NICE project (IST-2001-35293)

Natural Interactive Communication for Edutainment

NICE Deliverable D1.1-2a

Requirements and design specification for domain information, personality information and dialogue behaviour for the second NICE HCA prototype

19 April 2004

Author

Niels Ole Bernsen and the NISLab NICE team

NISLab, Odense, Denmark

Project ref. no.	IST-2001-35293
Project acronym	NICE
Deliverable status	Internal
Contractual date of delivery	29 February 2004 + agreed 6 weeks delay = 15 April 2004
Actual date of delivery	19 April 2004
Deliverable number	D1.1-2a
Deliverable title	Requirements and design specification for domain information, personality information and dialogue behaviour for the second NICE HCA prototype
Nature	Report
Status & version	Final
Number of pages	16
WP contributing to the deliverable	WP1
WP / Task responsible	NISLab
Editor	N/A
Author(s)	Niels Ole Bernsen and the NISLab NICE team
EC Project Officer	Mats Ljungqvist
Keywords	Edutainment, natural interactivity, conversational spoken dialogue systems, speech in computer games, multimodal interaction, human language technologies.
Abstract (for dissemination)	This deliverable D1.1-2a from the IST/HLT NICE (Natural Interactive Communication for Edutainment) project specifies the second NICE Hans Christian Andersen system prototype.

Table of Contents

1 Introduction

At the time of writing this report, in April 2004, the NICE HCA first prototype (PT1) has been developed and tested in near-to-full compliance with the specification presented in NICE Deliverable D1.1, *Requirements and design specification for domain information, personality information and dialogue behaviour for the first prototype* [Bernsen et al. 2002]. Two exceptions should be mentioned, however. Firstly, NICE HCA PT1 does not include semantic input fusion since this has not been possible to do so far. Secondly, PT1 already includes the natural language processing module planned for Month 32. The results of testing PT1 are reported in NICE Deliverable D7.2a, *Evaluation of the First NICE HCA Prototype* [Bernsen et al 2004a].

This document briefly describes the current status of the NICE HCA first prototype (PT1) (Section 2). We then describe the main problems that have been identified with the current system version and which should be addressed in developing PT2, the second NICE HCA prototype (Section 3). Based on Sections 2 and 3 as well as on the project's Technical Annex, Section 4 lists the main requirements to the extension and re-design of the NICE HCA system. Throughout, we put particular emphasis on NISLab's modules for natural language understanding, conversation management, and response generation. However, we also include a number of key requirements to our partners' modules as the satisfaction of those requirements appear to be important to achieving the overall objectives of NICE HCA PT2.

2 Status of the first NICE HCA prototype

Figure 2.1 shows the overall NICE HCA PT1 system architecture. Speech recognition and semantic input fusion are still absent from the system but there is an "input fusion" module which simply forwards linguistic and gesture input to the character module. Figure 2.2 shows the architecture of the natural language understanding module. Figure 2.3 shows the architecture of the HCA character module. Both the general NICE HCA system architecture and information flow, the NISLab modules just mentioned, and NICE HCA PT1 response generation are all described in more detail in recent publications, cf. [Bernsen 2003], [Bernsen and Dybkjær 2004], [Bernsen et al. 2004b], [Bernsen et al. 2004c], and [Corradini et al. 2004].

