*User Modeling and User-Adapted Interaction* **11:** 279–295, 2001.



# How Convincing is Mr. Data's[1] Smile: Affective Expressions of Machines

CHRISTOPH BARTNECK
*Eindhoven University of Technology, Den Dolech 2, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, E-mail: christoph@bartneck.de*

(Received: 15 July 2000; in final form 27 February 2001)

**Abstract.** Emotions should play an important role in the design of interfaces because people interact with machines as if they were social actors. This paper presents a literature review on affective expressions through speech, music and body language. It summarizes the quality and quantity of their parameters and successful examples of synthesis. Moreover, a model for the convincingness of affective expressions, based on Fogg and Hsiang Tseng (1999), was developed and tested. The empirical data did not support the original model and therefore this paper proposes a new model, which is based on appropriateness and intensity of the expressions. Furthermore, the experiment investigated if the type of emotion (happiness, sadness, anger, surprise, fear and disgust), knowledge about the source (human or machine), the level of abstraction (natural face, computer rendered face and matrix face) and medium of presentation (visual, audio/visual, audio) of an affective expression influences its convincingness and distinctness. Only the type of emotion and multimedia presentations had an effect on convincingness. The distinctness of an expression depends on the abstraction and the media through which it is presented.

**Key words.** abstraction, affective expressions, convincingness, distinctness, emotion, face, modality, music, speech

## 1. Introduction

The importance of emotions has been analyzed in numerous studies (Frijda, 1986; Lazarus, 1991), including several on the role of emotions in cognitive processes (Norman, 1981). Prior studies found that emotions play an important role in problem solving (Feist, 1994) and decision making (Barnes & Thagard, 1996) by providing information on the emotional desirability of the options available, therefore reducing and limiting reasoning to only those that induce positive feelings. Emotions also guide actions and control resources (Oatley & Jenkins, 1996). Emotions should play an important role in the design of interfaces (Picard, 1997a;

[1] Mr. Data is the android in the TV series *Star Trek—The Next Generation.* His goal is to become more human-like. He spends considerable effort to learn how to express emotions and to master humour. Since he is not able to really feel emotions this is a tough task for him. His struggle symbolises the efforts of scientists and engineers to design a convincing emotional machine.

Nielsen, 1994). People interact with machines as if they were social actors (Nass and Reeves, 1996). It is not unusual, for example, to hear people yelling at their computer just as if it might feel sorry and change its behavior (Picard, 1997b).

## 2. Theoretical Background

Two main viewpoints to describe emotions can frequently be found in the literature. One considers emotions as discrete categories (Ekman, 1973; Izard, 1977; Plutchik, 1980). The other characterizes emotions as points in a multidimensional space (Schlossberg, 1954; Osgood et al., 1957; Russel, 1979). Arousal and valence could, for example, define such a space. The two viewpoints are not as different as they might seem. The discrete categories, for example, can be described as clusters of points in the dimensional approach. Frijda (1986) argued that on the one hand the number of dimensions may prove to be large (Nowlis, 1966; Frijda, 1969; Smith & Ellsworth, 1985; Schiano et al., 2000), which moves the dimensional viewpoint toward the categorical. On the other hand, the discrete emotions vary along common dimensions (Izard, 1977) and can be ordered in terms of similarities and as pairs of opposites (Plutchik, 1980). This pushes the categorical viewpoint towards the dimensional.

Due to practical reasons, this study takes the categorical viewpoint on emotions. Many studies (Ekman et al., 1972) used the categories happiness, surprise, fear, anger, sadness and disgust. This study applies the same categories to take advantage of this solid theoretical framework.

## 3. How Do Humans Express Emotions?

Humans express their emotions through actions, which can be perceived through the visual, auditory and tactile modality. Body language, such as facial expressions and gestures, are the main elements perceived by the visual modality. Speech and music are the main elements perceived by the auditory modality. Actions perceived by the tactile modality (for example petting and punching) are, due to their little relevance for human-computer interaction, not in the scope of this study.

Expressing emotions is a natural act for humans. The ingenuous ease of it contrasts with the difficulty of describing it scientifically. Furthermore, the capability to express emotions can be refined through the performing arts, such as acting and singing. All music students spend hours with their teachers learning to play music not just as it is written in the score, but also in the appropriate emotion. Even students who have learned to do it are usually still unable to explain how they do it. Many studies have been performed to find out how humans express emotions. We summarize some of their results in the following paragraphs to provide the reader with a starting point for further literature review. We applied these results as the base for our design of affective expressions used as stimuli in the experiment described in Section 5.

### 3.1. Speech

Speech is a powerful method to communicate emotions. If your friend, for example, does not show up for a meeting with you, you can express your anger through a telephone call. You are restricted to speech, but your friend will most likely understand the emotional state you are in.

