

Face-to-Face Conversation with First-Generation Ancestors. Current Reality and Future Prospects

NIELS OLE BERNSEN AND LAILA DYBKJÆR

NISLab, Odense, Denmark

This paper addresses current realities and future prospects of getting to the core of immersion in the past, which, we argue, is to meet past people in person and engage them in conversation. We have built a system which enables children and adolescents to meet and have quasi-natural interactive conversation with a familiar person from the past. Based on description of the system for a cultural heritage audience, including results of a user test of the final prototype, we argue that first-generation ancestor technology for immersion in the past is already there for installation at cultural heritage sites. The paper finally discusses some main challenges to achieving fully immersive past reality comparable to meeting our contemporaries from different cultures face-to-face.

Categories and Subject Descriptors: H5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Artificial, augmented, and virtual realities*; H5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Natural language; Voice I/O; Interaction styles; Evaluation/methods*

General Terms: Natural interaction, multimodal systems, interactive cultural heritage

Additional Key Words and Phrases: Entertainment, education, edutainment, animated conversational characters, persons of the past

1. INTRODUCTION

Globalisation and airline de-regulation provide the opportunity to travel cheaply to virtually any point on the planet and we all feel the pull to go somewhere exotic relative to our habitat. The really rich even go into space. Sketchily, the pull comes from a desire to experience, first-hand, different climates, night skies, ecologies, landscapes and cityscapes, material culture highlights world-wide, food, and social environments. Everyone going abroad gets it all by just going. However, maybe the authors are not the only ones missing something essential, namely, to *meet the people* and get to know them in person on our own. This is often harder to achieve because of language differences and lack of opportunity. Still, though difficult, it is possible for those who work hard enough.

Although so basic as to escape notice, globe-trotting is limited by something far more absolute, i.e., that everything and everyone we see, hear, taste, smell, touch or meet in conversation is *contemporary*. The few remaining hunter-gatherers we might encounter may appear backwards to us but they are still our contemporaries. Familiar from the literature and movies, time travelling remains mankind's elusive dream. If time travels were possible, we might well decide to ditch our internal globe-trotter, stay where we are, and go back to our home-town at the time of our ancestors with a trembling desire to meet them. Why would we do that? Firstly, because this is personal, because we *are* our personal history and that of our family and our tribe; and secondly because, arguably, the

historical world in general is as fascinating as the contemporary world, the primary difference between them being one of accessibility. The gold standard of world accessibility, which might be called the *accessibility principle*, is *immersion*, it's being there and experiencing what's there with all our senses and sensibilities, and communicating with what's there using all our cognitive and empathic abilities. Given that, and despite the often missing personal encounters with people from other cultures, globe-trotting wins most of the time. All we have are remains of the past and all we have been able to do is to vest these with engaging drama to make up for what they lack in immersion. That's why the cultural heritage community spends much effort making the past more "alive" than it used to be, using detailed reconstruction, "sound and light" shows, people dressed up as from the past to perform their crafts, etc. Recently, these efforts are being complemented by virtual and augmented reality presence in past environments, abandoning the real remains for the illusion of actually being in the past or combining both.

This paper addresses current realities and future prospects of getting to the core of immersion in the past, which, according to the thrust of the argument above, is to meet past people in person and engage them in conversation. We have built a system which enables young people to have quasi-natural interactive conversation with a person from the past. Based on description of the system, including user test results, we argue that first-generation core past immersion is already there for installation at cultural heritage sites. We then discuss what it will take to achieve immersive past reality comparable to meeting contemporaries from different cultures face-to-face. Section 2 describes the experience of meeting our past character. Section 3 looks at the intellectual background for the system and describes the state of the art in terms of the application paradigms which converge in its construction. Section 4 presents results of a test of the final system prototype with representative users. Section 5 argues that the technology is ready for installation but that technology improvement on virtually all fronts would improve user experience and immersion. Section 6 concludes the paper.

2. CONVERSATION WITH HANS CHRISTIAN ANDERSEN

2.1 An Impressionistic View

Imagine that you go visit the Hans Christian Andersen Museum in his hometown, Odense in Denmark, and discover that the fairytale writer is back in person and happy to have conversation with you in English. Following the accessibility principle (Section 1), you skip the museum's Andersen memorabilia, the Andersen films shown at regular intervals,

the museum's new click-and-get-a-story graphical display, and the gift shop to go talk to the man himself. You might meet him in his study as shown in Figure 2.1.

In the course of conversation, you discover that 55-years old Andersen is somewhat confused about whether he is dead or alive (again?) and wonders if, nevertheless, he might



Figure 2.1. Andersen in his study. A picture of Thumbelina and a paper clipping of Andersen's are close to the camera.

finally find himself the wife he sadly missed in his life between 1805 and 1875. He realises that his memory is not what it was. For instance, although he does remember a goodly number of titles of his fairytales and snippets from some of them, he only remembers details of *The Little Mermaid*, *The Ugly Duckling* and *The Princess and the Pea*. Similarly, while he remembers a good deal of his childhood in Odense – his family, the games he

played, his school years and his hometown at the time – he remembers rather little of what happened after he left Odense at the age of 14 to go to Copenhagen to become famous. Most of what he remembers from adolescence and adulthood concerns his poverty in Copenhagen, the benefactors who helped him, his failures as would-be ballet dancer and actor, the travels to Rome and his failures to marry any of the women he loved. You also find that Andersen's personality is perfectly intact. He is interested in new technology, such as trains and photography, and loves travelling and beer. He is self-indulgent, proud of his literary achievements, slightly too satisfied to have met royalty and famous artists of his time, curious about what he has heard about his 200 years anniversary in 2005 to which he wasn't invited, which offends him no end, and *very* sensitive, to his looks, his height which, at 1.85, was 25 cm. above the male average at his time, his shoe size of 48, and more. His neurotic *Angst* of dogs and fires transpires as well.

Sensitivity is reflected in emotion, and Andersen turns sentimental when asked about his poor mother the washerwoman at Odense river drinking *Schnapps* to keep herself warm, or his grandfather who went insane and was put in an asylum; happy when extolling own achievements (only the Bible had more translations than my fairytales etc.); and angry when offended, which is easily done by just asking about his shoe size or listening to his reaction to suggestions in recent literature that he was homosexual. He is curious to know your name, age, gender, and nationality as well as about the games you like to play,

confiding that, at his age, he is mostly interested in making paper clippings. Hosting you in his study which is on display in Copenhagen and which we copied with various modifications, Andersen is happy to tell about the pictures on the walls and other objects, such as the women in his life, a picture of a locomotive, many fairytale illustrations, or his faithful travel bag. To make him tell a story, you hand-point to the object and say what you would like to know. Andersen turns towards the object and then turns back towards you before telling a story. He refuses, somewhat brusquely, to show you the other rooms in his apartment but you can enjoy the view from his windows onto the medieval centre of Copenhagen. To take a good look at all parts of Andersen's study, you can control his locomotion as well as ten or so different virtual camera angles onto the room and Andersen himself. When talking, Andersen makes hand and arm gestures and modestly changes his facial expression to reflect his current emotional state.