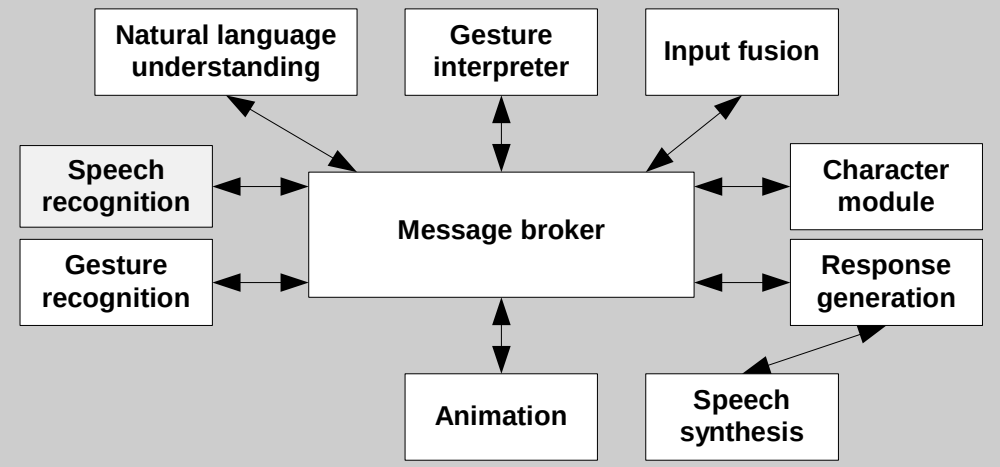

Figure 2.1. Overall NICE HCA system architecture.

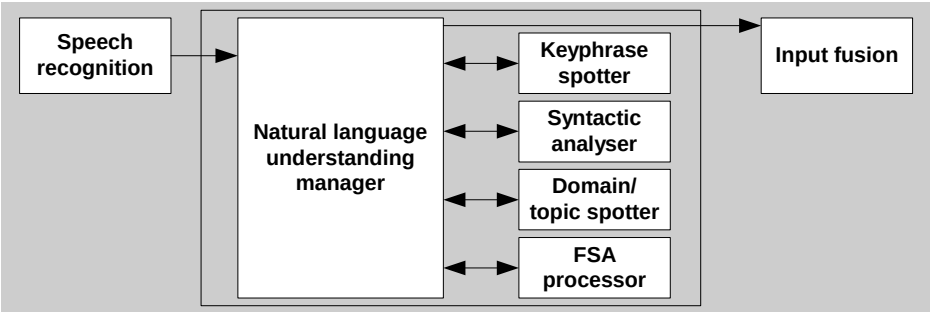

Figure 2.2. Natural language understanding module architecture.

The NICE HCA PT1 system was tested with target users in late January 2004. The NISLab bugs identified have now all been fixed. The user test system version did not include the HCA Character module sub-modules for handling non-communicative action and communicative functions (see Figure 2.3). These sub-modules have been completed, are being integrated, and will be tested in the coming weeks. Similarly, a version of the Scansoft speech recogniser is being integrated with the system while we are waiting to receive from partner Scansoft the first version of the Scansoft speech recogniser which has been trained with the data collected for the purpose, cf. NICE Deliverable D2.2a, *NISLab's Collection and Analysis of Multimodal Speech and Gesture Data in an Edutainment Application* [Bernsen et al. 2004d].