The most influential parameters for affective expressions in speech are pitch (level, range and variability), tempo and loudness. Many other studies used these parameters and Scherer (1979) summarized their results. Murray and Arnott (1992) conclude in their literature review that in general, the vocal effects caused by particular emotions are consistent between authors and between the different studies carried out, with only minor differences being apparent. However, Murray's and Arnott's quantification of these speech parameters are rather vague. A more concrete approach is the Affect Generator (Cahn, 1990) a software tool to synthesize affective speech. It allows settings on a scale from $-10$ to $+10$ for each of its parameters. Zero represents the parameter influences for neutral effect, while $-10$ and $+10$, the minimum and maximum influence, respectively. Unfortunately, this scale does not translate to results of other studies. It is only meaningful for this software tool.

A more general approach is to quantify parameters in percentage of the neutral setting. Mozziconacci (1998) quantified optimal pitch and tempo settings for certain emotions this way. However, calibrating the neutral setting remains difficult.

### 3.2. Music

Music is a difficult method to express emotions because culture (Davies, 1978; Crowder, 1984), skills of the performer (Bresin & Friberg, 1999; Juslin 1997a) and age of the listener influence the perception. The widely accepted association between mode (major and minor) and emotion (happy and sad) develops, for example, during childhood (Cunningham & Sterling, 1988; Geradi & Gerken, 1995; Kastner & Crowder, 1990).

Scherer and Oshinsky (1977) demonstrated that 66–75% of the variance in the affective attributes of music can be explained by manipulation of amplitude, pitch (level, variation and contour), tempo, envelope and filtration. Furthermore, they argue that their results overlap with the findings in affective expressions in speech. Juslin (1997b) summarized expressive principles which he obtained by a series of studies using several different instruments, performers and melodies.

### 3.3. Body language

Pantomimes use only facial expressions and bodily movements to express emotions. Their success is amazing considering the abstract vocabulary of movements available to them. The main components of body language are facial expressions, gestures and body movement. There is no difference in the relative importance of the components of body language (Ekman et al., 1980).

#### 3.3.1. *Facial Expression*

Expressing emotions through the face is so natural for humans that it takes a considerable amount of effort to mask them (Ekman et al., 1972). Keeping a 'poker face' in a critical situation is difficult. The main components used to express emotions are mouth, cheeks, eyes, eyebrows and forehead. Ekman and Frieser (1975) compiled archetypes of affective expressions in the human face. Humans do not need the high quality photos or photo-realistic computer renderings to perceive emotions in facial expressions. The study of Etcoff and Magee (1992) used drawings of the human face, generated by the caricature generator (Brennan, 1985). The drawing consisted of only 37 lines, but the subjects were still able to perceive the emotions accurately.

#### 3.3.2. *Gesture*

Ninety percent of the gestures only occur during speech (McNeill, 1992). They convey some information, but they are not richly informative and the information conveyed is largely redundant in the presence of speech (Krauss et al., 1991). Still, people pay attention to them (Nobe et al., 1997) and gestures certainly make speech more lively. An easy and precise vocabulary, such as notes for music, is, due to its variance and inconsistency, not available for gestures. However, McNeill (1992) grouped gestures into categories, such as Iconic, Metaphoric, Deictics and Beats.

#### 3.3.3. *Body Movement*

Most of the descriptive studies on affective body movement are informal (Frijda, 1986). Table I summarizes Frijda's analyses.

## 4. How Do Machines Express Emotions?

Machines are able to express emotions. Almost all experiments, which tested affective expressions, presented their stimuli to the subjects by using machines, such as speakers, tape recorders and computers. Only very few experiments used actors performing live in front of the subjects.

All affective expressions of machines are abstractions of human expressions. Even movies with actors talking to each other are not the real people and therefore an

*Table I.* Description of body movements for several emotions

| Emotion | Body movement |
|---|---|
| Fear | Forceful eye closure or staring at source, frowning by drawing the eyebrows together, bending the head, hunching the shoulders, bending the trunk and knees |
| Surprise | Widening of the eyes, brief suspension of breathing, general loss of muscle tone, mouth falls open |
| Anger | Teeth bared, fierce glance (fixed stare, eyes slightly widened, eyebrows contracted), clenching fists (optional), lips compressed,<br>Movements are vigorous and brisk, body tense |
| Sadness | Depressed corners of the mouth, lowered muscle tone, turning inward, weeping (optional) |
| Happiness | High frequency of unfounded and goalless changes in direction and the preponderance of movements orthogonal to the direction of locomotion, smiling, laughing (optional) |

Note: From Frijda (1986)

abstraction. The more abstract an expression is the more interpretation room towards the machine becomes available. However, machines do not have their own non-human emotions or the ability to express them. Humans would also not be able to understand non-human emotions without additional learning. This is not necessary for human emotions, because human–human interaction trained the user of an affective machine already. Therefore, machines should mimic human expressions or their abstractions to communicate emotions.

Many companies and researchers are working on the synthesis of affective expressions. The quality of synthesized facial expression is high (Pixar, 1998). A ready to use tool for synthesis of facial expression is the CSLU Toolkit (CSLU, 1999). It is software for speech recognition and synthesis, which includes an animated character, called Baldi. Massaro (1998) showed that humans perceive Baldi's affective facial expressions accurately. The results of the present study support his findings.

