Evaluation of Affective Interfaces

Kristina Höök
SICS
Box 1263
SE-164 29 Kista, Sweden
+46-8-6331500
kia@sics.se

ABSTRACT
One obvious challenge for affective interfaces is to find ways of checking whether the expressed emotions are understood by users, and whether the system can interpret user emotions correctly. Even more challenging is whether the overall usage scenarios are achieving their purpose of being e g engaging, fun, believable, or creating a relationship with the user, and how much of this can be attributed to the emotion modelling and expression. We propose a two-tiered evaluation model, similar to how adaptive interfaces are bootstrapped, in order to design and evaluate affective interaction. We exemplify this model through four studies of affective interfaces: a study of the *Agneta & Frida* system, two studies of the *Influencing Machine*, and one study of *SenToy*. We also discuss the issue of believability or naturalness of behaviour in these systems and claim that this is not necessarily the best goal for design.

Keywords
Affective interaction, synthetic characters, user evaluation

INTRODUCTION
Affective interfaces can be divided into those that express emotions, those that attempt to understand user emotions, and those that use affect as part of the system's "mind" – possessing emotions. Since the field is fairly new, user evaluation has not been done very often, and sometimes it is not even clear exactly what would be evaluated if we would attempt to make a user study. Researchers have been focused on issues of believability of characters, natural expressions, and not really concerned with whether the believability or naturalness does in fact contribute to the overall success of the system. Designers and researchers are often mislead into believing that human-human communication should be copied and used as the best model for interaction between human and machine. Studies in other fields, such as natural language interfaces, adaptive interfaces, intelligent user interfaces, etc., show that there are principles and peculiarities particular to the design of human machine interaction [Dahlbäck et al. 1993, Höök, 2000].

We believe that it is now time to move on and start evaluating affective interaction, and to involve users in the loop of designing them. In the context of the EC-funded SAFIRA project we noted that it was necessary to divide the user studies into two different levels: the first obvious challenge for affective interfaces is to find ways of checking whether the expressed emotions are understood by users, and whether the system can interpret user emotions correctly. It might be that a design of an affective interactive character is perfectly valid and well-suited to the overall goal of the system, but the facial emotional expressions of the character are hard to interpret. Thus the overall design fails anyway. Or the other way around, the emotional expressions might be easily understood by the user, but the design does still not achieve its overall goal of entertaining or aiding the user.

Thus once the interpretation loop is bootstrapped and working, the second, even more challenging goal for evaluation of affective interfaces, is whether the overall usage scenarios are achieving their purpose of being e g engaging, fun, believable, or creating a relationship with the user, and how much of this can be attributed to the emotion modelling and expression. In here, we aim to exemplify what kind of studies might be needed in this two-level design loop through shortly describing four studies of affective interfaces: a study of the Agneta & Frida system [Höök et al. 2000], two studies of the Influencing Machine [Sengers et al. 2002], and one study of SenToy [Andersson et al., 2002].

THE ROLE OF AFFECT IN INTERACTION
Before we can discuss what affect contributes with in interaction, we need to make a couple of notes on fuzzy terms that make readings in this area quite difficult. First of all, the term *affective* refers to something more than emotions, including personality factors and mood. What the terms emotion, personality or mood exactly means is not really agreed upon in this field. Some define emotions as a function that evaluates different alternative actions and makes it possible for an agent to choose between them. Others define emotions to be the innate, unconscious, basic emotions such as fear, sadness or happiness, and are exploring the effects of empathy. Others, again, try to package and understand the whole complex interaction between basic, innate emotions, bodily, somatic reactions, and higher-level cognitive reasoning about emotions and the relationship back and forth between the two.

Affective *interaction* in turn, is still a quite undefined term, including various aspects of using affect in the interaction between user and system. In Picard's book "Affective Computing" [Picard 1997], it included creating systems that: raise emotions in users, model users' emotional states, or actually 'possess' emotions and behaves as such. The focus was on how to utilise our understanding of emotions in order to produce better human-computer interaction, an interaction that would take user emotions into account.

Paola Rizzo [Rizzo 1999] provides an interesting analysis of a number of (implicit) hypotheses made in the affective interaction field, and convincingly shows that there is still much room for innovation that does not necessarily rely upon creating believable characters. It is not only through the use of cartoons that use facial-, bodily- or voice-based affective expressions that a system will be able to engage the user in an affective interaction. This might very well be accomplished through other means – especially as people seem so willing to ascribe human characteristics to any system [Reeves and Nass 1996].

