



# Description of the NICE Hans Christian Andersen System

Niels Ole Bernsen and Laila Dybkjær  
NISLab, University of Southern Denmark

June 2004



# Contents

- 1 Introduction .....5
  - 1.1 Purpose of the application .....5
  - 1.2 Input and output modalities .....5
  - 1.3 Target user group and physical use environment .....5
  - 1.4 Domains and tasks .....5
  - 1.5 System accessibility .....5
- 2 Technical issues .....6
  - 2.1 Platform, hardware requirements, and implementation language .....6
  - 2.2 Architecture .....6
- 3 Functionality .....7
- 4 Interface and usability .....7
- 5 Evaluation .....9
- 6 Conclusion .....12
- 7 References .....12





# 1 Introduction

This document describes the NICE Hans Christian Andersen (HCA) system following the SIMILAR Usability SIG system description structure. The purpose of the structure is to (i) ensure similarly structured system descriptions at similar levels of detail in order to (ii) provide sufficient information for undertaking usability evaluation comparisons of the systems in question, and for documenting current practice.

The NICE HCA system is being developed as part of the European Human Language Technologies NICE project (2002-2005) on Natural Interactive Communication for Edutainment. We describe in this report the implemented first prototype (PT1) of the system.

## 1.1 Purpose of the application

The main goal of the NICE HCA system is to demonstrate natural human-system interaction for edutainment, in particular involving children and adolescents, by developing natural, fun and experientially rich communication between humans and embodied historical and literary characters.

## 1.2 Input and output modalities

The user communicates with HCA using spontaneous speech and 2D gesture. 3D animated, life-like embodied HCA communicates with the user through speech, gesture, facial expression, body movement and action. Communication takes the form of spoken conversation. The language is English

## 1.3 Target user group and physical use environment

Target users are 10-18 years old children and teenagers. The primary use setting of the HCA system is in museums and other public locations. Here users from many different countries are expected to have conversation with HCA for an average duration of, say, 5-15 minutes.

## 1.4 Domains and tasks

The system may be partly viewed as a new kind of computer game which integrates spoken conversation into a professional computer games environment and aims to entertain through emulated human-human conversation. However, the system also has an educational purpose which is being pursued by providing ample correct factual information through story-telling and otherwise, and both visually and orally. The system is *not* a task-oriented system but is defined through the domains of conversation it enables.

## 1.5 System accessibility

The system has been demonstrated at various occasions. It is a research prototype under development and we do not provide freely available demos for download. A short demo video can be downloaded from <http://www.niceproject.com/about/>.

## 2 Technical issues

This section describes technical aspects of the NICE HCA system in terms of platform, hardware requirements, implementation language, and architecture.

### 2.1 Platform, hardware requirements, and implementation language

The NICE HCA system runs on a Windows 2000 platform. It has not been tested on any other platform and for the moment there are no plans for testing the system on other platforms.

In order to run the system one needs a powerful computer with 500-1000 Mb RAM and a good graphics card like G-Force 4.

The implementation languages used in the HCA system are mostly Java, C++ and Sicstus Prolog. Regarding the NISLab modules (cf. Figure 2.1), the natural language understanding module is developed in C++. The character module is developed in C++ and also draws on an access database. The response generation module is developed in C++ and Prolog.

### 2.2 Architecture

The system's event driven, modular, asynchronous architecture is shown in Figure 2.1. The modules are: a speech recogniser from partner Scansoft (not used in PT1); a gesture recogniser based on the free OCHRE neural networks Java software; gesture interpretation developed by partner LIMSI; input fusion from partner LIMSI (not in PT1); off-the-shelf speech synthesis, including time calculation for animation tags; character animation and virtual world simulation from partner Liquid Media; and natural language understanding, character modelling, and response generation from partner NISLab. The modules communicate via a central message broker, publicly available from KTH. The broker is a server that routes function calls, results and error codes between modules. The Transmission Control Protocol (TCP) is used for communication. The broker coordinates input and output events by time-stamping all messages from the modules as well as associating them to a certain dialogue turn. The behaviour of the broker is controlled by a set of message-passing rules, specifying how to react when receiving a message of a certain type from one of the modules.

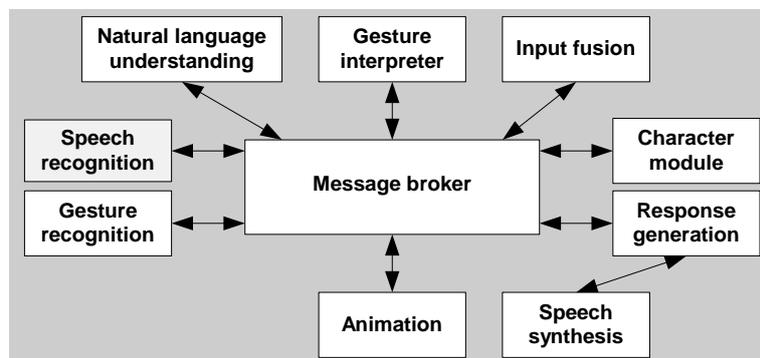


Figure 2.1. General NICE HCA system architecture.

