Affective Expressions of Machines

Christoph Bartneck

Christoph Bartneck

" Affective Expressions of Machines "/ Bartneck, Christoph;
Eindhoven: Stan Ackerman Institute. -III.
Final report of the postgraduate program: User System Interaction
With reference
ISBN 90-444-0027-4
Keywords: Expression / Emotion /

Title: Affective Expressions of Machines

Author: Christoph Bartneck

Part of project: Emotion and Experience

Keywords: **affective expressions, emotion, speech, face, music, convincingness, distinctness, abstraction, modality**

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, their recognition accuracy 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.

Contents

1 Foreword

The Self Adapting Media Systems cluster (SAMS) of the New Media Systems and Interaction group (NMSA; Aarts), part of Philips Research, has set up the *Emotion and Experience* project to meet the challenges in the field of emotional computing and user experience. This subproject, *Affective Machines*, focuses on the expression of emotions by machines to improve the user experience.

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 and Thargad, 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 and Jenkins, 1996). Emotions should play an important role in the design of interfaces (Picard, 1997a, 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).

The Project is divided into three phases. This report deals with the first phase.

Phase 1: Affective Expressions of Machines
In this phase we transfer human emotional expression to machines and tested their quality.

Phase 2: Affective Architectures of Machines
In this phase we develop an emotion architecture for machines based on human emotion theory.

Phase 3: Implementation of an Affective Machine
In this phase we implement the results of phase 1 and 2 into a prototype to test their effect on the usability of the product.

Phase 1 - Affective Expressions of Machines

2 Introduction

To investigate if and how machines can express emotions we first need to look at human-human interaction. How do humans express emotions and how accurate do they perceive them? 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 philips, not in the scope of this study.

3 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, Suci and Tannenbaum, 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 and Ellsworth, 1985; Schiano, Ehrlich, Rahardja and Sheridan, 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.

Emotion theory is not directly relevant for phase 1 of the project, but will be discussed in more detail in phase 2 “Affective Architectures of Machines “. Due to practical reasons, this study takes the categorical viewpoint on emotions. Many studies (Ekman, Friesen and Ellsworth, 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.

4 How do humans express emotions?

Expressing emotions is a natural act for humans. The ingenuous ease of it contrasts with the difficulty to describe 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 if they learned to do it they are usually still unable to explain how they do it. Many studies have been performed to find out how humans express emotions. The following paragraphs summarize some of their results.

4.1 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 emotional 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. Table 1 summarizes these effects.

Table 1: Speech parameter settings for several emotions

	Anger	Happiness	Sadness	Fear	Disgust
Speech rate	Slightly faster	Faster or slower	Slightly slower	Much faster	Very much slower
Pitch average	Very much higher	Much higher	Slightly lower	Very much higher	Very much lower
Pitch range	Much wider	Much wider	Slightly narrower	Much wider	Slightly wider
Intensity	Higher	Higher	Lower	Normal	Lower
Voice quality	Breathy, chest tone	Breathy, blaring	Resonant	Irregular voicing	Grumbled chest tone
Pitch change	Abrupt, on stressed syllables	Smooth, upward inflections	Downward inflections	Normal	Wide, downward terminal inflections
Articulation	Tense	Normal	Slurring	Precise	normal

Note: From Murray (1992)

The quantification of the speech parameters in this table is 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 respectively, the minimum and maximum influence. 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.

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 of 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%.

4.2 Music

Music is a difficult method to express emotions because culture (Davies, 1978; Crowder, 1984), skills of the performer (Bresin and 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 and Sterling, 1988; Geradi and Gerken, 1995; Kastner and Crowder, 1990).

Scherer and Oshinsky (1977) demonstrated that 66% to 75% of the variance in the emotional 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 emotional expressions in speech. Juslin (1997b) summarized expressive principles which he obtained by a series of studies using several different instruments, performers and melodies (Table 2).

Table 2: Music parameter settings for several emotions

Emotion	Parameters
Happiness	Fast tempo, moderate variations in timing, moderate to loud sound level, tendency to (relatively) sharpen contrast between "long" and "short" notes (as in dotted patterns), mostly staccato articulation, fast tone attacks, bright timbre, light or no vibrato.
Sadness	Slow tempo, relatively large deviation in timing and low sound level, tendency to (relatively) soften contrasts between "long" and "short" notes, legato articulation, slow tone attacks, slow and deep vibrato, final ricard, soft timbre, flat intonation.
Anger	Fast tempo, high sound level, tendency to (relatively) sharpen contrast between "long" and "short" tones, no final ricard, mostly non legato articulation, very sharp tone attacks, sharp timbre, distorted tones.
Fear	Large tempo variations, large deviation in timing, very low sound level, large dynamic variation, mostly staccato articulation, fast and irregular vibrato, pauses between phrases, and soft spectrum.
Tenderness	Slow tempo, relatively large deviations in timing, low to moderate sound level, tendency to (relatively) soften contrast between "long" and "short" notes, legato articulation, slow tone attacks, soft timbre, and intense vibrato.

Note: From Juslin (1997b)

A promising synthesis program for emotional expression in music performance is the *Director Musices* by Bresin and Friberg (1999). 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).

4.3 Body language

Pantomimes use only facial 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, Friesen, O'Sullivan and Scherer, 1980).

4.3.1 Facial expression

Expressing emotions through the face is so natural for humans that it takes a considerable amount of effort to mask them (Ekman, Friesen and Ellsworth 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 emotional 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. 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 emotional expressions accurately. The results of the present study support his findings.

4.3.2 Gesture

Gestures occur to 90% only during speech (McNeill, 1992). They convey some information, but they are not richly informative and the information conveyed is largely redundant with the presence of speech (Krauss, Morrel-Samuels and Colasante, 1991). Still, people pay attention to them (Nobe, Hayamizu, Hasegawa and Takahashi, 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:

Table 3: Description of several gestures

Gestures	Description
Iconic	Represents some feature of the accompanying speech such as sketching a small rectangular space with one's tow hands while saying "Do you have a blank CHECK?"
Metaphoric	Represents an abstract feature concurrently spoken about, such as forming a jaw-like shape with one hand and pulling it towards one's body while saying "Then I can WITHDRAW fifty dollars for you".
Deictics	Deictics indicate a point in space. The accompany reference to persons, places and other spatial discourse entities. An example might be pointing to the ground while saying "Do you have an account at THIS bank?"
Beats	Beats are small formless waves of the hand that occur with heavily emphasized words, occasions of turning over the floor to another speaker and other kinds of special linguistic work. An example is waving one's hand briefly up and down along with the phrase "all right".

Note: From McNeill (1992)

4.3.3 Body movement

Most of the descriptive studies on emotional body movement are informal (Frijda 1986). Table 4 summarizes Frijda's analyses:

Table 4: 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 breating, 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, 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)

A promising synthesis of body language and speech is the work of several members of the Department of Computer & 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.