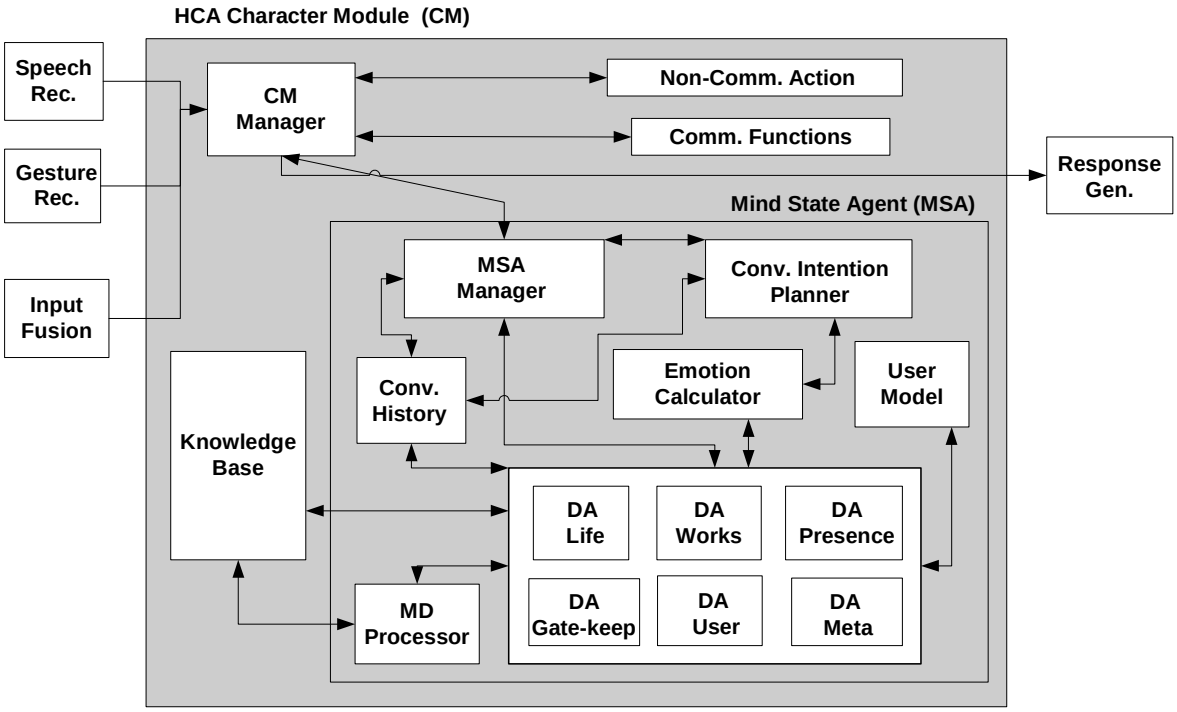

Figure 2.3. HCA character module. DA is domain agent. MD is mini-dialogue.

3 Issues identified in the user test

Analysis of the user test results provided extensive information on the strengths and weaknesses of the NICE HCA PT1 system, cf. [Bernsen et al 2004a]. In the following, we highlight the most important issues which, according to the user test analysis, need to be addressed in the development of PT2.

3.1 Natural language understanding (NLU)

The user test demonstrated that our newly developed NLU module is on the right track and does not have to be fundamentally re-designed. The NLU module parses the spoken input at an appropriately shallow level for non-task-oriented, conversational applications. The main issues in the further development of natural language understanding (NLU) functionality evidenced by the user test, are:

1. the semantic equivalence class problem;
2. increased lexical and grammatical coverage;
3. integration with semantic input fusion;
4. co-reference handling.

(1), the semantic equivalence class problem, is the following. In PT1, the NLU produces, among other things, a semantic representation of the user's spoken input. The semantic representation is matched against possible output in the knowledge base, cf. Figure 2.3. However, since several different semantic input representations may match the same output, these semantic input representations may be described as constituting a semantic equivalence class. The open-endedness-in-principle of semantic equivalence classes (a) creates a knowledge base maintenance problem which should be solved in a different way, cf. Section 4, and (b) makes it difficult or impossible to significantly increase the capacity for conversation planning, cf. Section 3.2.

(2), increased lexical and grammatical coverage, is a rather self-evident requirement to PT2. It must be kept in mind, however, that we are not building a commercial conversational system. So, lexical and grammatical coverage extensions must be kept in check by more important priorities to do with optimising PT2's generic processing mechanisms.

(3), integration with semantic input fusion, is a rather self-evident requirement to PT2. In PT1, we have no way of handling user input which, e.g., combines speech and gesture in asking "Who is this person [pointing to a picture on the wall in HCA's study]?".

(4), although the amount of co-reference resolution needs identified in the user test corpus is limited, co-reference remains a serious challenge in conversational systems. In PT2, ways must be found to (a) handle speech/gesture co-reference in cooperation between the NLU and the input fusion module, and (b) handle the spoken discourse co-reference phenomena which, e.g., arise when a user in in-depth conversation on HCA's family asks about his mother "Did you love her?".