In terms of information flow, the speech recogniser sends an n-best set of hypotheses (only in PT2) to natural language understanding which sends a 1-best hypothesis to input fusion. Similarly, the gesture recogniser sends an n-best hypothesis set to the gesture interpreter

which consults the animation module as to which object the user may have indicated. In PT1, the input fusion module simply forwards an n-best list of pairs of (recognised pointable object + gesture confidence score) from the gesture interpreter and/or a 1-best natural language understanding output to the character module which takes care of input fusion, when required. The character module sends a coordinated verbal/non-verbal output specification to the response generator which splits the output into synchronised text-to-speech and animation. For a more detailed description of the NICE HCA architecture, see [Bernsen et al. 2004a].

### 3 Functionality

The NICE HCA system is not task-oriented, i.e. there are no particular tasks which the user is meant to solve together with the system. Rather the system is domain-oriented, i.e. the user can address, in any order, any part of any domain or topic within HCA's knowledge domains, using spontaneous spoken mixed-initiative dialogue, and pointing gesture. In PT1, the domains are: HCA's fairytales, his childhood life in Odense, his physical presence and his study, the user, and HCA's role as "gate-keeper" for access to the fairytale world. In addition, HCA has a 'meta' domain in order to be able to handle repair meta-communication during conversation. When the conversation is about HCA's study, the user may use 2D gesture input to indicate an object which HCA might want to tell a story about.

In PT1, HCA has fairly limited knowledge about his domains of conversation. His output behaviours are being composed on-line from approximately 300 response templates and 100 primitive non-verbal behaviours. We have implemented his domains breadth-first in order to maximally explore the issues involved, rather than first implementing a single domain in depth. The cover story is that HCA is back! However, he still has to re-learn much of what he once knew. If the user would do him the favour of visiting him later, he is convinced that he will have become much more of what he once was. In addition to the very true information provided by the cover story, the story may help convince users that HCA is not (yet) a full virtual person and make them behave accordingly. HCA does not tell the cover story up front to new users and does not, more generally speaking, instruct users on how to interact with him or inform them of what he is able to have conversation about. Rather, users will be told his cover story if they either explicitly ask what HCA knows about or can do, or if they show too much interest in things he does not know about (yet).

The system is in principle always in one of three states producing either non-communicative action output when HCA is alone in his study, communicative function output when HCA is listening, or paying attention, to a visitor's contribution to the conversation, or communicative action when HCA produces a conversational contribution. However, as long as the recogniser is not connected to the rest of the system the listening behaviour cannot be realised.

HCA has an internal emotional state space model which is updated each time a user input evokes emotions in HCA.

### 4 Interface and usability

Figures 4.1 and 4.2 show 55-year old HCA surrounded by artefacts in his study. The study is a rendering of HCA's study on display in Copenhagen, modified so that he can walk around freely. Pictures relating to HCA's knowledge domains have been hung on the walls. The user may point to them and ask questions about them; HCA can tell stories about the pictures and also about some other objects in his room, such as his travel bag, which is visible to the right in Figure 4.2.

Lacking locomotion autonomy in the first prototype, HCA's locomotion is controlled by the user who is also in control of four different virtual camera angles onto his study. Figures 4.1 and 4.2 show two of these angles. A third angle is a close-up of HCA's desk and the pictures above the desk (visible to the left in Figure 4.2). A fourth angle is a bird's-eye view on the study where HCA's position in the room is marked.



**Figure 4.1.** HCA in his study.



**Figure 4.2.** HCA in his study with his back to the user.

In PT1, the animation engine only allows HCA to display one movement at a time, which means that he can, e.g., blink but not move his mouth at the same time. Basically, this means that he looks somewhat rigid because he is quite far from behaving non-verbally like a human being when speaking.

The system does not require any particular skills from the user apart from being able to use a pointing device and speak. Thus, no user training is foreseen and there is no system manual. If the user says something that HCA does not understand, this will be apparent from his answer, cf. the utterance marked with an \* in Figure 4.3. Figure 4.4 provides another conversation snippet from the user test, cf. Section 5. For more details about how HCA is modelled as a life-like conversational character, see [Bernsen and Dybkjær 2004b].