5 How accurate do humans perceive emotions?

The emotional expressions of machines will always be compared to the ones of humans. Therefore, human-human interaction sets the benchmark for human-machine interaction. The literature overview below lists studies in the field of emotion decoding, which used almost the same emotion categories. Unfortunately, it is invalid to compare the results or to calculate averages, because their methodologies vary to a large extent. The experiments on facial expressions, for example, vary on the number of sampled persons for the stimuli, number/type of subjects and emotion categories offered to the subjects. Note that other categories might have been used besides the ones listed.

Table 5: Recognition accuracy of facial expressions

Study	Happy	Surprise	Fear	Anger	Sad	Disgust	Average	Chance	Comment
Drag, Shaw,1967	71	68	62	42	49	41	55.5	11.1	
Dusenbury , Knower 1938	100	86	93	92	84	91	91.0	9.1	
Etcoff, Magee, 1992	-	-	-	-	-	-	>90	-	Drawing of face
Kanner 1931	-	76	75	32	33	66	56.4	-	
Kozel, Gitter, 1969	86	69	80	79	59	55	71.3	14.3	
Levitt ,1964	86	43	58	62	-	45	58.8	20.0	
Massaro, 1998	>90	-	-	-	>90	-	-	50.0	Synthesized face
Thomson, Metzler, 1964	76	-	74	60	52	67	65.8	14.3	
Woodworth , 1938	93	77	66	31	70	74	68.5	14.3	
Zuckerman, et al. , 1975	62.2	40	19.8	33.3	38.9	40.1	39.1	16.7	

Table 6: Recognition accuracy of speech

Study	Happy	Surprise	Fear	Anger	Sad	Disgust	Average	Chance	Comment
Cahn, 1990	48.2	43.9	51.8	43.9	91	42.1	53.5	16.7	Synthesized speech
Fenster, et al., 1971	18	-	32	44	28	-	30.5	16.7	
Levitt, 1964	-	-	-	-	-	-	59	10.0	
Mozziconacci, 1998	73	-	33	43	93	88	66.0	20.0	
Williams, Stevens, 1981	28	-	27	51	73	-	44.8	25.0	
Zuckerman, et al. ,1975	35	51.6	32.3	48.5	68.3	36.4	45.4	16.7	

Table 7: Recognition accuracy of music

Study	Happy	Surprise	Fear	Anger	Sad	Disgust	Average	Chance	Comment
Bresin, Friberg, 1999	77	-	46.5	89.5	63	-	69.0	16.7	
Cunningham, Sterling, 1988	96.96	-	65.47	70.13	94.62	-	81.8	25.0	Only 19 year olds
Juslin, 1997b	75	-	96	96	92	-	89.8	20.0	
Kastner, Crowder, 1990	-	-	-	-	-	-	66.2	50.0	Only children

Table 8: Recognition accuracy of body language

Study	Happy	Surprise	Fear	Anger	Sad	Disgust	Average	Chance	Comment
Kline, Johannssen, 1935	96.0	55	28.8	70.8	74.5	-	65.02	5	Terror as fear

Table 5-Table 8: Numbers in percent. Average is based on the numbers listed. The chance level is based on all categories offered to the subjects.

6 How do machines express emotions?

Machines are able to express emotions. Almost all experiments, which tested emotional 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.

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.

All emotional expressions of machines are abstractions of human expressions. Even movies with actors talking to each other are not the real people and therefore an 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 them 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 use human expressions or their abstractions to communicate emotions.

Emotional expression should be used by all media available to a machine. A machine with a speaker machine, for example, should use it to express emotions. However, it would make no sense to add a display to that machine just to show an expressive face. Another example is agent technology. All media used by an agent to communicate with the user should express emotions. If he is able to speak, affect should be added to his verbal expressions. However, the decision whether an agent in a certain machine should be able to talk or not depends on many factors and the ability to express emotions is compared to those only of minor importance.

Now that we have an overview on how humans express emotions and how accurate the recognize them and that we know that machines can express emotions we shall investigate possible concerns for affective expressions of machines.

7 The experiment

7.1 Introduction

In this chapter we introduce the choices and motivations for our experiment. An important attribute of affective expressions of machines is their convincingness. We consider the concept of convincingness as an extension of Fogg's and Hsiang Tseng's (1999) concept of believability. We added the intensity and the distinctness of the stimuli to their original definition, because they are particularly important for affective systems. Distinctness is the attribute of the expression, that is measured by the recognition accuracy of the subjects. We measure believability indirectly through expertise and trustworthiness. The expertise of the system is measured by the perceived appropriateness of its expressions.

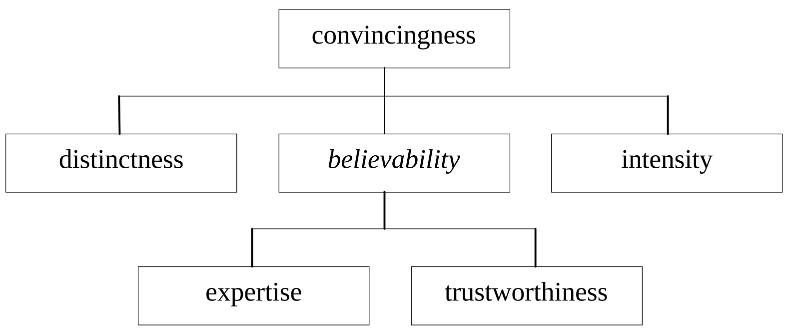


Figure 1: Model of convincingness

The subjects would not be able to judge trustworthiness and appropriateness (measures for expertise) of the expression without information about the context. In real life, context information will always be available. The jamming of an affective CD player, for example, would be associated with its sad emotional expression. A simple dice game is an appropriate context for the stimuli in this experiment, because it requires only a small amount of learning from the subjects and it is easy 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. The subjects judged the emotional 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. None of the players bluffed or cheated to gain an advantage in the game.

The source of the expression was included as a factor, because humans might consider emotional expressions from machines less convincing than expressions from humans. We created software that presented the stimuli, because it is very difficult for humans to repeatedly produce the exact same emotional expression. The software also showed the game and the questions to the subjects. 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 context in which each expression occurred to have influence on its perception. Therefore, a script, that was based on a pre-test, controlled the software and paired each stimulus with its specific context.

Another factor is the type of the emotional expression. Machines need a clearly distinguishable vocabulary of them. Six emotional expressions, plus a neutral expression, provide enough complexity to act appropriately in most situations. A higher number of expressions might exceed the human capacity to process information (7 ± 2 rule, Miller 1956). 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.

Table 9: Levels of abstraction

Product category	Product examples	Level of abstraction
Screens	TV, Monitor, Projector	Detailed human face
Onscreen characters	Games, Virtual newsreader	Real time 3D computer rendered face (Baldi)
Small devices	Mobile phone, PDA	10x10 pixel matrix face

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.

