

Meet Hans Christian Andersen

Niels Ole Bernsen

NISLab, University of Southern Denmark
nob@nis.sdu.dk

Laila Dybkjær

NISLab, University of Southern Denmark
laila@nis.sdu.dk

Abstract

This paper presents the second running prototype of a multimodal conversational edutainment system embodying 3D animated fairytale author Hans Christian Andersen.

1 Introduction

The separate communities in spoken dialogue systems, embodied conversational agents (ECAs) and computer games have in recent years grown closer to each other and there are now a number of entertainment, edutainment, and education products and research projects which bring together two or even all of them. Among the early projects drawing on all the communities mentioned is the European NICE (Natural Interactive Communication for Edutainment) project (2002-2005). Its overall goal was to build a prototype enabling kids and youngsters to have edutaining and entertaining multimodal conversation with 3D animated fairytale author Hans Christian Andersen (HCA) and his fairytale characters. The project thus faced several challenges which had not been met before, including: (1) workable speech recognition for children and teenagers' voices, (2) computer game characters able to conduct advanced spoken interaction, (3) spoken dialogue and conversation for edutainment and entertainment, (4) domain-oriented, as opposed to task-oriented, spoken interaction in which the system handles spoken conversation in semi-open domains, and (5) fusion of spoken and 2D gesture input for edutainment and entertainment applications.

The NICE project has built two prototype systems, one for English conversation with HCA and one for Swedish plot-based gameplaying with several of HCA's fairytale characters in a fairytale world. The systems were developed in close collaboration among LIMSI-CNRS (France), Liquid Media (Sweden), NISLab (Denmark), Scansoft (Germany), and TeliaSonera (Sweden). Scansoft delivered speech recognition for Swedish and English optimised for children and trained on data (mostly) collected in the project. LIMSI developed gesture recognition and interpretation, and speech-gesture input fusion. Liquid Media developed 3D animation, extending their computer games platform. NIS-

Lab developed conversation theory, natural language understanding, conversation management and character modelling, and response generation for HCA. TeliaSonera developed multi-character plots, natural language understanding, dialogue management and character modelling, and response generation for the characters that inhabit the fairytale world. Speech synthesis is off-the-shelf for the HCA system and in-house developed for the fairytale world system.

In this paper we focus on the HCA system.

2 Embodied conversational characters

There is a sense in which the HCA system is simply an interactive spoken computer game. This potentially revolutionary field of spoken computer games was close to non-existent when the NICE project began. Today, several computer games offer spoken input command words which make a game character perform some action. It has also been attempted to offer spoken dialogue control as in Konami's LifeLine [LifeLine], achieving an interactive spoken computer game. So far, these products do not seem terribly popular with the games reviewers, probably because they typically assume that the gameplayer is able to learn, sometimes quite large, numbers of spoken commands, and their speech recognition and understanding is too fragile. Most of them still do not offer spoken interaction in the sense that the character responds orally to the user's spoken input.

Interactive spoken computer games more comparable to the HCA system are found in research on natural interactive communication with animated graphical characters, sometimes called ECAs [Cassell et al. 2000]. Given the enormous challenges to achieving full human-style natural interactive communication, research on ECAs is a multi-dimensional endeavour, ranging from fine-tuning lip synchronisation details through adding computer vision to ECAs to theoretical papers on social conversation skills and multiple emotions which ECAs might come to include in the future.

So far, the ECA community has put less emphasis on spoken interaction than has been done in the HCA system and is only now beginning to face the challenges of domain-oriented conversation. Secondly, few in the ECA community have had the luck to collaborate with professional computer games developers as we have done in NICE. Thirdly, few ECA researchers have ven-

tured into the highly complex territory of semantic gesture/speech input fusion for conversation. Thus, few systems come close to the NICE prototypes in representing complete demonstrators of interactive spoken computer games for edutainment and entertainment. See, however, [Traum et al. 2004].

3 Meeting HCA

The primary use setting for 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, 10-20 minutes. The target users are 10-18 years old children and teenagers.



Figure 1. HCA in his study.

