit the market, there was the world of adventure games. Adventures such as *Hunt the Wumpus* (Ahl, 1979) and *Colossal Caves Adventure* (Hafner & Lyon, 1996) were among the first text interactive fiction games which lured players into immersing themselves in a magical world. They both featured a competent text parser that could understand advanced commands from the player who could control the main character's movements with the arrow keys on the keyboard (Montfort, 2003; Wolf & Perron, 2003). This game concept was later extended to include animated color graphics, a pseudo 3D-perspective of the world game, music playing in the background, and sound feedback and turned into the game *King's Quest I: Quest for the Crown*, a story based on classic fairy-tale elements where a knight has to save a kingdom in distress by recovering a set of lost treasures. A game like this had never been made before and marked the beginning of a new era in interactive graphical games (Levy, 2001).

Real-life training of people is a highly challenging and demanding process that can be effectively improved with the deployment of special-purpose software. New developments in the fields of speech recognition, natural language processing, and computer graphics have given rise to the emergence of more sophisticated game-like computer interfaces and simulations to learning systems (Rieber, 1996). Also called serious games, these recent interfaces have attracted the interest of educators, military and professional practitioners who are interested in the possible benefits of bringing gaming and simulation together by immersing real people into possible critical situations that they should be prepared to deal with. 3D virtual environments of a large-bodied aircraft cargobay and turbine engine blades coupled with interaction modalities based on a head-mounted display and a 6 degrees-of-freedom mouse have been utilized for the training of inspection methods for aviation maintenance technicians (Washburn et al., 2007). The *Mission Rehearsal Exercise* (Gratch & Marsella, 2005) is a system designed to teach leadership skills in high-stress social and emotionally charged situations. *Carmen's Bright IDEAS* is an interactive health intervention system designed to train communication skills and social competence of medical personnel interacting with mothers of pediatric cancer patients (Marsella & Gratch, 2003). In *FearNot* (Aylett et al., 2005) children can watch bullying incidents that take place between scripted agents in a virtual school and are asked to act in support to the victimized characters. The system aims to enable children to develop strategies to cope with bullying situations through empathic interaction with the synthetic characters that inhabit the virtual school. While most research on embodied conversational characters has concentrated on the graphical representation and conversational capabilities of virtual agents, (Oviatt et al., 2004) investigated the question of whether auditory embodiment can provide cues that influence user behaviors and ultimately affects the learning performance of children interacting with a cartoon-like character to learn about marine biology. The *Colorado Literacy Tutor* (Cole et al., 2005) is a set of software tools that aims to improve literacy and student

achievement in any subject area by helping pupils learn to read fluently, to acquire new knowledge through understanding of texts, and to appropriately express their ideas in writing. (Massaro et al., 2006) developed a speech and language tutor centered about a talking head as conversational agent for children with language challenges. Synthetic characters have also increasingly been used in storytelling and tutoring applications for children (Ryokay & Cassell, 1999; Robertson & Oberlander, 2002; Vaucelle, 2002).
Both the immediacy of the interaction with interactive characters and the encapsulation of people into a gaming environment add a natural and entertaining experience to the user and can be geared toward a specific learning objective in a way that is consistent with the constructivist theory of education. Users can perform complex activities such as driving a virtual vehicle or navigating through a 3D photorealistic artificial world populated by autonomous characters that can interact and engage in social interaction with human users and/or other in-world avatars according to patterns governed by artificial intelligence programs designed to achieve specific learning objectives.
A series of cognitive processes like e.g. active discovery, analysis, problem-solving, memorization, conversation, visual and emotional stimulation, interpretation and/or physical activity that occurs while using these game-like interfaces deeply contributes in rooting learning in internal brain circuits and ultimately supports the learning process. The high degree of interactivity results in users actively engaged in communication with the virtual world and its inhabitants and is seen as an important factor for effective learning (Stoney & Wild, 1998). Moreover, besides facilitating learning, most of these interfaces are also designed to participate towards the educational goal in a cooperative manner so as to reflect the observation that children collaborate with peers naturally and often rely on each others support during learning processes. Game-based learning has been used quite a lot in adult learning programs too. Business strategy games have been used for many years in the management and financial areas[1] (Prensky, 2000) as well as more recently to introduce computer science programming assignments (Giguette, 2003).
The benefits of using graphical characters, by contrast to plain learning applications, lie in the distinctive use of (sometimes stylized) face and gestures to reflect interpersonal attitudes, deliver communicative content, and provide feedback to which users naturally pay a great deal of attention (Knapp, 1978; Fabri et al., 2002).
Digital learning environments such as computer games, simulations, and embodied conversational characters have all the potential to provide a cognitive bridge between actual experiences and abstractions which is crucial for teaching children to deal with complex problem solving and comprehension issues. The big challenge in educational software for children is to understand how to utilize the available technology in order to engage them directly in collaborative interactions in a way as to benefit their cognitive development.

3. Game-like interface for children edutainment

3.1 System overview

Our current real-time game scenario consists of a player interacting with a full-body single embodied character (Figure 1, left) impersonated by the fairy-tale author Hans Christian Andersen (HCA). Interaction takes place in an entertaining and educational manner within

[1] See for instance www.learningware.com, www.games2train.com, www.socialimpactgames.com, or www.corporatelearningforum.com

a 3D graphical world via spoken dialogue as well as pen gestures. Several other characters can be added to the system. However, since we did not create a large enough knowledge base large enough for all characters, they would currently interact the same way as the HCA virtual character does. Typical input gestures are ink markers like lines, points, circles, etc. entered at will via a mouse-compatible input device or using a touch-sensitive screen.

Fig. 1. (*left*) HCA full-body conversational character in his study; (*right*) Cloddy Hans is one of HCA's fairy tale characters that can be encountered in the fairy world.

Some objects in the author's study have been designed to resemble events experienced by the character and/or works that he created during his real life. For instance, a picture of the Colosseum in Rome hanging over his desk serves as a visual link to his visit to Italy and more specifically the Italian capital city. Similarly, books are stored in shelves while a small set of the writer's personal objects of the writer, such as his umbrella and walkstick, are placed at different locations within the study. These objects had a central role in the writer's life and thus offer a topic of conversation to the user and form the basis for multimodal interaction with the character. Object behaviors are used as visual feedback in deictic utterances as well as for their selection and manipulation. Apart from them, the system offers other domains of discourse including: the writer's fairy tales, his life, his physical presence in the study, and the domain of solving meta-communication problems that occurs during speech/gesture interaction. In order to reinforce the learning experience and make the interaction even more entertaining, in a companion system (Boye & Gustafson, 2005), the user is also granted access to a 3D fairy tale world populated by HCA's fairy tale characters (Figure 1, right). The user can wander about, manipulate objects and collect information useful to solve tasks, which arise while exploring the fairy world, such as e.g. passing a bridge guarded by a witch. For the user to have the impression that she is interacting with distinct, believable agents, each virtual character has its own proper appearance, voice, actions, and personality.

Users perceive the world around them through a first-person perspective. They can explore HCA's study, talk to him, in any order, about any topic within HCA's knowledge domains, using spontaneous speech and mixed-initiative dialogue, change the camera view, refer to and talk about objects in the environment, and also point at or gesture to them. HCA reacts emotionally to the user input by displaying emotions and by employing a meaningful combination of synchronized verbal and non-verbal behaviors. He can get angry or sad because of what the user says, or he gets happy if the user, for instance, likes to talk about his fairy tales.

3.2 Agent architecture

A software system that is supposed to behave in a human-like manner needs to be able to perform a large set of tasks, both externally (talking, gesturing, moving about etc.) and internally (interpreting sensory data, evaluating user's input, monitoring plan execution, etc.). The flexible and responsive nature of multi-agent architectures in which agents communicate, cooperate, coordinate and negotiate to meet particular goals under specified timing constraints naturally lends itself to such an application.

The theory and development of software agents has been an active field of research for a few decades. Several working definitions have been proposed and eventually a consensus was reached in (Jennings at al., 1998) where an agent is deemed as a computer system operating in a certain environment and capable of flexible autonomous actions towards its design objectives. Several models for agent communication have been put forward, the agent broker and agent to agent being the two most representative (Cheyer & Martin, 2001; DiPippo et al., 1999). The agent to agent model is a completely distributed framework where each agent knows the name of any other agents with which it might need to communicate. In the agent broker model instead, a special agent is tasked with finding agents to fulfill services required by other agents requesting that specific services. To that extent, this model relies on a central facilitator, the broker agent that administers communication among agents. Those, in turn need to register with the facilitator in order to advertise the services they offer.