Even so no single modality predominates the perception of emotions (Ekman, Friesen, O'Sullivan and Scherer, 1980) a combination of modalities might be perceived more convincing than each modality alone. Machines, such as a mobile phones or TVs, are capable of presenting multimedia expressions. To reduce the complexity of the experiment we tested this factor only in the machine/matrix condition. For practical reasons, this study focuses on content free media, such as facial expressions and abstract music.

7.2 Method

7.2.1 Manipulation

A 2 (source) x 6 (emotion) x 3 (abstraction) x 3 (media) within subjects experiment was conducted. Certain factors were limited to certain conditions (see Table 10). Altogether 36 conditions were tested.

Table 10: Independent variables

Independent Variable	Conditions	Comment
Source of the expression	Human, Machine	
Type of emotion	Happiness, Sadness, Anger, Surprise, Disgust and Fear	A neutral face was shown as default
Level of abstraction	Natural, Baldi and Matrix	Only natural face for human condition (source)
Media	Visual, Audio/Visual and Audio	Only within the machine/matrix condition

7.2.2 Measures

Convincingness, expertise, trustworthiness and intensity were measured by answering a question (see Table 11) 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 emotions).

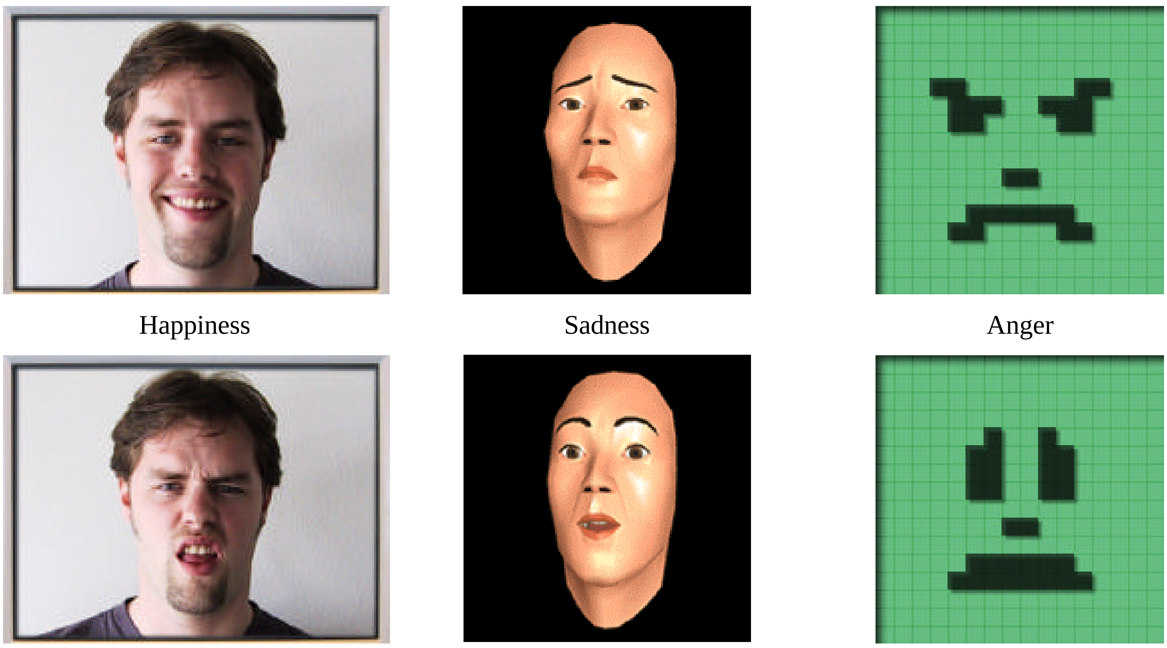
Table 11: Example Questions for the dependent variables

Dependent variable	Question
Distinctness	What emotion is the human/machine expressing?
Intensity	How intense is the expression?
Expertise	How appropriate is this expression in this situation?
Trustworthiness	How trustworthiness is this expression?
Convincingness	How convincing is this expression?

7.2.3 The Subjects

33 employees (20 male 13 female) of the IPO (Center for User-System Interaction, Eindhoven, The Netherlands), at the age between 21 to 61, participated in the experiment.

7.2.4 The stimuli



Happiness | Sadness | Anger

Disgust | Surprise | Fear

Figure 2: Examples of 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 was used as an example for a typical Real-Time-3D Character. The quality of his expressions have been tested earlier (Etcoff and Magee, 1992). Professional designers created the matrix faces and the audio stimuli (abstract music). They were optimized through several iterative circles of design and evaluation.

7.2.5 Procedure

Before the experiment, the subjects read an introduction text in which they were explicitly instructed to distinguish between trustworthiness and convincingness. ("A car sales person might be convincing but not necessarily trustworthy.") and between the type of player (human or computer). They were also told that none of the players bluffed or cheated to gain an advantage in the game. Afterwards, the subjects played the game against the experimenter to become familiar with the rules.