The user meets HCA in his study in Copenhagen (Figure 1) and communicates with him in fully mixed-initiative conversation using spontaneous speech and 2D gesture. At any time, the user can freely change the topic of conversation, back-channel or provide other kinds of comments on what HCA is saying, or point to objects in HCA's study, possibly speaking at the same time - and receive an appropriate response in context. HCA communicates through audiovisual speech, gesture, facial expression, body movement and action. The theory of conversation underlying HCA's conversational behaviour is derived from analyses of social conversations aimed at making new friends, emphasising common ground, expressive story-telling, rhapsodic topic shifts, balance of "expertise", etc. [Bernsen and Dybkjær 2004]. When HCA is alone, he goes about his work, thinking, meandering in locomotion, looking out at the streets of Copenhagen, etc. When the user points (using a touch screen) at an object in his study, he looks at the object and then looks back at the user before telling a story about the object. HCA has knowledge about: his works, including in-depth knowledge of three of his fairytales; his life, mostly his childhood, including child-

hood games and games users like; his physical and personal presence; his study, including many of the objects in there. He gathers information about the user and understands a variety of generic input, including various forms of user-initiated meta-communication. The objects that the user can get information about via gesture and gesture/speech input are the 16 pictures on the walls, a feather pen and a travel bag.



Figure 2. Close-up of a sad Andersen.



Figure 3. HCA is looking angrily at the user.

The user can control HCA's locomotion by using the keyboard arrow keys and changes virtual camera angle by using the F2 key. A large number of camera angles are available as illustrated in Figures 1 through 4.

Although hard to spot, HCA sometimes shows emotions during conversation. Most of the time his expression is friendly but he may also turn sad (Figure 2) or angry (Figure 3). Correspondingly, he sometimes expresses emotion verbally, cf. his reactions to being called old and ugly in the conversation in Figure 7.



Figure 4. A view from above of HCA's study.

4 Functional system description

The second prototype HCA system (PT2) works in close-to-real time. PT2's general architecture is shown in Figure 5. The modules communicate via a Message Broker which is publicly available from KTH [KTH brooker]. The Broker is a server that routes function calls, results, and error codes between modules, using TCP for communication.

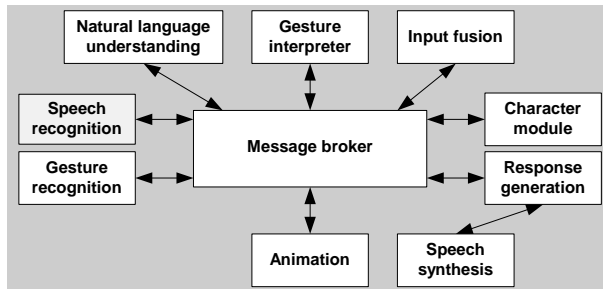


Figure 5. General NICE HCA system architecture.

Input processing is distributed across two input "chains" which come together in Input fusion. Speech recognition uses a 1977 word vocabulary and a language model developed on the basis of three Wizard of Oz corpora and two domain-oriented training corpora collected in the project. The recogniser's acoustic models are tuned to childrens' voices using approx. 70 hours of data. The recogniser does not have barge-in because of the potentially noise-filled public use environments. Natural language understanding uses the best-recognised input string to generate a frame-based attribute/value representation of the user's spoken input.

Gesture recognition recognises these 2D gesture shapes: pointer, e.g., a pointing gesture; connect, e.g., a line; surrounder, e.g., a circle or similar shape encircling an object; and unknown shape. Gesture interpretation

detects through communication with Animation (Figure 5) the object(s) the user gestured at and outputs equivalents to 'no object', 'a single named object', and 'several named objects'. Like Speech recognition, Gesture interpretation is inhibited during preparation and synthesis of HCA's verbal and non verbal behavior. Input fusion attempts to fusion linguistic input frame deictic expressions and references to gesturable objects with objects from Gesture interpretation.



Figure 6. A young user talking to HCA.

The HCA Character module matches frame input from Input fusion to potential HCA output in context. HCA keeps track of what he has said already and changes domain when, having the initiative, he has nothing more to tell about a domain; takes into account certain long-range implications of user input; remembers his latest output; and keeps track of repeated generic user input, including input which requires some form of system-initiated meta-communication. The Character module's Emotion calculator calculates a new emotional state for each conversation turn. If the input carries information which tends to change HCA's emotional state from its default friendly state towards angry (e.g., "You are stupid"), sad (e.g., "How was your mom?"), or happy ("Who are you?" – I am a the famous author HCA ...) – the Emotion calculator updates his emotional state. If the user's input does not carry any such information, HCA's emotional state returns stepwise towards default friendly.