The widely used Open Agent Architecture (Martin et al., 1999) relies on this latter model and is also the inspiring model of the architecture of our choice. We have been using the agent architecture developed by colleagues in our companion project[2]. It is simple to use, easy to implement, and lightweight. For platform independence and to facilitate debugging, agent communication occurs with text only over standard TCP/IP. A central facilitator routes messages among registered agents. It also knows which servers are deployed and how to start them to allow automatic restart in case of unexpected server crash. Agent to agent communication to bypass the broker is also possible and is even enforced whenever the data exchanged is binary given that the facilitator can deal with text only messages.

As a whole, the HCA system is realized as an event driven, modular, asynchronous multi-agent architecture. Several single agents take care of different aspects of the interaction with the user: a speech recognizer senses the user input, a gesture recognition agent interprets ink entered by users, the input fusion agent ensures modality fusion, a response generation module deals with speech synthesis and graphical animations and a dialogue manager (DM) manages the conversation with children as it evolves. Resorting to an agent architecture allows the different developers involved to focus on a specific well-defined functionality. In this way, the architecture makes it possible to create a bigger application from a set of agents that were not necessarily designed to work together. This also facilitates a wider reuse of the expertise embodied by each single agent, their maintenance and debugging.

In our system, the broker coordinates input and output events by time-stamping all module messages and associating them to a certain conversation turn. The behavior of the broker is controlled by message-passing rules, specifying how to react when receiving a message of a certain type from one of the modules. Despite a facilitator-centered configuration, the

[2] www.speech.kth.se/broker/

information flow typically occurs in a pipeline-like manner. As depicted in Figure 2, any time an input is sensed, the n-best hypothesis lists from either the speech recognizer or the gesture recognizer or both are sent to the natural language understanding (NLU) module and the gesture interpreter, respectively. The gesture interpreter consults the animation module to figure out which on-screen objects the user has referred to while gesturing. Output from those two agents is then forwarded to the gesture/speech input fusion module which, in turn, provides input for the dialogue manager (DM) which is responsible for the management of the interaction with the user. It has, among others, to plan for the next response to the user, to update the characters' emotional state, and to keep track of the dialogue history. Eventually, the response generator, informed by the DM, coordinates a text-to-speech message to play back synchronized with the rendering of the corresponding character animation.

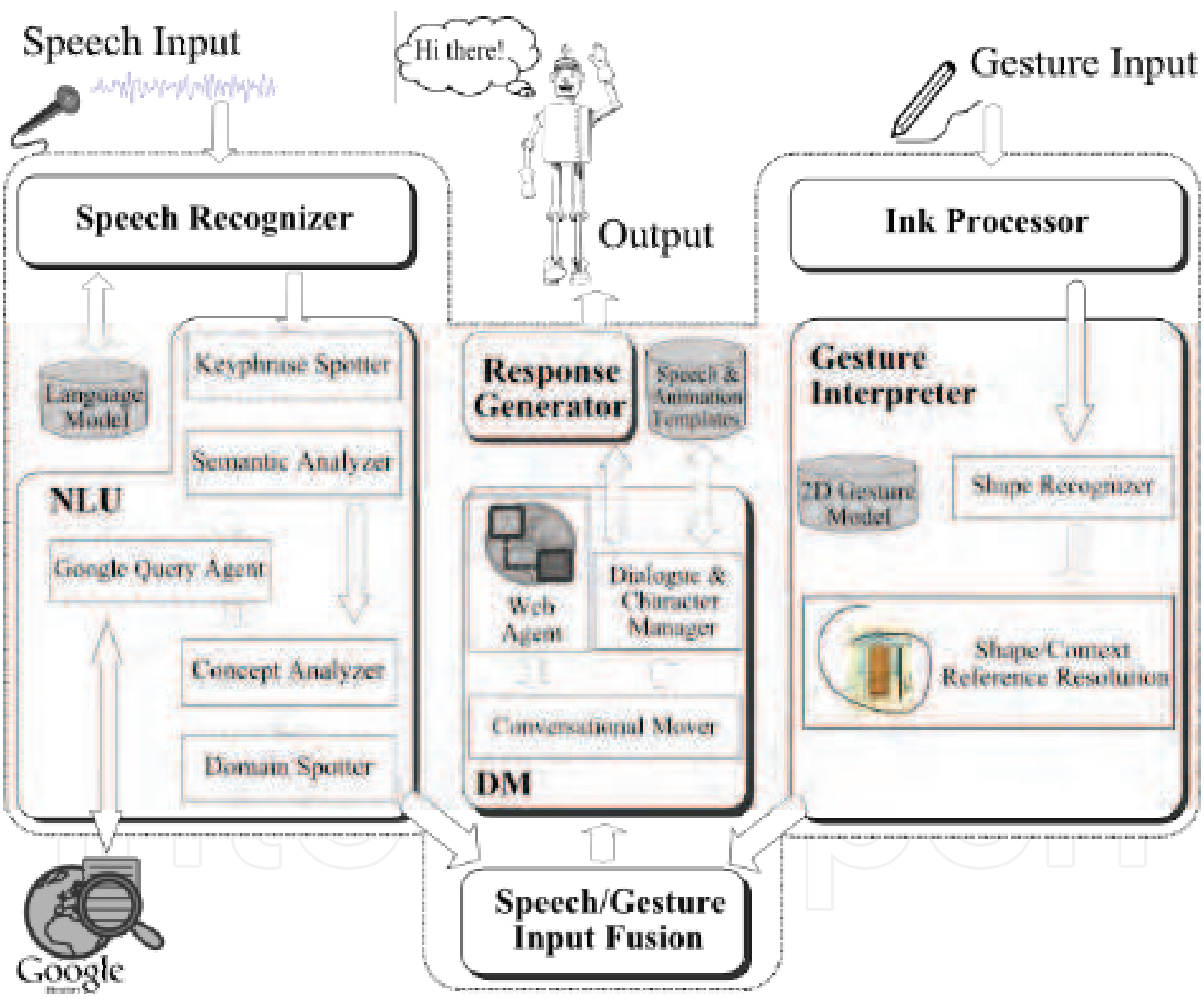


Fig. 2. Detailed view of the whole system architecture and information processing flow.

An ontology is used as common knowledge representation formalism shared among the system modules to create a domain independent architecture. In this way, moving to another character only requires a modification of the ontology-based knowledge representation. We described the input fusion, response generator and dialogue manager modules in details in (Corradini et al., 2003; Corradini et al., 2005a; Corradini et al., 2005b).

The next subsections focus on and address some issues encountered while dealing with the speech modalities of users, and notably children, during interaction with the system.

3.3 Children spoken language recognition: issues

Despite the growing number of kids accessing speech operated applications, spoken dialogue systems developed so far have an inherent problem that directly transfers in the development of our conversational prototype: they have been mainly designed for adult users. While the state of the art in automatic speech recognition and synthesis is still not completely satisfactory for the adult population, the endeavor of enabling speech technologies for children represents even a greater research challenge. In fact, past investigations have shown that children's voices are more variable in terms of acoustic characteristics and prosodic features, are more disfluent when compared to adult speech (Darves & Oviatt, 2002; Oviatt et al., 2004) and change developmentally (Yeni-Komshian et al., 1980; Oviatt & Adams, 2000). Shy and introvert children can be hard to engage in interaction with a conversational character and are reluctant to speak or they speak low in volume if at all (Darves & Oviatt, 2004). A study on a reading tutor for preschool children showed that off-the-shelves speech recognizers perform poorly unless a new acoustic model created from the speech of children in the target age range is employed. By explicitly accounting for common mispronunciations, speech recognition rose to an astounding 95% rate (Nix et al., 1998). Research also indicated that young people tend to employ partly different strategies when interacting with dialogue systems than adults do (Coulston et al., 2002; Oviatt et al., 2004). For instance, younger children use less overt politeness markers and verbalize their frustration more than older children (Bell & Gustafson, 2003). Moreover, children seem to adapt their response latencies and the amplitude of their speech signal to that of their conversational partners. Differently from adults, children do not often modify the lexicon and syntax of an utterance (Bell & Gustafson, 2003). Moreover, in case of communication problems while interacting with conversational agents, research indicates that kids tend to repeat critical original utterances verbatim with just a few modifications of certain phonetic cues, notably by increasing the tone and volume of their voice (Bell & Gustafson, 2003).