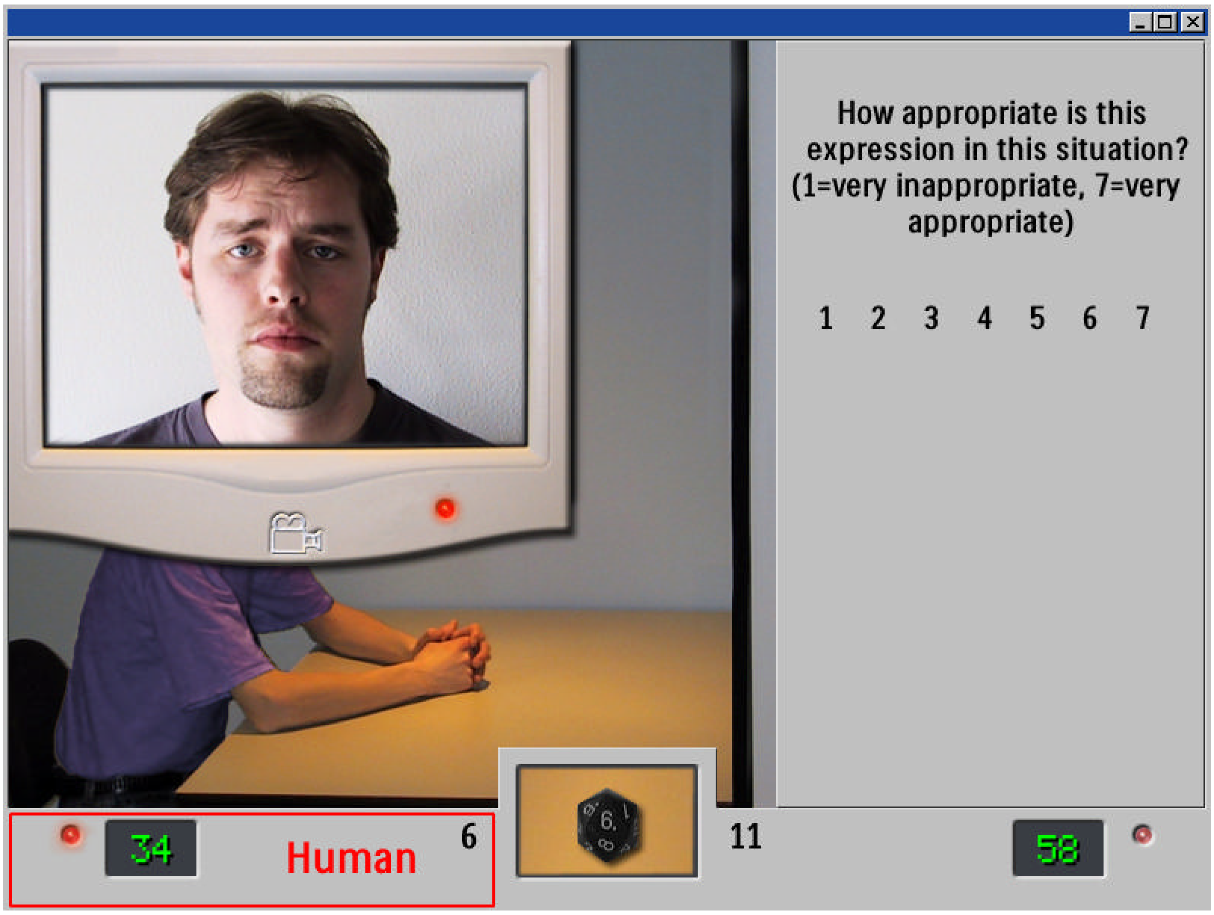


Figure 3: Screenshot of the software. The opponent rolled 11. The human player rolled a 6 and therefore lost the round.

Then, the subjects performed 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 about the process and the software. Afterwards the experimenter left the room. The subjects played 6 games, each consisting of 30 rounds. The subjects had to answer one question per round by clicking with the mouse on a response button such as a 1-7 scale or the list of emotions. The core experiment took 45 minutes to complete with a pause of 5 minutes in the middle. The subjects received small presents for their participation.

7.2.6 Apparatus

A lap-top with a 14” screen (800x600 pixels) was used to run the software. The stimuli were presented in a screen area (160x160 pixels) at the top-left, the questions and possible answers were presented in a screen area (300x600 pixels) at the right. A set of stereo-speakers were connected to the lap-top to play the audio stimuli.

8 Results

ANOVA's were conducted on all dependent measures. Furthermore, a multiple regression analyses and several t-tests were performed on certain measures. The α level was set to 0.05 for all tests.

Table 12: 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 α=.05

Table 12 presents the correlation matrix for variables predicting convincingness. 84.1% of the variance in convincingness can be predicted from distinctness, intensity, trustworthiness and expertise. Distinctness is only weakly correlated (r=.380) to convincingness and is not a significant (sig.=.107) predictor. Both, convincingness (r=.874) and intensity (r=.736) are strongly correlated to trustworthiness. Intensity is not a significant (sig.=.462) predictor for convincingness when trustworthiness is already considered in the analyses (collinearity). Trustworthiness alone predicts 75.6 % of the variance in convincingness.

The type of emotion has significant (F[5,160]=29.696, p<.001) influence on convincingness. Surprise and happiness were more convincing (t[32]=3.974, p<.001) than sadness, disgust and anger which were more convincing (t[32]=3.562, p=.001) than fear.

Trustworthiness depends significantly on the type of emotion (F[5,160]=27.43, p<.001) expressed. Surprise and happiness were slightly more trustworthy (t[32]=2.588, p=.014) than disgust and sadness. Disgust and sadness were more trustworthy (t[32]=2.959, p=.006) than and anger, which was more trustworthy (t[32]=2.987, p=.005) than fear.

Expertise is influenced significantly by the type of emotion (F[5,160]=20.035, p<.001) expressed. The scores for surprise, happiness, sadness and disgust were higher (t[32]=3.848, p=.001) than for anger and fear.

Intensity depends significantly on the type of emotion (F[5,160]=6.258, p<.001) expressed.
Surprise and disgust were slightly more intense (t[32]=3.848, p=.001) than happiness, sadness, anger and fear.

Distinctness is significantly influenced by the type of emotion (F[5,160]=17.011, p<.001) expressed. The scores for sadness (90%) were above (t(32)=4.478, p<.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, fear (70%) and disgust (68%).

