Bartneck, C. (2003). Interacting with an Embodied Emotional Character.
Proceedings of the Design for Pleasurable Products Conference (DPPI2004), Pittsburgh pp. 55-60. | DOI: 10.1145/782896.782911

Interacting with an Embodied Emotional Character

Christoph Bartneck
christoph@bartneck.de

Philips Research
Prof. Hostlaan 4
5656 AA Eindhoven
The Netherlands

Eindhoven University of Technology
Den Dolech 2
5600 MB Eindhoven
The Netherlands

ABSTRACT
A salient feature of the ambient intelligent home of the future will be the natural interaction between the home and its inhabitants through speech. An embodied home character is necessary to ensure a natural dialogue by continuously providing intuitive feedback in the form of conversational and emotional body language. This study experimentally investigates the influence of the character's embodiment (screen character and robotic character) and its emotional expressiveness on the enjoyability of the interaction. The presence of emotional expressions significantly increased the enjoyability of the interaction with the robotic character. The embodiment had no significant influence on the enjoyability. However, in the robotic character condition a social facilitation effect and a high forgiveness for speech recognition errors was observed.

Categories & Subject Descriptors:
J.4 [Computer Applications]: Social and Behavioral Sciences --- Psychology

General Terms: Human Factors, Design

Keywords: Emotion, embodied, character, robot, enjoyability, ambient intelligent home.

INTRODUCTION
Many companies, universities and research institutes are working on the home of the future. Besides Microsoft and IBM, Philips Research is one of the key players and made recently a considerable step forward by opening a first prototype, called "Home Lab". One of the central concepts in the idea for the HomeLab is Ambient Intelligence [1].

A key component of ambient intelligence is the natural interaction between the home and the user. The most natural human interaction is speech and therefore the ambient intelligent home should also use speech technology. Not only the TV and video recorder might be voice controlled, but also most other electronic devices such as ovens, the lights and heating. The house itself might become voice controlled.

DPPI'03, June 23-26, 2003, Pittsburgh, Pennsylvania, USA.

One of the difficulties of communication between two partners, such as the ambient intelligent home and its user, through speech is that none of them can ever be sure that the information transmitted will be perceived successfully and understood correctly. Therefore, the ambient intelligent home requires and anthropomorphic character. Instead of talking to an empty room and receiving voice messages through a central speaker system the user would talk to one central anthropomorphic character. This character could employ the dialogue control acts simultaneously to the speech acts, so that the user has natural and constant feedback about the status of the information transmission. Such an character would also provide the possibility to amplify the meaning of a certain message by employing gestures, body posture and emotional facial expressions.

Another key component of the ambient intelligent home is that the devices and computers themselves will move into the background, becoming invisible to the user. The advantage of a home character would be that the user could address all these devices and computers through one central and visible entity that would execute the given commands.

FOCUS OF THIS STUDY
This study focuses on emotional facial expressions of an embodied character because they provide natural and continuous feedback to the user about the status of the communication and status of the ambient intelligent home.

The embodiment is another important aspect of the home character since it provides the physical basis for its emotional expression. The two types frequently found in literature and products are screen characters and robotic characters. The perhaps best-known and possibly most disliked screen character is Microsoft's Paperclip, a little help agent for the Microsoft Office software [9]. The other type are robots, such as Kismet [4] which are not yet commercially available. Little is known about how the different embodiments influence the human-character interaction. Several characters have been proposed, but no comparative study is available yet.

The largest effect that the emotional expressiveness and the embodiment of the character might have on the interaction between a user and the home may be found in the enjoyability of the interaction. Enjoyability is one of the subcategories of user satisfaction, which is again a subcategory of usability, as defined in ISO 9241. The other

Bartneck, C. (2003). Interacting with an Embodied Emotional Character.
Proceedings of the Design for Pleasurable Products Conference (DPPI2004), Pittsburgh pp. 55-60. | DOI: 10.1145/782896.782911

categories of usability, efficiency and effectiveness, might be affected as well, but the enjoyability aspect of user satisfaction appears to be the most promising one since even the most critical studies [7] about the benefits of emotional characters for human-machine interaction consider the increase of enjoyment as a possible contribution. This leads to the two main research questions of this study:

1. Will the user perceive the interaction with a character that uses emotional expression more enjoyable than with a character that does not use emotional expressions?

2. Will the user perceive the interaction with a robotic character more enjoyable than with a screen character?

To be able to investigate these research questions it was necessary to first create a home character and an interaction context in which the user can interaction with this character.