The modifications we made to Andersen's real study are: (i) insertion of pictures of



Figure 2.2. Andersen gesturing in front of his writing desk, back to a mediaeval street in Copenhagen. Pictures show the Swedish singer Jenny Lind, Coliseum in Rome and the ugly duckling, among other things.

women in his life, of his benefactors, and from his fairytales; (ii) addition of large double doors meant to open onto a fairytale world in which you could meet Cloddy Hans, Thumbelina and other characters; (iii) replacement of Andersen's sitting writing desk with the standing desk of a famous Dane contemporary, the philosopher Søren Kierkegaard (1813-1855). Kierkegaard's desk (Figure 2.2) is on display at the Copenhagen City Museum.

This was done because our animation project partner objected to having to animate Andersen performing the perfectly ordinary processes of sitting down and standing up, arguing that this was a waste of resources; and (iv) addition of more floor space for Andersen to walk around while working on his own or talking to visitors. All these modifications of the photorealistic virtual reality graphics representing Andersen's study are ad hoc and reversible, and reflect the fact that strict historical similitude was not key in this particular project. Few have noticed the errors we made in the modification process, but one trains expert did point out that the picture we put up of an old locomotive shows one that was only put into service in the 1890s, i.e., well after Andersen's death.

2.2 Andersen in Conversation

Andersen is polite and attentive, happy to tell stories about his life and fairytales, himself and his study. He greets you and takes an early opportunity to ask questions about you in order to get to know his visitor. You can always take the initiative and change the topic of conversation. If you do that early on, he will follow your lead and postpone his remaining questions about you until an opportunity turns up later in the conversation. If you do not take the lead, he will do so by offering to talk about one of the domains he is knowledgeable about, such as his fairytales, taking one step into the domain at a time and always offering to address a particular class of topics or tell a story before actually telling the story. He will let you know in some way early on that there is still a lot he cannot remember from what he once knew. He works hard to remember more and hopes that you will re-visit him at some later time when his memory has improved. If, at some point, you are at a loss as to what to say to continue the conversation, he might offer to tell you in more detail about which topics he remembers within some particular domain. Or you might ask him about that, asking, e.g., what he knows about or would like to talk about. In response, he will not present a lengthy oral list of topics but simply illustrate where you might go together in conversation. Like so much else of Andersen's conversational contributions, this information is context-dependent so that, e.g., early information about what he knows about will be pretty general and explain the domains of knowledge he has, whereas when, later on, conversation may have entered the domain of his literary works and, in particular, his fairytales, he would explain what he currently knows about these.

Like humans, Andersen never tells you the same story twice at his own initiative. If asked the same question twice, he will typically modify and re-phrase his response the second time around. Humans often do that and, like humans, Andersen is sensitive to your expressed interests in his stories. For instance, if you tell him that you are not interested in his fairytales, he will be slightly saddened but still respectfully refrain from offering fairytale stories later on during conversation. Some topics, like that of his beloved grandfather, are too emotional for him to pursue at length and this he will admit. Otherwise, he continues conversation in spite of any emotional ups and downs caused by the topics addressed, that is, unless you persist in offending him. Following mild reproaches, such as "There was no need to say that" and subsequent attempts to steer the conversation into smoother waters, his annoyance will have grown to anger and he will eventually tell you that he must end the conversation because he has other important things to attend to.

Like humans, Andersen is sensitive to the semantics and pragmatics of the diversity of little interjections that his visitors might produce in conversation. For instance, if he

tells a particularly rich story, such as the one about his father who enrolled in the army to take part in the Napoleonic wars, and you respond by saying, e.g., “Cool!” or “That’s amazing!” – he will acknowledge by saying, e.g., “It is, isn’t it” or similar before carrying on. If you say instead, as humans do to confirm to the speaker that they are listening but without expecting explicit acknowledgement, e.g., “Yes”, “I see” or “No”, Andersen will not acknowledge but continue the conversation. But of course, if he asks you a yes/no question and you reply by “Yes” or “No” or their linguistic equivalents in context, then he will draw the full conversational implications of the choice you just expressed.

Since there is still a lot that he does not know but which his real-but-deceased predecessor knew, and since Andersen-resurrected has no practical way of telling you *exactly* what he does and doesn’t know, it is probably inevitable that you will step outside the bounds of his current knowledge at some point. For instance, it is common in conversation to address some topics in more depth than others, e.g., because we get curious to know more about what our interlocutor just said. So if you ask Andersen, e.g., if he wrote any scary fairytales and he responds by mentioning some of those, you might want to know *why* he wrote scary fairytales in the first place. Andersen will reply that he does not know why and then carry on the conversation.

In most cases, however, when you overstep his knowledge bounds, he will plainly not understand what you say, and this, just as plainly, is *not* what humans do. In most cases, a human interlocutor would understand what was said but realise not to know or remember the answer. Still, we sometimes fail to *hear* what was said or we hear it all right but fail to understand what was *meant*. In such cases, we typically say “Say again?” or use one of the linguistic equivalents for making the other repeat or re-phrase the previous utterance. Andersen does likewise. So if he thinks that he failed to *hear* what was said, he asks the interlocutor to repeat, and if he thinks he failed to *get the point* he also asks for repetition, the difference being that, in the first case, he says that he did not hear what was said, in the other, that he did not understand. If the same happens the next time around and again, Andersen realises that this is out of knowledge bounds and tries to gently steer the conversation towards topics he actually knows about. For instance, the second time he might ask you to re-phrase, the third he might say that this is probably about something he is not yet familiar with, and the fourth time he might suggest a change of topic to something he knows about, possibly offering to explain what he knows about as well.

Figure 2.3 shows a transcription which combines two recorded conversations Andersen made with children during our second-prototype user test. The figure illustrates most of the issues described above.

Interlocutor	Hans Christian Andersen
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?
i'm 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 because 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 don't 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 moonwalk	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 mother's 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: feather-Pen	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 kilometre 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 2.3. Transcribed conversation with Andersen.

In fact, the only items introduced above that are absent from the conversation in Figure 2.3 are: Andersen's handling of interjections by either acknowledging or ignoring them, depending on their contents; Andersen refusing to talk about a particularly emotive topic any more; his stopping of the conversation due to overwhelming anger; and his handling of 'why' questions.

There is a general question to be asked about the conversation in Figure 2.3. The question is not easy to uniquely phrase but some phrasings might be: Is the conversation “natural”? Is it in the style of human-human conversation? Is it convincing given the domains and topics, the status and role of interlocutors, the system’s edutainment purpose, and the ever-changing discourse context? The importance of answering this question is that the answer will determine the extent to which we may claim to have achieved conversational technology which could be ported to the building of systems for conversation with virtually any individual from history. Before addressing the question in Section 4, we provide some background information in the following section.

3. BACKGROUND, THEORY AND STATE OF THE ART

A system like the Andersen system is not constructed out of thin air but based on many different strands of background and theory which we will now look at.