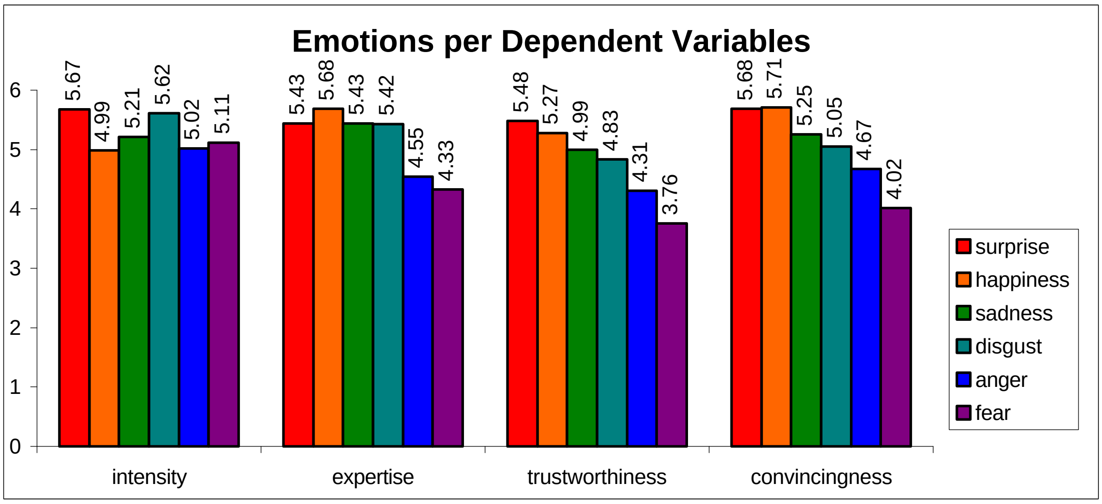


Figure 4: Emotions per dependent variables

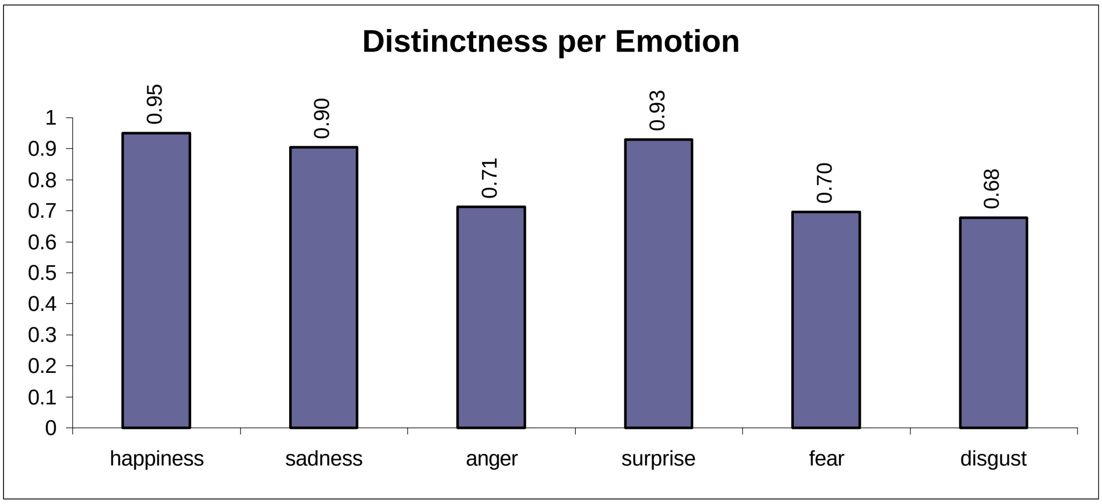


Figure 5: Distinctness per emotion

Knowledge about the source of the emotional expression has no significant ($F[1,32]=.379$, $p=.542$) influence on its convincingness. Only the scores for distinctness ($F[1,32]=4.238$, $p=.048$) and trustworthiness ($F[1,32]=5.092$, $p=.031$) were influenced a little.

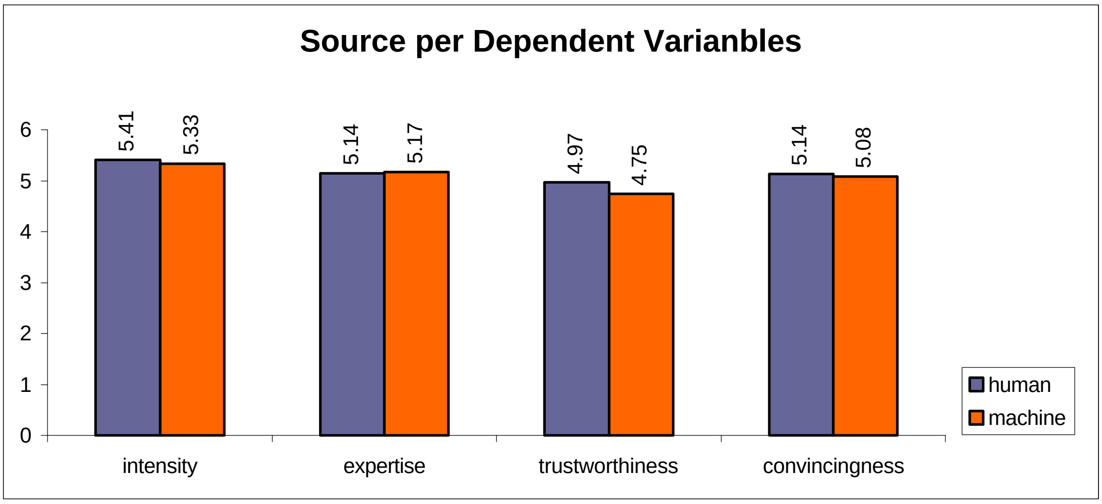


Figure 6: Source per dependent variables

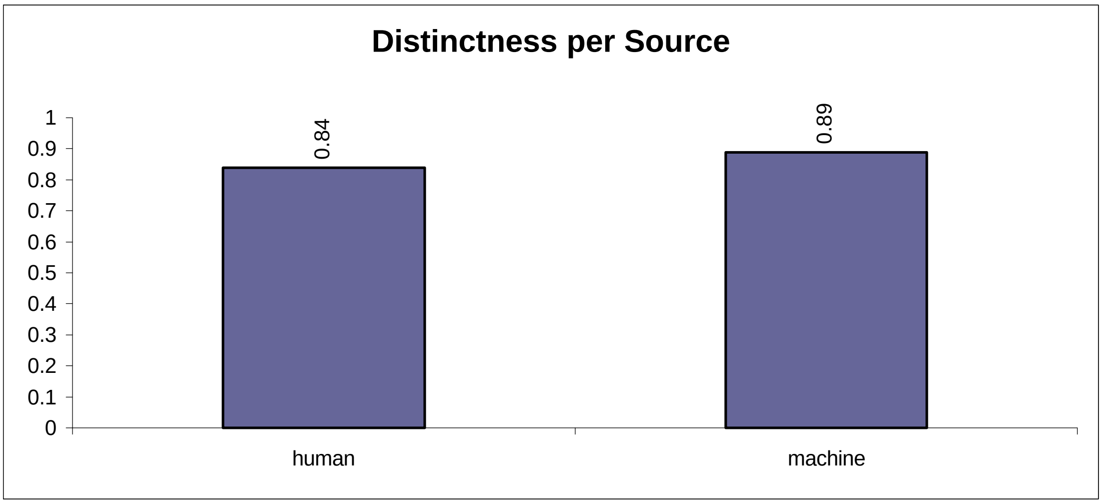


Figure 7: Distinctness per source

The abstraction of an emotional expression has no significant F(2,64)=.008, p=.992 influence on its convincingness. Only the scores for distinctness (F(2,64)=20.873, p<.001) were influenced significantly. The scores for the Baldi faces (94%) were higher (t[32]=2.262, p=.031) than for the natural faces (89%), which were above (t[32]=4.455, p<.001) the ones for the matrix faces (77%).