INTERACTION CONTEXT

It was not possible to directly test a character in an ambient intelligent home context, because such a home was not yet available during the execution of this study. Therefore, an interaction context had to be created that closely resembles the interaction context of an ambient intelligent home and thus allowed to study the resulting behavior of the home character and the user. Little is known about how user will interact with an ambient intelligent home, but several scenarios were created that helped to envision plausible interactions [3]. The analyses of these scenarios resulted in the definition of a cooperative negotiation as the interaction context for this study. Previous studies [11] successfully used a negotiation task for a user evaluation of an emotional interface character and hence no obvious reason is apparent that would speak against its usage in this study.

The negotiation task

Various precautions were taken to ensure that the negotiation had an appropriate level of difficulty, that participants did not have a personal preference for any of the items in the negotiation and that the value structure of the negotiation items allowed for joint gains and therefore give the negotiation a cooperative nature. Stamps were chosen as negotiation items. They were randomized and without noticing both parties had exactly the same starting conditions. The stamps were displayed on the negotiation board (see Figure 1) and the value for each stamp for the user was displayed below each stamp.

The negotiation board

The negotiation board was displayed on a Wacom LCD Touch Screen (see Figure 1). The board was divided into two sides. The top side was assigned to the character and the bottom side to the user. Each side had 30 deposit pads on which 20 stamps were placed. Furthermore each side had two trading pads on which the negotiators could place the stamps they would like to trade. Each deposit and trading pad had a value label at its bottom, showing the value of each stamp to the participant. Action indicators displayed the last action of each negotiator, such as making a full offer or refusing a deal. These indicators also helped the participants to remember what actions are possible. A turn indicator on the left side showed whose turn it was. The participant's side at the bottom of the screen had an additional score field to display the current score, which is the sum of values of all stamps in his or her possession.

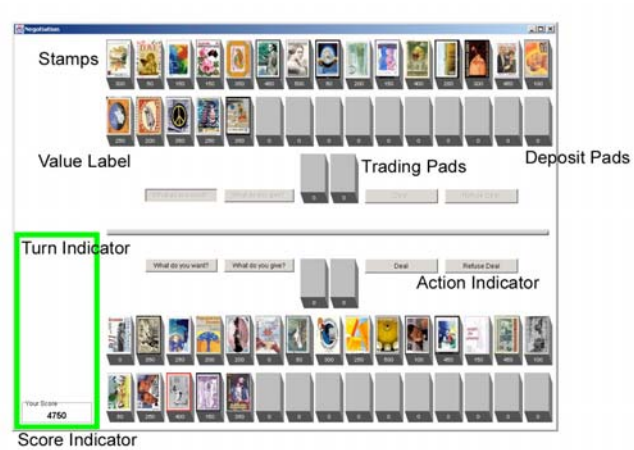

Figure 1: The negotiation board

IMPLEMENTATION OF THE CHARACTERS

The design of a home character has to consider many aspects of which only a few are mentioned below. A more detailed discussion is available elsewhere [3].

Match of the character's abilities and appearance

The match is particularly important since the appearance of the character will automatically evoke certain expectations of the user. If, for example, a character has a highly anthropomorphic body, the user might expect to be able to talk to it. The visual appearance characters has reached the level of humans [2], but not its speech recognition and synthesis. No interface character has passed the Turing test yet, because their conversational skills are still very limited. To lower the user's expectations of the conversational skills of the home character, an animal character was used. Animal characters are a good representation for the home character since the communication abilities of animals match the technical abilities of the current speech based system. A dog for example, is able to understand simple human speech, but still makes mistakes. This mirrors the quality of current speech recognition software.

Besides the limitations of the speech recognition accuracy it is also not yet possible to systematically express emotions convincingly through synthesized speech. It would be unconvincing if the character showed a certain facial emotional expression and talked with a neutral voice [10]. It might even be interpreted as lacking sympathy for the user. Therefore the character in this study did not use speech synthesis. The animal form of the character matches its missing ability, since animals are not able to talk either.

Bartneck, C. (2003). Interacting with an Embodied Emotional Character.
Proceedings of the Design for Pleasurable Products Conference (DPPI2004), Pittsburgh pp. 55-60. | DOI: 10.1145/782896.782911

In the negotiation task the user communicated with character by speaking out the desired negotiation action. If the user wanted to, for example, make a full offer he or she would have to say to the character: ”make deal”. The available actions and their corresponding phrases were displayed on the action indicators on the negotiation board. An “ehmm?” sound was played if the speech recognizer did not understand the participant’s utterance. The character in return would not speak, but point out its actions on the action indicators and use its facial expression to signal the status of the negotiations process.