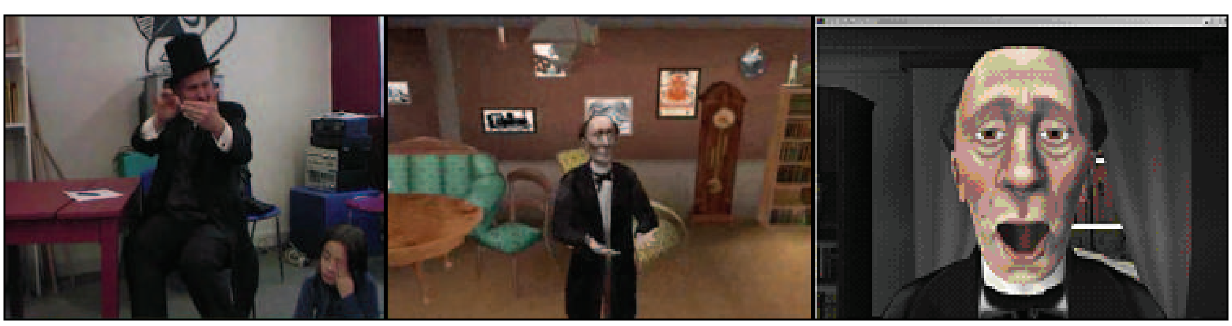


Fig. 3. (*left*) A human actor impersonating HCA interacting with school pupils in the writer's native town of Odense; (*middle*) snapshots of an animation; (*right*) face expressing surprise.

These research findings motivated us to collect a corpus of children conversational data. In fact, the few existing corpora of children speech turned out to be not usable in our system for none of them was in Danish and moreover consisted of either prompted speech or monologues of children recounting stories (D'Arcy et al., 2004; Eskenazi, 1996; Gerosa & Giuliani, 2004; Hagen et al., 1996).

We transcribed and analyzed several hours of collected video and audio-taped conversation of young subjects involved in a series of interactive sessions in both *Wizard of Oz* studies and in an after-school class where they played with a real human actor impersonating Hans Christian Andersen (Figure 3, left). The video data was partly used to generate the graphical animations (Figure 3, middle and right). The audio data from these interactive sessions was instead used to create two corpora of children-computer spoken conversation containing spontaneous dialogue data in English and in Danish, respectively. A similar task was also carried out by our project partners for the Swedish language (Bell et al., 2005). The corpora were then used for the creation and training of dedicated acoustic models for the speech recognizer. The deployment of such acoustic models from the speech of children in the target age range of our system immediately boosted the recognition rate of our speech recognizer and confirms the experimental results reported in (Nix et al., 1998).

3.4 Children conversation with the virtual character

Beside differences in the speech signal, there are additional distinctions between adults and children that directly influence and make the development of automatic spoken systems for children difficult. Their behavioral patterns of interaction with a computer are different from those of adults because they are still learning linguistic rules of social communications and conversation. Moreover, there are significant differences in those patterns even among children according to their age range, gender, and the socio-economic and ethnic backgrounds. Children's behavioral patterns are quite different from those of adults in terms of attention and concentration as well. Preschoolers are generally able to perform an assigned task for not longer than about half an hour (Bruckman & Bandlow, 2002). In (Halgren at al., 1995) it was found that children tend to click on visible feature just to see what happens as reaction to their actions. If an action gave rise to some feedback event that they judged interesting (like a nice sound or an animation), many kids kept on clicking to experience the feedback over and over again. In a similar work, (Hanna et al., 1997) discovered that if a funny noise was used as an error message several children repeatedly generated the error just to hear it again.

There are still many additional general issues of technical nature that need to be addressed and solved before computer interfaces can properly become conversational and multimodal. Question-answering systems, command and control dialogues, task-oriented dialogues and frame-based dialogues (Allen et al., 2001; Rudnicky et al., 1999; Zue et al., 2000) are subclasses of practical natural dialogue for which very robust and successful language processing methods have been already proposed. Their main limitation - its fixed context - is simultaneously its greatest strength since it allows building very robust and feasible spoken dialogue systems. However, they are a simplification of real human conversational behavior for they control and restrict the interaction rather than enrich it. By contrast to task-oriented and information spoken dialogue system, we propose a domain-oriented conversation that has no task constraints and can be enriched by either accompanying or complementary pen-gestures. The user is free to address, in any order, any topic within HCA's knowledge domains, using spontaneous speech and mixed-initiative dialogue, and pen markers to provide context to the interaction.

We dedicated a great deal of attention in defining proper design strategies that motivate children, keep them engaged for a certain period of time, and make them produce audible speech that can be reasonably processed by a speech recognizer. To reflect the finding that they tend to use a limited vocabulary and often repeat utterances verbatim, we created a database of possible replies for our back-end that lexically and grammatically mirror the expected input utterance. In other words, we decided that the parser for the user input utterances should also be capable of parsing output sentences i.e. the sentences produced by the conversational agent. Moreover, we never aimed at nor did we need a parser capable of full linguistic analyses of the input sentences. The analysis of data collected in *Wizard of Oz* studies and other interactive adult-children interactive sessions showed that most information could be extracted by fairly simple patterns designed for a specific domain and some artificial intelligence to account for the context at hand.

The key idea underlying our semantic analysis is the principle of compositionality for which we compose the meaning of an input sentence from both the meanings of its small parts and based on the relationships among these parts. The relatively limited grammatical variability in children' language and their attitude of repeating (part of) sentences, made it possible for us to build a very robust language processing systems based on patterns and finite state automata designed for each specific domain. This strategy proved sufficient for the understanding of most practical children spontaneous dialogues with our system and empirically confirms both the practical dialogue hypothesis for which '*..the conversational competence required for practical dialogues, while still complex, is significantly simpler to achieve than general human conversational competence..*' (Allen at al., 2000) as well as the domain-independence hypothesis which postulates that practical dialogues in different domains share the same underlying structures (Allen at al., 2000).

Technically, the NLU module consists of four main components: a key phrase spotter, a semantic analyzer, a concept finder, and a domain spotter. Any user utterance from the speech recognizer is forwarded to the NLU where a key phrase spotter detects multi word expressions from a stored set of words labeled with semantic and syntactic tags. This first stage of processing usually is helpful to adjust minor errors due to misrecognized utterances by the speech recognizer. Key phrases are extracted, and a wider acceptance of utterances is achieved. The processed utterance is sent on to the semantic analyzer. Here, dates, age, and numerals in the user utterance are detected while both the syntactic and semantic categories for single words are retrieved from a lexicon.

In fact, relying upon these semantic and syntactic categories, grammar rules are then applied to the utterance to help in performing word sense disambiguation and to create a sequence of semantic and syntactic categories. This higher-level representation of the input is then fed into a set of finite state automata, each associated to a predefined semantic equivalent according to data used to train the automata. Anytime a sequence is able to traverse a given automaton, its associated semantic equivalent is the semantic representation corresponding to the input sentence. At the same time, the NLU calculates a representation of the user utterance in terms of dialog acts. At the next stage, the concept finder relates the representation of the user input, in terms of semantic categories, to the domain level ontological representation. Once semantic categories are mapped onto domain level concepts and properties, the relevant domain of the user utterance is extracted. The domain helps in providing a categorization of the character's knowledge set. The final output in form of concept(s)/subconcept(s) pairs, property pairs, dialog act and domain is sent on to other system components that deal with the current dialogue modeling. More details about the processing steps of this module along with few explanatory examples can be found in (Mehta & Corradini, 2006).