3.1 Motivation and Rationale

The idea to build what became Andersen came from our experience from building increasingly complex spoken dialogue systems since 1991. A *spoken dialogue system* is not multimodal like the Andersen system but uses speech input and speech output to enable spoken dialogue with users. Today, spoken dialogue systems proliferate in different languages and typically enable users to accomplish a particular, well-defined task through spoken dialogue, such as obtaining flight, traffic, or weather information, making hotel reservation, negotiating destination with a spoken navigation system in the car, or being connected through an automated switchboard, cf. the examples in [Delgado and Araki 2005]. Our idea was to break with this *task-oriented system* paradigm and demonstrate a system able to conduct “real” spoken conversation like the conversations humans have when there is no immediate task to be solved and no work to be done but, rather, serendipitous time to talk to other people to get to know them better, exchange views about anything of common interest, etc. We call this new paradigm *domain-oriented conversation* because the goal is to develop human-style conversation about particular domains.

We knew that domain-oriented conversation would be technically hard to do, so the next step was to try to make sure not to aim too far beyond the state of the art in a single project. In particular, as we did not know whether domain-oriented conversation could be built at all, it was clear that our virtual human would have to be exposed to new interlocutors at relatively short-duration intervals. This happens in museums where visitors typically come to spend a limited amount of time wishing to experience all or most of what’s

on display. So we decided to build a famous person from our cultural heritage who should be able to conduct domain-oriented conversation for, say, 5-15 minutes. Moreover, people from the past are famous because of having done particular kinds of things, such as Ghandi who led India to independence, Newton who created the foundations of mechanics, or Andersen who wrote nearly 200 fairytales which still enthuse children and adults across the world. By focusing on one such character, we had a natural delimitation of the domains of conversation which visitors would expect to address.

Although spoken conversation has been studied for decades by linguists, conversational analysts, anthropologists, psychologists and others, no analysed corpus shows conversation between a past historical character and people from the present. Yet we would

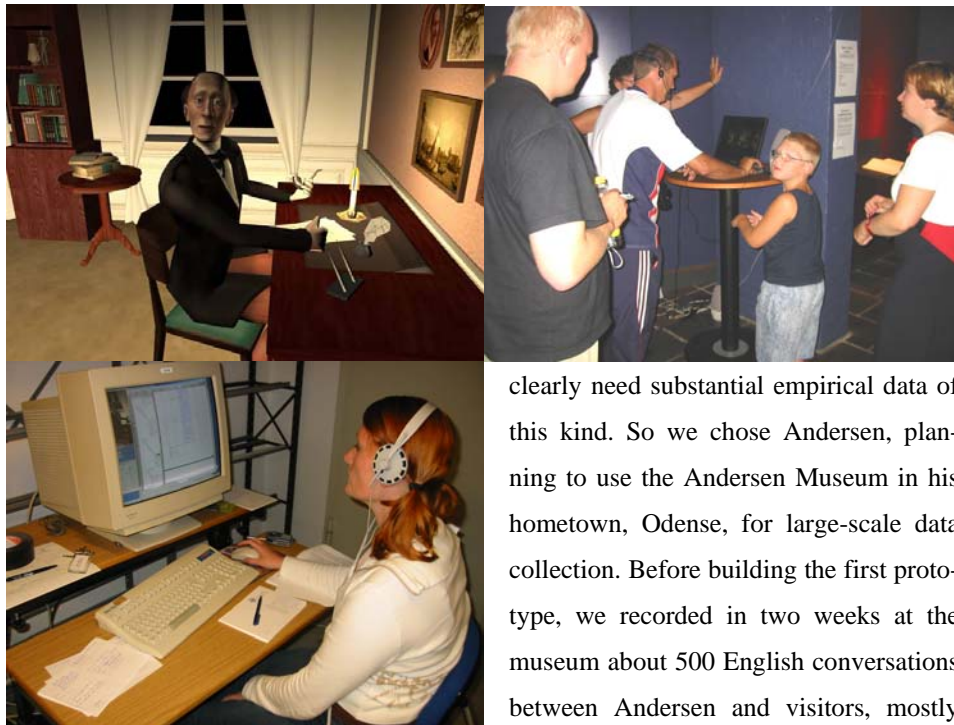


Figure 3.1. Andersen simulated in the museum.

(a) Andersen, (b) a crowd of visitors, (c) a wizard.

clearly need substantial empirical data of this kind. So we chose Andersen, planning to use the Andersen Museum in his hometown, Odense, for large-scale data collection. Before building the first prototype, we recorded in two weeks at the museum about 500 English conversations between Andersen and visitors, mostly children, from 29 countries. Andersen's conversational contributions were simu-

lated by human "wizards" sitting in the museum's basement and following the detailed system output specification in carrying out conversations via wireless connection to a laptop in the museum. The simulators are called wizards because the simulation method is known as Wizard of Oz [Bernsen et al. 1998]. The rich data generated, and other corpora from the project are described in [Bernsen et al. 2006] and available from [<http://www.niceproject.com/data/>]. Figure 3.1a shows the early, modestly animated Andersen used in the simulation. He sits on the laptop screen surrounded by museum visi-

tors in Figure 3.1b. Figure 3.1c shows a wizard. The on-screen hypertext version of Andersen's output specification enables the wizard to use max. two mouse clicks to locate and read aloud or otherwise produce specified output in context whatever the user says in conversation. As one of the wizards was female, we used voice distortion to avoid the oddity of Andersen speaking with a female voice.

The choice of Andersen suggested aiming at young users rather than adults, so we chose the target user group of 10-18-year olds. The risk inherent to that choice was that current speech recognition is less effective for children's voices, which implied a need to collect large amounts of children's speech data to improve speech recognition. On the positive side we assumed that, compared to adults, children might be more tolerant to infelicities in the system's conversation as long as interaction is fun and educational.

How should users interact? Spoken conversation was crucial, of course, but speech is only one modality of information exchange among others used in conversation. Natural interactive conversation involves two-way exchanges using audiovisual speech, i.e., synchronised speech, lip and mouth movements, gesture of different types, facial and gaze expression, head and body posture, and possibly object handling and locomotion as part of conversation. In general, from a system development point of view, it is currently much harder to achieve *recognition and understanding* of these modalities than to *generate* them. Today's computer game and film animators are extremely skilled at generating humans and their communicative behaviours in life-like, photorealistic 3D graphics, and good-quality speech synthesis is available as well. Yet the large majority of these fantastic products are still presented to passive spectators in the cinema or are interactively operated in primitive fashion by using a mouse or controller and a keyboard. So while we could aim for reasonable quality of natural interactive output graphics and speech, we had to be prudent in selecting the input modalities. We chose *spontaneous input speech*, i.e., the user can say whatever comes to mind during conversation, unhampered by having to memorise command-and-control keywords or pre-designed phrases or sentences that the system can recognise and understand; and *2D pointing gesture*. Pointing gesture is one among several different types of gesture used by people in conversation and is a useful complement to speech for spatial reference. While human gesture, pointing or otherwise, is typically 3D, i.e., performed in 3D space, 2D gesture is performed onto a surface, such as a tactile screen. Thus, although behaving natural interactively in conversation himself, Andersen is essentially blind, being sensitive only to speech and deictic touch. However, he emulates being able to see the objects the users touch in his study.