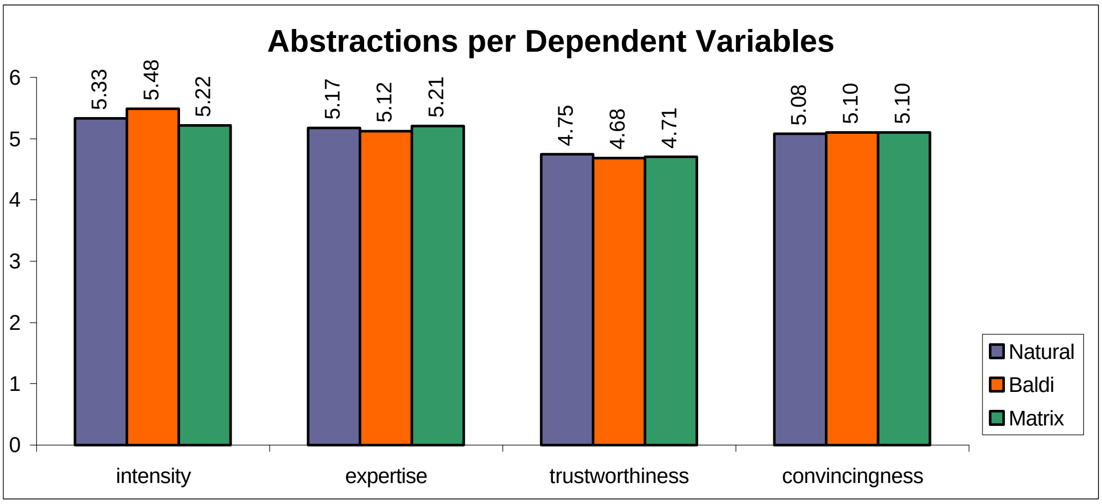


Figure 8: Abstraction per dependent variables

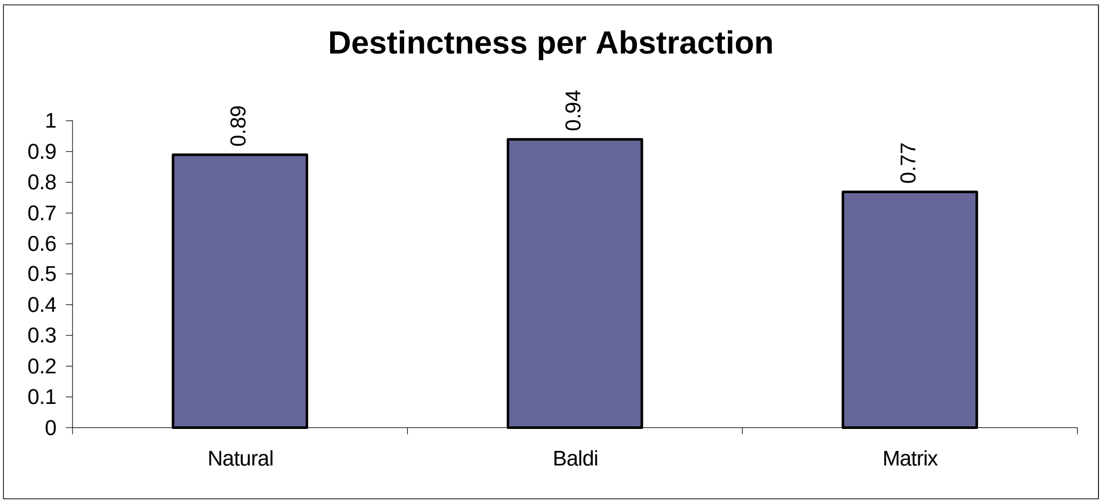


Figure 9: Distinctness per abstraction

The media of an emotional expressions has significant (F[2,64]=4.332, p=.017) influence on its convincingness. Visual and audio/visual expressions were slightly more convincing (t(32)=2.089, p=.045) than audio expressions.

Trustworthiness is significantly influenced (F[2,64]=7.535, p=.009) by the media of the expression. Audio/Visual expression were (t(32)=2.545, p=.016) more trustworthy than the audio expressions alone and only little more convincing than the visual expressions alone.

Intensity significantly depends (F[2,64]=9.349, p<.001) on the media of the expression. Visual and audio/visual expressions were (t[32]=2.623, p=.013) more intense than audio expressions alone.

Expertise and Distinctness (77% visual, 75% audio/visual, 68% audio) were not significantly influenced.

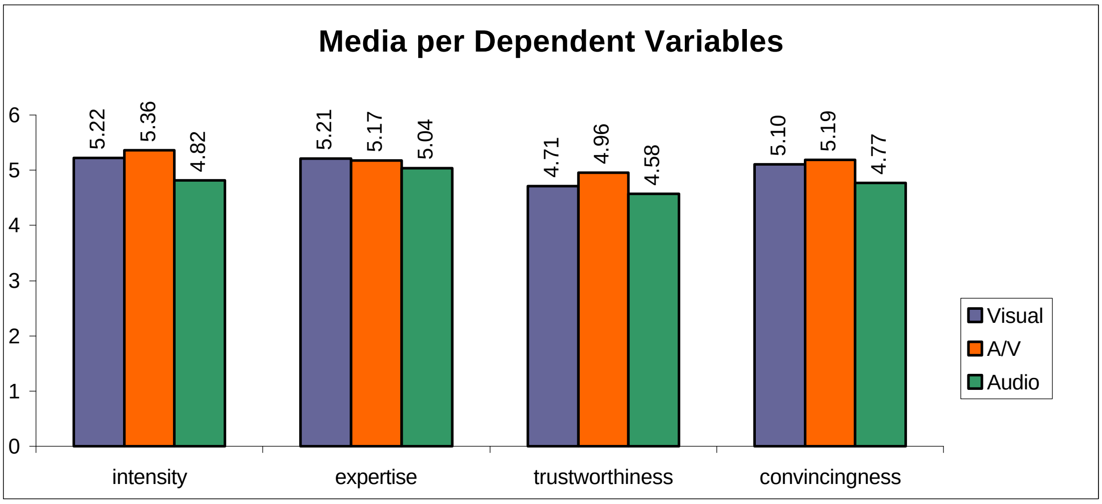


Figure 10: Media per dependent variables

We did not find a significant gender difference.

We compared the distinctness scores in the experiment for each stimulus with their results in the pretest. The final experiment provided context information and the pretest did not.

The context has significant ($F[1,53]=15.844$, $p<.001$) influence on the distinctness of the expressions. The distinctness scores for the audio stimuli increased ($t[53]=2.289$, $p=.026$) from 55% to 68%. The scores for the matrix faces raised ($t[55]=3.191$, $p=.002$) from 62% to 77% and the scores for the Baldi faces increased ($t[28.365]= 4.497$, $p<.001$) from 73% to 94%. Only the distinctness scores for the natural face did not change significantly.

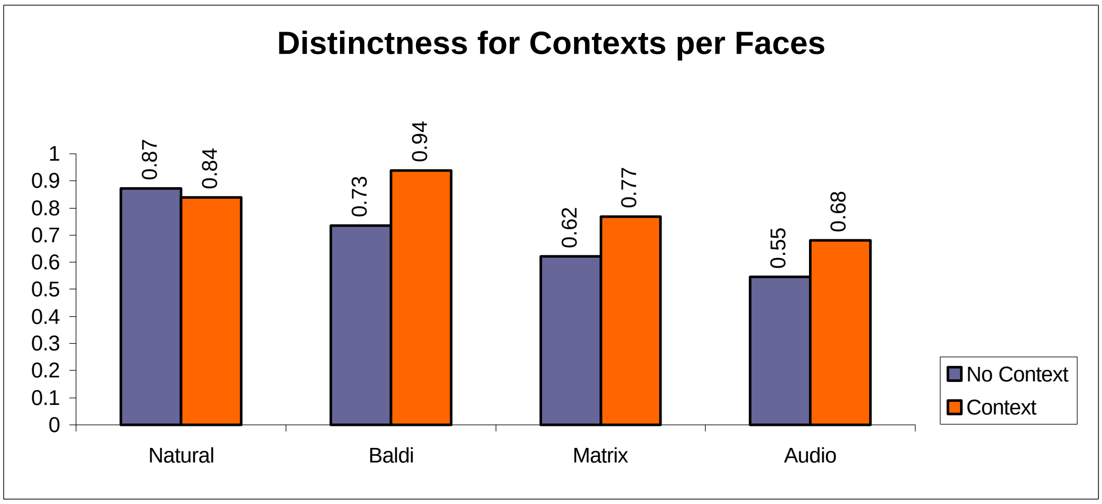


Figure 11: Distinctness per faces

9 Discussion

Distinctness is, against our expectations, not a significant predictor for convincingness. 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 high R Square (0.756) 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. It merges convincingness and trustworthiness into a new variable (convincingness') and leaves out distinctness. It also solves the collinearity problem for intensity and trustworthiness.