3.2 Conversation management

The user test demonstrated that the general design and architecture of our newly developed HCA character module is on the right track and does not have to be fundamentally revised. At more detailed levels, however, some amount of re-design and re-implementation is to be expected due to the issues discussed below. Since we are addressing the unsolved problem of how to emulate human-human conversation, it came as no surprise when the user tests revealed a number of shortcomings in the way in which conversation is currently managed by

the character module. The main issues in the further development of conversation management functionality evidenced by the user test, are:

1. too limited mini-dialogue flexibility;
2. insufficient conversation planning;
3. more articulate abstraction hierarchies in HCA's knowledge representation;
4. lack of capability of user claims evaluation; and
5. extended domain coverage.

(1), too limited mini-dialogue flexibility, is apparent when, e.g., the user tries to change topic or domain in a mini-dialogue. The user should be able change topic or domain at any time.

(2), insufficient conversation planning, is apparent when, e.g., HCA continues to ask the user questions about a fairytale which HCA was already told that the user is not familiar with. PT2 needs significantly improved capacity for conversation planning, including the keeping track of which topics and their implications have been addressed in the conversation already.

(3), more articulate abstraction hierarchies in HCA's knowledge representation, is instrumental to achieving (1) and (2) above. The more hierarchically articulate HCA's knowledge representation is, the better he can control the relevance of his output, whether the output must be provided to a user who "breaks out of" or "breaks into" a mini-dialogue (1), or HCA must identify a to-the-point continuation of the conversation (2).

(4), lack of capability of user claims evaluation, is apparent when, e.g., the user falsely describes some knowledge item of HCA's, such as his age or the main character in a fairytale. HCA should be able to evaluate the truth of user claims concerning knowledge in his possession.

(5), extended domain coverage, is a rather self-evident requirement to PT2. It must be kept in mind, however, that we are not building a commercial conversational system. So, lexical and grammatical coverage extensions must be kept in check by more important priorities to do with optimising PT2's generic processing mechanisms.

3.3 Response generation

As was the case with respect to the PT1 natural language understanding and character modules, the user test demonstrated that the design and architecture of our newly developed response generation module is on the right track and does not have to be fundamentally revised. In fact, the main requirements to the further development of response generation for PT2 concern extensions which depend on the provision of more expressive rendering functionality. Also, a more extensive use of emotion values should be taken into consideration.

4 Requirement and design specification for PT2

The user test issues described in Section 3 constitute one important source for specifying the requirements to NICE HCA PT2. There are two other sources as well, however. The second source is that part of the generic requirements stated in the NICE project Technical Annex (TA) which have not been addressed in developing PT1. The third source consists of requirements or potential requirements to non-NISLab modules which may not have been described explicitly in the NICE TA but which have emerged in the course of the first two years of the project.

In the following, we present and comment on the PT2 specification requirements module-by-module, cf. Figures 2.1 through 2.3. In some cases, a requirement still needs to be confirmed by the NICE partner concerned. This is stated when the requirement is presented.

Given the fact that we are now deeply into the process of design analysis and have initiated rapid prototyping for PT2, we have combined PT2 requirements and design specification below.

The specification to follow is clearly a highly ambitious one. We should therefore add that we definitely intend to develop to all requirements not marked *"desirable"* below. However, as we are into a rather advanced research agenda, we need to state the qualification that, due to unexpected problems arising in the coming months, we may have to abandon some requirement or part of some requirement.

As a final note, one requirement (15) in Section 4.4 needs confirmation by partner LIMSI. Similarly, several of the requirements in Sections 4.6 and 4.7 are being discussed with partner Liquid Media at the moment.

4.1 Speech recognition

1. Trained versions of the Scansoft speech recogniser must be added to the NICE HCA system.
2. Push-to-talk and timeout functionality.