On the one hand, the proposed NLU is not capable of capturing fine distinctions and subtleties of language since it cannot produce a detailed semantic representation of the input utterance. One the other hand, it is not possible to create a system grammar that covers all possible variations and ambiguities of the natural language used by children in our data set. Altogether, as evinced in the system evaluation (see section 4), our shallow parsing approach which employs the use of semantic restrictions in the grammar (captured by a series of rules) to enforce semantic and syntactic constraints has proved a feasible and robust trade-off approach.

3.5 Out of domain conversation

During a set of usability test sessions, we realized that children frequently ask out-of-domain questions that are usually driven by external events or characters which are popular at the time of the interaction. For instance, in early sessions children frequently asked about the *Lord of the Rings* while this subject was completely ignored in later studies where e.g. *Harry Potter* was a much more common topic of discussion (Bernsen et al., 2004).

We were thus confronted with the difficult and ambitious objective of developing conversational agents capable of addressing everyday general purpose topics. In fact, we cannot expect conversational characters to conduct a simulated conversation with children that exclusively revolves around the agent's domains of expertise. Such a situation, coupled with the decreasing capability of children to focus on a specific subject for prolonged period of times (Bruckman & Bandlow, 2002), would make any interface pretty boring and ultimately conflict with the educational objectives.

The synthetic character should be endowed with the capability of reaching out into topics that could not be covered by the developers during the creation of the system. Previous systems have typically used simplistic approaches of either ignoring or explicitly expressing inability to address out of domain inputs. We could avoid in advance or limit situations where children ask questions related to an unconstrained range of utterances by keeping the conversational flow on a specific, well defined (from the system's perspective) track and leave room for as less opportunities as possible for the human interlocutor to take the initiative (Mori et. al., 2003). However, maintaining full control of the interactive session is a strategy that conflicts with the mixed initiative nature of our system. Another approach is to engage users in small talk when they go out of topics (Bickmore & Cassell, 1999) yet the range of discussion topics is still limited since it is dependent on the amount of templates that can be created off-line. We wanted to reduce the authorial burden of content creation for different general purpose discussion topics. In (Patel et. al., 2006) an approach to handle out of domain input through a set of answers that explicitly state that the character either doesn't know or doesn't want to reveal the answer is presented. This approach is in general better than saying something completely absurd, however this strategy is more suitable for training simulations where the goal of the system is to keep the conversation on track so as to achieve the training goal. For our domain where the goal of our agent is to provide an appropriate educational reply along with a rich social experience to kids, that strategy does not work either. *Façade* (Mateas & Stern, 2004), an interactive drama domain, uses various deflection strategies to bring back the discussion onto the main conversation as well as to limit the depth in which players can drill down on any one topic. These strategies present an interesting solution to avoid out of domains input for a story based domain. An ongoing story provides the user with enough narrative cues to integrate the deflection output used by characters into the ongoing narrative flow. Differently from this latter work, in our approach, we wanted to address the general purpose topics apart from the domain topics rather than deflecting them to bring the conversation back onto the domain topics.

As we have seen in the previous section, in our implemented system the NLU module has generic rules for detecting dialog acts present in the user utterance. These dialog acts provide a representation of user intent like types of question asked (e.g., asking about a particular place or a particular reason), expression of opinion (like positive, negative or generic comments), greetings (opening, closing) and repairs (clarification, corrections, repeats). These dialog acts are reused across different domains of conversation. Moreover, generic rules are used to detect the domain independent properties (e.g., dislike, like, praise, read, write etc). The NLU categorizes the word(s) that are not processed internally into an unknown category. The longest unknown sequence of words is combined into a single phrase. These words are then sent to a web agent that uses Google's directory structure to find out whether the unknown words refer to a name of a movie, game, or a famous personality and the corresponding category is returned to the NLU. The web agent eventually finds a quick and concise output using three freely available open-domain Question-Answering systems: AnswerBus (Zheng, 2002), Start (Katz, 1997), and AskJeeves[3] or the web page at specific game and movies websites[4]. The web agent employs a set of heuristics, such as removing output with certain stop words, to pick one single reply. Once a sentence is selected, we remove control/graphical characters to get a plain string that can be played by the TTS component. We also make a first attempt at categorizing the retrieved information in order to generate appropriate non-verbal behaviors synced up with spoken utterances (Mehta & Corradini 2008).

4. System evaluation

4.1 Are animated characters effective?

To date there is no clear answer to this question. The evaluation of the effectiveness of including conversational animated characters in user interfaces is a complex and arguable task. In (Dehn & van Mulken, 2000) a review of several interfaces with synthetic agents seems to indicate that there is little or no improvement in user performance. Nonetheless, the authors of that review also suggest to take this conclusion very carefully on the ground that the systems analyzed could not be compared consistently due to the different evaluation methods employed.

Despite ambiguous or inconclusive results and the lack of experimental evidence, we argue that animated agents enhance the user experience first and foremost because they allow for a simulated face-to-face communication that is the most effective mean of communication as well as method of instruction among humans. Moreover, animated agents have the potential of increasing user motivation, stimulating learning activities, enhancing the flow of information, and fulfilling the need for personal relationship in learning (Gulz, 2004).

It is however extremely difficult to assess pedagogical benefits of character enhancement and then to generalize the results. As noted in (Cole at al., 2004) the ideal evaluation of computerized learning environments would consist of repeated interaction with the

[3] www.askjeeves.com
[4] www.game-revolution.com and www.rottentomatoes.com

animated agents over long periods of time to validate the observations on the basis of factors such as e.g. the nature of the task, the personal characteristics of the users, and the believability of the graphical agent.

4.2 Setting the stage

We ran many pilot studies involving children in the attempt to discover the main factors that contribute in creating better computer games with an educational objective in the foreground. How computers are able (or perceived) to play, the degree of challenges, entertainment and interaction they offer, the amount of new knowledge assimilated, and the believability of the game characters, seem to be important factors.

We report here on a study with thirteen young subjects evenly split between males and females (6 and 7 subjects, respectively) recruited in local schools in the city of Odense in Denmark. Each user session had a duration of approximately 50-60 minutes including an exploratory phase with the interface and a post-session informal discussion with each participant. The average age was 13.1 years (12.8 for males and 13.3 for females).

All pupils were Danish native speakers with advanced skills in speaking English. Fifty-three percent of them (100 percent of males and 43 percent of females) declared themselves as being a frequent (i.e. more than 1 hour/week) videogame and/or console player, with a peak of 45 hours/week spent in gaming by a male teenager. 38.5 percent of the participants (28.6 percent of females and 50 percent of males) had been exposed before to computing systems able to process speech and/or gesture; all of them were acquainted with an earlier version of our system. When asked about their favorite games, children said that they like to play with games of any genre, ranging from shoot-'em-up (66.6 percent of males and 0 percent of females), action, platform, to sports and strategy games (40 percent of males and 50 percent of females). With regard to pre-interaction knowledge about the writer, his life, his fairy tales and the historical period he lived in, 53.8 percent of the children (42.8 percent of females and 66.7 percent of males) declared to have a fair to very good knowledge of these historical and literacy facts and events. Despite surprising at first, this high level of knowledge is due to the fact that Odense is Hans Christian Andersen's hometown. In local schools he is often subject of discussion and several cultural events organized by the Odense municipality are often related to its world-renowned citizen.

Fig. 4. (*left*) A child interacting with the system; (*right*) hand gesturing on a touch sensitive screen to operate a virtual object within HCA's study.

To be able to play with the system, each subject had to wear a microphone headset to enter spoken utterances. They could choose among a touch screen, a mouse and a keyboard for entering ink gesture markers. Initially, the participant was given a 15 minutes session to get accustomed with the system. During this time an assistant was present to help out in case of questions about system functioning. At the end of the introductory session, after a short break, each subject was given a set of tasks to carry out during an additional interaction session lasting for approximately 20 minutes without any external human assistant support (Figure 4). We video and audio taped each session while system events were all automatically logged into XML files for further dialogue analysis. Players were allowed to break up the game at any time for any reason. At the end of the interaction each participant was interviewed according to a set of predefined questions. Informal discussions also typically occurred. Eventually each child was handed out (without being told in advance) a theater ticket as a reward for the time spent in the interaction. The questions were used to survey four main aspects, namely user's gaming habits, system interaction capabilities, system's educational and entertainment values, as well as open-ended questions for the subject to provide us with valuable insights and suggestions for creating a better system.

