The NICE Fairy-tale Game System[1]

Joakim Gustafson, Linda Bell, Johan Boye, Anders Lindström and Mats Wirén
TeliaSonera AB, 12386 Farsta, Sweden
firstname.lastname@teliasonera.com

Abstract

This paper presents the NICE fairy-tale game system, in which adults and children can interact with various animated characters in a 3D world. Computer games is an interesting application for spoken and multimodal dialogue systems. Moreover, for the development of future computer games, multimodal dialogue has the potential to greatly enrichen the user's experience. In this paper, we also present some requirements that have to be fulfilled to successfully integrate spoken dialogue technology with a computer game application.

1 Introduction

The goal of the NICE project is to allow users of all ages to interact with lifelike conversational characters in a fairy-tale world inspired by the Danish author H C Andersen. To make these characters convincing in a computer game scenario, they have to possess conversational skills as well as the ability to perform physical actions in an interactive 3D world.

What primarily distinguishes the NICE fairy-tale game system from other spoken dialogue systems is that the human-computer dialogue takes place within the context of an interactive computer game. However, spoken and multimodal dialogue is not supposed to be just an 'add-on' to the game, but the user's primary means of progression through the story. The rationale for this is the great potential for more natural interaction we see in making methods from multimodal dialogue systems available in controlling gameplay. Potentially, spoken and multimodal interaction will make it possible to create a more engaging and immersive experience, or even facilitate the development of new kinds of computer games.

Secondly, what makes NICE differ from typical spoken dialogue systems is the attempt to move away from strictly task-oriented dialogue. Instead, the interaction with the characters is domain-oriented. This means that the dialogue concerns different subplots in the fairy-tales, but without a clear goal-orientation and without other demands than it being entertaining to the user. Furthermore, social interaction plays an important role in the fairy-tale world where the game takes place. By engaging in socializing with the animated characters, the user will find out things necessary to overcome various obstacles and enable progression through the story.

Thirdly, a feature that differentiates NICE from other systems is that the main target user group of the system is children and young users. Previous studies have indicated that children employ partly different strategies when interacting with dialogue systems than adults do, and that there are also differences between age groups. For instance, younger children use less overt politeness markers and verbalize their frustration more than older children do (Arunachalam et al. 2001). It has also been shown that children's user experience is improved if they can communicate with a system with a 'personality' and that they benefit from being able to choose from several input modalities (Narayanan and Potamianos 2002). Furthermore, since many young people have a lot of experience with computer games, the believability of the dialogue characters and natural expressions will be critical aspects for the system's success.

Thus, computer games provide an excellent application area for research in spoken dialogue technology, requiring an advance of the state-of-the-art in several fronts. Perhaps more importantly, game players will have a lot to gain from a successful incorporation of spoken dialogue technology into computer games. Today's computer games are limited by the user's input options, which are often restricted to direct manipulation and simple commands. In the development of the next generation of computer games, we believe that multimodal dialogue has the potential to greatly enrichen the user's experience. For instance, spoken interaction makes it possible to refer to past events and objects currently not visible on the screen. Social interaction, which is already part of popular games such as SIMS, can be improved with spoken dialogue. Furthermore, speech and multimodal interaction supports cooperative games, where the user and character works together in solving a mutual problem.

[1] The work described in this paper was supported by the EU/HLT funded project NICE (IST-2001-35293), www.niceproject.com

2 Spoken dialogue systems

Spoken dialogue systems have so far mostly been designed with an overall goal to carry out a specific task, e.g. accessing time table information or ordering tickets (e.g. Zue et al. 1991; Aust et al. 1995). With task-oriented systems, it is possible to build domain models that can be used to predefine the language models and dialogue rules. The existence of predefined tasks makes it rather straight-forward to evaluate the performance of the dialogue system.

Recent developments have made it possible to modify and extend the goals of spoken dialogue systems. Explorative dialogues, in which users are encouraged to browse through information without pursuing a specific task, have been presented by (Cassell et al. 1999; Bell et al. 2001). These dialogues still contain tasks to be solved during the interaction, e.g. giving constraints or receiving information about objects. However, explorative dialogue systems cannot be evaluated using merely the number of turns between different user interactions. A user who continues speaking with the system for a long time may do so because she is finding a lot of interesting information.