The quality of synthesized speech is far behind compared to the developments in synthesized facial expression and body language. Toy Story and all other computer-animated movies up to this point are good examples for this. They all successfully used computer-generated characters, but they all fall back to real actors for the voices. The most promising synthesis of emotions in speech is the Affect Generator by Cahn (1990) mentioned above. She successfully applied 17 parameters, which resulted in a recognition accuracy of 78.7%.

The Director Musices by Bresin and Friberg (1999) is a promising synthesis program for affective expression in music performance. It is a rule-based software tool for automatic music performances. By altering 17 parameters they have been able to reach a recognition accuracy of 64% (14% chance level).

A promising synthesis of body language and speech is the work of several members of the Department of Computer and Information Science at the University of Pennsylvania (Cassel et al., 1998). They implemented a system which automatically generates and animates conversations between multiple human-like agents with appropriate and synchronized speech, intonation, facial expression and hand gestures.

Already today, products which express emotions are available. Sony's entertainment robot 'Aibo' (Sony, 1999) is able to express six emotions and their blends. Therefore, it is uninteresting for this study to ask if machines can express emotion. More important is the question if there is a difference in the perception of emotions expressed by either a machine or a human and what attributes of the emotional expression are most important for the perception.

## 5. The Experiment

### 5.1. Introduction

Most of the previous studies and implementations described above concentrated their evaluations on the distinctness of the affective expression (Bartneck, 2000), which they measured through the recognition accuracy of the subjects. In this experiment we expand the evaluation with another important attribute: the convincingness of the affective expression. Moreover, we investigate possible differences in the perception of affective expressions of humans and machines. Are affective expression of machines as convincing as affective expressions of humans? What factors influence convincingness? These questions are particular important for the development of affective embodied agents.

We consider convincingness as a synonym for believability. We use this term to prevent confusion with Fogg's and Hsiang Tseng's (1999) model of computer believability. They described a model of believability in which the perceived trustworthiness and expertise of a system predicts its believability. They defined trustworthiness as the perceived goodness or morality of the system and expertise as the perceived knowledge and skill of the system.

We took their model as the basis for our model of convincingness of affective expression. However, applying their definition of expertise to affective expressions is not easy. We consider the perceived appropriateness of affective expressions as a measure for affective expertise of the system. A system that displays the right emotion at the right time (appropriate) has expertise in the field of affective expressions. Furthermore, we assume that the intensity and distinctness of an affective expression has influence on its convincingness. This leads us to a first model of convincingness (see Figure 1).

To judge trustworthiness and appropriateness (measures for expertise), it is necessary to know the context in which the specific emotion is expressed. In real life, context information will always be available. The jamming of an affective CD player, for example, would be associated with its sad affective expression. For our experiment we chose the context of a simple dice game, because it required only a small amount of learning from the subjects and it was easy for the subjects

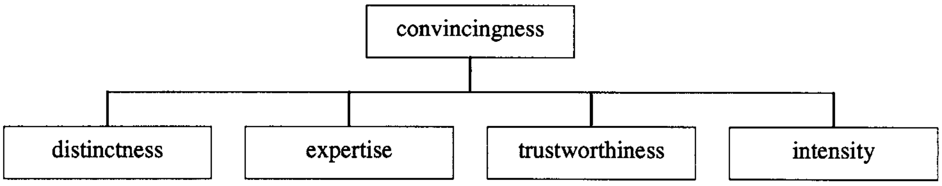


*Figure 1.* Model of convincingness.

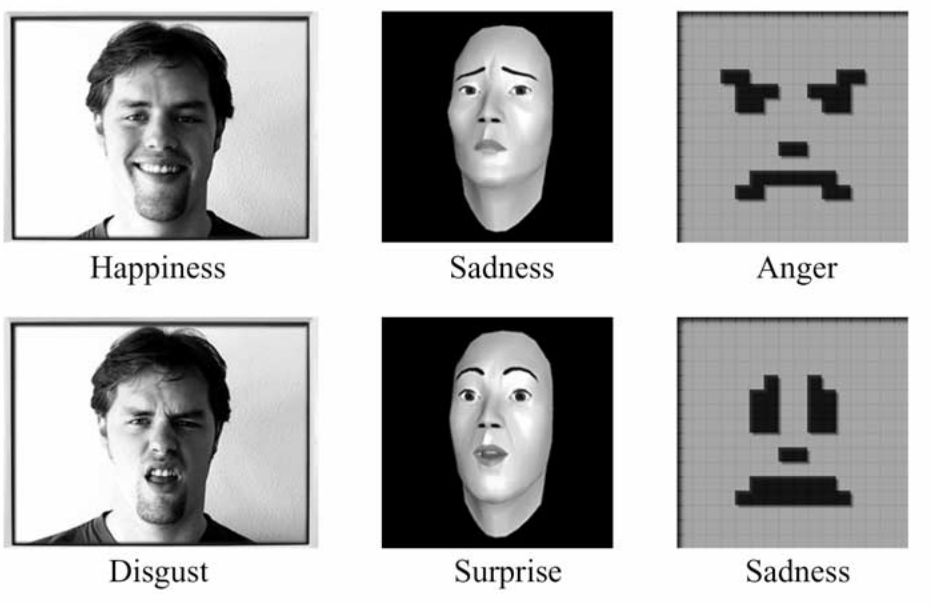


*Figure 2.* Examples of the stimuli.

to evaluate. The subjects, however, did not participate in the dice game because their own emotional state would influence their perception. Therefore, they only observed the game.