The character evaluated the negotiation situation from a rational and emotional point of view. The latter was based on a simplified OCC model [12]. Depending on the outcome of these evaluations the character made its next game move and expressed its emotional state through its face. The character was able to express happiness, sadness and anger by changing the form of its eyebrow and lip.

Hardware

The main design objectives for the form of the hull was to maximize its cute appearance by applying form factors of infants, such as a big head, big eyes and round shapes (see Figure 2). The hull is made from polyurethane and is therefore very soft.

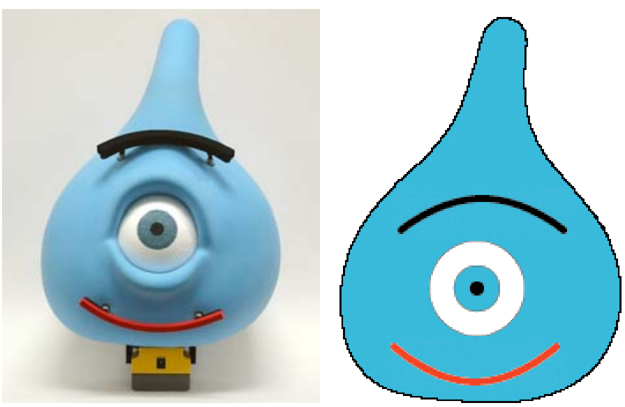

Figure 2: eMuu, robotic and screen based

Software

The software for the two Lego Robotic Command Explorers (RCX) was written in JAVA. The software running on the PC was also written in JAVA and used IBM’s ViaVoice for the speech recognition. A detailed technical description of the characters is available elsewhere [3]. A frontal photograph of the robotic eMuu was the base for the design of the screen character. Its appearance was designed in plain 2D without any spatial clues, such as bevels or shadows.

THE EXPERIMENT

An experiment was conducted in a usability lab at the Eindhoven University of Technology to answer the research questions mentioned before.

Measurements

Enjoyability

None of the previous studies in the area of interface characters focused their evaluation on the enjoyability of the interaction. This research area is only at its beginning and no validated instrument for measuring the enjoyability of interacting with characters is available at the moment.

Therefore, a questionnaire developed by Hoonhout [8] was used to measure enjoyability. The questionnaire was in Dutch, which ensured that the participants, who were all native Dutch, had an optimal chance to understand the concepts in the questionnaire. The complete questionnaire will be published in an upcoming paper [8]. The 7-point scale questionnaire consisted of 59 items in 8 concepts. An example in the English translation is given for each concept in the following list. Please notice that positive and negative items were used for each concept.

1. Enjoyable, entertaining - I think that using this set-up was boring.

2. Intrinsically motivating, engaging - When you are busy with this set-up, you want to keep on working with it.

3. Challenging, stimulating, exciting - I think that it was a challenge to work with this set-up.

4. Exploration, curiosity, fantasy, novelty, surprise - This invites you to find out what you can do with it.

5. Concentration, attention focus, immersiveness - When I started using this set-up, I found it difficult to stop working with it.

6. User control, assurance, efficacy - I felt comfortable when I worked with this set-up.

7. Satisfaction, usability, utility - I am satisfied with this set-up.

8. Pride - I would not use this set-up in the future.

In the following text these eight concepts are referred to by their underlined word and written in the *italic style*. The complete questionnaire is referred to as the “enjoyability profile”.

Negotiation performance of the user and character

Positive moods tend to increase negotiator’s tendencies to select a co-operative strategy [6] and enhance their ability to find joint gains [5]. Therefore it was expected that the participants who enjoy the interaction with the character would achieve higher joint gains than participants who do not. Thus, the joint gain is an indicator for the enjoyability of the interaction. The joint gain was calculated by adding the user’s and character’s score. The joint gain, the user’s and the character’s score were automatically logged in the experiment and were analyzed at a later state. In the following text the negotiation performance is referred to by “joint gain”, “user’s score” and “character’s score”.

Speech recognition accuracy

The participant interacted with the home character through speech. However, the speech recognition accuracy may vary between participants and introduce a bias. To control this factor the recognition accuracy was automatically logged and analyzed at a later stage.

Manipulation

A mixed 2x2 between/within participants experiment was performed (see Table 1). The between participant factor was the embodiment, which was either a screen character or a robotic character. The within participant factor was the emotional expressions which were either present or not.