U: what is your favourite fairy tale  
HCA: I like several of my fairy tales very much. One of my favourites is the ugly duckling. The ugly duckling is not accepted by the other ducks and animals because it looks different. It has a hard time of suffering and troubles. Eventually it wishes only to die till it discovers that it is in fact a beautiful swan. The morale of the fairy tale is that if you are different then you will get bullied.  
U: how old are you  
HCA: I am fifty-five years old. And I am in a good shape occasionally I ride on horses and go for long walks and when I travel I even sometimes hike.  
U: what is your interests  
\*HCA: Did I already tell you that only the bible has been translated into more languages than my fairy tales.  
U: really

**Figure 4.3.** Excerpt of a conversation from the user test.

HCA: If you know it what do you think of the character the little mermaid?  
U: she is very beautiful  
HCA: Can you tell me about the fairy tale The Little Mermaid?  
U: there was a little mermaid who could not wait till she turned fifteen so she could get up of the water  
HCA: What do you think of the prince the character from the fairy tale The Little Mermaid?  
U: she is very very curious  
HCA: Can you tell me about the prince the character from the fairy tale The Little Mermaid?  
U: he is a young prince who sees the little mermaid and falls in love with her  
HCA: I hear that the present crown prince Frederik has now found his princess. What do you think about it?  
U: I think it is very good for him and I think mary will be a good queen

**Figure 4.4.** Another excerpt of a conversation from the user test.

## 5 Evaluation

The first HCA system prototype was tested in January 2004 with 18 users (nine girls and nine boys) from the target user group of 10-18 years old kids and teenagers. This is so far the only user test of the implemented system which has been performed. In the user test setup, only the recogniser was replaced by a wizard who typed what the user said. The rest of the system was running. The speech recogniser still needed to be trained on 40-50 hours of speech data recorded with mostly non-native English speaking children.

Users arrived in parallel, so there were two test rooms, two wizards, and two interviewers. In one room, the user had a mouse and a touch screen for gesture input while in the other room only a mouse was available as pointing device. In the room with the touch screen, the user could also watch HCA on a 42" flat-panel screen. An observer was present in this room as well, cf. Figure 5.1.



**Figure 5.1.** User interacting with HCA. Observer in foreground.

Each user test session had a duration of 60-75 minutes. A session included conversation with HCA in two different conditions followed by an interview. In the first, 15-minutes condition, the users only received basic instructions on how to operate the system, i.e. to speak using the headset, control HCA's movements, control the virtual camera angles, and gesture using mouse or touch screen. In the second condition, the user received a set of 13 brief conversation scenarios, such as "Find out if HCA has a preferred fairytale and which it is", "Make HCA tell you about two pictures and two other objects in his study", and "Tell HCA about games you like or know". The user fully decided on the order and number of scenarios to carry out.

All interactions were logged, audio recorded, and video recorded. In total, approximately 11 hours of interaction were recorded on audio, video, and logfile, respectively. In addition, 18 sets of structured interview notes were collected.

The interviews each took 15-20 minutes and they have been the main source for our evaluation of PT1 together with the logfiles. Figure 5.2 shows the evaluation criteria used to usability evaluate HCA PT1 together with comments and an evaluation score per criterion. More details on the test are available in [Bernsen and Dybkjær 2004a]

Criterion	Evaluation	Score 1-5
Basic usability criteria		
Speech understanding adequacy	No speech recognition in PT1 Natural language processing in PT1: limited but better than basic	As planned 3 acceptable for PT1.
Gesture understanding adequacy	Further improvement needed	3 basic for PT1
Combined speech/gesture understanding adequacy	No semantic input fusion module in PT1	1
Output voice quality	Mostly OK, intelligible, not unpleasant, modest syllable swallowing	4 good for PT1
Output phrasing adequacy	Mostly OK, no user remarks	4 good for PT1
Animation quality	Further improvement needed in rendering capabilities and output design, cf. above	3 acceptable for PT1
Quality of graphics	Rather good, only a (true) user remark on too dark graphics due to the study light sources	4/5 very good for PT1
Ease of use of input devices	Microphone, mouse, touch screen, keyboard:	4/5 very good for