PREVIOUS WORK

There are very few user studies of the short-term and even fewer of the long-term effects of affective interaction. On the other hand, designers of artefacts, artists, musicians, writers, people in advertising, and more recently web-designers and game designers, have played around with evoking emotions for ages. What differs here is the *interaction* between the artefact aimed at raising emotions or expressing emotions and the viewer/listener/reader/user reactions and (affective) actions at the interface. The user will be involved in the loop in a more active manner.

A lot of the work on affective interfaces is focused on implementing affective interaction through interactive characters, but affective interaction may also be realised in various other ways. In many affective interaction scenarios (besides interactive characters), the goal is to entertain. The HCI community has only recently started to debate how to take those characteristics into account when performing usability studies or providing input to design. These aspects are sometimes referred to as hedonic usability factors [Hassenzahl et al. 2000] or pleasure-based human factors[1].

Studies of Interactive Characters

Interface characters have been much criticised and debated in the HCI community [Shneiderman 1997, Lanier 1996, Suchman 1997]. They are said to violate good usability principles, to obscure the line of responsibility between human and machine, and to confuse both designers' and users' understanding of the computer's abilities and inner models of events. The proponents, on the other hand, regard these parameters as opportunities rather than reasons to avoid characters in the interface [Höök 2000].

The studies that have been done show effects on motivation [Mulken et al. 1998, Lester at al. 1997] – the presence of a character makes the interaction perceived as more easily comprehended and more compelling even if the objective measurements on learning and memory do not show any difference.

As pointed out by Andrew Stern [Hayes-Roth et al. 1998], the designer of the Catz and Dogz system, the artistic design and practical understanding of the creating of synthetic characters is crucial in determining the success of a system. A similar point is made in [Elliott & Brzeinski 1998] when they cite [Lester et al. 1997]:

> *"Lester gives the examples of, on the one hand, a humorous, lifelike, joke-cracking, character that ultimately impedes problem solving through his distracting presence; and on the other, a dull assistant that always operates appropriately but yet fails to engage the student. When communications from an agent must be coordinated to be both engaging and purposeful issues in timing, and the multi-layering of actions arise."*

Another effect of synthetic characters is the ways in which they tend to raise expectations of anthropomorphism of the system [Reeves & Nass 1996]. Such *anthropomorphic effects* seem to have many dimensions. On the one hand the user may expect the system to be intelligent and cognitively potent. [Brennan & Ohaeri 1994] showed that users talked more to the anthropomorphic interface. [King and Ohya 1995] showed that users attributed more intelligence to anthropomorphic interfaces. [Koda & Maes 1996] showed that realistic faces are liked and rated as more intelligent than abstract faces. Opponents of synthetic characters argue that raised anthropomorphic expectations may lead to frustration in the user when the system cannot meet the expectations [Shneiderman 1997]. For instance, the presence of a talking face might influence the user to expect the system to possess natural language and dialogue competence, which no system of today can live up to. The general conclusion is that the more 'natural' the interface, the higher expectations on intelligence in the system.

Criteria for Success

While more traditional user interaction has found their usability criteria (such as efficiency related to users' tasks,

[1] A workshop on Pleasure-Based Human Factors was held in Copenhagen 2000. The workshop was subsequently held in Singapore under the name Affective Human Factors Design.

number of errors, learnability) affective interaction still needs to explore which criteria will determine whether the affective aspects of the interaction in fact does contribute to the success of the system.

Furthermore, affective interaction systems must, similar to intelligent user interfaces [Höök 1998] be evaluated in two steps. First, we must make sure that the end users understand the emotions expressed by the system, or that the emotions expressed by users are understood by the system. Second, we need to check whether this in fact leads to the desired effects on the overall interaction. It might be that a design of an affective interactive character is perfectly valid, but the facial emotional expressions of the character are hard to interpret. Thus the overall design fails anyway. Or the other way around, the emotional expressions might be easily understood by the user, but the design thus still not achieve its overall goal of entertaining or aiding the user.

Thus, for any affective interface we would like to pose three different questions:

1) Do users understand which emotions are expressed by the system, be it through speech, child-like drawings, facial expression, or bodily behaviour?
2) Does the system correctly interpret users' emotional state expressed through a plush toy, art post cards, or the dialogue with the system?
3) If the answer is "yes" to both the previous questions, we can go further and check whether the overall goals with the affective parts of the system contribute to the interaction?