We created software that showed the game, the stimuli and the questions to the subjects on a computer screen (see Figure 3). The subjects judged the affective expression of one player. The opponent of this player was sitting behind a wall, invisible to the subjects. This setup ensured that the subject could not sympathize with one player, due to the gender, attractiveness or type (human or machine). Moreover, by focusing on one player the subjects did not need to constantly re-evaluate the situation from opposing points of view. A certain result in the game would be an advantage for one player and naturally a disadvantage for the other. It was clarified that none of the players bluffed or cheated since such a behavior would be useless

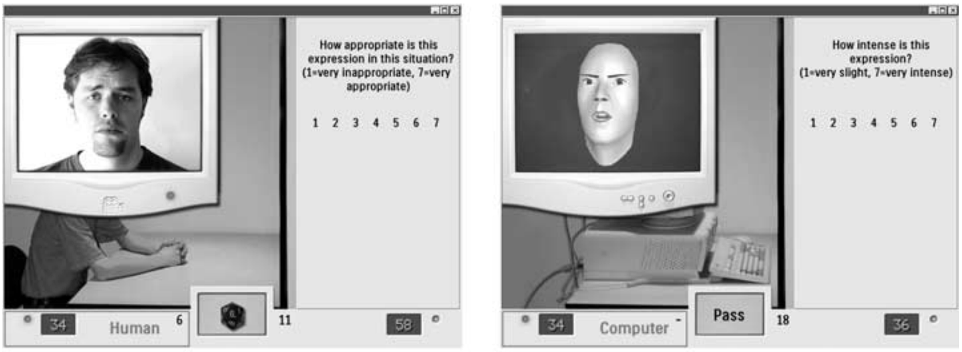


*Figure 3.* Screenshot of the software. Left: Human condition. Right: machine condition. The opponent rolled 11. The human/machine player rolled a 6 and therefore lost the round.

and against the rules of the game. None of the players would gain an advantage in the game by doing so. Keeping a poker face did not help the players to win.

The source of the expression was included as a factor, because humans might consider affective expressions from machines less convincing than expressions from humans. We used software to present the stimuli, because it is very difficult for humans to repeatedly produce exactly the same affective expressions. To distinguish the two conditions for the source of the expression we labeled the player either 'Human' or 'Machine'.

Furthermore, we used different background pictures. In the human condition a person was sitting at the table and in the machine condition a computer was placed on the table (see Figure 3). We expected the situation in which each expression occurred to have an influence on its perception. Therefore, a script, that was based on a pre-test, controlled the software and paired each stimulus with its specific situation in the game (context).

Another factor is the type of the affective expression. Machines need a clearly distinguishable vocabulary of them. Six affective expressions, plus a neutral expression, provide enough complexity to act appropriately in most situations (see Figure 2). A higher number of expressions might exceed the human capacity to process information ($7 \pm 2$ rule, Miller, 1956). The neutral expression was shown by default and changed into an affective expression for evaluation by the subjects. However, the expressive abilities of the machine might be limited. A mobile phone, for example, has only a small LCD display. It is impossible to present a human face in all its details on it. Therefore, it is important to test if the abstractions of an expression are convincing as well. We tested 3 levels of abstraction, which were based on typical applications in the area of consumer electronics.

Humans would use their own face and not an abstraction of it to express an emotion. Therefore, we only need to test one abstraction level in the human condition of the source. This will set the benchmark to which the machine's expressions will be compared (see Table III).

*Table II.* Levels of abstraction

| Product category | Product examples | Level of abstraction |
|---|---|---|
| Screens | TV, Monitor, Projector | Natural human face |
| Onscreen characters | Games, Virtual newsreader | Baldi (Real time 3D computer rendered face) |
| Small devices | Mobile phone, PDA | Matrix face ($10 \times 10$ pixel) |

Even so, no single modality predominates the perception of emotions (Ekman et al., 1980) a combination of modalities might be more convincing than each modality alone. Machines, such as mobile phones or TVs, are capable of presenting multimedia expressions. To reduce the complexity of the experiment we tested the media factor only in combination with the matrix face (see Table III). We presented either only the matrix face or the audio or the combination of both. The subjects were asked to evaluate the visual and audio stimuli as one expression if they appeared simultaneously. For practical reasons, this study focuses on content free media, such as facial expressions and abstract music.

### 5.2. METHOD

#### 5.2.1. *Manipulation*

A 2 (source) $\times$ 3 (abstraction) $\times$ 3 (media) $\times$ 6 (emotion) within subjects experiment was conducted. Certain factors were limited to certain conditions (see Table III and Section 5.1 for an explanation). Altogether 36 conditions were tested.