3.2 Theory

Non-work-or-task-related conversation comes in many different varieties, including confessionals, life-crisis talks, small-talk, confrontations, seductions, fun-and-nonsense exchanges, big boss/employee conversations, interviews with actors looking back, learning-from-the-other talks, gossiping, talks about life, death and all, and grandparent/grandchild conversations. We chose to develop for a variety of the kind of conversation we have when we meet a stranger, start to talk and maybe end up becoming friends [Gabor 2001], taking into account the age difference between Andersen and his target users as well as the assumption that museum visitors are there because they wish to learn about Andersen.

What characterises conversations of this kind is, we specified: (1) Initially, the interlocutors search for *common ground*, such as shared background, interests, or knowledge. (2) The conversation is replete with *story-telling*, primarily by Andersen but also by the user. Andersen is a partner in conversation, not a story-telling machine which just needs to be triggered to go, and the user has stories that he wants to hear, too. Conversation is (3) *rhapsodic*, i.e., highly tolerant to digression, the partners having no need to exhaust some domain or topic before addressing another, so they may change domain or topic at any time. (4) Conversation is *friendly and polite* by default and a main goal for Andersen is that the visitor will find it worthwhile and edutaining, and wish to come back at some later time. (5) *Initiative is fully mixed*, i.e., just as Andersen is no story-telling machine driven by request, he must not be the one who drives the conversation forward. He should *both* be capable of the latter with timid visitors who don't push to take the initiative themselves *and* able to have conversation with visitors who arrive with a loaded agenda of issues they wish to talk about. The museum simulations (Section 3.1) show that children differ very much in this respect and rather independently of age, gender and English-speaking skills. (6) Andersen should master *turn-taking* to let the visitor have the initiative as long as s/he wants to *and* be able to take the initiative when the user does not wish to keep it. Mastering turn-taking also means being able to distinguish between contributions which (i) imply that the user takes the initiative, (ii) require acknowledgement but do not imply change of initiative, and (iii) merely signal that the user is following what Andersen is saying and does not expect acknowledgement.

Note that the above, high-level characterisation of a particular type of conversation is not a complete theory of Andersen's conversation. It only highlights some main challenges addressed by the system. A complete theory would explain a wealth of other, and typically more specific mechanisms which guide the construction of Andersen's conversation, such as that his contributions should be coherent and consistent, a question must

be followed by an answer, the answer must address the point made in the question, and much else, some of which is illustrated in the comments to the example in Figure 2.4.

3.3 State of the Art

In state-of-the-art terms, the Andersen system is at the cross-section of progress in (1) spoken dialogue systems towards spoken entertainment and education, (2) animated conversational characters, (3) spoken computer games, (4) multimodal systems, and (5) interactive cultural heritage technology. The relationship between these five strands is easily summarised. (1) through (4) represent *research and technology convergence* towards natural interactive systems which are inherently multimodal. (5) is a key application domain because the past is half of everybody's life, it's not just whence we came considered as a topic of detached study, but who we are. *Natural interactive systems* represent the long-term goal of building systems that communicate in all the same ways in which people communicate with one another, and about all the same things. This is such an obvious convergence goal for hitherto separate research fields that convergence had to happen.

Spoken dialogue systems have begun to transcend their traditional task- and work-orientation [Bernsen et al. 1998] to address a wider potential. Although we know of no domain-oriented system other than Andersen, spoken tutoring systems are emerging, particularly in the US [Roberts 2000, Clark 2001, Litman and Silliman 2004]. Systems may include a talking head, teach skills in a dedicated virtual environment or support learning of abstract subjects, such as qualitative physics. Closer to Andersen is the work done for the US military on animated conversational characters for training field commanders in handling potential civil unrest and negotiation with non-military actors, such as doctors [Hill et al. 2003, Traum et al. 2005]. Common to these developments is that the new system generation focus on improving system reasoning capabilities and sense of situation and context in order to improve situated spoken communication.

Animated conversational characters emerged as an interactive paradigm in the 1990s, combining interests in graphics rendering of communicating faces, cartoon figures and photorealistic embodied characters, intelligent information presentation [Maybury 1993], and affective computing [Picard 1998]. Spoken dialogue was there early on but tended to differ from work done in the spoken dialogue systems community by focusing less on core task-oriented functionality and more on exploratory issues, such as turn-taking in the REA (Real-Estate Agent) system [Cassell et al. 2000]. Today, convergence between spoken dialogue and animated characters is well advanced and characters are being used for many different purposes, such as learning, e-commerce or as personal assistants.

Spoken computer games in the sense of computer games enabling spoken dialogue with the computer game characters is a relatively recent development. The CHIMP project had goals similar to ours, i.e., to enable children to communicate with animated characters using speech and 2D gesture in a game application [Narayanan et al. 1999]. While Andersen may be viewed as a conversational spoken computer game, the games industry has begun by including speech in a different way, such as for keyword-based command and control of computer game characters' actions. Konami's LifeLine [<http://www.gamespot.com/ps2/action/operatorsside/review.html>] offers spoken dialogue control and hence achieves an interactive spoken computer game. Early systems on the market do not appear to have received enthusiastic acclaim, partly because the players have difficulty remembering the words they are allowed to use and partly because the speech recognition technology tends to be too fragile. Yet it seems obvious that ongoing technology convergence towards natural interactive systems could accelerate progress by harnessing the enormous potential of current computer games technologies.

The term *multimodal systems*, if not conceptually then in actual fact, is a catch-all for systems which break with the paradigm of graphical user interfaces (GUIs) through which we interact using mouse, keyboard and screen text and graphics. Conceptually, GUI-based systems are multimodal systems as well [Bernsen 1994, 2002] but the importance of the rapidly growing multimodal systems field is to explore, improve and combine processing of the many *non-GUI* modalities required for exchanging information with natural interactive systems. For instance, to make Andersen *see* his interlocutor, we need image processing for identifying and tracking the user and for recognising the user's facial expression, 3D gestures, eye gaze, head and body posture, physical actions and locomotion. Moreover, to make Andersen *understand* multimodal communication, such as a user's expression of emotion through combined speech prosody, facial expression, and gaze, we need powerful mechanisms for real-time *multimodal fusion* of information represented in those modalities. Andersen already performs limited multimodal fusion of speech and 2D gesture input, but this is merely the tip of a major and complex challenge to the development of more natural interactive systems [Martin et al. 2006].

Jointly, the four strands show that many different communities are converging towards building more human-like conversational characters which could eventually fulfil our desire to meet past people in person. In the *cultural heritage* field, the quest for more immersive user experience has so far been pursued primarily through advanced (output) information representation, often using virtual reality technology, whereas new interactive input modalities are only slowly being introduced. A precursor of Andersen's is the

Swedish August system which, at an exhibition in Stockholm, used the expressive talking face of author August Strindberg to have dialogue with visitors about restaurants in Stockholm, the Royal Technical University, himself and his work, and generally to create a fun and entertaining atmosphere [Gustafson et al. 1999]. Complementary to Andersen, another direction of work is computer vision-based augmented reality systems which track the user's pointing hand, enabling visitors to get information about artefacts on virtual display [Malerczyk et al. 2005]. New input modalities are also appearing in cultural heritage search, providing new ways to search archives for professionals and ordinary users, including the disabled, for instance via haptic interfaces, spoken question-answering, and image processing technologies for search in still images and video data.