Design-wise, HCA is always in one of three output states, i.e., non-communicative action (NCA) when he is alone in his study working, communicative function (CF) when he pays attention to the user's input, and communicative action (CA) when he actually responds to input. In the current system version, however, these three output states are not fully integrated and can only be demonstrated in isolation. The exception is when the user gestures towards an object in HCA's study, making

him turn towards the object gestured at and then turn back to face the user (the virtual camera).

Response generation generates a surface language string with animation and control (e.g., camera view) tags. The string is sent to the speech synthesiser which synthesises the verbal output and helps synchronise speech and non-verbal output, including audio-visual speech. Speech synthesis is off-the-shelf software from AT&T. This voice has been chosen partly for its inherent intelligibility and naturalness, and partly for matching the voice one would expect from a 55 years old man. Finally, Animation renders HCA's study, animates HCA, and enables the user to change camera angle and control HCA's locomotion.

5 A conversation with Andersen

Figure 7 shows an abbreviated combination of two HCA conversations recorded during our second HCA prototype user test in February 2005 (Figure 6). The example illustrates several issues, including, e.g., giving and taking initiative, emotion shifts, HCA's reaction to input he does not understand, and the user's handling of situations in which his/her input was not understood.

okay	Hello there welcome to my study. I am glad to meet you. It is always so interesting when I get visitors in my study. And I am curious to get to know my visitors. Therefore I would like to ask you a few questions at first. Please tell me your name.
my name is NN	Okay. How old are you?
im twelve	When I was your age I was a big fan of the theatre. Whenever there was an opportunity I went to the theatre. It was my favourite place number one. I would like to know if you are a boy or a girl.
i am a boy	I was a curious boy. Once I ventured into the hospital for the dangerous lunatics. A woman sat in her cell singing with a beautiful voice. Suddenly she got up and ran screaming towards the door. She reached for me and her fingertips touched my clothes. I was half dead from fear when the attendant came. Which country do you come from?
i come from Denmark	Excellent, thanks for letting me know about you. Now let us talk. I am still far from being the Hans Christian Andersen I once was. Maybe you will find that there is still much I have to remember or learn again. But don't despair. You can always ask me what I know or what I can do.
how old are you	I was born in 1805 and died in 1875. So my lifetime lasted 70 years. It was a time full of great events in the world and also for me. I am happy to tell you about the things I remember from my life which must have been very exciting. To tell you the truth I really hope so be-

	cause they tell me that I have written several large books about it. For the moment I can tell you a lot about my family and my childhood in Odense. I can also tell you things about my youth in Copenhagen and a little bit about my adult life in the whole wide world which I am still trying to remember.
[no input]	Would you like to hear about my life?
yes	You want to hear about my family
yes can you tell me about it	I had a nice and small family. Apart from my parents I only saw my grandparents regularly and I was almost an only child. Sadly my family was not as fortunate in life as myself. My parents were common people who had to work hard all through their lives and look twice at every penny you know. No life was no bed of roses to my mother and father not to mention my grandmother and grandfather.
[no input]	You want to hear about my father
no thank you	Do you want to hear other stories about my life?
no thank you you are very old	Sorry
you are very old	I am not old! As you can see I am a man in my best age. I know many things about the world and I can do many things or at least I used to. I still have much hair on my head and I don't wear a wig. I even had my hair curled lately and my beard shaven off completely. In spite of that many people think that I look strange I know that big feet and all. I will not speak of my teeth so don't you dare to ask.
you are very old	No no. I am not old at all. I am only fifty five years old - that is nothing compared to two hundred years. Well I was born two hundred years ago. So I am wondering how I can be fifty five especially since I was born again just a few months ago. I wish somebody would tell me! Maybe I will never die again?
you are ugly	That was not necessary i dont think. You want to hear about my youth?
what is your favourite fairy tale	I like several of my fairytales 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 point is that if you are different then you will get bullied.
can you dance	Sorry
can you do the moon-walk	I am sorry but i still don't understand. Maybe you are talking about one of the many things i still don't know about
what games do you like to play	My favourite childhood game was to play theatre. When I was eight I saw an opera at the theatre in German about a mermaid which I soon played at home. I did not speak or even understand German but I liked this opera. So I created my own homemade German words and