*Table III.* The 36 conditions of the experiment

| Source | Abstraction | Media | Emotion |
|---|---|---|---|
| Human | Natural face | Visual | happiness, sadness, anger, surprise, fear and disgust |
| Machine | Natural face | Visual | happiness, sadness, anger, surprise, fear and disgust |
| Machine | Baldi face | Visual | happiness, sadness, anger, surprise, fear and disgust |
| Machine | Matrix face | Visual | happiness, sadness, anger, surprise, fear and disgust |
| Machine | Matrix face | Audio | happiness, sadness, anger, surprise, fear and disgust |
| Machine | Matrix face | Audio and Visual | happiness, sadness, anger, surprise, fear and disgust |

#### 5.2.2. *Measures*

Convincingness, expertise, trustworthiness and intensity were measured by answering:

- How convincing is this expression?
- How appropriate is this expression in this situation?

- How trustworthy is this expression?
- How intense is this expression?

on a 1–7 scale (e.g. 1 = very unconvincing, 7 = very convincing). The distinctness of an expression was measured by the recognition accuracy of the subjects (forced choice between the 7 categories).

#### 5.2.3. *The Subjects*

Thirty-three employees (20 male 13 female) of the Technical University of Eindhoven, at the age between 21–61, participated in the experiment.

#### 5.2.4. *The Stimuli*

Three actors produced facial expressions that we photographed with a digital camera. They were asked to imagine an event in which each emotion was felt strongly. In a pre-test we analyzed the distinctness of their expressions and for the final experiment the expressions of the most successful actor were used. Baldi (CSLU, 1999) was used as an example for a typical Real-Time-3D Character. The quality of his expressions have been tested earlier (Etcoff & Magee, 1992). Professional designers created the matrix faces and the audio stimuli based on the information in Sections 3 and 4. They were optimized throuh several iterative circles of design and evaluation prior to this study. The audio stimuli consisted of short pieces of abstract music, similar to the beeps of R2D2 in the movie Star Wars.

*Table IV.* Pearson correlation coefficients for variables predicting convincingness across all conditions

| | Convincingness | Distinctness | Intensity | Expertise |
|---|---|---|---|---|
| Distinctness | 0.380* | — | | |
| Intensity | 0.677 | 0.280* | — | |
| Expertise | 0.787 | 0.377 | 0.418 | — |
| Trustworthiness | 0.874 | 0.180* | 0.736 | 0.666 |

* not significant $\alpha = 0.05$

#### 5.2.5. *Procedure*

Before the experiment, the subjects read an introduction text about the experiment and the dice game. They were explicitly instructed to distinguish between the trustworthiness and convincingness of the emotional expression (original text: ‘A car salesperson might be convincing, but not necessarily trustworthy.’) and between the type of player (human or computer). It was clarified that none of the players bluffed or cheated since such a behavior would be useless and against the rules of the game. None of the players would gain an advantage in the game by doing so. After reading the instructions the subjects played the game against the experimenter to become familiar with the rules.

Then, the subjects observed 4 training games with the software to get used to the interface. The software showed the questions and recorded the answers. In these training games they were confronted with all stimuli and all questions. In a short pause before the start of the experiment, the experimenter answered questions the subjects might have had about the process and the software. Afterwards the experimenter left the room. The subjects observed 6 games, each consisting of 30 rounds. One emotional expression occurred per round to which the subjects had to answer one question by clicking with the mouse on a response button such as a 1–7 scale or the list of emotions. The expressions of the players appeared either before (e.g. fear) or after (e.g. happiness) throwing the dice. The core experiment took 45 min to complete with a pause of 5 min in the middle. The subjects received small presents for their participation.

### 5.2.6. *Apparatus*

A lap-top with a 14" screen (800 × 600 pixels) was used to run the software. The stimuli were presented in a screen area (160 × 160 pixels) at the top-left, the questions and possible answers were presented in a screen area (300 × 600 pixels) at the right. A set of stereo-speakers were connected to the lap-top to play the audio stimuli.

## 6. Results

This paper focuses on the results for the model of convincingness and the scores for distinctness and convincingness. A more detailed description of this study is available elsewhere (Bartneck, 2000). Analyses of variance (ANOVA) were conducted on all dependent measures. Furthermore, a multiple regression analyses and several $t$-tests were performed on certain measures. The $\alpha$ level was set to 0.05 for all tests. Convincingness was measured on a 7-point scale ranging from $1 =$ very unconvincing and $7 =$ very convincing and distinctness was measured by the recognition accuracy (Table V–Table VIII).

*Table V.* Average convincingness and distinctness scores for each emotion across all conditions

| Emotion | Convincingness | Distinctness |
|---|---|---|
| Surprise | 5.68 | 93% |
| Happiness | 5.71 | 95% |
| Sadness | 5.25 | 90% |
| Disgust | 5.05 | 68% |
| Anger | 4.67 | 71% |
| Fear | 4.02 | 70% |

Table IV presents the correlation matrix for variables predicting convincingness across all conditions. A variance of 84.1 in convincingness can be predicted from distinctness, intensity, trustworthiness and expertise. Distinctness is only weakly correlated ($r = 0.380$) to convincingness and is not a significant (sig = 0.107) predictor. Both, convincingness ($r = 0.874$) and intensity ($r = 0.736$) are strongly correlated to trustworthiness. Intensity is not a significant (sig = 0.462) predictor for convincingness when trustworthiness is already considered in the analyses (collinearity). Trustworthiness alone predicts 75.6% of the variance in convincing- ness.