4. WHAT THE USERS THINK

We have tested two successive Andersen prototype versions, PT1 and PT2, with representative users in the lab. In the PT1 user test all system components except the speech recogniser were running. Speech recognition was simulated by wizards who typed what the users said into the system. The test showed that we seemed to be on the right track with the story-telling get-to-know-each-other approach to conversation (Section 3.2) and that Andersen's conversational skills needed significant improvement in flexibility and context-awareness [Bernsen and Dybkjær 2004]. This was expected because the purpose of PT1 [Bernsen et al. 2004] was to demonstrate system proof-of-concept rather than any substantial level of sophistication in the individual components. The young users were more appreciative of Andersen's stories about himself and his life than of what he told about his fairytales. As 17 of the 18 test users were Danish school kids, we explain this observation by the fact that Danish kids have good knowledge about Andersen's fairytales and are therefore unlikely to learn much about these from a revived Andersen who only remembers the basics on his fairytales. The children were less knowledgeable about Andersen's life and personality and felt that they learned a lot about these domains from talking to him. All but a single user felt that they had little difficulty speaking English with him. However, many were not familiar with the English titles of his fairytales and several felt that it was sometimes hard to produce the real-time conversational contributions needed to take and keep initiative in conversation.

4.1 Second Prototype User Test

The second prototype, PT2, was tested with 13 Danish school kids. All system components were running and the test was done in much the same way as the PT1 test, with two different test conditions, the first one based on free-style interaction, the second on a

handout listing 11 generally phrased proposals for what the user could try to find out about Andersen's domains, make him do, or explain to him [Bernsen and Dybkjær 2005]. The conversation in Figure 2.4 was recorded in the second test condition. Figure 4.1 shows a PT2 user. Each complete test session took 60-75 minutes, including post-test



Figure 4.1. A user in action.

interview. One important difference between the two tests was that in the PT1 test, all users had a mouse for gesturing onto objects in Andersen's study and half of the users also had the option of using a tactile screen instead. The result was that the users, most of whom play computer games from several to many hours a week, tended to use the mouse a lot, clicking on

everything in sight with little or no regard for what they were talking to Andersen about at the time. This is very far from the way most people use (3D) pointing gesture in human-human conversation. Another consequence was that the PT1 post-test interviews showed that the users who tried it had a highly ambivalent attitude to the value of using a tactile screen. We therefore decided to give all users a tactile screen and no mouse in the PT2 test.

4.2 User Interviews

We asked a total of 31 questions in the PT2 post-test interviews. Eight initial questions addressed the user's identity, background, computer game experience and experience in talking to computers. We had no substantial input on the final (standard) question about whether the user had any other comments. This leaves 22 questions about the system itself and how it was to interact with it, which are shown in abbreviated form in Figure 4.2 which presents a quantified summary of the interview results. Bernsen and Dybkjær [2005] explain the method used in rating user answers and discuss the answers to each individual question. We now look at the larger picture emerging from the interviews.

Pointing Gesture and Speech. As regards *pointing gesture input*, PT2 users were quite positive about using the tactile screen. Moreover, also in marked contrast to the way users did mouse-pointing with PT1, the transcribed PT2 test conversations show that users (i) always pointed in a contextually meaningful way, i.e., by pointing to some object and waiting for Andersen's response, and (ii) only in a couple of cases spoke in a semantically independent manner while pointing, like the girl who said, trying to offend Ander-

sen, “I don’t like your hair” whilst pointing to his feather pen. We conclude that not only do the users *find* it natural to use the tactile screen for pointing while speaking, they also *behave* perfectly naturally when doing 2D gesture pointing during conversation.

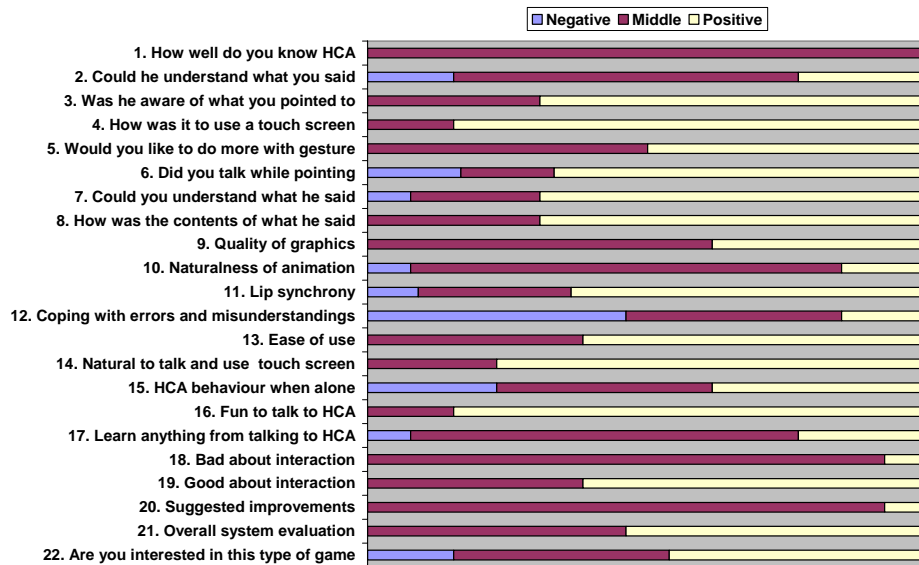


Figure 4.2. Summary of interview results from the second prototype user test. HCA is Andersen.

Graphics and Animation. The quality of the graphics and of lip synchronisation was viewed as reasonably good. The naturalness of animation, on the other hand, received critical comments from most users, the key target being Andersen’s walk, weird antics, and bumping into walls. The antics problem is one of script contents design, while the bumping into walls shows that Andersen has no motion planning. No user commented on Andersen’s facial display of emotion which, admittedly, could have been made more visible. All of these problems can be remedied by more substantial effort in graphics and animation. The system’s graphics and animation is described in [Corradini et al. 2004].

Speech Understanding. We found, as in the PT1 interviews, that Danish kids understand spoken English amazingly well. The answers concerning Andersen’s understanding and his ability to cope with errors and misunderstanding reflect several issues of crucial importance to conversation with people from the past: (1) Andersen does not understand words outside his 2000 word form *vocabulary*. (2) his speech recognition is *not top-tuned* to children’s voices, so even well-pronounced within-vocabulary words may be mis-recognised. (3) recognition often fails of *mispronounced and indistinctly pronounced* words and phrases. (4) recognition has difficulty with *disfluent and noisy* input, such as hesitations, false starts like “Could you say ... tell me about your school please”, and background noise. (5) *ungrammatical* input stands less chance of being correctly recog-

nised than grammatical input. (6) the recogniser has little chance of capturing *lengthy and verbose* input, such as +20 words input. Finally (7), the recogniser performs worse, not better, if, upon having been misunderstood, the user changes *speech manner* and speaks more slowly or more loudly to be understood. With children having English as second language, the frequency of mispronounced, indistinctly pronounced, hesitant and disfluent, and ungrammatical input is higher than with native English-speaking children.