We do not consider barge-in a top priority at this stage. The complexity of allowing barge-in is expected to be significant. It can be done, no doubt, but we consider that the project resources are better spent on addressing other challenges first.

3. Fast-track communication from the speech recogniser to the Communicative Functions module (Figure 2.3).

This is necessary for enabling HCA to show attention to the user's input.

4.2 Natural language understanding

4. The natural language understanding (NLU) module will output concepts.

Based on our PT2 design analysis, we are specifying a notion of a concept which will be basic to PT1, solving the semantic equivalence issue described in Section 3.1. Concepts are attribute-value structures which will constitute the internal semantic "lingua franca" of PT2. Concepts are expected to be in one-to-one correspondence with the output which the system can generate.

5. *Desirable:* several concepts per user input utterance (turn) will be processed.
6. The processing of users' spoken input will be informed by knowledge of HCA's most recent output.
7. Deliver deictic expressions to input fusion.

8. *Desirable:* key spoken discourse anaphor handling.

We need to experiment with solutions to this problem. As a first step, we will evaluate the extent to which the NLU module's knowledge of HCA's most recent output, including its domain(s) and topic(s), will help resolve co-reference.

9. Increased lexical and grammatical coverage in coordination with the planned extensions of HCA's domain knowledge.

4.3 Gesture processing

10. Reliable gesture recognition and interpretation processing, solidly based on task analysis and post-test user behaviour analysis for the following two user tasks: (a) point to or otherwise indicate (encircling, X-ing, etc.) an object in HCA's study in order to get information about it, (b) play a game with HCA in which the user must identify objects in HCA's study fitting a description provided by HCA.
11. More referenceable objects in HCA's study (compared to PT1).

4.4 Input fusion

12. Rapid prototyping in order to quickly have a first full-gesture-chain input processing solution, including input fusion.

In the absence of input fusion in PT1, we need several PT2 input fusion iterations to get things right.

13. Adequate solutions to any timing issues arising from the need to potentially fusion gesture input and natural language input. The timing solutions must not jeopardise the system's real-time performance.

Obviously, there is little point in having a real-time system, such as PT1, if input fusion imposes significant designer-determined waiting delays.

14. Adequate and reliable input fusion for the two tasks mentioned in (10) above, fusioning natural language module output concepts with gesture information.
15. No forwarding of input fusion tasks to the character module, all tasks must be solved by the input fusion module. The input fusion module should either deliver a 1-best semantic fusion solution to the character module or report to the character module the nature of any problem, of referential ambiguity or otherwise, it may have. Character module meta-communication will then take care of the problem.

4.5 Character module

16. Handle input concepts including variable substitution.

This will simplify and add intelligence to the response generation process.

17. Inform NLU about HCA's most recent output.

This will improve the NLU's handling of "no-domain, no-topic, no-back-channelling" input, such as "Yes, I like it".

18. Communicative functions (Figure 2.3) are selected by the character module and the actual output behaviour is built by the response generator.

HCA can be perceived as paying attention to the user's input process.

19. Non-communicative action (Figure 2.3) is selected by the character module and the actual output behaviour is built by the response generator.

HCA can be perceived to be doing meaningful things in his study even when he is not engaged in conversation.

20. Domain and topic-specific ontologies for knowledge and domain representation.

PT1 uses knowledge base ontologies already. However, we need a more articulate ontology structuring for HCA's domains and topics in order to support flexible conversation management.

21. Ontology-based capability for evaluating user claims.

This is one of the advantages of using articulate ontologies, cf. (20).

22. Conversation intention planning based on highly articulate sectorial hierarchies organised as relevance networks.

This is a crucial set of mechanisms for ensuring conversational coherence: relevance, logicality, absence of obscurity, etc.

23. More articulate, hierarchically organised script and plan rules than in PT1.

This is another crucial set of mechanisms for ensuring conversational coherence: domain and topic shifts when appropriate, first-things-first when HCA enters a new domain, avoidance of undue repetition, etc.