Table 1: Coding of the conditions

	Emotion not present	Emotion present
Screen Character	S	SE
Robotic Character	R	RE

There might have been individual differences between the participants in the negotiation task itself. Therefore a baseline (N) was used in which the participants negotiated without the presence of a character. The data from this baseline condition is used as a Covariant in the analyses of the influence of the embodiment on the measurements. It eliminates the influence that the negotiation task itself might have had on the measurements. The participants were randomly assigned to the cross-balanced presentation orders of the four conditions. The speech recognition accuracy and the negotiation performance are two factors that could not be controlled and might have had an influence on the other measurements. Therefore, they were considered independent variables and covariants for certain statistical tests.

Participants

53 participants, 34 male and 19 female at the age from 18 to 53 took part in the study. All of them were native Dutch. None of the participants had knowledge about collecting stamps. None of them were professional business/sales persons or received negotiation training.

Procedure

First, the experimenter welcomed the participants. Then they read an introduction explaining the purpose and process of the experiment and the negotiation task. The experimenter trained the participants to correctly pronounce the phrases that they could use in the negotiation game. Furthermore, they were given tips, such as not to increase their loudness or separate the words in case the speech recognizer was not able to understand them.

The participants performed a training session of the negotiation tasks. Afterwards, they could ask questions about the process and the rules of the negotiation before the experimenter left the room. In each of the three parts of the experiment, the participants interacted with the system for ten minutes and afterwards filled in a questionnaire. After the three parts the experimenter entered the room, thanked the participant for his or her effort and handed out a little gift.

Equipment

The negotiation game was presented on a Wacom PL-400 flat panel touch screen. The participants used a pen to click on the stamps and their voice to select an action (see Figure 3). A Philips SBC-ME400 uni directional microphone and IBM’s ViaVoice for the speech recognition was used. The participants filled in the questionnaire on a separate laptop.

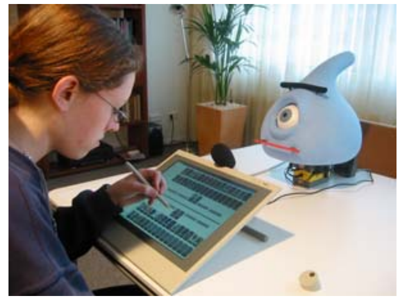

Figure 3: Experiment setup.

Data preparation

The score for every participant for each of the 59 items in the questionnaire were transformed to z-scores. This normalisation created comparable, individual scores that were re-scaled to have a mean of 0 and a standard deviation of 1.

Given the relatively recent development of the questionnaire, a reliability analysis was performed to test its internal consistency. This analysis was performed on the data of the neutral condition, because this condition was presented to all participants first. This condition could have not been influenced by the other conditions and therefore all 53 participants were included in the analysis.

Items that considerably reduced Cronbach’s Alpha, a coefficient of reliability, were excluded from the further analyses. Given the high Alpha values for the eight concepts, ranging from 0.69 to 0.88, and the high overall Alpha of 0.96, the questionnaire can be considered sufficiently reliable.

The mean of all remaining items belonging to one concept was calculated for each of the eight concepts in the enjoyability profile. Since the relative importance of each concept for enjoyability is not known, the best approach is to assume that they are all of equal importance.

The speech recognition accuracy was calculated by dividing the number of correct recognitions by the total number of utterances. The different degrees of freedom in the various statistical tests are based on missing data for certain participants.

Results

The recognition accuracy effect

An ANOVA was performed to test if there was a difference in the speech recognition accuracy in the two embodiment conditions. There was a significant difference within the robot condition (R,RE) (F[2,21]=56.136, p<.001). The recognition accuracy in the neutral (N) condition (70%) was significantly higher (t[23]=6.512, p<.001) than the accuracy in the character (R) condition (56%), which was again significantly higher (t[24]=3.654, p=.001) than in the emotional character (RE) condition (43%). There was no significant difference within the screen conditions (S,SE).

An ANOVA was performed to test the difference in speech recognition accuracy between the screen and robot conditions. There was a significant difference in the emotional character condition (SE,RE) (F[1,50]=5.661, p=.021). The recognition accuracy was significantly higher in the screen (SE) condition (58%) than in the robot (RE) condition (43%) (t[50]=3.903, p=.021).