Yet another type of dialogue system aims to present its users with an engaging and entertaining experience, without the presence of an external predetermined task. Conversational kiosks, such as August (Gustafson and Bell 2000) and MACK (Cassell et al. 2002), encourage users to engage in social dialogues with embodied characters. Such dialogues are amenable to handling by a correctly designed dialogue system, since they primarily bring up features from the shared context.

3 Interactive storytelling

Interactivity has been defined as “a kind of drama where the audience can modify the course of the actions […] thus having an active role” (Szilas 1999). In interactive scenarios, the user helps the story unfold and may affect its course depending on his or her active participation. It has been argued that interactive storytelling will change computer entertainment by introducing better narrative content and allowing users to interfere with the progression of the storyline (Cavazza et al. 2002). However, Young (2001) suggests that the drama manager of the system should put a limit to the user’s actions by not allowing interference that violates the overall narrative plan. Most interactive games developed so far allow users to intervene in the storytelling by acting on physical objects on the screen using direct maniputation (Young 2001; Cavazza et al. 2002). Moreover, some systems allow users to interact with characters by means of written text input (Mateas and Stern 2002). In addition, Cavazza et al. (2002) explored using a speech interface that handled isolated utterances from the user.

4 The NICE fairy-tale game scenario

The overall goal of the project is to provide users with an immersive dialogue experience in a 3D fairy-tale world, see Figure 1. To this end, we have chosen to make spoken and multimodal dialogue the user’s primary vehicle of progressing through the story. It is also by verbal and non-verbal communication that the user can gain access to the goals and desires of the fairy-tale characters. This will be critical as the characters will ask the users to help them in solving problems. These problems either relate to objects that have to be manipulated or information that has to be retrieved from other fairy-tale characters.

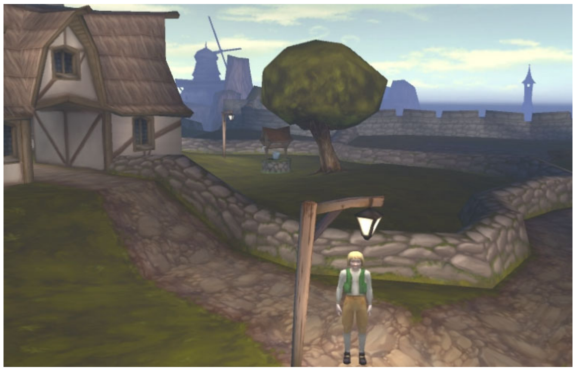

Figure 1. Cloddy Hans in the fairy-tale world.

The fairy-tale domain was chosen because of its classic themes and stereotypical characters, well-known to most adults as well as children. Some of these familiar characters are shown in Figure 2.

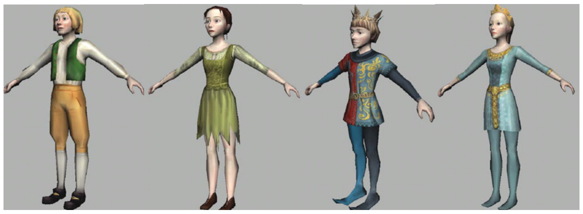

Figure 2. The fairy-tale characters.

To facilitate the progression through the story, we introduce Cloddy Hans, the user’s faithful assistant. Cloddy Hans’s character is conveyed to the users in the following way: he is a bit slow to understand, or so it seems. He sometimes appears hard of hearing and only understands spoken utterances and graphical gestures at a rather simple level. Cloddy Hans does not take a lot of initiatives, but is honest and anxious to try to help the user. In spite of his limited intellectual and perceptual capabilities, he may sometimes provide important clues through sudden flashes of insight.

The user can ask Cloddy Hans to manipulate objects by referring to them verbally and/or by using the mouse. To understand the reason for not allowing users to directly manipulate objects on the screen, we have to recall what distinguishes NICE from other games, namely, spoken multimodal dialogue. We thus want to ensure that multimodal dialogue is appreciated by the user not just as an 'add-on' but as the primary means of progressing in the game. Our key to achieving this is to deliberately limit the capabilities of the key actors — the user and Cloddy Hans — in such a way that they can succeed only by cooperating through spoken multimodal dialogue. In other words, the user is intelligent but cannot himself affect objects in the world; Cloddy Hans on the other hand is a bit slow but capable of physical action according to what he gets told (and he may occasionally also provide tips to the user).