4.3 Results from the interviews

Two persons independently evaluated the questionnaires. User interviews were transcribed and mapped onto numerical values on a Likert scale from 1 to 5. For instance, when looking at the subjective entertainment degree experienced by the user, we mapped sentences such as e.g. '*I had no fun at all*' and '*the interaction was very entertaining, amazing!*' to 1 and 5, respectively. Inter-rater reliability for second scoring of the questionnaire data was 94%. Data analysis over the single categories revealed numerical value distributions of sufficient regular shape.

Thus, despite the limited sample size, the obtained results can be shown in terms of statistical measures like the mean and the standard deviation. These values for a few categories, each characterized by a reasonably symmetric distribution of and no outliers among its numerical values, are:

Interface easy of use (difficult = 1, very easy = 5):	mean = 3.9	stdev = 0.28
Graphics and quality of animations (bad = 1, great = 5):	mean = 3.38	stdev = 0.75
Agent's understanding skills (very poor = 1, great = 5):	mean = 3	stdev = 0.57
Entertaining degree (not at all = 1, very exciting =5):	mean = 3.77	stdev = 0.44
Degree of learning (none = 1, much = 5):	mean = 3.08	stdev = 0.64
Use of gestural input (superfluous = 1, very useful = 5)	mean = 3.98	stdev = 0.22
System's overall rate (very bad = 1, great = 5):	mean = 3.62	stdev = 0.87

In other words, the system was overall rated fairly well. It was perceived as exiting and funny, with a reasonable degree of added educational value. With regard to the educational content, most of the users did not indicate what exactly they have learnt, yet when they did they mostly referred to the writer's life and family while stating that they already knew a great deal about his fairytales and therefore there was nothing new to learn about this topic. In the light of that, more specific questions on what aspects of the writer subjects have learnt about while playing should be considered in future studies.

The interaction with the character is driven primarily by the speech modality however a small set of pen gestures is available to operate on objects in the room as well. Interestingly,

53.8% of the subjects (50% males and 57.1% females) stated that they liked the gesture modality and/or wanted to do more with it. Despite gestures were not used extensively by the subjects, we hypothesize that they ease shy users into the conversation (shy users generally start with clicking on a picture and then just wait for something to happen; rarely they ask about it). Gestures may help breaking the initial hesitance on the part of the user and help to establish a relationship with the interactive character, which forms the basis of a smooth overall conversation.
From a dialogue management point of view we were interested in evaluating aspects like conversation success, domain coverage, robustness, etc. Table 1 depicts the average number of turns over each domain as well as their percentage of domain coverage during interaction sessions analyzed for the usability study.

Domain Name	Average # of Turns	Percentage
Fairy Tales	8.2	9.6
Life	6.9	8.1
Physical Presence	4.8	5.7
Study	13.3	15.6
User	7.7	9.0
Generic	44.2	51.9
All Domains	**14.2**	**100**

Table 1. Domain coverage.

The study domain relates to information about the objects in HCA's study, so every time the user points at something in the study the study domain is triggered. This can be a good indicator of the multi-modality input behavior of the users. The generic domain is the one most addressed by kids confirming empirical evidence regarding their attention and concentration difficulties (Bruckman & Bandlow 2002) and ultimately pointing out the need of a reliable mechanism that makes out-of-domain conversation possible. The generic domain contains also meta-communication turns which were triggered e.g. anytime a low confidence score occurred in the speech or gesture recognizer or the NLU. In a study with 186 input sentences we analyzed our approach in dealing with out-of-domain questions. The results are depicted in Table 2.

Question Type	*Coverage*	*Correct Answer*	*Some Answer*	*Wrong Answer*	*No Answer*
When	8 (4.3%)	2	1	2	3
Who	36 (19.4%)	14	8	3	11
What	65 (34.9%)	51	1	3	10
Keyword(s) list	51 (27.4%)	15	2	2	32
Other	26 (14.0%)	8	2	2	14
Total	**186 (100%)**	**90 (36.4%)**	**14 (7.5%)**	**12 (6.5%)**	**70 (37.6%)**

Table 2. Results of handling out-of-domain questions.

The evaluation study provides also empirical evidence that the interaction with the character is driven primarily by the speech modality despite the availability of 2D pen gestures to operate on objects and entities in the three dimensional virtual room. On average, about 6.3% of the actual turns were gesture only, 80.7% speech only and 13% displayed a multimodal speech-gesture content. Despite this latter figure may seem to be low at a first glance, it should be noted that not all turns necessarily required gestural input (e.g. when the user asked about the age, name, etc.). By comparing the set of potential multimodal situations occurring during the interactions as identified by the human transcribers with the set of the actual multimodal situations (i.e. these covering 13% of the user study interaction), we had an astonishing 96.4% overlap. These correct multimodal turns typically occurred anytime speech was accompanied by deictic words to refer to objects or entities in the virtual world. The 3.6% agreement discordance between the actual multimodal situations and the ideal case, was mostly due to anaphoric expressions used to refer to entities in the game that were talked to in the previous turn(s). In other words, while children considered speech as the main communicative modality, the study provides empirical evidence of a balanced use of modalities and a preference of gesture for manipulable objects and entities.

Interestingly, 53.8% of the subjects (50% males and 57.1% females) stated they liked the gesture modality and/or wanted to do more with it. Despite gestures were not extensively used by the subjects, we hypothesize that they ease shy users into the conversation. Shy users were indeed also those who displayed most interaction patters like scribbling or random on-screen clicking just to see if they get any feedback. In those situations, despite there is no clear cut to define when a user turn starts and a computer turn ends, we recognized some 6.3% of the total turns as being characterized by being gesture only patterns. This behavior pattern is common among users, and thus we believe that gestures may help breaking the initial hesitance on the part of the user and help form a relationship with the interactive character, which forms the basis of a smooth overall conversation.

We haven't performed any data correlation analysis because of the limited number of subjects and thus the lack of a large set of data. A very preliminary examination about the correlation between entertainment and favorite game genre and gameplay expertise, respectively, proved itself inconclusive.

4.4 Comments from the subjects

In this subsection, we report a set of quotes from children subjects together with the results of the user studies that highlight the interface aspects in relation to the pedagogical goals, the way of interaction, the graphical design, as well as desirable improvements of our game-like interface.

Educational Content

Children highlighted that their interaction with the agent either extended ".. *information about his life is more fun than about his fairy tales. The user knows his fairy tales but not his life* .." or brushed up ".. *I haven't really learnt anything than I didn't already know but it helped me recall a number of things* .." their knowledge about HCA.

We actually did not expect much of an increase in knowledge about HCA because this is also a subject that they learn in great detail and in different courses at school and with after-school activities offered by the Odense municipality. Boys expressed twice time more than girls that they had increased their set of knowledge after the interactive session. Girls were more likely to highlight already existing knowledge on the subject.

Nonetheless, the user study indicates that children believe that they increased their knowledge after playing with our system. This ultimately tells us that we are on the right track to achieve the educational objectives that we envisioned at early stages of development and supports the belief that animated agents in virtual environments provides an interactive experience that helps children learning.