Let us look at how to counter these problems. As for (1) *vocabulary*, we could probably augment the system's vocabulary to 4-5000 words or more without significant damage to recognition, if we collect much more data for language modelling. Strong language modelling is essential to good speech recognition because it provides the recogniser with a good statistical model of which words are most likely to follow a certain word. Many cultural heritage applications could be built with a vocabulary of 4-5000 words. (2) *better recogniser tuning* is a matter of using more children's speech data. This is a sparse resource today. Alternatively, the cultural heritage application could target adults, reducing the problem. The *pronunciation, disfluency, ungrammaticality, verbosity and speech manner* problems (3, 4, 5, 6 and 7) can be significantly reduced by (i) instructing users in how to speak to the system, (ii) aiming at native speakers, and (iii) having adult users. Research indicates that adults are more flexible than children in revising their speech strategies in case of misrecognition and misunderstanding, e.g., through simpler language and shorter utterances [Bell 2000]. In summary, using existing technology and approaches, it is possible to obtain significantly higher scores on the crucial questions of understanding and error handling abilities than we did in the PT2 user test.

Fun and Learning. The users unambiguously found talking to Andersen to be fun, felt that they learned from the conversation, and were generally positive about the contents of the conversations. This is evidence of positive reception of a system of this kind. Note, however, that it is common to find *positive bias* in users of new technology, probably because its novelty elicits a positive response all by itself, independently of the system's merits as calmly evaluated aspect by aspect. Still, in a museum environment which includes an Andersen-style system, chances are that many children might evaluate their conversation with the character among their top experiences from visiting the museum.

Overall. On the issues of *what is good or bad and in need of improvement*, criticism not made earlier included inconsistency between the user's and Andersen's control of his locomotion, and between camera angle and his turning towards an object pointed at. These are real design problems of character autonomy vs. user character control. Also, Andersen should have more knowledge, better walk, less antics, improved understanding,

ask users more questions, and have improved prosody. Several praised his story-telling, “easy English” and good voice. The system was generally regarded as being easy to use.

In their *overall evaluation*, the users scored the system at 1.5 on a scale from 1.0 (great) through 2.0 (interesting) to 3.0 (rather negative). Ten users were interested in spoken computer games for some or all gaming purposes.

5. FUTURE CHALLENGES AND PERSPECTIVES

In this section we take stock of the technology, asking two questions: (1) What can it do for cultural heritage sites to give access to past people, their views of themselves and their times? (2) What are the challenges to creating more complete access to past people?

5.1 Is the Technology Ready?

We believe that Andersen technology is ready for *first-generation ancestor* development. The claim rests on three premises: (1) that, assuming good casting of the character and its virtual environment, historical accuracy, and a conversation model appropriate for the purpose, the technology enables 5-15 minutes of conversation with virtually any character and for most user groups. (2) that the limitations to the Andersen system which have been noted above can be remedied through state of the art methods and commercial-quality development. The result will be a character which understands *more* input *better* than Andersen, *knows* far more than he, *visually behaves* satisfactorily as a 3D graphics embodied, animated character, and is more *robust* than he so that there is no need for a technician on stand-by. Otherwise, the character will conduct conversation Andersen-style. (3) that it is believed that systems of this kind could provide significant enrichment of our access to the past in museums and other cultural heritage locations.

Some cultural heritage sites might want Andersen, but most would want different characters, like Newton, Ghandi, all past US presidents, Aristotle, or the ordinary Viking whom you might meet in his hometown in-between the long sea voyages that made Vikings both famous and infamous. These demands can be met by replacing Andersen *contents* by *different* contents without modifying the core technology. Note, however, that these are what is commonly called *content-rich* applications which require substantial resources for creating historically correct contents for the system to use.

The technology can be adapted in different ways for a given character and audience. 2D pointing gesture may not be needed. Some applications may not need fully mixed-initiative conversation but will rather be user-driven question-answering systems having substantial knowledge but offering less sophisticated spoken interaction. Others might take the visitor on a system-directed *conversational tour* but allow little user initiative

other than selecting among the stories offered. Compared to Andersen, all of these systems would be less complex. Yet another variation within the limitations of the technology would be *tutorial systems* for children in which the character asks questions and gives them problems to solve related to the character's historical role and achievements, other exhibits at the cultural heritage site, etc.

5.2 Future Challenges

Between first-generation ancestors and past people who are nearly as accessible as our contemporaries (Section 1) lies a huge research agenda which, to our knowledge, has not been developed into a roadmap complete with progress time estimates. Let's now look at challenges and perspectives for the next 10-15 years, including (1) historical research and character rendering, (2) improved user input, and (3) improved conversation.

Historical Veracity. In a cultural heritage context, historical veracity would seem mandatory for all or most ancestor applications. While this is of course primarily a matter of contents and applied historical research, interesting technical challenges seem to appear as well. Research into human history is commonly regarded as the least "scientific" among the historical sciences which also include, e.g., geology, astronomy and evolutionary biology. A cultural heritage ancestor will never be the final one, not just because new research results keep turning up but also because we change attitude towards past people and events, and our interests in the past change so that we want to focus on other things than did previous generations. This implies that, eventually, *many different* Andersen, Ghandis etc. will be developed, some of them probably very different from one another because of representing rather different interpretations of the past.

Ancestor-building probably will influence historical research. Consider Newton, for instance. Do we know the exact quality of his voice, the details of his manner of speaking, his vocabulary, grammar and dialect, his facial expression patterns, the frequency and manner of his gestures, his gait, temper, bluntness or politeness, and the extent to which he revealed emotions in conversation, his attitudes towards numerous issues beyond the core science, his world view and its differences from ours? Probably not to the extent we would wish were we to build him! Perhaps the sources could tell more about these characteristics if a special effort were made to look for them? Suppose that we know most of these things in detail about some well-documented person. To which extent should we strive to incorporate them in the ancestor? Doing so poses interesting and unsolved technical challenges of detailed modelling of a person's natural behaviour and thinking. Andersen certainly does not meet these requirements. We were happy to find a

male voice from AT&T which everybody agrees is suitable for a 55-years old gentleman telling stories, but his voice, language, facial expressions, gestures and gait probably are far from being authentic in addition to being rather primitive. Moreover, people change throughout their lives and these changes are reflected in the way they communicate and in what they communicate about. We opted for Andersen at 55 simply to render him in his “best age” and with lots of life, fairytales and experience to address in conversation.

Separating what is authentic from what’s not is a general issue in cultural heritage displays. In classical sculpture restoration, such as of Michelangelo’s damaged Pietá in St. Peters in Rome, one approach is to clearly mark all non-authentic additions so that it remains possible to distinguish authentic parts from reconstructed and possibly historically inaccurate parts. Should this approach be used for ancestor characters as well? The screen might, e.g., have more or less detailed iconography informing visitors what in the character’s current behaviour is authentic and what is not certified as being authentic.