The fairy-tale game will start with an introductory dialogue, in which the user meets Cloddy Hans in H C Andersen's fairy-tale laboratory, see Figure 3. The simple task the user and Cloddy have to solve together is to take fairy-tale objects from a shelf and put them in the appropriate slot in a fairy-tale machine. Each slot is labelled with a symbol, which denotes the type of object supposed to go there, but since Cloddy Hans is not very bright, he needs help understanding these labels.

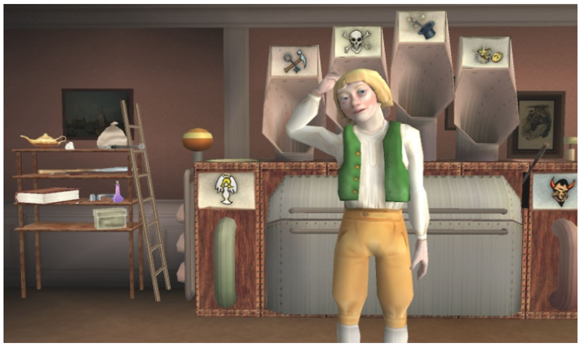

Figure 3. Cloddy Hans in the fairy-tale lab

The initial scenario is a 'grounding game' set in the context of a narrow task. In other words, its real purpose is a training session in which the user and Cloddy Hans agree on what different objects can be used for and how they can be referred to. This process also lets the player find out (by trial-and-error) how to adapt in order to make it easier for the system to understand him or her. Moreover, Cloddy Hans sometimes explicitly instructs the user. For example, one lesson might be that it is sometimes more efficient to use multimodal input instead of just spoken utterances.

The subsequent game in the fairy-tale world depends on what objects have been chosen by the user in the initial scenario. The advantage of this is that the objects are already grounded; for example, a sack of gold will be visually recognized by the player and there is an already agreed way of referring to it.

5 System characteristics

The game scenario as presented in the preceding section puts a number of requirements on the system. The scenario involves several animated characters, each with its own intended distinct personality. These personalities must be made explicit for the game player, and manifest themselves on all levels: from the appearance of the characters, their gestures and voices, choice of words, to their long-term behavior and overall role in the fairy-tale world. Furthermore, the characters need to be responsive, and be able to engage in conversation which makes sense to the player of the game.

On the surface level, then, we need to have beautifully crafted animated characters and environments (these have been designed by the computer-game company Liquid Media). Each character must have its own voice that conveys the nature of that character's personality, and be able to use prosodic cues to signal mood and emotions. To this end, a unit-selection speech synthesizer has been developed. Cloddy Hans has been given a slow, deep voice that goes along with his intended dunce personality. His repertoire of gestures and his style of walking also amplifies the impression of a slow-witted but friendly person.

On the input side, we need to recognize continuous, unconstrained speech for users of all ages. Previous studies have shown that children's speech is associated with elevated error rates (Potamianos et al. 1997; Oviatt and Adams 2000), making it necessary for Scansoft to retrain the NICE recognizer's acoustic models. In addition, we need to take into account the disfluent speech patterns that are likely to arise, most probably because the users are unused to the situation or distracted by the virtual environment. On the other hand, not all input needs to be adequately interpreted. Much of the socializing utterances from the user can be handled in a satisfactory way by using shallow methods. Furthermore, the interpretation of the goal oriented interactions is simplified by the fact that the system knows which objects are visible on the screen and, more importantly, since it already knows what problems the fairy-tale characters has asked the user to help them to solve. Finally, the user also has the possibility of referring to objects using a pointing device. The software for the interpretation of this graphical input has been developed by LIMSI.

The above characteristics have led us to design the system's interpretation of user input in the following way. The system is implemented as a set of eventdriven processes that communicate via message-passing. The architecture is essentially an extension of the one described in (Bell et al. 2001). This architecture allows, among other things, for highly flexible turn-taking. When the user speaks, the system first tries to categorize the utterance as either social (needing only shallow interpretation) or goal-oriented (needing further analysis).