It is not really possible to separate the affective parts of the interaction from the rest of the design, but our task here is to try and focus on aspects of the overall goal that relates to the affective parts of the system.

All three systems we describe below are semi-autonomous or autonomous agents. For agents interacting with end users, we have the following crucial design problems:

- Agents need to display behaviour and affective expressions in such a manner that the user *understands* them. This means that they cannot always act in the most efficient rational way but instead they might have to act in ways that conveys to the user what is going on [Sengers 1998]. This also holds for agents that work in multi-agent systems but where their results have to be communicated to a user in the end.
- Agents need to be *timely*. When an emotion is displayed to the user it has to come at the right point in time, and last for an appropriate length [Ruttkay et al. 2000]. If an affective response from the user is the aim, then the interaction has to be carefully paced so that the user can follow it without being bored or puzzled.
- Agents sometimes need to have interesting *personalities*. Only then will their emotional behaviour be comprehensible and interesting to the user. Conveying the personality might be difficult if the interaction with the user is limited. This is where idle behaviour or interaction between several agents can come into play. When several agents interact, they can take the opportunity to show more of their personality traits.
- For some affective agent situations, it is necessary to create a *narrative context* (a situation, an interaction history) in order to understand the emotional behaviour.
- If the agent is used for a longer time span with a user, *different personalities and attitudes* might be needed in order to fit the needs of different users [Nass & Gong 2000].

FOUR STUDIES OF AFFECTIVE INTERFACES

We have performed four different studies of affective interfaces, and each illustrate a step in evaluating the overall effects of affective interaction:

- The study of SenToy illustrates how a study in an early stage of the design cycle can help bootstrap the design of affective input (performed through gestures with a toy). It also shows the need to differentiate between "natural behaviours" and how users really will interact with computer systems.
- The two studies of the Influencing Machine show the importance of first making sure that the affective output from the system is understood by users, before checking if the overall interaction idea is succeding.
- Finally, the study of Agneta & Frida shows the difficulty of measuring success of an affective system aimed at entertaining and the importance of clearifying the goal of the system.

Since we did the Agneta & Frida study before the other three studies, we shall start by describing it and the inspiration we gained on study methods.

Agneta & Frida Study

Our first study was of the Agneta & Frida system. These two animated female characters – mother and daughter – sit on the desktop, watching the user's browser more or less like watching television. They make humours and sometimes nasty comments of the web pages, the user actions, and sometimes just randomly talk to one-another. Initial testing helped us find the right timing for the jokes – a crucial aspect of humour is to deliver at the right moment.

In the proper study of Agneta & Frida, we measured how many times users smiled or laughed, the amount of time they spent with the system, their mood before and after using the system, and their responses to questionnaire ques-

tions after their session with Agneta & Frida. Interestingly, none of these correlated. This might be because the measurements were bad and fuzzy, or because people are generally known to behave in a socially desirable way, i.e. according to what they believe the experimenter desires. Since experimenter and designer in our study could be considered to be one and the same by some subjects, it is likely that such 'politeness' effects were present when answering some of the questionnaire-questions. As the experimental leaders left the room when they interacted with the system, hopefully some of these effects are minimised.

But another way of explaining the non-correlation, is to assume that the variables simply measure different things. We believe that although all of them try to capture the overall experience of the system, they may, in fact, measure different *aspects* of this experience.

For example, facial expressions may provide indications of the immediate appreciation of the system. Facial expression will perhaps show the instantaneous reactions to the jokes, but not the retrospect evaluations of Agneta & Frida. The post-usage replies, on the other hand, might reflect subject's 'after-thoughts' about the system, which may be influenced by moral and ethical preferences. Such retrospect judgements may be influenced by moral, ethics, and more official views of what humour and entertainment should or should not be.

The mood measurement – which lands somewhere in-between the instantaneous reactions during use and the post-usage replies – will again measure something else. Since it showed that the Agneta & Frida subjects were in a better mood compared to the without Agneta & Frida subjects, it provides us with some evidence that the system positively influences users' experience of the system on an emotional level.