24. Any-time-anywhere break-in and break-out of mini-dialogues.

The user can change domain and topic at any time and at any point in the conversation.

25. More extensive meta-communication handling compared to PT1.

This is a main PT2 challenge as well. We need to see how far we can get in PT2 as regards the handling of semantic and referential ambiguity, user clarification handling, system clarification handling, out-of-domain input handling, obscure input handling, etc.

26. Smart handling of back-channelling.

The decision to de-prioritise barge-in is expected to reduce, but not entirely remove, the problem of handling back-channelling. We need to find generic ways to address the remaining issues.

27. Extended domain/topic coverage, including: more fairytales, something on HCA's other works, more on HCA's person and physical presence, something on HCA's teenage and adult life, more on games, new topic on modern inventions.
28. *Desirable:* more use of conversation history patterns compared to PT1.
29. *Desirable:* machine learning of games and/or modern inventions ontologies.
30. *Desirable:* extended user modelling and use of models compared to PT1.
31. *Desirable:* sense user interest in a domain/topic.
32. *Desirable:* praise users when deserved.

4.6 Response generation

33. Decrease the proportion of verbal output hard-coding.
34. A more non-verbally expressive and varied HCA: more non-verbal behaviours per utterance compared to PT1.

This depends on partner Liquid Media. In particular, it is important the HCA be able to produce several concurrent (i.e. temporally overlapping or simultaneous) non-verbal behaviour elements. In PT1, he can only produce a single non-verbal behaviour element at a time.

35. Fine-tuned, emotional state-dependent, non-verbal emotion expression using, i.a., amplitude and speed for all CA output.

This partly depends on partner Liquid Media.

36. Emotional state-dependent levels for when HCA stops smiling, looks sad, looks angry, etc.

This requires scripting rather than tuning as in (35).

37. Fine-tuned, emotional state-dependent, verbal emotion expression.
38. Solution to the issue of how to reconcile the user's control of HCA's locomotion with HCA self-control.

This depends on partner Liquid Media. An obvious possibility is to prevent the user from controlling HCA's locomotion, making him autonomous.

39. Solution to the issues of how to make HCA turn towards the speaking user/turn towards the user-gestured object given the fact that, currently, the user can control four different virtural camera angles.

This depends on partner Liquid Media. An obvious, but potentially counter-productive, possibility is to prevent the user from controlling the camera angles.

4.7 Rendering

See also Section 4.6.

40. Debugging of HCA's study so that he will not go through the walls or stand in the middle of his furniture.
41. Lip synchronisation.
42. Goal-directed HCA body movement (walking, pointing, looking, ...).
43. Additional primitive non-verbal behaviours compared to PT1.
44. A slightly more friendly-looking default HCA.
45. More light in HCA's study.

4.8 Speech synthesis

So far, we have tested two different TTS systems for HCA. In the user test, we used Scansoft's RealSpeak. This TTS system produced reasonable quality output speech except for some amount of syllable-swallowing. However, as every single user in the user test noticed, it is a major problem that this system only has a female voice! For this reason, we have subsequently replaced RealSpeak by Microsoft's free TTS system. This TTS system has a less natural voice than RealSpeak but has a good scheme for dealing with visemes.

46. Evaluate quality and functionality of Microsoft's synthesiser by comparison with other products on the market in order to improve, if possible, the intelligibility and naturalness of HCA's voice.

4.9 Inter-NICE system connectivity

47. Enable users to go from HCA's study into the fairytale world.

4.10 Maintenance and portability

A major requirement (and desideratum) to the NICE HCA system is to make the system as easy as possible to maintain and port to different characters.

48. Ease of contents maintenance (addition, removal) across speech recognition, natural language understanding, character module, knowledge base, and response generation.
49. Ease of portability to entirely different conversational characters, such as Newton or Gandhi.
50. Make the knowledge base the central repository for maintenance, so that all system contents modifications will automatically be inherited by the relevant modules.