A linear regression analyses was performed to test if there was a correlation between the recognition accuracy and the enjoyability profile across all conditions. *Challenging*, *novelty*, *user control* and *usability* were significantly, but weakly correlated with the recognition accuracy. The Pearson Correlation coefficient ranged from –0.26 for challenging to 0.16 for user control. The R^2 values were very low for all cases, ranging from 0.06 for *novelty* to 0.02 for *user control*. This indicates that the suggested model does not seem to fit. Scatter plots revealed that the scores for *challenging* and *novelty* were high at low recognition accuracy and low for high recognition accuracy. The opposite effect was observed for *user control* and *usability*. Their scores were high at high recognition accuracy and low at low recognition accuracy.

The second linear regression investigated the correlation between recognition accuracy and the enjoyability profile depending on their type of embodiment and presence of emotional expressions. *Usability*, *user control*, *novelty* and *challenging* were significantly, but weakly correlated with the recognition accuracy in the across all screen conditions (N, S, SE). The significance disappeared for *user control* and *usability* in the presence of the character (S, SE), but remained for *challenging* and *novelty*. Scatter plots revealed the same direction of the correlation as mentioned before. No significant correlation could be found in the robotic conditions (R,RE).

The emotional expressions of the characters effect

Table 2 presents the mean values for the enjoyability profile in the four conditions. An ANOVAs was performed to test the influence of the addition of emotions to the characters. Emotional expressions had no influence (F[7,18]=1.959, p=.119) on the enjoyability profile in the Screen Character condition (S,SE), however in the robot condition (R,RE) it had a significant influence (F[7,21]=2.749, p=.34).

T-Tests were performed to check differences on the individual concepts. In the robot condition the score for *novelty* was significantly higher (t[27]=3.451, p=.002) in the emotional character (RE) condition (0.132) than in the character (R) condition (-0.093). The score for *concentration* was significantly lower (t[27]=2.083, p=.047) in the emotional character (RE) condition (0.1428) than in the character (R) condition (0.2842). In the screen condition the score for *novelty* was significantly higher (t[24]=2.274, p=.032) in the emotional character (SE) condition (-0.021) than in the character (S) condition (-0.1415). The score for *usability* was significantly lower (t[24]=2.635, p=.015) in the emotional character (SE) condition (-0.1074) than in the character (S) condition (-0.0675).

Table 2: Mean values for the eight concepts in the different conditions (S, R, SE, RE).

	Screen Character		Robot Character	
	Character	E. Character	Character	E. Character
Entertaining	0.01	0.07	0.09	0.10
Motivating	-0.20	-0.11	-0.20	-0.32
Challenging	0.14	0.10	0.06	0.13
Novelty	-0.14	-0.02	-0.09	0.13
Concentration	0.19	0.14	0.28	0.14
User Control	-0.04	-0.02	0.00	-0.15
Usability	0.07	-0.11	-0.08	0.00
Pride	-0.06	-0.04	-0.08	-0.12

ANOVAs and t-Tests were performed to investigate the influence of the emotional expressions on the user's score. No significant difference was found.

The embodiment of the characters effect

ANOVAs were performed to test the influence of the embodiment on the user's score. First, the influence on all conditions was tested. There was a significant influence (F[1,104]= 6.745, p=.01). The user scored significantly higher (T[104]= -2.59, p=.011) in the robot (6618 euro) condition (R,RE) than in the screen (6237 euro) condition (S,SE).

Next the influence of the embodiment was investigated separately for the non-emotional (S, R) and emotional characters (SE, RE). There is a significant difference (F[1,51]= 6.04 , p=.017) in the non-emotional character condition (S,R). The users in the robot condition (R) scored (6680 euro) significantly (t[51]= -2.45, p=.017) higher than in the screen (S) condition (6176 euro). In the emotional character conditions (SE,RE) only a trend is visible. The users scored higher in the robot (RE) condition (6555 euro) than in the screen (SE) condition (6298 euro).

ANCOVA were performed to investigate the influence of the embodiment (screen or robot) on the enjoyability profile. The corresponding depending variable in the neutral condition (N) was used as the covariant. Only *novelty* was influenced significantly (F[1,103]=4.137, p=.04). The same effect was observed when looking at the character and emotional character conditions separately. Only *novelty* was influenced significantly (F[1,50]= 3.919, p=.05) in the emotion character conditions (SE,RE). However, a t-Test revealed that the means (SE=-0.02, RE=0.13) were not significantly different (t[51]=1.237, p= .22).