*Table VI.* Average convincingness and distinctness scores for each source across all the natural face conditions (see Figure 3)

| Source | Convincingness | Distinctness |
|---|---|---|
| Human | 5.14 | 84% |
| Machine | 5.08 | 89% |

The type of emotion has significant ($F[5, 160] = 29.696$, $p < 0.001$) influence on convincingness. Surprise and happiness were more convincing ($t[32] = 3.974$, $p < 0.001$) than sadness, disgust and anger which were more convincing ($t[32] = 3.562$, $p = 0.001$) than fear.

Distinctness is significantly influenced by the type of emotion ($F[5, 160] = 17.011$, $p < 0.001$) expressed. The scores for sadness (90%) were above ($t[32] = 4.478$, $p < 0.001$) the ones for anger (71%). There was no significant difference within the 'higher' scores, sadness (90%), happiness (95%) and surprise (93%) and within the 'lower' scores, disgust (68%) anger (71%) and fear (70%).

*Table VII.* Average convincingness and distinctness scores for each visual abstraction level of the machine condition

| Abstraction | Convincingness | Distinctness |
|---|---|---|
| Natural | 5.08 | 89% |
| Baldi | 5.10 | 94% |
| Matrix | 5.10 | 77% |

Knowledge about the source of the affective expression has no significant ($F[1, 32] = 0.379$, $p = 0.542$) influence on its convincingness. Only the scores for distinctness ($F[1, 32] = 4.238$, $p = 0.048$) were influenced a little (human 84%, machine 89%). However, this small difference is negligible.

The abstraction of an affective expression has no significant ($F[2, 64] = 0.008$, $p = 0.992$) influence on its convincingness. Only the scores for distinctness ($F[2, 64] = 20.873$, $p < 0.001$) were influenced significantly. The scores for the Baldi faces were higher ($t[32] = 2.262$, $p = 0.031$) than for the natural faces, which were above ($t[32] = 4.455$, $p < 0.001$) the ones for the matrix faces.

*Table VIII.* Average convincingness and distinctness scores for each media in the matrix face condition

| Media | Convincingness | Distinctness |
|---|---|---|
| Visual | 5.10 | 77% |
| Audio/Visual | 5.19 | 75% |
| Audio | 4.77 | 68% |

The medium used to express an emotion has significant ($F[2, 64] = 4.332$, $p = 0.017$) influence on its convincingness. Visual and audio/visual expressions were slightly more convincing ($t[32] = 2.089$, $p = 0.045$) than audio expressions. Distinctness was not significantly influenced.

*Table IX.* Pearson correlation coefficients for variables predicting convincingness'

| | Convincingness′ | Intensity |
|---|---|---|
| Intensity | 0.732 | — |
| Expertise | 0.747 | 0.418 |

## 7. Discussion

Distinctness is, against our expectations, not a significant predictor for convincingness. The reason for this outcome can probably be found in our methodology. It was impossible for the subjects to evaluate their choice, because we did not provide them with feedback about the correctness of their interpretation. Therefore, they rated the convincingness of the expressions independent of whether they interpreted the emotion correctly or not. They could make up their own interpretation of why this expression makes sense in this context. To confirm this finding we would need to perform a control experiment in which we provide both, matching and mismatched information about the type of the emotion. Even though distinctness is not a predictor for convincingness, communication would fail between the machine and the user if the expression is frequently misinterpreted. The expression would convince the user of the wrong circumstances.

The subjects were explicitly instructed to distinguish between trustworthiness and convincingness (text from instruction: 'A car sales person might be convincing but not necessarily trustworthy'). The strong correlation between trustworthiness and convincingness and the finding that trustworthiness alone predicts 75.6% of the variance in convincingness suggests that the difference between these two concepts is very small. The subjects might have even treated the words as synonyms. Therefore we would like to propose a new model for convincingness (see Figure 4). It merges convincingness and trustworthiness into a new variable (convincingness) and leaves out distinctness. It also solves the collinearity problem for intensity and trustworthiness.

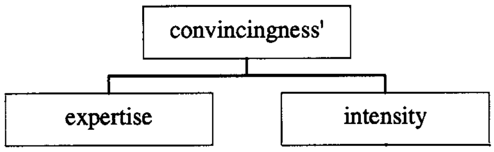


*Figure 4.* New model of convincingness.

We calculated the convincingness′ score for each subject by taking the average of his/her convincingness and trustworthiness scores. A variance of 75.6% in convincingness′ can be predicted from intensity and expertise and both are significant predictors (sig $< 0.001$, see Table IX).

Fogg's and Hsiang Tseng's (1999) model defined believability by its components trustworthiness and expertise. Our data suggest that the concepts of believability (synonym for convincingness) and trustworthiness are not distinct enough to be evaluated separately.