Interaction with the characters

Altogether children enjoyed the interaction with the character which´they thought ".. *it was really cool ..*" and was "*.. good enough but he* (HCA) *is not the most polite person around ..*" Some reports point out a few cases where the character did not act upon the wish and expectation of the player. They felt that their interaction was "*.. frustrating when he did not answer my questions ..*" and HCA "*.. didn't understand everything. One has – by trial and error – to find a formulation which he can understand to bring the system is on the right track ..*" We further examined also the goodness of the technical system in term of reliability and accuracy of all its single components with particular emphasis on the speech processing and gesture processing modules. This technical evaluation revealed expected shortcomings on the side of the speech recognizer (Mehta & Corradini 2006) which are however out of our control. The lack of barge-in capabilities in our current prototypes was highlighted in a few comments such as in "*.. it would be good if HCA stopped talking when asked ..*"

Graphics and Character Believability

Children appreciated the life-like animations and graphical appearance of the character judging that "*.. The good graphics also makes it* (the system) *entertaining ..*" However, the repertoire of HCA's actions was sometimes perceived as rather limited. A few subjects felt that "*.. maybe he should also be able to do more things such as smoke his pipe ..*"

Suggestions for Improvements

Children found the overall system interesting and useful, "*.. it is different, more lively, to be told (about HCA life, fairy tale, family etc.) rather than just to read about it all ..* " and "*.. it was entertaining to hear what he told ..*" Most children expressed the wish that they would definitely prefer to use the software compared to a classroom session. They thought that "*.. it was more fun than learning the same at school ..*" Interestingly, the system has potential also in teaching new words or expressions to children interested in learning a foreign language as stated by "*.. his vocabulary is fine; I learnt some new words in English ..*"

Subjects were asked what features of the current prototype needed to be improved. Excerpts from children quotes on this issue highlighted the current limitations of the game in terms e.g. of "*.. missing actions. Maybe there are not so many 12 year old kids who are interested in HCA. It is not really what you would really like to go home and play with. Maybe better suited for smaller children ..*" as well as regarding the lack of a clear underlying storyline as highlighted in "*.. HCA's life story should be told up-front. It helps create a context and makes easier to understand the pictures ..*"

One child reported that "*.. users should be allowed to visit other parts of his house ..*" and brought up the issue of having a small number of places currently available for the user to explore and experiment with. As a consequence not every youngster was keen to play with our system on a daily basis. As a boy participant put it: "*.. I would not spend hours on such game every day. There are not so many challenges ..*"

We need to address the wish expressed by a couple of pupils to "*.. add more new things one can point to and get a story about. There could also be stories spanning two pictures where the view angle changes automatically when HCA starts talking about the second picture ..*" Comments like "*.. it would be desirable to have more things to point to with creative stories attached to which could even be a bit surprising ..*" seem to indicate the wish for more manipulable objects. At the same time, however, other partecipants were pretty happy about the current amount and

behavior of the existing ones as it can be inferred from the comment "*.. the use of pictures one can point to is creative, it would have been boring with a book to browse instead ..*"
The addition of more sound or music to make the interface more funny and attractive was also suggested through the opinion expressed as "*.. it could be funny to have music played in certain situations for instance when you click on a picture or HCA crashes against the wall..*"

5. Discussions and conclusions

Play is more than just entertainment for children. It is a fundamental activity that supports them in developing communication skills, managing feelings and emotions, learning the foundations of social rules, and abstracting concepts. The efforts to exploit the motivation and engagement that computer games naturally offer have recently given rise to a tremendous interest in the use of game-like applications for training and learning. Such kind of applications shifts the player into the participant role and acts as a catalyzer that turns an interactive session into a learning-by-doing experience. Differently from the traditional teacher-based learning paradigm, such a constructivist approach places the learner at the center of the learning process.
It is also indisputable that computers are compelling for children and adolescents. By giving them the control on the pace and kind of actions, they can repeat any activity as often as they like and experiment with variations. Hence, appropriate software can engage children in creative play, problem solving, and conversation with positive effects on their cognitive and social learning and development (Clements 1994; Haugland & Shade 1994).
Technology for children broadly falls into two categories: educational products and digital entertainment. Edutainment is what results in blending these two genres and it is also the framework of the system we have developed. We have created an aesthetically elegant, entertaining and intellectually challenging interactive architecture for young people of age ranging from 10 to 18 years to play and interact with a synthetic conversational agent impersonating the Danish historical luminary Hans Christian Andersen.
The conceptual goal of the project was to allow children and teenagers to collect information representing an organic history and a coherent body of knowledge through conversation and narrative in a funny way. The underlying idea was that a combination of an educational informative system and a gaming environment into a single application offers new opportunities towards more effective and rewarding learning experience. Technically, the task of building game-like interfaces populated by conversational characters represents a tremendous challenge for the research community and involves several large research questions: how to deal with children spoken language, how to deliver the appropriate behavior and information over different modalities in an interesting and engaging manner in every given dialogue situation, how to present a wide spectrum and depth content structure, how to keep up with a dialogue over virtually any topic without interrupting the flow of conversation in case of misunderstanding or out of domain topics, and many more. At the same time, we had to face usuability issues related to the target users of the system. For instance, we had to account for the importance of emotions in a learning process. Depressed or anxious children cannot assimilate new knowledge and learn as effectively. Therefore the assessment and/or display of emotions play an important role and help in improving the effectiveness of computer-based learning environments populated with

virtual agents. We have learnt that anthropomorphism in the interface is not a benefit in itself unless it is coupled with proper expressive, interactive and conversational capabilities; a finding confirmed also by previous research (Cassell at al., 2000). Our research seems to confirm the ability of animated characters to engage and motivate children especially when they can communicate with a system capable of displaying emotions and personality and when they can choose from several input modalities (Narayanan & Potamianos, 2002).

The set of user studies that we ran with school pupils showed a high degree of satisfaction in our system in terms of graphical appearance, ease of use, interaction modalities as well as pedagogical goals. While the world of computer games mirrors the dominance of male programmers and designers in the production lines and is geared for a male audience (Bae et al., 2004, Bryce & Rutter, 2003), our system seems to hint towards interactive game-like interfaces that appeal both girls and boys and therefore has a potential to reduce gender bias towards digital technologies. This partially confirms findings of other studies on gaming and gender (Brunner et al., 1998, Gorriz & Medina, 2000, Lucas & Sherry, 2004) regarding girls' preference for role playing games and narrative. On the long term, we therefore see a significant potential in the use of such kind of interactive systems for the promotion of gender balance enrollment in computer science and information technology courses which are currently characterized by a large gender imbalance (Vegso, 2005).

From the subjects' wish lists, we notice a demand for cunning, challenging and adapting environments. Participants want more fun, interaction, action and competition. They wish for more clickable objects and entities, mobility and actions.

We believe that our system can be useful for developing and designing conversational systems with improved error handling of speech recognition and language processing for children as well as adults. To that extend we plan to examine users' error handling strategies in real human-human dialogues and transfer the findings to human-computer and more specifically to children-computer interaction computer dialogue systems. By responding with a blend of text-to-speech and nonverbal behaviors the animated character can provide an empirical foundation for developing effective adaptive strategies.

Additional benefits of software systems of this kind to pedagogy and education are their availability over the time as a sort of digital tutor on-demand and the possibility to individualize the interaction.

More studies, with more subjects and over longer periods of time should be nonetheless carried out before drawing any strong conclusions of general validity (Cole et al., 2004). We are aware of the fact that usability studies conducted with subjects that already participated in previous field tests should be taken with care for the past experience could have some influence on what the subjects later chose as discussion topics with the character and how they chose to interact with it.

On a conclusive note, we recognize that despite learning being a highly challenging and demanding process it can be improved with the deployment of special-purpose software instruments. We believe that computers supplement childhood activities and materials, such as art, books, exploration with writing materials, and dramatic play but cannot replace them. Educational software can offer highly valuable opportunities for collaborative play, learning, and creation but it can also be misused. Eventually, the final decision about whether and how educational software tools can be used lies exclusively in the hands of the educators and their professional judgments.

Finally, we think that such a complex system could have been implemented only within the framework of a multi-agent architecture. This gave the developers of each single agent a large freedom in the choice of programming language, developing environment, etc. and made the system highly modular thus ultimately easily expandable and easy to debug. Nonetheless, there still are some technical issues to be resolved regarding such kind of software architectures. 3D graphical conversational agents and other emergent technologies demand more research with focus on interactive and autonomous systems. Recent developments in multi-agents research have come closer to the goal of building intelligent systems of general competence. Nonetheless, several areas such as design description, security, reusability, and implementation must be investigated further before agents can be universally accepted as a reliable and robust framework.

6. Acknowledgements

We would like to thank Marcela Charfuelan, Dymtro Kupkin, Holmer Hemsen and Mykola Kolodnytsky for design and programming support and Svend Killerich for data entry.