	users generally quite positive	PT1
Frequency of interaction problems, spoken part	A larger number of bugs, primarily loops, found than was expected. A total of 13.3% of the output was found affected by bugs. The non-bugged interaction, on the other hand, showed better performance than expected.	Bugged interaction: 2 barely adequate for PT1 Non-bugged interaction: 3/4 acceptable for PT1
Frequency of interaction problems, gesture part	Some bugs, an algorithm problem, a stack problem, no waiting function	3 basic for PT1
Frequency of interaction problems, graphics rendering part	Two serious generic bugs found: users get lost in space outside HCA's study, HCA immersed in furniture	2 barely adequate for PT1
Sufficiency of domain coverage	Approx. 300 spoken output templates and 100 primitive non-verbal behaviours: further improvement needed	3/4 acceptable for PT1.
Number of characters the user interacted with in the fairy tale world	N/A HCA's study is distinct from the fairytale world	N/A
Number of objects the subject(s) interacted with through gesture	21 pointable objects in HCA's study: in general, the users pointed to most of them.	3 acceptable for PT1
Navigation in the fairy tale world	N/A HCA's study is distinct from the fairytale world	N/A
Number of topics addressed in the conversation	All generic topics (approx. 30), not all topic details	As expected
<b>Core usability criteria</b>		
Conversation success	Most users pointed out that HCA's responses were sometimes irrelevant. Due to loops and core research difficulties.	3/4 acceptable/ good for PT1
How natural is it to communicate via the available modalities	Very positive user comments overall	4/5 very good for PT1
Output behaviour naturalness	Very complex criterion, hard to score. Still, users were surprisingly positive.	3/4 quite acceptable for PT1
Sufficiency of the system's reasoning capabilities	Capabilities are basic at this stage	3 acceptable for PT1
Ease of use of the game: How well did users complete the scenario tasks?	Difficulties mainly due to loops and conversation management	3 acceptable for PT1
Error handling adequacy, spoken part	Limited in PT1. User test data and speech recogniser addition needed for identifying problems and designing improvements	2 acceptable for PT1
Error handling adequacy, gesture part	No error handling involving gesture	1
Scope of user modelling	User age, gender and nationality collected, age used	3
Entertainment value	User test very positive	4 good for PT1
Educational value	User test very positive	4 good for PT1
User satisfaction	User test very positive	4 good for PT1

**Figure 5.2.** Evaluation criteria applied to the HCA PT1 system.

The answers collected from the 18 users who participated in the user test were, even surprisingly, encouraging. Overall, the users found that the technology is on the right track and represents a first glimpse of entirely new spoken computer games technology which could significantly improve the entertainment and educational value of computer games as well as attracting a new group of users who have not been so interested in traditional computer games. More information about the evaluation of the interview data can be found in [Bernsen and Dybkjær 2004c].

Work on user test evaluation continues. In particular, substantial efforts are being put into how to annotate and score user-HCA conversations in accordance with the theory of domain-oriented conversation for entertainment underlying the application. New metrics are being developed for conversation success, symmetry in conversation, etc., see also [Bernsen et al. 2004b].

## 6 Conclusion

The first NICE HCA prototype was well received by the users who found it entertaining and fun to have conversation with life-like HCA using speech and 2D gesture input. What we have to do in developing the second NICE HCA prototype is to (i) improve the system's capabilities in various ways, many of which were pointed out by the users. In particular, the flexibility of spoken conversation management should be improved, the language understanding capabilities of the system should be improved, the graphics should be thoroughly debugged, and more expressive non-verbal behaviour should be developed. Secondly (ii), we should augment HCA's story-telling repertoire, particularly with respect to his knowledge about himself and his life, as well as about objects in his study, further increasing, if possible, the edutainment qualities of the system in the process.

The second HCA prototype (PT2) is now under development. In PT2 there is particular emphasis on increased conversational coherence and flexibility. The design and development is inspired by the data collected in the user test, cf. Section 5, and data collected in an earlier, fully simulated Wizard of Oz setup of the system, cf. [Bernsen et al. 2004b]. The second prototype will be ready by the end of 2004.

## 7 References

The references below can be found at the NICE website at <http://www.niceproject.com/>.

Bernsen, N.O., Charfuelàn, M., Corradini, A., Dybkjær, L., Hansen, T., Kiilerich, S., Kolodnytsky, M., Kupkin, D., and Mehta, M.: First Prototype of Conversational H. C. Andersen. Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 2004), Gallipoli, Italy, 2004a, 458-461.

The paper describes in some detail the architecture of the NICE HCA first prototype.

Bernsen, N.O. and Dybkjær, L.: Evaluation of the First NICE HCA Prototype. NICE Deliverable D7.2a, Part 1, April 2004a.

Evaluation report describing the first user test of the first HCA prototype carried out in January 2004.

Bernsen, N.O. and Dybkjær, L.: Domain-Oriented Conversation with H.C. Andersen. Proceedings of the Workshop on Affective Dialogue Systems, LNAI 3068, Springer Verlag, Kloster Irsee, Germany, June 2004b, 142-153.

Paper describing our approach to making HCA life-like, capable of domain-oriented conversation, and affective.

Bernsen, N.O. and Dybkjær, L.: Evaluation of Spoken Multimodal Conversation. Proceedings of the International Conference on Multimodal Interfaces (ICMI), 2004c (to appear).

Paper describing HCA PT1, focusing on multimodal conversation and user test evaluation results on multimodal conversation.

Bernsen, N. O., Dybkjær, L. and Kiilerich, S.: Evaluating Conversation with Hans Christian Andersen. Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'2004), Vol. III, Lisbon, Portugal, May 2004b, 1011-1014.

The paper presents an analysis of data from a large-scale in-field Wizard of Oz simulation of the HCA system.