In this study, the type of emotion has the strongest influence on convincingness. The two 'positive' emotions happiness and surprise are rated highest on almost all variables. Anger and especially fear were rated lowest. Highly abstracted faces were as convincing as natural faces. Only the distinctness of an expression was influenced by its abstraction. Interestingly, the Baldi face (94%) scored higher than the natural face (89%). The distinctness of synthetic facial expression has reached the level of natural faces. Both scores are rather high compared to results of other studies (Bartneck, 2000). However, most of those studies did not provide context information with their stimuli. The source of the affective expression had no influence on its convincingness. This result is in line with the media equation (Nass & Reeves, 1996).

These results are particularly important for the development of affective embodied agents. We showed that affective expressions of machines, including agents, are perceived as convincing as affective expressions of humans. However, no affective embodied agent has passed the Turing test yet. The conversational skills of today's agents are limited. To be convincing, their embodiment should match their skills. An agent with inferior speech recognition, for instance, should not use a natural human face as its embodiment. The natural human face would raise conversational expectations that the agent would not live up to. Therefore the agent should use an abstracted embodiment. We showed that such abstracted embodied agents are able to express convincing affective expressions.

Our improved model of convincingness suggests that an affective embodied agent should express the right emotion at the right time with the right intensity. Further research is necessary to enable them to do so.

## 8. Conclusions

We created a literature review of affective expressions for speech, music and body language by summarizing results of previous studies on the quality and quantity of their parameters and successful examples of synthesis.

We applied this knowledge to create a vocabulary of facial and audio expressions that we improved through several iterative circles prior to this study. We proved this vocabulary to be more distinct than most previous designs (Bartneck, 2000).

The affective expressions of machines are as convincing as expressions of humans. Our results support the work of Nass and Reeves (1996). We showed that abstracted expressions are as convincing as natural human faces. Their distinctness, however, decreases with a higher level of abstraction. At a certain point, communication would fail due to frequent misinterpretations of the expressions. This problem can be avoided by leaving out less distinct emotion categories, such as fear.

Our initial model of convincingness based on Fogg's and Hsiang Tseng's model of believability does not fit affective expressions and therefore we proposed a new model that is based on appropriateness and intensity of the expression. Both, the influence of the context and the relation between gradients of intensity and appropriateness are interesting subjects for further research.

In short, the vocabulary of affective expressions is working, but further research on the grammar and the etiquette is necessary.

## Acknowledgements

This study was conducted at Philips Research Laboratories, Eindhoven, The Netherlands. The author would like to thank Don Bouwhuis and Gerard Hollemans for their help and support. This study would have been impossible without the support of Elmo Diederiks.

Correspondence concerning this article should be addressed to Christoph Bartneck, IPO, Den Dolech 2, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, christoph@bartneck.de or at Philips Research, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands.

## References

Bartneck, C.: 2000, *Affective Expressions of Machines*, Technical University of Eindhoven, The Netherlands. Also available at http://www.bartneck.de/work/aem.pdf

Barnes, A. and Thagard, P.: 1996, Emotional decisions. *Proceedings of the Eighteen's Annual Conference of the Cognitive Science Society, University of California*, pp. 426–429.

Brennan, S. E.: 1985, The caricature generator. *Leonardo* **18**, 170–178.

Bresin, R. and Friberg, A.: 1999, Synthesis and decoding of emotionally expressive music performance. *Proceedings of the IEEE 1999 Systems, Man and Cybernetics Conference (SMC'99)*, Tokyo, pp. 317–322.

Cahn, J. E.: 1990, The generation of affect in synthesized speech. *Journal of American Voice I/O Society,* **8**, 1–19.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, W., Douville, B., Prevost, S. and Stone, M.: 1998, Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In: M. Maybury and W. Wahlster (eds.), *Readings in Intelligent User Interfaces*, Morgan Kaufmann Publishers.

Crowder, R. G.: 1984, Perception of the major/minor distinction: I. Historical and theoretical foundations. *Psychomusicology*, **4**, 3–10.;

CSLU, 1999, CSLU Toolkit. http://cslu.cse.ogi.edu/toolkit/

Cunningham, J. G. and Sterling, R. S.: 1988, Development change in the understanding of affective meaning in music. *Motivation and Emotion,* **12**, 399–413.

Davies, J. B. (ed.): 1978, *The Psychology of Music*. Stanford, Stanford University Press.

Ekman, P. (ed.): 1973, *Darwin and Facial Expression: A Century of Research in Review*. New York: Academic Press.

Ekman, P., Friesen, W. V. and Ellsworth, P. (eds.): 1972, *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. New York: Pergamon Press.

Ekman, P., Friesen, W.V., O'Sullivan, M. and Scherer, K.: 1980, Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, **38**, 270–277.

Ekman, P. and Frieser, W. (eds.): 1976, *Pictures of Facial Affects*. Palo Alto: Consulting Psychologist Press.

Etcoff, N. L. and Magee, J. J.: 1992, Categorical perception of facial expressions. *Cognition*, **44**, 227–240.

Feist, G.: 1994, The affective cpmsequences of artistic and scientific problem solving. *Cognition and Emotion*, **8**, 489–502.

Fogg, B.J. and Hsiang Tseng: 1999, The elements of computer credibility. *Proceedings of CHI 99*, Pittsburgh, pp. 80-87.