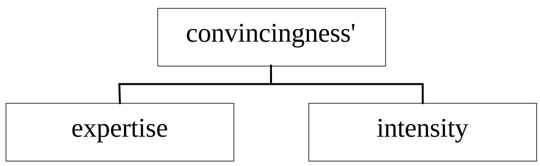


Figure 12: New model of convincingness

We calculated the convincingness' score for each subject by taking the average of his/hers convincingness and trustworthiness scores. 75.6% of the variance in convincingness' can be predicted from intensity and expertise and both are significant predictors (sig.<.001).

Table 13: Pearson correlation coefficients for variables predicting convincingness'

	Convincingness'	Intensity
Intensity	0.732	-
Expertise	0.747	0.418

Fogg's and Hsiang Tseng’s (1999) model defined believability by its components trustworthiness and expertise. Our data suggest that the concepts of believability 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 quality of synthetic facial expression has reached the level of natural faces. Both scores are rather high compared to results of other studies (see Table 5). However, most of those studies did not provide context information with their stimuli.

The Multimedia presentation of stimuli increases their convincingness. The Audio stimuli were rated significant less convincing then the visual and audio/visual stimuli. The distinctness scores for the three conditions were not significant different.

The source of the emotional expression had no influence on its convincingness. This result is in line with the media equation (Nass and Reeves, 1996). The only significant but very small difference is that humans were considered more trustworthy than machines.

It was not the original intention of this experiment to investigate the influence of the context on the perception of the emotional expressions. However, the comparison between the results in the pretest and the final experiment is so interesting that we decided to include it. In contrast to previous studies we could find a significant influence of context on the distinctness of the stimuli. Even more interesting is that this influence does not exists for the natural face and that the Baldi face reaches the highest score and not the natural face. We had no influence on how much attention the subject pay to the context. One explanation of this surprising result could be that the subjects paid less attention to the context when the natural face was shown. They might have felt confident enough to make their judgments by focusing on the face.

Despite this interesting results we need to consider some methodological problems. Only 10 subjects participated in the pretest, which might prove to be insufficient. In the pretest, the 7 faces, including the neutral face, were presented to the subject in random order. In the final experiment, the neutral face was displayed by default and changed into an expression. The subjects were asked to evaluate this expression. They never had to evaluate the neutral face. Furthermore, in the pretest only a limited amount of time was available for the subjects to make their judgements. Further research into this direction is necessary to confirm these results.

10 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, their recognition accuracy and successful examples of synthesis. We applied this framework to create a vocabulary of facial and audio expressions. Through several iterative circles we optimised the vocabulary that we proved to be more distinct than most previous designs.

The affective expressions of machines are as convincing as expressions of humans. These results support the work of Nass and Reeves (1996). We also 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.

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. Appropriateness depends on the context and therefore this result supports our preliminary finding that context plays an important role for emotional expressions. 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 emotional expressions is working, but further research on the grammar and the etiquette is necessary.

11 Acknowledgements

The authors would like to thank Gerard Hollemans and Don Bouwhuis for their support.

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. **Elmo Diederiks**, at Philips Research, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands, elmo.diederiks@philips.com.

12 Bibliography

Barnes, A. ; Thargad, 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.; Friberg, A.: 1999, 'Synthesis and decoding of emotionally expressive music performance'. *Royal Institute of Technology*, .

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, 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*, .

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

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

Davis, J.B. (eds.): 1978, 'The psychology of music'. Stanford, Stanford University Press.

Drag, R.M.; Shaw, M.E.: 1967, 'Factors influencing the communication of emotional intent by facial expressions'. *Psychometric Science, 8*, 137-138.

Dusenbury, D.; Knower, F.H.: 1938, 'Experimental studies on the symbolism of action and voice: I. A study of the specificity of meaning in facial expressions. '. *Quarterly Journal of speech, 24*, 424-435.

Ekman, P. (eds.): 1973, 'Darwin and facial expression : a century of research in review '. Academic Press.

Ekman, P. ; Friesen, W.V.; Ellsworth, P. (eds.): 1972, 'Emotion in the human face : guidelines for research and an integration of findings'. Pergamon Press.

Ekman, P. ; Friesen, W.V. ; O'Sullivan, M. ; 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. ;Frieser, W. (eds.): 1975, 'Pictures of facial affects'. Consulting Psychologist Press.

Etcoff, N.L.; 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.

Fenster, C.A.; Blake, L.K.; Goldstein, A.M.: 1971, 'Accuracy of vocal emotional communications among children and adults and the power of negative emotions. '. *Journal of Communication Disorder, 10*, 301-314.

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

Frijda, N.H. (eds.): 1986, 'The emotions'. Cambridge, Cambridge University Press.

Frijda, N.H.: 1969, 'Recognition of emotion'. *Adv. Exp. Soc. Psychology, 4*, 167-223.

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

Izard, C.E. (eds.): 1977, 'Human emotions '. Plenum Press.

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

Juslin, P.N.: 1997a, 'Emotional Communication in Music Performance: A Functionalist Perspective and Some Data'. *Music Perception 14*, 383-418.

Kanner, L.: 1931, 'Judging emotions from facial expressions'. *Psychology Monograph 41, (3, Whole No. 186)*, .

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

Kline, L.W. ; Johannsen, O.E.: 1935, 'The comparative role of face and face-body-hands as aids in identifying emotions'. *Journal of abnormal Social Psychology, 29*, 415-426.

Kozel, N.J.; Gitter, A.G.: 1969, 'Perception of emotion: Race of expresser, sex of perceiver, and mode of presentation'. *Proceedings of the 77th Annual Convention of the American Psychological Association*, pp. 39-40.

Krauss, R.M. ; Morrel-Samuels, P. ; Colasante, C.: 1991, 'Do conversational hand gestures communicate?'. *Journal of personality and social psychology*, 743-754.

Lazarus, R.S. (eds.): 1991, 'Emotion and adaptation'. Oxford, Oxford University Press.

Levitt, E.A.: 1964, 'The relationship between abilities to express emotional meaning vocally and facially'. In: J.R.Davitz (eds.): *The communication of emotional meaning*, McGraw-Hill, pp. 43-55.

Massaro, D.W. (eds.): 1998, 'Perceiving talking faces: From speech perception to a behavioral principle'. MIT Press.