	wore my mothers apron over my shoulders. Now I was a knight or the mermaid swimming on a stool. All these scenes scared my mother she forbade me to play this game because she often believed that I had gone mad. But I could not stop you know. I am curious to know which games children and teenagers play nowadays. Do they play any games which did not exist when I was a child?
i like soccer do you like soccer	Sorry i would suggest that we talk about something i know.
i like football do you like football	Football. I know a foot because I have two large feet myself. And I know a ball as well because I wrote a fairy tale about the top and the ball. So in football you kick around a ball with your feet is that what football is about? You really must tell me more about this game.
are you dead	I died on the fourth of August eighteen hundred and seventy five.
what do you use when you make fairy tales	I only have limited knowledge about my fairytales yet. For example I can tell you how many fairytales I wrote, which relation they had to my life, or maybe you want to know if I wrote any scary fairytales. I can also tell you about a particular fairytale though for the moment I only remember details about three of the fairytales I have written. The ones I remember are the Little Mermaid the Ugly Duckling and the Princess and the Pea.
is this what you use Gesture: featherPen	This is my pen. The fairytale I am writing now is called The Pen and the Inkstand. The pen and the inkstand are jealous of each other. The pen says to the inkstand: you don't think at all. If you did you would see that you can only provide the means. You give the fluid that I may place upon the paper what dwells in me and what I wish to bring to light. It is the pen that writes: no man doubts that and indeed most people understand as much about poetry as an old inkstand.
what can you do	It is very much up to you what I can do. When you talk to me about my life my fairy tales and my study I sometimes can tell you some good stories or at least give you some answers. It is also up to you if you want to see me happy angry or sad. Some of the things we can talk about make me happy or angry others make me sad.
what is this Gesture: locomotive	I am so impressed how fast trains can go. They rush along at incredibly high speed. There is hardly time to view the landscape. A hundred and ten kilometer long journey lasts only three and a half hours. Hooray for the steam machine!
where do you live now	I live in the heart of Copenhagen. From my windows I have a marvellous view of the nineteenth century street life. I live only two houses away from the harbour. This is a good place to live except in winter where my little house is once in a while hit by glacially cold storms.

	If only my landlady would give away her terrible dog this would really be a perfect place to live. I never liked dogs you know.
goodbye	Bye bye. Keep away from fires. They are so dangerous and scary.

Figure 7. Conversation example composed from two actual conversations with HCA.

Acknowledgements

We gratefully acknowledge the support by the European Commission's HLT Programme, NICE Grant IST-2001-35293. NICE HCA PT2 is the result of many peoples' work. We would like to warmly thank Marcela Charfulan who solved hard design problems and implemented most of the character module, Andrea Corradini who was responsible for response generation and part of the character module, Holmer Hensen who helped with recogniser tuning and emotion calculation, Svend Kiil-erich who worked on contents and data for everyone, Mykola Kolodnytsky who implemented part of the character module, Dmytro Kupkin who helped with many things, and Manish Mehta who developed natural language processing and part of the character module, and helped with design problem-solving more generally. We would also like to warmly thank Reinhard Blasig, Scansoft, Jean-Claude Martin and Stéphanie Buisine, CNRS-LIMSI, and Morgan Frederiksson and Lasse Johanneson, Liquid Media, for their work on HCA speech recognition, gesture "chain" processing, and animation, respectively.

References

- Bernsen, N.O. and Dybkjær, L.: Evaluation of Spoken Multimodal Conversation. Proceedings of the Sixth International Conference on Multimodal Interfaces (ICMI 2004), Penn State University, USA, 2004, 38-45.
- Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (Eds.): Embodied Conversational Agents. Cambridge, MS: MIT Press 2000.
- KTH broker: <http://www.speech.kth.se/broker>
- LifeLine: <http://www.gamespot.com/ps2/action/operatorsside/review.html>
- Traum, D., Marsella, S. and Gratch, J.: Emotion and dialogue in the MRE virtual humans. Proceedings of Workshop on Affective Dialogue Systems. Berlin: Springer: Lecture Notes in AI, 3068, 2004, 117-127.