Frijda, N. H. (ed.): 1986, *The Emotions*. Cambridge, Cambridge University Press.

Frijda, N. H.: 1969, Recognition of emotion. *Advances in Experimental Social Psychology*, **4**, 167–223.

Gerardi, G. M. and Gerken, L.: 1995, The development of affective responses to modality and melodic contour. *Music Perception*, **12**, 279–290.

Izard, C. E. (ed.): 1977, *Human Emotions*. New York: Plenum Press.

Juslin, P. N.: 1997a, Emotional communication in music performance: a functionalist perspective and some data. *Music Perception*, **14**, 383–418.

Juslin, P. N.: 1997b, Perceived emotional expression in synthesized performances of a short melody: capturing the listener's judgment policy. *Musicae Scientiae*, **2**, 225–256.

Kastner, M. P. and Crowder, R. G.: 1990, Perception of major/minor: IV. Emotional connotation in young children. *Music Perception*, **8**, 189–202.

Krauss, R. M., Morrel-Samuels, P. and Colasante, C.: 1991, Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, 743–754.

Lazarus, R. S. (ed.): 1991, *Emotion and Adaptation*. Oxford, Oxford University Press.

Massaro, D. W. (ed.): 1998, *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. MIT Press.

McNeill, D. (ed.): 1992, *Hand and Mind: What Gestures Reveal about Thought*. Chicago, University of Chicago.

Miller, G.A.: 1956, The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review,* **63**, 81–97.

Mozziconacci, S. (ed.): 1998, *Speech Variability and Emotions: Production and Perception*. Ph.D. thesis, IPO, Technical University of Eindhoven.

Murray, I. R. and Arnott, J. L.: 1992, Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustic Society of America*, **93**, 1097–1108.
Nass, C. and Reeves, B. (eds.): 1996, *The Media Equation*. Cambridge, SLI Publications, Cambridge University Press.
Nielsen, J. (ed.): 1994, *Usability Engineering*. San Diego: AP Professional.
Nobe, S., Hayamizu, S., Hasegawa, O. and Takahashi, H.: 1997, Are listeners paying attention to the hand gestures of an anthropomorphic agent? An evaluation using a gaze tracking method'. In: I. Wachsmuth and M. Fröhlich (eds.), *Gesture and Sign Language in Human-Computer Interaction*, Springer.
Norman, D. A.: 1981, Twelve issues for cognitive science. In: D. Norman (ed.), *Perspectives on Cognitive Science*, Norwood: Ablex Publishing, pp. 265–295.
Nowlis, V.: 1966, Research with the mood adjective check list. In: S. S.Tomkin and C. E. Izard (eds.), *Affect, Cognition and Personality*, New York, Springer, pp. 352–389.
Oatley, K. and Jenkins, J. M. (eds.): 1996, *Understanding Emotions*. Oxford: Blackwell.
Osgood, C. E., Suci, G. J. and Tannenbaum, P. H. (eds.): 1957, *The Measurements of Meaning*. Champaign: University of Illinois Press.
Picard, R. (ed.): 1997a, *Affective Computing*. Cambridge: MIT Press.
Picard, R.: 1997b, Does HAL cry digital tears? Emotion and computers. In: D. Stork (ed.): *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge: MIT Press.
Pixar, 1998, Geri's game. http://www.pixar.com/shorts/shorts.html
Plutchik, R. (ed.): 1980, *Emotion: A Psycho Evolutionary Synthesis*. New York, Harper and Row.
Russel, J. A.: 1979, Affective space is bipolar. *Journal of Personality and Social Psychology*, **37**, 345–356.
Scherer, K. R.: 1979, Nonlinguistic vocal indicators of emotion and psychopathology. In: C. E. Izard (eds.), *Emotions in Personality and Psychopathology*, New York, Plenum.
Scherer, K. R. and Oshinsky, J. S.: 1977, Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, **4**, 331–346.
Schiano, D. J., Ehrlich, S. M., Rahardja, K. and Sheridan, K.: 2000, Face to InterFace: Facial affect in (Hu)Man and Machine. *CHI 2000 Conference Proceedings*, The Hague, pp. 193–200.
Schlossberg, H.: 1954, Three dimensions of emotion. *Psychological Review*, **61**, 81–88.
Smith, C. A. and Ellsworth, P. C.: 1985, Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, **48**, 813–838.
Sony, 1999, Aibo. http://www.world.sony.com/aibo/frame13.html

## Authors' Vita

**Christoph Bartneck** is a Ph.D. student in User-System Interaction at the Technical University of Eindhoven, The Netherlands. He received his B.A. in Industrial Design/Information Science from the FH Design and Media of Hannover, Germany in 1997, and his M.T.D. degree in User-System Interaction from the Technical University of Eindhoven in 2000. His primary interests lie in the areas of animated emotional agents and robotics, although previous work has included projects in website design and CD-ROM development for Eagle River Interactive (USA), LEGO A/S (Denmark) and the Technology Centre Hannover GmbH (Germany). His paper summarises his thesis work on Affective Expressions of Machines.