51. Possibly include a maintenance tool.

Requirements (48) through (50) act as key constraints in our ongoing analysis, specification and design work. For obvious reasons, any successful solution to user-character domain-oriented conversation that we might demonstrate in the course of the NICE project, will multiply its value if we succeed in demonstrating easy maintenance and portability as well.

5 Status of NICE HCA PT2 development

As remarked in Section 2, the non-communicative action and communicative function character module sub-modules have been completed and are currently being integrated into the character module, and the (non-trained) Scansoft speech recogniser is being integrated as well.

The work on PT2 requirements specification reported above has been ongoing for two months. Contents analysis and information gathering for PT2 has been ongoing for two months as well. Conversation management design analysis began in January 2004. So far, we have design-specified a general, domain-independent algorithmic solution to the issues posed by the crucial requirements (20) through (24) in Section 4, cf. also requirements (48) and (49). A first, very simple test implementation is underway.

6 References

[Bernsen et al. 2002] Niels Ole Bernsen, Johan Boye, Svend Kiilerich and Ulrich Lindahl: Requirements and Design Specification for Domain Information, Personality Information and Dialogue Behaviour for the First NICE Prototype. *NICE Deliverable D1.1*. NISLab, Denmark: September 2002. 30 pages.

[Bernsen 2003] Bernsen, N. O.: When H. C. Andersen is not talking back In Rist, T., Aylet, R., Ballin, D. and Rickel, J. (Eds.): *Proceedings of the Fourth International Working Conference on Intelligent Virtual Agents (IVA'2003)*, Kloster Irsee, Germany, September 2003. Berlin: Springer Verlag 2003, 27-30.

[Bernsen and Dybkjær 2004] Niels Ole Bernsen and Laila Dybkjær: Domain-Oriented Conversation with H.C. Andersen. To appear in *Proceedings of the Workshop on Affective Dialogue Systems*, Kloster Irsee, Germany, June 2004.

[Bernsen et al 2004a] Niels Ole Bernsen, Laila Dybkjær, Stéphanie Buisine and Jean-Claude Martin: Evaluation of the First NICE HCA Prototype. *NICE Deliverable D7.2a*. NISLab, Denmark: April 2004. 30 pages.

[Bernsen et al. 2004b] Niels Ole Bernsen, Marcela Charfuelàn, Andrea Corradini, Laila Dybkjær, Thomas Hansen, Svend Kiilerich, Mykola Kolodnytsky, Dmytro Kupkin and Manish Mehta: Conversational H.C. Andersen. First Prototype Description. To appear in *Proceedings of the Workshop on Affective Dialogue Systems*, Kloster Irsee, Germany, June 2004.

[Bernsen et al. 2004c] Niels Ole Bernsen, Marcela Charfuelàn, Andrea Corradini, Laila Dybkjær, Thomas Hansen, Svend Kiilerich, Mykola Kolodnytsky, Dmytro Kupkin and Manish Mehta: First Prototype of Conversational H.C. Andersen. To appear in *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 2004)*, Gallipoli, Italy, May 2004.

[Bernsen et al. 2004d] Bernsen, N. O., Dybkjær, L., and Kiilerich, S.: NISLab's Collection and Analysis of Multimodal Speech and Gesture Data in an Edutainment Application. *NICE Deliverable D2.2a*. NISLab, Denmark, April 2004. 30 pages.

[Corradini et al. 2004] Andrea Corradini, Morgan Fredriksson, Manish Mehta, Jurgen Königsmann, Niels Ole Bernsen, and Lasse Johannesson: Towards Believable Behavior Generation for Embodied Conversational Agents. To appear in *Proceedings of the Workshop on Interactive Visualisation and Interaction Technologies, IV&IT 2004*, Krakow, Poland, June 2004, in conjunction with the International Conference on Computational Science 2004 (ICCS 2004).