Adding Input Modalities. As discussed in Section 3.1 Andersen is essentially blind. To overcome the blindness of historical characters who were not blind or hard-of-seeing at revival age, camera-based input and advanced image processing is needed. Let us look at some of the advantages. First, audiovisual speech recognition is an active research area today because indications are that recognition can be improved by combining speech signal recognition with mouth and lip information. Secondly, input image processing would enable replacement of 2D tactile screen pointing by 3D pointing input, making pointing fully natural in principle. Thirdly, the character would recognise and interpret additional types of gesture which, like pointing, are frequent in conversation, such as linguistic expression-like *emblems* (the V sign, thumbs-up, etc.) or rhythmic *baton* gestures used, i.a., to stress points made in speech [McNeill 1992]. Fourthly, image processing can accomplish a number of basic tasks, such as inform the character when a user arrives and leaves and how many users are currently visible, track each of them, and report the user’s body and gaze orientation. Fifthly, image processing could monitor the user’s face for expressions of emotion and attitude. Finally, image processing can be used to identify and track objects which form part of conversation, such as when the user points to an image on the (real or virtual) wall or a certain part of it, anachronistically demonstrates an iPod to the character, etc. In 15 years from now, we can expect to see all of this combined in research systems and some of it in commercial systems.

Part of the problem and future promise of adding more natural interactive input modalities to ancestors is that information in different modalities needs to be *combined* by the system in order to achieve understanding of some aspect of the user’s behaviour. This

is done in audio-visual speech recognition where the system's perception of what the user says is a joint product of the acoustic speech signal and the visually perceived mouth and lip movements. Combined speech and gesture, as in Andersen, is another example. For a third, consider the combination of information from user speech content, speech prosody, facial expression, and body posture in order to determine the user's current emotional or attitudinal state. Although this so-called *multimodal fusion* is becoming an active research area, the problems addressed are difficult ones both theoretically and technically, and our guess is that we will only see slow but steady progress over the next 15 years.

Improving Conversation. No matter how successful Andersen might appear in the conversation shown in Figure 2.4, and no matter how much we improve his speech recognition, vocabulary, language understanding, and domain knowledge representation using state of the art methods, he is still far from fully mastering the subtle ping-pong of situated spontaneous *spoken* human-human conversation. One of the things Andersen is missing is an *observation-based user model* which would allow him to build, gradually during conversation with a particular user, a model of the user's domain-related knowledge, preferences, communication skills, and more, and use the information dynamically in conversation. Andersen has little of that, such as listening to the user's preferences about what *not* to talk about, following which he refrains from entering those topics or domains at his own initiative later on. On-line user modelling has become an active research topic with implementations in task-oriented spoken dialogue systems [Bernsen 2003, Komatani et al. 2005]. Its value in domain-oriented conversation basically is to enable the character to *listen* just as well as the character can *express* herself or himself. Correspondingly, a character *self* component might handle all input to do with the character itself, its personality (knowledge, interests, emotions), physical appearance, and behaviour in communication. Andersen has several functions of this kind, one being responsible for updating his emotional state, another for answering questions about his personality and physical appearance. However, it is easy to think of a more unified, consistent and robust approach to character self-knowledge than this.

Another aspect which Andersen is (mostly) missing is the ability to *go in depth with a particular topic* of conversation no matter if the topic is first introduced by Andersen or the user. While it is relatively simple to make a character tell a story, the attached hook is the user's expectation to be able to explore the story further by asking why, when, where, who, and how questions. This is a hard problem. A related issue is to know when to *stop* going in depth. Currently, when discussing games the user likes to play, Andersen is prepared, if allowed by the user, to ask the user twice to mention and then explain a game

the user likes. This is, of course, an arbitrary and un-situated design decision which prevents Andersen from showing the subtle sense of situation that humans possess. Also, one of the moments in the simulated conversations recorded in the museum (Section 3.1) was when a child felt so immersed in conversation that it *volunteered* information to Andersen, such as “I live in the country myself”. This happens too infrequently in the PT2 user test conversations and, when it happens, Andersen fails to get it.

Small-talk has received considerable research attention [Cassell and Bickmore 2002]. Andersen steers into his knowledge domains right away, if the user lets him, with no small-talk – but should he? An advantage of Wizard of Oz is the possibility to explore design options beyond the specification upon which the wizards base their conversational contributions. A simple approach is to make the wizard *improvise* new domains of discourse, new styles of conversation etc. In the museum, the wizards improvised small-talk, among other things. Andersen would welcome a visitor as usual but, before launching his priority quest for user information, he made grandfather-style small-talk like “Are you on summer vacation?”, “Is it nice?”, “I suppose that you come visit me with your family?”, or talk about today’s weather. The wizards would only small-talk when they could tell that they were speaking to a child. We stopped the small-talk experiment when it became apparent that the children found it awkward to have small-talk with Andersen. An appropriate explanation for this finding still escapes us.

Finally, let’s return to the issue of character *knowledge* which has been treated so far as a resource that can be extended indefinitely through state of the art methods of (ontology-based) knowledge representation. Ideally, we would install a first version of ancestor X at cultural heritage site Y, record conversations with visitors, and then bootstrap X’s knowledge to include frequently addressed topics which X failed to deal with. We did some of that between PT1 and PT2. Workable as this solution is, it is also costly. It would be far easier if we could, e.g., harness the virtually infinite knowledge resources of the web! A colleague made an experiment of this kind in collaboration with a visiting student, the idea being to use existing freeware website Q&A (Question-Answering) systems for Andersen’s purposes. We have also taken steps to make Andersen *accessible world-wide*. Once robustly implemented, there is no reason why it should not be possible to access an ancestor for conversation over the Internet. Since web-based speech recognition technology is still fragile, our idea is to enable *typed input* conversation with ancestors who would reply in speech and graphics as usual. All that is needed is a typed-text interface whose detailed design is interesting but essentially straightforward.

6. CONCLUSION

Andersen demonstrates, we argue, that ancestor technology is sufficiently mature for cultural heritage installation in a range of languages and capable of enabling *first-generation* domain-oriented conversation with virtually anyone from the past. Assuming a speech recognition success rate of +85%, which would seem achievable today, and sufficient care for historical, aesthetic and technical design details, the system might become a successfully deployed world-first in demonstrating the perspectives in meeting and having conversation with people from the common history of mankind. Once installed, we recommend the bootstrapping approach for incrementally improving the ancestor's knowledge (Section 5.2). While still at the research stage, the recommended revision of the architecture of the Andersen system which takes into account the needs of multimodal interaction (Section 5.2) is feasible today for systems at Andersen's level of multimodal input complexity. Cultural heritage sites appear to represent a uniquely promising ambience for early installation of the technology because visitors only expect to spend limited time at each exhibit. This is very different from computer games which are expected to yield an average of, say, 30 hours of exciting and immersive interaction. We have tried to suggest a variety of uses to which the technology might be put, including character X – the Q&A expert, X - the story-telling tour guide, X – the history tutor, as well as fully mixed-initiative systems like Andersen. All systems can be furnished with typed text input as an alternative to speech input, enabling ubiquitous conversation with the ancestor.