CONCLUSIONS

The speech recognition software used for this experiment showed two practical problems. First, the speech recognition in the robotic conditions (R,RE) decreased and second, there was a considerable difference in recognition accuracy between the participants. The decline in speech recognition accuracy in the robot conditions was most likely caused by the noise of the robot's motors. In the character condition (R) the motor for the panning of the head was randomly switched on and off. In addition to this, the motors of the eyebrow and lip were switched on and off irregularly in the emotional character condition (RE). It is to be expected that such irregular sounds are more disturbing for speech recognition software than monotonous noise, such as a constant hum. The considerable difference in speech recognition accuracy between participants might be related to the quality of the speech recognition program.

Two linear regression tests were performed to check if the speech recognition accuracy had an effect on the enjoyability profile. The weak correlations and very low R2 values suggest that this influence is negligible. The speech recognition problems were not strong enough to compromise the validity of the experiment.

Adding emotions to the character affected measurements in the robot conditions. The expressions did not influence the user's score even though the expression provided useful information for the negotiation task. A possible reason for this could be that the time the participants interacted with the character might have been too short to acquire a good understanding. Getting to know a human and learn his preferences and expressions takes considerably longer than the time available in the experiment. However, the emotional expressions certainly increased the *novelty* aspect. The results of the experiment suggest that the participants enjoyed the interaction with an emotional expressive character more than with non-expressive character, in particular if it is a robotic character.

The embodiment of the character (screen or robot) had no influence on the enjoyability profile, but affected the user's score. In the robot condition the participants acquired higher scores than in the screen condition. However, the joint gain, which is the sum of the user' score and the character's score, remained unaffected. This leads to the interpretation that the well-known social facilitation effect, which says that the mere presence of another person will motivate the people to exert more effort in a task was observed. The robot character appears to have a stronger social facilitation effect than the screen character and therefore the participants put more effort into the negotiation. Another interesting effect of the embodiment is that even though the speech recognition accuracy in the robot condition was below the screen conditions, and even decreased from the robot character condition (R) to the emotional robot condition (RE), it did not affect the rating on usability and user control. The participants appeared to be forgiving for the speech recognition errors in the robotic condition. Considering all the measurements this study comes to the conclusion to reject the hypothesis that user will perceive the interaction with a robotic character more enjoyable than with a screen character. The robotic character embodiment, however, has other interesting effects, such as the social facilitation effect and a high forgiveness of the participants for the speech recognition errors. These effects alone might make it worthwhile to use a robotic character for the ambient intelligent home.

ACKNOWLEDGMENTS

This study was supported by: Stan Ackerman Institute, J.F.Schouten Institute, Advance Telecommunication Research Institute International and Philips Research.

REFERENCES

1. Aarts, E., Harwig, R., & Schuurmans, M. (2001). Ambient Intelligence. In P.Denning (Ed.), *The Invisible Future* (pp. 235-250). New York: McGraw Hill.
2. Bartneck, C. (2001). How convincing is Mr. Data's smile: Affective expressions of machines. *User Modeling and User-Adapted Interaction*, 11, 279-295.
3. Bartneck, C. (2002). eMuu - an embodied emotional character for the ambient intelligent home. Ph.D. thesis Eindhoven University of Technology.
4. Breazeal, C. (1999). Robot in Society: Friend or Appliance? In *Agents99 workshop on emotion-based agent architectures* (pp. 18-26). Seattle.
5. Carnevalve, P. J. D. & Isen, A. M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational behavior and human decision process*, 37, 1-13.
6. Forgas, J. P. (1998). On feeling good and getting your way: Mood effects on negotiator cognition and bargaining strategies. *Journal of personality and social psychology*, 74, 565-577.
7. Hollnagel, E. (1999). Keep cool: the value of affective computer interfaces in a rational world. In *Proceedings of HCI International '99* (pp. 676-685). Munich.
8. Hoonhout, H. (2001). Development of a rating scale to determine the enjoyability of user interactions with consumer devices Eindhoven: Philips Research internal report.
9. Microsoft (2000). Microsoft Agents. http://www.microsoft.com/msagent [On-line].
10. Nass, C. & Gong, L. (2000). Is Maximization or Consistency the More Social? The Case of Synthesized Voices and Faces. Unpublished.
11. O'Reilly, W. S. N. (1996). Believable Social and Emotional Agents. Ph.D. Carnegie Mellon University, Pittsburgh, PA.
12. Ortony, A., Clore, G., & Collins, A. (1988). The Cognitive Structure of Emotions. Cambridge: Cambridge University Press.