What aspect of experience is most important – and thus determining the appropriate method of measurement – is of course dependent on the design goals. If we aim to entertain for a one-time usage situation, then maybe it is more important that subjects smile a lot; if we want subjects to return to the space, then their post-usage evaluation should be emphasised. The fact that many users were disturbed by Agneta & Frida – but still enjoyed their company – indicated that we failed to create a feeling of flow or relaxed relationship to the space. If that is our design goal, than other means should be used.

Our results point at the difficulty of gathering facial expressions and using those as a means to measure subjects' affective reactions towards computer systems. Users' physical reactions of interactions with systems are not necessarily good predictors of users' inner mental states. It is possible to see that the user gets aroused, but not whether it is a positive arousal or not, nor whether it reflects a mental state that the system should adapt to [Riseberg et al. 1998]. In order to pinpoint finer distinctions in the emotional reactions, we have to consider the users interpretation, understanding, attitudes, and expectations of computer culture. The experience of jokes and irony, for instance, will be determined by personal expectations, but also by social and cultural context. Our views on humour are reflections of our personality and who we want to be in the eyes of others.

Studying the Influencing Machine

We took many of the experiences from the Agneta & Frida study with us when we studied the Influencing Machine designed by Phoebe Sengers and collegues [Sengers et al. 2002]. First of all, we made the subjects enter the study room in groups rather than one by one. Facial expressions and discussions among the subjects are much more interesting to study with a group of users as opposed to single users infront of a screen. Second, we asked more questions about the subjects background and attitudes. Thirdly, we kept the interviews after their session much more open-ended to allow for them to express various views and ideas, rather than a simple "yes, I like it" or "no I don't" in a questionnaire.

Agneta & Frida and the Influencing Machine are quite different systems. The Influencing Machine grew out of the affective computing field, but takes on a different stance. Affective computation generally focuses on the informatics of affect: structuring, formalizing, and representing emotion as informational units. Through the Influencing Machine Sengers and colleagues proposed instead an enigmatics of affect, a critical technical practice that respects the rich and undefinable complexities of human affective experience. The interactive installation allows users to insert postcards into a mailbox, where the postcards represent their affective input influencing the machine. The machine responds through displaying its emotional state through emotionally expressive sound and child-like drawings displayed on the wall. The Influencing Machine bridges the subjective experience of the user and the necessary objective rationality of the underlying code. It functions as a cultural probe, reflecting and challenging users to reflect on the cultural meaning of affective computation.

The question is: how would we check whether the machine did in fact incur reflection? Should we look upon Sengers and colleagues as artists and place the machine in a museum as an art installation and then consider it done – "message delivered" – or should we go with the human computer interaction research paradigm and perform a usability study? If we had chosen the first alternative, we would have missed out on one of the important and difficult characteristics of the provocation, namely *interaction.* Interaction is a process where the machine and the user takes turn in creating dialogue acts, creating the content together. This differs somewhat, though not entirely, from an art exhibition, where the artist is reaching out with a message, the audience is interpreting, but there is very little turn-taking between the two after that. On the other hand, if the second method is used, a usability study, we would not be capturing the effects of the provocation; we would only be able to

capture whatever the users did not understand of the interaction with the machine.

Still, our machine, let us call it the Influencing Machine, needs to be trimmed, it needs to be tested to make sure that it in fact works as intended, otherwise it will not serve is purpose. This is why, in the end, we decided to make the Influence Machine go through two cycles of "user studies" – but to make these quite different from normal usability studies.

The first Influencing Machine study

Our first study of the machine was done at a very early stage in the development cycle. The Influencing Machine did not have any sound system at this point. The evaluation was explorative in nature, as our main goal was to feed back into the design process.

Users were brought in small groups (six groups with in total 12 subjects) into a room with the Influencing Machine. Users were told that the installation had something to do with emotions, and were then allowed to play with the system as long as they liked. On average, they spent about 20 minutes in the room.

Generally speaking, users were first curious, then became frustrated. Often this frustration stemmed from not being able to control the machine. They had a great deal of trouble figuring out the relationship between postcard and drawings. For some users this became a barrier that stopped their interest in the machine. Some users found the Influencing Machine drawings too simple and drawn too slow. The mailbox itself was liked. Unfortunately, the bar code reader in the mailbox made a beep whenever a postcard was inserted. This led subjects to think of the mailbox as a machine rather than a form of communication with a semi-living being.

A complication is the frustration that users often developed with lack of control. Many users got irritated and frustrated when they could not figure it out. Certainly this is an affective reaction, but not one intended. These thoughts and observations led to a number of system design changes performed by Sengers and colleagues.