7. References

Ahl, D. H. (ed.) (1979) *More BASIC Computer Games*. New York: Workman Publishing

Allen, J., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L. & Stent,, A. (2000) An Architecture for a Generic Dialogue Shell. *Journal of Natural Language Engineering*, 6(3)

Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L. & Stent, A. (2001) Towards Conversational Human-Computer Interaction, AI Magazine, 22(4):27-37

Anolli, L., Mantovani, F., Balestra, B., Agliati, A., Realdon, O., Zurloni, V., Mortillaro, M., Vescovo, A. & Confalonieri, L. (2006) The Potential of Affective Computing in E-Learning: MYSELF project experience

Aylett, R. S., Paiva, A., Woods, S., Hall, L. & Zoll, C. (2005) Expressive Characters in Anti-Bullying Education. Animating Expressive Characters for Social Interaction, Canamero, L. and Aylett, R. (eds.), John Benjamins Publisher

Bae, Y., Choy, S., Geddes, C. Sable, J. & Snyder, T. (2004) Trends in educational equity of girls and women. *Technical Report 2005016*, National Center for Education Statistics, U.S. Department of Education

Bahrick, L. E., Lickliter, R. & Flom, R. (2004) Intersensory redundancy guides infants' selective attention, perceptual and cognitive development. *Current Directions in Psychological Science*, 13

Bell, L. & Gustafson, J. (2003) Child and Adult Speaker Adaptation during Error Resolution in a Publicly Available Spoken Dialogue System. *Proceedings of Eurospeech*, Geneva, Switzerland

Bell, L., Boye, J., Gustafson, J., Heldner, M., Lindström, A. & Wirén, M. (2005) The Swedish NICE Corpus – Spoken dialogues between children and embodied characters in a computer game scenario. *Proceedings of Interspeech*, Lisabon, Portugal

Bernsen, N. O., Dybkjær, L. & Kiilerich, S. (2004) Evaluating Conversation with Hans Christian Andersen. *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pp. 1011-1014

Bickmore, T. & Cassell, J. (1999) Small talk and conversational storytelling in embodied conversational interface agent. *AAAI fall symposium on narrative intelligence*, pp. 87-92

Boye, J. & Gustafson, J. (2005) How to do dialogue in a fairy-tale world. *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue,* Lisbon, Portugal, September 2-3

Bruckman, A. & Bandlow, A. (2002) HCI for Kids. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, Jacko, J. and Sears, A. (eds.), Lawrence Erlbaum and Associates

Brunner, C., Bennett, D. & Honey, M. (1998) Girl games and technological desire. *From Barbie to Mortal Kombat: Gender and computer games*, Cassell, J. and Jenkins, H. (eds.), MIT Press, Cambridge

Bryce, J. & Rutter, J. (2003) Gender dynamics and the social and spatial organization of computer gaming. *Leisure Studies*, 22:1-15

Cassell, J., Sullivan, J., Prevost, S. & Churchill, E. (eds.) (2000) *Embodied Conversational Agents*. MIT Press, Cambridge, MA, USA

Cheyer, A. & Martin, D. (2001) The Open Agent Architecture. *Journal of Autonomous Agents and Multi-Agent Systems*, 4(1):143-148

Clements, D.H. (1994) The uniqueness of the computer as a learning tool: Insights from research and practice. *Young children: Active learners in a technological age*, Wright, J.L. and Shade, D.D. (eds.), pp. 31–50. Washington, DC: NAEYC

Cole, R., van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S., Wade-Stein, D., Ward, W. & Yan, J. (2004) Perceptive Animated Interfaces: First Steps towards a New Paradigm for HCI. *Proceedings of the IEEE*, 91(9):1391-1405

Coulston, R., Oviatt, S. L. & Darves, C. (2002) Amplitude Convergence in Children's Conversational Speech with Animated Personas. *Proceedings of the 7th International Conference on Spoken Language Processing*, Vol. 4, pp. 2689-2692

Corradini, A., Mehta, M., Bernsen, N.O., Martin, J.-C. & Abrilian, S. (2003) Multimodal Input Fusion in Human-Computer Interaction on the Example of the on-going NICE Project. *Proceedings of the NATO-ASI conference on Data Fusion for Situation Monitoring, Incident Detection, Alert and Response Management*, pp. 223-234

Corradini, A., Mehta, M., Bernsen, N.O. & Charfuelan, M. (2005a) Animating an Interactive Conversational Character for an Educational Game System. *Proceedings of the ACM International Conference on Intelligent User Interfaces*, pp. 183-190

Corradini, A., Mehta, M. & Charfuelan, M. (2005b) Interacting with an Animated Conversational Agent in a 3D Graphical Setting. *Proceedings of the Workshop on Multimodal Interaction for the Visualization and Exploration of scientific data*, pp. 63-70

D'Arcy, S. M., Wong, L. P. & Russell, M. J. (2004) Recognition of read and spontaneous children's speech using two new corpora. *Proceedings of the 9th International Conference on Spoken Language Processing*

Darves, C. & Oviatt, S. L. (2002) Adaptation of Users' Spoken Dialogue Patterns in a Conversational Interface. *Proceedings of the 7th International Conference on Spoken Language Processing*, Vol. 1, pp. 561-564

Darves, C. & Oviatt, S. L. (2004). Talking to digital fish: Designing effective conversational interfaces for educational software. *From brows to trust: Evaluating Embodied Conversational Agents*, Pelachaud, C, and Ruttkay, Z. (eds.), Kluwer Academic

Dehn, D.M. & van Mulken, S. (2000) The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, Vol. 52, pp. 1-22

DiPippo, L. C., Hodys, E. & Thuraisingham, B. (1999) Towards a real-time agent architecture - A Whitepaper. *Proceedings of the 5th International Workshop on Object-Oriented Real-Time Dependable Systems*, pp. 59-64

Eskenazi, M. (1996) KIDS: A database of children's speech. *Journal of the Acoustical Society of America*, Vol. 100

Fabri, M., Moore, D.J. & Hobbs, D.J. (2002) Expressive Agents: Non-verbal Communication in Collaborative Virtual Environments, *Proceedings of International Conference on Autonomous Agents,* Bologna, Italy

Foreman, J. (2003) NEXT-Generation Educational Technology versus the Lecture. *Educause Review*, July/August 2003, pp. 13-22

Gerosa, M. & Giuliani, D. (2004) Investigating automatic recognition of non-native children's speech. *Proceedings of the International Conference on Spoken Language Processing*, pp. 1521-1524

Gratch, J. & Marsella, S. (2005) Lessons from Emotion Psychology for the Design of Lifelike Characters, *Applied Artificial Intelligence*

Gogate, L. J., Walker-Andrews, A. S. & Bahrick, L.E. (2001) The Intersensory Origins of Word Comprehension: an Ecological-Dynamic Systems View. *Development Science*, 4(1):1-37

Gorriz, C. M. & Medina, C. (2000) Engaging girls with computers through software games. *Communications of the ACM*, 43(1):42–49

Giguette, R. (2003) Pre-games: Games Designed to Introduce CS1 and CS2 Programming Assignments. *Proceedings of the ACM SIGCSE*, pp. 288-292

Goleman, D. (1995) *Emotional intelligence*. Bantam Books, New York

Gulz, A. (2004) Benefits of Virtual Characters in Computer Based Learning Environments: Claims and Evidence. *International Journal of Artificial Intelligence in Education*, Vol. 14, pp. 313–334, IOS Press

Hafner, K. & Lyon, M. (1996) *Where Wizards Stay Up Late*. Simon & Schuster Trade, 1996

Hagen, A., Pellom, B. & Cole, R. (2003) Children's speech recognition with application to interactive books and tutors. *Proceedings of the IEEE ASRU Workshop*

Halgren, S., Fernandes, T. & Thomas, D. (1995). Amazing AnimationTM: Movie Making for Kids Design Briefing. *Proceedings of the ACM CHI*, Denver, CO, USA

Hanna, L., Risden, K. & Alexander, K. (1997) Guidelines for Usability Testing with Children. *Interactions*, 4(5):9-14

Haugland, S.W. & Shade, D.D. (1994) Software evaluation for young children. *Young children: Active learners in a technological age*, Wright, J.L. and Shade, D.D. (eds.), pp. 63–76. Washington, DC: NAEYC

Jennings, N.R., Sycara, K. & Wooldbridge. M. (1998) A Roadmap of Agent Research and Development. *Autonomous Agents and Multi-Agent Systems*, Vol. 1, 275-306, Kluwer Academic Publishers

Katz, B. (1997) Annotating the World Wide Web using natural language. Proceedings of RIAO Conference on Computer Assisted Information Searching on the Internet