Finally, the long-term behavior of a character is decided by its set of internal goals and rules. A goal is essentially a predicate (that can be either true or false) concerning of the state of the virtual world. For instance, a character may have a goal to acquire a certain object or visit a certain place. If a given goal is not fulfilled (the predicate is false), the character will try to fulfill it. To this end it will use its set of rules, that define actions and dialogue acts that are likely to contribute to reaching the goal.

6 Evaluation issues

Task-oriented spoken dialogue systems are usually evaluated in terms of objective and subjective features. Objective criteria include the technical robustness and core functionality of the system components as well as system performance measures such as task completion rate. Subjective usability evaluations estimate features like naturalness and quality of the interactions, as well as user satisfaction reported in post-experimental interviews. However, many of these measures are simply not relevant for entertainment-type applications, where user satisfaction increases rather than decreases with task completion time. It can even be difficult to define what the completion of the task would be. In practice, computer games are usually evaluated by professional game reviewers and by the users in terms of number of copies sold.

In the evaluation of the NICE fairy-tale game sales figures will not be possible to use, and several of the traditional objective measures are less relevant due to the domain. Instead, subjective measures involving features like “narrative progression”, “character believability”, and “entertainment value”, will be used. They will be obtained off-line, by interviewing the users after their interactions and asking them to fill out questionnaires. Users will be asked how they perceived the quality of the actual interaction, as well as the personality of the fairy-tale characters. Expert evaluators, who will be able to replay the user interactions and inspect the system logs, will also be employed. Examples of evaluation questions to the experts include: “Do the characters display meaningful roles and believable personalities that contribute to the story?”, “Do they succeed in signaling their level of understanding”, “To what extent is the user able to affect the plot”?

In order to be able to replay the user interactions with the fairy-tale system, all communication between the system modules are logged with time stamps. This will be a valuable tool both in the iterative system development and for system evaluations. At present, we are in the process of collecting data with the introductory game scenario. The data collected will be used to develop the subsequent scenarios in the fairy-tale game.

References

Arunachalam, S., D. Gould, E. Andersen, D. Byrd and S. S. Narayanan. (2001). Politeness and frustration language in child-machine interactions. *Proceedings of Eurospeech*: 2675-2678.

Aust, H., M. Oerder, F. Seide and V. Steinbiss (1995). The Philips automatic train timetable information system. *Speech Communication* **17**(3-4): 249-262.

Bell, L., J. Boye and J. Gustafson (2001). Real-time handling of fragmented utterances. *Proc. NAACL 2001 workshop on Adaptation in Dialogue Systems*.

Cassell, J., T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson and H. Yan (1999). Embodiment in conversational interfaces: Rea. *Proceedings of CHI*: 520-527.

Cassell, J., T. Stocky, T. Bickmore, Y. Gao, Y. Nakano, K. Ryokai, D. Tversky, C. Vaucelle and H. Vilhjlmsson (2002). MACK: Media lab Autonomous Conversational Kiosk. *Imagina 02*. Monte Carlo.

Cavazza, M., F. Charles and S. J. Mead (2002). Character-based interactive storytelling. *IEEE Intelligent Systems, Special issue on AI in Interactive Entertainment*: 17-24.

Gustafson, J. and L. Bell (2000). Speech technology on trial - Experiences from the August system. *Natural Language Engineering* **6**(3-4): 273-286.

Mateas, M. and A. Stern (2002). Architecture, authorial idioms and early observations of the interactive drama Facade. *Technical report CM-CS-02-198*.

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

Oviatt, S. and B. Adams (2000). Designing and evaluating conversational interfaces with animated characters. *Embodied Conversational Agents*. J. Cassell, J. Sullivan, S. Prevost and E. Churchill. MIT Press.

Potamianos, A., S. Narayanan and S. Lee (1997). Automatic speech recognition for children. *Proceedings of Eurospeech*. **5**: 2371-2374.

Szilas, N. (1999). Interactive drama on the computer: beyond linear narrative. *AAAI 1999 Fall Symposium on Narrative Intelligence*.

Young, R. M. (2001). An Overview of the Mimesis Architecture: Integrating Intelligent Narrative Control into an Existing Gaming Environment. *Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*.

Zue, V., J. Glass, D. Goodline, H. Leung, M. Phillips, J. Polifroni and S. Seneff (1991). Integration of speech recognition and natural language processing in the MIT voyager system. *Proc. ICASSP'91*. Toronto.