Users were confused about the emotional meaning of the imagery. The addition of the sound system helps to clarify the agent's interpretation of input cards and its emotional state. In addition, an internal emotional display was developed showing the level of each of the internal emotions. Although they were reluctant to show these internals, by offering the user an opportunity to understand how the agent is designed to feel, users can and do engage in critical reflection on whether they believe that the drawings actually express the stated internal emotion state. This display can be set in a state were it will fade away over time, supporting users through their initial exploration without constraining further interaction.

Users were also confused about the nature of influencing versus controlling the system. With the above improvements to emotional expression, including direct sound feedback instead of mechanical Mailbox beeping for changes in emotion, users would hopefully have a better understanding of how they affect the system. At the same time, this concept is subtle and runs counter to users' everyday experiences with computers; it may simply be in its nature that it is hard for users to understand.

Finally, users were sometimes bored by the drawings themselves. Speeding up the drawings, reducing the persistence of behaviours so that new forms appear more quickly, and adding some more complex drawings will probably raise user interest. Also, transitions between drawings need to be handled more gracefully. In the old version, the system draws for a while and then clears the screen and starts over. The graphics was reimplemented to remove these rough breaks by layering over one another and gradually fading away.

The second Influencing Machine study

The second study was performed in a similar fashion to the first study. The results from this second study showed that the design changes did indeed achieve the desired result; users were more positive, less confused, and more of them did understand the point and were willing to discuss the intended provocation than in the first study. The replies to the interview questions and the interactions the groups did with the machine indicated that the group who had the emotional display on did more easily grasp that the machine expressed emotions and could be influenced.

Subjects were more inclined to form theories of what was going on inside the Influencing Machine and we got more positive comments about the drawings and the overall experience. The subjects from the second study also used the Influencing machine twice longer in average than the subjects from the first study. But there were still those subjects who experienced frustration and who were less inclined to "get the point". Seven of the nine groups had different theories that they tested during their session with the machine. They tried to make the machine respond in a special way through putting a certain card or a specific category of cards inside the machine, for example they tried to use only dark-coloured cards in order to see the response from the machine. The groups that tested different theories during the session seamed to have more fun during the session than the other two groups, but after a while most of them got somewhat frustrated when the response from the machine was not what they expected:

> "I would really like to know what it was all about. To start with you felt positive enthusiasm, but then you start to wonder when nothing works. I saw no logic in it. But it was fun, though it cools off after a while, you get somewhat irritated just because you do not know what is going on. Then you become systematic so that you just add people or just really colourful cards, or the same artist, but I

> did not find any difference. No new symbols or colours or anything." (Group #8)
>
> "Simple, but one is made in such as way that one will be looking for a purpose (a meaning) to what one does, but we discovered by and by that it did not matter what we did, the machine worked on anyway. That is how it felt, but it was beautiful." (Group #1)

This might be an effect of the setting – the users perhaps sees themselves as "people testing a scientific system" and this in turn alters the way they use the system, i.e. the experimental set-up itself affects the way people interpret and react.

The three groups that consisted of elderly people (60+) seemed to be more frustrated by the fact that they could not control the machine and fully understand how the machine worked. We got the impression that the elderly more than the younger participants felt a little stupid, when they could not control the technology.

A Wizard of Oz study of SenToy

Finally, the fourth study we did was the first step towards bootstapping the design of an affective input device – the SenToy. SenToy is a doll with sensors that allows users to (partly) control their avatars in an adventure game. A Wizard of Oz study was used early in the design process to find the best relationship between user movements of SenToy and the resulting affective expressions and movements of their avatar.

In a Wizard of Oz study, the user is made to believe that they are interacting with a system, while in reality they are interacting with a human Wizard, sitting behind the screen pretending to be the system. This study was performed with dolls that did not have any sensors at all, but where the Wizard interpreted users' actions with the doll and made the avatar express the corresponding emotion. Since subjects divided their visual attention between the doll and the screen with their avatar, subjects sometimes missing the actual performance of an emotion of the avatar's face or body as they were focusing on the doll and moving the doll. The Wizard adjusted to this problem, delaying until the subject had finished their movement with the doll, or sometimes, even making the avatar perform the action twice.