The rest is large perspectives for access to the past comparable to meeting contemporary people from other cultures, and more research to improve all aspects of immersive access until we reach the point at which our favourite character from the past could truly say “I don't think that more is known about me”.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support by the European Commission's HLT (Human Language Technologies) Programme, Grant IST-2001-35293, for the NICE (Natural Interactive Communication for Edutainment) project in which the Andersen system was developed. Special thanks are due to Scansoft who developed basic speech recognition, Liquid Media who developed graphics and animation, LIMSI-CNRS who developed gesture processing, and colleagues at NISLab who worked on natural language understanding (Manish Mehta), the Character module (Marcela Charfuelàn), Response generation (Andrea Corradini), contents (Svend Kiilerich), web knowledge (Manish Mehta), knowledge base, typed interface and troubleshooting (Mykola Kolodnytsky), and more (Michel Genereux, Thomas K. Hansen, Holmer Hensen, Dymytro Kupkin, Torben Madsen).

REFERENCES

- BELL, L. 2000. *Linguistic Adaptation in Spoken and Multimodal Dialogue Systems*. Licenciate Thesis. Royal University of Technology, Speech, Music and Hearing, Stockholm, Sweden.
- BERNSEN, N. O. 2003. On-line user modelling in a mobile spoken dialogue system. In *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech)*, H. BOURLARD, Ed. Bonn: International Speech Communication Association (ISCA), Vol. I, 737-740.

- BERNSEN, N. O. 1994. Foundations of multimodal representations. A taxonomy of representational modalities. In *Interacting with Computers 6(4)*, 347-71.
- BERNSEN, N. O. 2002. Multimodality in language and speech systems - from theory to design support tool. In *Multimodality in Language and Speech Systems*, B. GRANSTRÖM, D. HOUSE AND I. KARLSSON, Eds. Dordrecht: Kluwer Academic Publishers, 93-148.
- BERNSEN, N. O., CHARFUELÂN, M., CORRADINI, A., DYBKJÆR, L., HANSEN, T., KIILERICH, S., KOLODNYTSKY, M., KUPKIN, D., AND MEHTA, M. 2004. First prototype of conversational H. C. Andersen. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI)*, M. F. COSTABILE, Ed. New York: Association for Computing Machinery (ACM), 458-461.
- BERNSEN, N. O., DYBKJÆR, H. AND DYBKJÆR, L. 1998. *Designing Interactive Speech Systems. From First Ideas to User Testing*. Springer Verlag.
- BERNSEN, N. O., AND DYBKJÆR, L. 2004. Evaluation of spoken multimodal conversation. In *Proceedings of The Sixth International Conference on Multimodal Interaction (ICMI)*, New York: Association for Computing Machinery (ACM), 38-45.
- BERNSEN, N. O., AND DYBKJÆR, L. 2005. User evaluation of conversational agent H. C. Andersen. In *Proceedings of the 9th European Conference on Speech Communication and Technology (Eurospeech)*, Lisbon, Portugal, 2473-2476.
- BERNSEN, N. O., DYBKJÆR, L., AND KIILERICH, S. 2006. H. C. Andersen conversation corpus. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, Genova, Italy, 505-510.
- CASSELL, J., AND BICKMORE, T. 2002. Negotiated collusion: Modeling social language and its interpersonal effects in intelligent agents. *User-Modeling and Adaptive Interfaces 12*, 1-44.
- CASSELL, J., SULLIVAN, J., PREVOST, S., AND CHURCHILL, E. 2000. *Embodied Conversational Agents*. Cambridge, MA: MIT Press.
- CLARK, B., FRY, J., GINZTON, M., PETERS, S., PON-BARRY, H., AND THOMSEN-GRAY, Z. 2001. A multimodal intelligent tutoring system for shipboard damage control. In *Proceedings of International Workshop on Information Presentation and Multimodal Dialogue (IPNMD-2001)*, Verona, 121-125.
- CORRADINI, A., BERNSEN, N. O., FREDRIKSSON, M., JOHANNESON, L., KÖNIGSMANN, J., AND MEHTA, M. 2004. Towards believable behavior generation for embodied conversational agents. In *Proceedings of the International Conference on Computational Science (ICCS), Workshop on Interactive Visualisation and Interaction Technologies (IV&IT)*, Heidelberg: Springer Verlag: Lecture Notes in Computer Science, Vol. 3038, 946-953.
- DELGADO, R. L.-C., AND ARAKI, M. 2005. *Spoken, Multilingual and Multimodal Dialogue Systems. Development and Assessment*. Hoboken, NJ: John Wiley.
- GABOR, D. 2001. *How to Start a Conversation and Make Friends*. London: Sheldon Press.
- GUSTAFSON, J., LINDBERG, N., AND LUNDEBERG, M. 1999. The August spoken dialogue system. In *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech)*, Budapest, Hungary, 1151-1154.
- HILL, R.W., GRATCH, J., MARSELLA, S., RICKEL, J., SWARTOUT, W., AND TRAUM, D. 2003. Virtual humans in the mission rehearsal exercise system. *Künstliche Intelligenz, Special issue on Embodied Conversational Agents, AI 4(03)*, 5-10.
- KOMATANI, K., UENO, S., KAWAHARA, T. AND OKUNO, H. G. 2005. User modeling in spoken dialogue systems to generate flexible guidance. *User Modeling and User-Adapted Interaction 15(1-2)*, 169-183.
- LITMAN, D., AND SILLIMAN, S. 2004. IT-SPOKE: An intelligent tutoring spoken dialogue system. In *Proceedings of the Human Language Technology Conference (HLT/NAACL), Companion Proceedings*, Boston, MA, 233-236.
- MALERCZYK, C., DÄHNE, P., AND SCHNAIDER, M. 2005. Exploring digitized artworks by pointing posture recognition. In *Proceedings of the 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, Pisa, Eurographics Association, 113-119.
- MARTIN, J.-C., BUISINE, S., PITEL, G., AND BERNSEN, N. O. 2006. Fusion of children's speech and 2D gestures when conversing with 3D characters. *Special Issue of Signal Processing 86(12)*, Elsevier, 3596-3624.
- MAYBURY, M. T. 1993. *Intelligent Multimedia Interfaces*. AAAI Press.
- MCNEILL, D. 1992. *Hand and Mind*. University of Chicago Press.
- NARAYANAN, S., POTAMIANOS, A., AND WANG, H. 1999. Multimodal systems for children: building a prototype. In *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech)*, Budapest, Hungary, 1727-1730.
- PICARD, R.W. 1998. *Affective Computing*. Cambridge, MA: MIT Press.
- ROBERTS, B. 2000. Coaching driving skills in a shiphandling trainer. In *Proceedings of the AAAI Fall Symposium on Building Dialogue Systems for Tutorial Applications*. North Falmouth, MA, 150.
- TRAUM, D., SWARTOUT, W., GRATCH, J., AND MARSELLA, S. 2005. Virtual humans for non-team interaction training. In *Proceedings of the Autonomous Agents and Multi-Agent Systems (AAMAS) Workshop on Creating Bonds with Humanoids*, Utrecht, The Netherlands.