Knapp, M.L. (1978) *Nonverbal Communication in Human Interaction*. 2nd edition, Holt, Rinehart and Winston Inc., New York

Kort, B., Reilly, R. & Picard, R. W. (2001) An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion, *Proceedings of International Conference on Advanced Learning Technologies*

LeDoux, J. (1998) The emotional brain: The mysterious underpinnings of emotional life. Weidenfeld & Nicholson, London, UK

Levy, S. (2001) *Hackers: Heroes of the Computer Revolution*. Penguin Putnam, New York

Lucas, K. & Sherry, J. L. (2004) Sex Differences in Video Game Play: A Communication-Based Explanation. *Communication Research* 31(5):499-523

Mantovani, F. (2001). VR learning: Potential and Challenges for the Use of 3D Environments in Education and Training. *Towards Cyber-Psychology: Mind, Cognitions and Society in the Internet Age*, Chapter 12, IOS Press

Marsella, S. & Gratch, J. (2003) Modeling coping behavior in virtual humans: don't worry, be happy. *Proceedings of the AAMAS*, pp. 313–320, ACM Press

Martin, D.L., Cheyer, A. & Moran. D.B. (1999) The Open Agent Architecture: A Framework for Building Distributed Software System. *Applied Artificial Intelligence*, (13):91-128

Massaro, D.W., Liu, Y., Chen, T.H. & Perfetti, C. (2006) A Multilingual Embodied Conversational Agent for Tutoring Speech and Language Learning. *Proceedings of the 9th International Conference on Spoken Language Processing*, pp. 825-828

Mateas, M. & Stern, A. (2004) Natural Language Undertanding in Façade: Surface-text Processing, Lecture notes in computer science, Vol. 3105, Springer Verlag, pp. 3-13

Mayer, R. E. (2001) *Multimedia Learning*. Cambridge University Press, New York, 2001

Mayer, R. E. & Moreno, R. (2003) Nine Ways to Reduce Cognitive Load in Multimedia Learning. Educational Psychologist, 38(1):43-52

Mehta, M. & Corradini, A. (2006) Understanding Spoken Language of Children Interacting with an Embodied Conversational Character. *Proceedings of the Combined Workshop on Language-Enabled Educational Technology and Development and Evaluation of Robust Spoken Dialogue Systems* at the ECAI, Riva del Garda, Italy, pp. 51-58

Mehta, M. & Corradini, A. (2008) Handling Out of Domain Topics by a Conversational Character. *Proceedings of the 3rd ACM International Conference on Digital Interactive Media in Entertainment & Arts*, Athens, Greece

Montfort, N. (2003) *Twisty Little Passages: An Approach To Interactive Fiction*. The MIT Press

Moreno, R. & Mayer, R. E. (2002) Verbal Redundancy in Multimedia Learning: When Reading Helps Listening. *Journal of Educational Psychology*, 94(1):156-163

Mori K., Jatowt A. & Ishizuka M. (2003) Enhancing Conversational Flexibility in Multimodal Interactions with Embodied Lifelike Agents. *Proceedings of the ACM International Conference on Intelligent User Interfaces*, pp. 270-272

Narayanan, S. & A. Potamianos (2002) Creating conversational interfaces for children. *IEEE Transactions on Speech and Audio Processing*, 10(2):65-78

Nix, D., Fairweather, P. & Adams, B. (1998) Speech recognition, children, and reading. *Proceedings of the ACM CHI*, Los Angeles, CA

Oviatt, S. L. & Adams, B. (2000) Designing and evaluating conversational interfaces with animated characters. *Embodied Conversational Agents*, Cassell, J,. Sullivan, J., Prevost, S. and Churchill, E. (eds.), MIT Press, Cambridge, MA, USA, pp. 319-343

Oviatt, S. L, Darves, C. & Coulston, R. (2004) Toward Adaptive Conversational Interfaces: Modeling Speech Convergence with Animated Personas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 11, Nr. 3, pp. 300-328

Patel, R., Leuski, A. & Traum, D. (2006) Dealing with Out of Domain Questions in Virtual Characters. *Proceedings of the International Conference on Intelligent Virtual Agents*

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D. & Strohecker, C. (2004) Affective Learning - A Manifesto. *BT Technology Journal*, 22(4):253-269

Prensky, M. (2000) Digital Game-Based Learning. McGraw-Hill.

Ryokai, K. & Cassell, J. (1999) StoryMat: A Play Place with Narrative Memory. *Proceedings of the ACM International Conference on Intelligent User Interfaces*

Rieber, L. P. (1996) Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development* 44(2):43-58

Robertson, J. & Oberlander, J. (2002) Ghostwriter: Educational Drama and Presence in a Virtual Environment. *Journal of Computer Mediated Communication*, Vol. 8

Rudnicky, A., Thayer, E., Constantinides, P., Tchou, C., Shern, R., Lenzo, K., Xu W. & Oh, A. (1999) Creating natural dialogs in the Carnegie Mellon Communicator system. *Proceedings of Eurospeech*, Vol. 4, pp. 1531-1534

Savidis, A., Grammenos, D. & Stephanidis, C. (2007) Developing inclusive e-learning and e-entertainment to effectively accommodate learning difficulties. *Journal on Universal Access to the Information Society*, 5:401–419, Springer Verlag

Squire, K. D. (2005) Changing the game: What happens when videogames enter the classroom? *Innovate*, Vol. 1, Nr. 6

Steffe P. L. & Gale J. (eds.) (1995) *Constructivism in Education*. Lawrence Erlbaum Associates

Stoney, S. & Wild, M. (1998) Motivation and interface design: maximising learning opportunities. *Journal of Computer Assisted Learning*, Vol. 14, pp. 40-50

Vaucelle, C. (2002) DollTalk: A computational toy to enhance children's creativity. *Proceedings of the ACM CHI*, pp. 20-25

Vegso J. (2005) Interest in CS as a major drops among incoming freshmen. *Computing Research News*, 17(1):17–18

Washburn, C., Stringfellow, P. & Gramopadhye, A. (2007) Using Multimodal Technologies to Enhance Aviation Maintenance Inspection Training, Proceeding of International Conference on Human Computer Interaction, Beijing, P.R. China, July 24-27

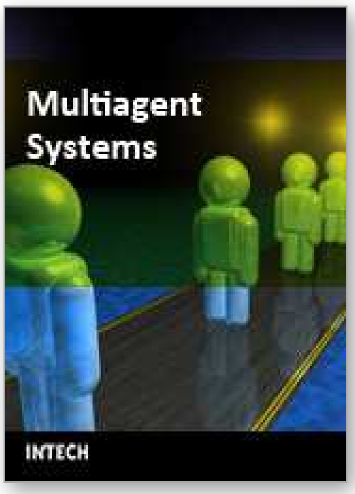
Wickens, C. C. (2002) Multiple Resources and Performance Prediction. *Theoretical Issues in Ergonomics Science*, *3* (2)159-177

Wolf, M. J. P. & Perron, B. (2003) *The Video Game Theory Reader*, Routledge

Yeni-Komshian, G., Kavanaugh, J. & Ferguson, C. (eds.) (1980) *Child Phonology*, Volume 1: Production, Academic Press, New York

Zheng, Z. (2002) AnswerBus Question Answering System. *Proceedings of the Human Language Technology Conference*

Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T. & Hetherington, L. (2000) Jupiter: A Telephone-based Conversational Interface for Weather Information. *IEEE Transactions on Speech and Audio Processing*, 8(1):100-112



Multiagent Systems

Edited by Salman Ahmed and Mohd Noh Karsiti

ISBN 978-3-902613-51-6
Hard cover, 426 pages
Publisher I-Tech Education and Publishing
Published online 01, January, 2009
Published in print edition January, 2009

Multi agent systems involve a team of agents working together socially to accomplish a task. An agent can be social in many ways. One is when an agent helps others in solving complex problems. The field of multi agent systems investigates the process underlying distributed problem solving and designs some protocols and mechanisms involved in this process. This book presents an overview of some of the research issues in the field of multi agents. It is a presentation of a combination of different research issues which are pursued by researchers in the domain of multi agent systems as they are one of the best ways to understand and model human societies and behaviours. In fact, such systems are the systems of the future.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821