The study showed that there are behaviours that most users will easily pick up to express emotions, but that these are not necessarily linked to any "natural behaviour". First, users will not behave in the same way when expressing emotions through a doll rather than through their own bodily behaviours. There are numerous reasons for this, among them the cultural notions for how dolls and cartoon characters behave when expressing emotions. Secondly, we needed to put users in a loop where they are given feedback from the system through how the avatar reacts. Users will *learn* how to create the right behaviour through watching the face of the avatar when they perform actions on the SenToy. Thus there is room for "unnatural" learnt behaviours.

The study also told us which dimensions of movements (distance to screen, movements of limbs, etc.) that work. The study also provided input on which properties of the SenToy are important to consider, such as the softness, size and facial expression. Wizard of Oz studies have previously been used for natural language interface [Dahlbäck et al. 1993] and intelligent agent design [Maulsby et al. 1993] and we showed that it can effectively be used also in the domain of affective input design.

DISCUSSION OF RESULTS

Taken together this four different studies shows the importance of bootstrapping affective interaction and making sure that the affective expressions or affective input opportunities are understood before the overall system can be evaluated. The studies also reveal some important issues to be dealt with once this bootstrapping has been done and the system is evaluated against its overall purpose. In particular, we find that the field of affective computing often make simplistic statements where it is claimed that users will more easily bond with an affective system, become more efficient if not stressed or disturbed at the right moment, etc. The Agneta & Frida study and the Influencing Machine study show how complex the reactions are to these interfaces and how much depends on the users' background, age, attitudes and interest – to some extent this is different from normal usability issues.

The studies also point at the difficulties of performing these studies in laboratory environments. In a laboratory study users will assume that there is a task to be solved and with affective interaction there is not necessarily any such goal-oriented task.

The possibility of bringing in several subject at the same time proved to be successful. The subjects who came in groups to the Influencing Machine studies formed more theories and had richer interaction than those who entered alone.

Still even from laboratory studies as the ones discussed here, some lessons for how to design the systems can be extracted. First of all, all the studies confirm that issues of timing are crucial. Agneta & Frida's jokes have to arrive at the right moment, the Influencing Machine has to be influenced at the right level and draw its drawings fast enough in order for the interaction to work, and finally, the avatar reactions to SenToy has to be delayed or prolonged enough for the user to both handle the doll and watch the avatar on the screen.

For SenToy, many lessons were learnt before the costly process of creating a doll with sensors was started. All four studies, but in particular the SenToy study, definitiely show that it will be a mistake to only aim for "naturalness" in the affective expressions. Most interactive agents and affective interfaces are interesting in that they are different from how

we behave in human-human relationships, but still similar enough for us to recognice them and have fun with them.

We need to do more of these open-ended explorative studies, early on in the design process, before we can start doing the studies that really matter: namely those that show that affect in interaction does indeed contribute something different from other kinds of design. In this process, we need to more openly discuss which measurements will indeed be related to the overall goals of the entire system. The non-correlation between measurements in Agneta & Frida shows how difficult it is to separate an understanding of what kind of experience we want to evoke from users' attitudes and values. It also shows that we need to be clear of what kind of experience it is that we want to give the user: a short-term fun thing, a post-usage positive attitude, a provocation that continues even after using the system as for the Influening Machine, or what?

Also, measuring bodily reactions as number of smiles or sweat proves to be less useful measurements of overall experience.

While we have not presented a complete framework for how to bootstrap design and evaluate affective characters, we believe that our studies could be the inspiration to taking some more steps in this direction. In particular, we hope to encourage taking users into the loop when designing the interaction cycle wrt timing, narrative context, understanding of affective input and output, and being more open to the effects of users' attitudes and cultural values.

ACKNOWLEDGMENTS

Thanks to Gerd Andersson, Pia Mårtensson, Per Persson, Ana Paiva and Phoebe Sengers who worked with me on the studies discussed here. Thanks also to the anonymous subjects who participated in the studies. The work was done within the EC-funded SAFIRA (Supporting Affective Interactions for Real-time Applications) project.

REFERENCES

1. Andersson, G., Höök, K., Mourão, D., Paiva, A., Costa, M., Using a Wizard of Oz study to inform the design of SenToy, accepted as an exhibit at Designing Interactive Systems, DIS'02, to be held in London, ACM, June 2002.
2. Brennan, S.E., Ohaeri, J.O. Effects of Message Style on Users' Attributions toward Agents. Conference Companion, CHI'94, Boston (1994)
3. Dahlbäck, N., Jönsson, A., and Ahrenberg, L. Wizard of Oz studies—Why and how. *Proc. ACM International Workshop on Intelligent User Interfaces* ACM, 1993.
4. Elliot C., Brzeinski, J.: Autonomous Agents as Synthetic Characters. Communications of the ACM (1998).
5. Hassenzahl M., Platz A., Burmester M., Lehner K.: Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. Proceedings of CHI'2000. ACM, The Haague (2000)
6. Hayes-Roth B., Ball G., Lisetti C., Picard R.W., Stern A.: Panel Session: Affect and Emotion in the User Interface at the Intelligent User Interface Conference (IUI'98), San Francisco. ACM Press (1998)
7. Höök K.: Evaluating the Utility and Usability of an Adaptive Hypermedia System. Journal of Knowledge-Based Systems 10(5) (1998)
8. Höök K.: Steps to take before IUIs become real. Journal of Interaction with Computers 12(4) (2000)
9. Höök K., Persson P., Sjölinder M.: Evaluating Users' Experience of a Character-Enhanced Information Space. Journal of AI Communications 13(3) (2000) 195-212
10. King W.J., Ohya J.: The representation of agents: a study of phenomena in virtual environments. Proc. of the 4th IEEE International Workshop on Robot and Human Communication RO-MAN'95 Tokyo, Japan (1995)
11. Koda T., Maes P.: Agents with Faces: The Effects of Personification of Agents. Proceedings of HCI'96, London, UK (1996)
12. Lanier J.: My problems with agents. Wired (1996)
13. Lester J., Converse S., Stone B., Kahler S., Barlow T.: Animated Pedagogical Agents and Problem-Solving Effectiveness: A Large-Scale Empirical Evaluation. In Proceedings of the Eighth World Conference on Artificial Intelligence in Education, Amsterdam, the Netherlands. IOS Press (1997) 23-30.
14. Maulsby, D., Greenberg, S. and Mander, R. Prototyping an intelligent agent through Wizard of Oz. *In Proc of the CHI'93*, The Netherlands, ACM Press, 1993.
15. Mulken S. van, André, E., Müller, J. The Persona Effect: How Substantial Is It?. Proceedings of HCI'98, Sheffield, UK (1998) 53-66
16. Nass C., Li G.: Speech interfaces from an evolutionary perspective. Communications of the ACM 43(9) (2000) 36-43
17. Picard, R.W.: (1997). Affective Computing. MIT Press (1997)
18. Reeves B., Nass C.: The Media Equation: How People Treat Computers, Television, and New media Like Real People and Places, Cambridge University Press (1996)
19. Riseberg, J. Klein, R. Fernandez, and R. W. Picard (1997), Frustrating the User on Purpose: Using Biosignals in a Pilot Study to Detect the User's Emotional State, *CHI 98*, late-breaking short paper.
20. Rizzo P.: Emotional Agents for User Entertainment: Discussing the Underlying Assumptions. International Workshop on Affect in Interactions: Towards a New Generation of Interfaces, 21-22 October 1999, Siena, Italy, AC'99, Annual Conference of the EC I3 Programme (1999)

21. Ruttkay Z.M. et al.: A facial repertoire for avatars, in Proceedings of the “Learning to Behave: Interacting Agents” workshop, CELE-Twente Workshops on Natural Language Technology, Twente University, October 2000, (2000)

22. Sengers P.: Anti-Boxology: Agent Design in Cultural Context. Carnegie Mellon University Department of Computer Science, PhD Thesis (1998)

23. Sengers, P., Liesendahl, R., Magar, W., Seibert, C., Müller, B., Joachims, T, Geng, W., Mårtensson, P., and Höök, K. The Enigmatics of Affect, accepted as a full paper at Designing Interactive Systems, DIS’02, to be held in London, ACM, June 2002.

24. Shneiderman B. Direct Manipulation for Comprehensible, Predictable and Controllable User Interfaces. Moore J., Edmonds E., Puerta A. (eds.): Proceedings of 1997 International Conference on Intelligent User Interfaces, Orlando, Florida. ACM (1997)

25. Suchman L. A.: From Interactions to Integrations. Howard S., Hammond J., Lindegaard G. (eds.): *Proceedings of Human-Computer Interaction INTERACT’97*. Chapman & Hall (1997)