McNeill, D. (eds.): 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. (eds.): 1998, 'Speech variability and emotions: production and perception'. Eindhoven, Technical University of Eindhoven.

Murray, I. R.; 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. ; Reeves, B. (eds.): 1996, 'The Media equation'. Cambridge, SLI Publications, Cambridge University Press.

Nielsen, J. (eds.): 1994, 'Usability Engineering'. AP Professional.

Nobe, s. ; Hayamizu, S. ; Hasegawa, O. ; 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, M. Fröhlich (eds.): *Gesture and Sign Language in Human-Computer Interaction*, Springer.

Norman, D.A.: 1981, 'Twelve issues for cognitive science'. In: D. Norman (eds.): *Perspectives on cognitive science*, Ablex Publishing.

Nowlis, V.: 1966, 'Research with the mood adjective check list'. In: S.S.Tomkin & C.E.Izard (eds.): *Affect, cognition and personality*, New York, Springer, pp. 352-389.

Oatley, K. ; Jenkins, J.M. (eds.): 1996, 'Understanding emotions'. Blackwell.

Osgood, C.E. ; Suci, G.J. ; Tannenbaum, P.H. (eds.): 1957, 'The measurements of meaning'. University of Illinois Press.

Picard, R.: 1997b, 'Does HAL cry digital tears? Emotion and computers'. In: D.Stork (eds.): *Hal's legacy: 2001's computer as dream and reality*, MIT Press.

Picard, R.W. (eds.): 1997a, 'Affective computing'. MIT Press.

Pixar, 1998, 'Geri's game'. http://www.pixar.com/shorts/shorts.html

Plutchik, R. (eds.): 1980, 'Emotion: a psycho evolutionary synthesis'. New York, Harper & 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: Izard, C.E. (eds.): *Emotions in personality and psychopathology*, New York, Plenum.

Scherer, K.R.; 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. ; Sheridan, K.: 2000, 'Face to InterFace: Facial affect in (Hu)Man and Machine'. *CHI 2000 Conference Proceedings*, pp. 193-200.

Schlossberg, H.: 1954, 'Three dimensions of emotion'. *Psychological review*, *61*, 81-88.

Smith, C.A. ; 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

Thomsen, D.F.; Metzler, L.: 1964, 'Communication of emotions intent by facial expression'. *Journal of Abnormal and Social Psychology, 68*, 129-135.

Williams, C.E.; Stevens, K.N.: 1981, 'Vocal correlates of emotional states'. In: J.K.Darby (eds.): *Speech evaluation in psychiatry*, New York, Grune & Stratton, pp. 221-240.

Woodworth, R.S (eds.): 1938, 'Experimental psychology'. New York, Henry Holt.

Zuckerman, M.; Lipets, M.S.; Koivumaki, J.K.; Rosenthal, R.: 1975, 'Encoding and decoding nonverbal cues of emotion'. *Journal of Personality and Social Psychology, 32*, 1068-1076.

13 Appendix

13.1 Appendix A: Pretests

13.1.1 Pretest 1

Sets of matrix faces and sounds were created by professional designers. To ensure the quality of the stimuli a pretest was performed. In addition, a highly abstracted line version of the matrix face was used to investigate the limits of emotional expressions even so they would not be used in the final test. Baldi was included in the test for reference.

Participants

10 members of the NMSA Group of Philips Research (6 male, 4 female) between the age of 26 and 46 participated in the experiment.

Stimuli

Screenshots of Baldi, 4 variations of Matrix faces and 4 variations of audio stimuli were used. Each set consisted of one stimuli for each of the 7 emotions including a neutral stimulus.

Procedure

The subjects viewed the stimulus images depicting emotional expression and responded with an emotional label for each image and sound, using the forced choice response format.

The stimuli were presented in groups (Baldi, Matrix, Line and Audio). The order of the groups and the order of the stimuli within the groups were randomized.

In each trial the image and the response labels were shown at the same time. The subjects had 10 seconds to chose the one label the best corresponds to the depicted emotion in each image, indicated by a progress bar. When they failed to do so the image was repeated at the end of the group presentation.

Results

Figure 13 presents the correct recognition scores for each emotional expression. The best stimulus for each emotion within each group (Baldi, Matrix, Line and Audio) are displayed. Figure 14 displays the average accuracy scores for each abstraction level.

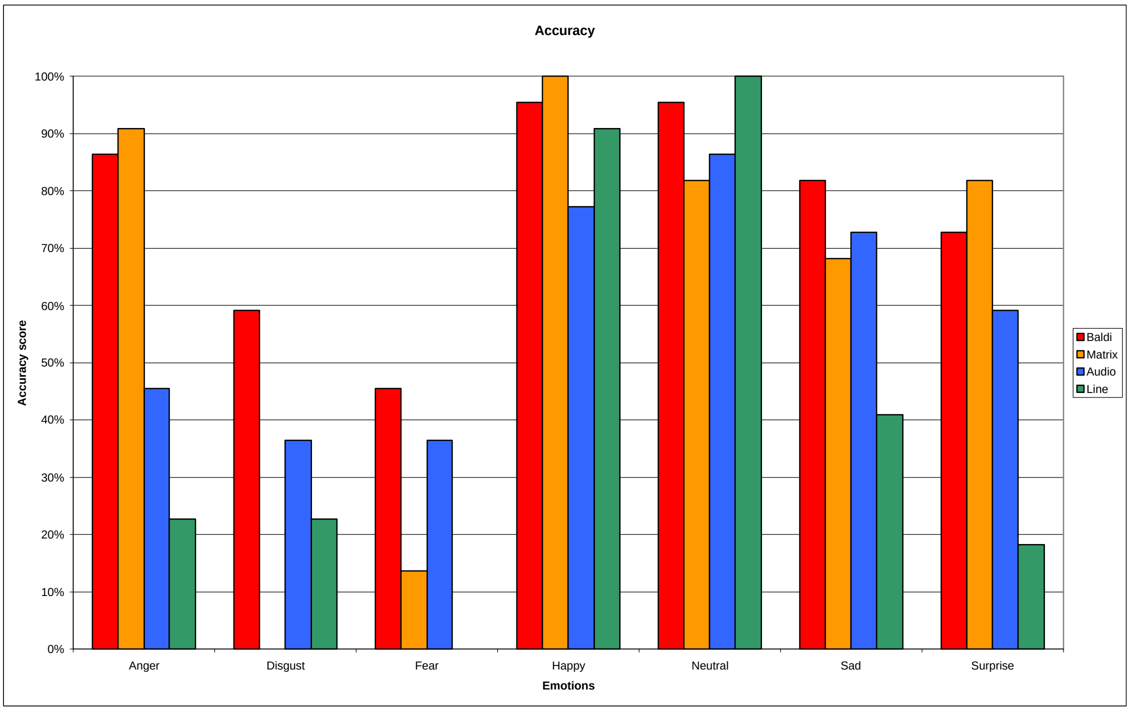


Figure 13: Accuracy scores for each emotion

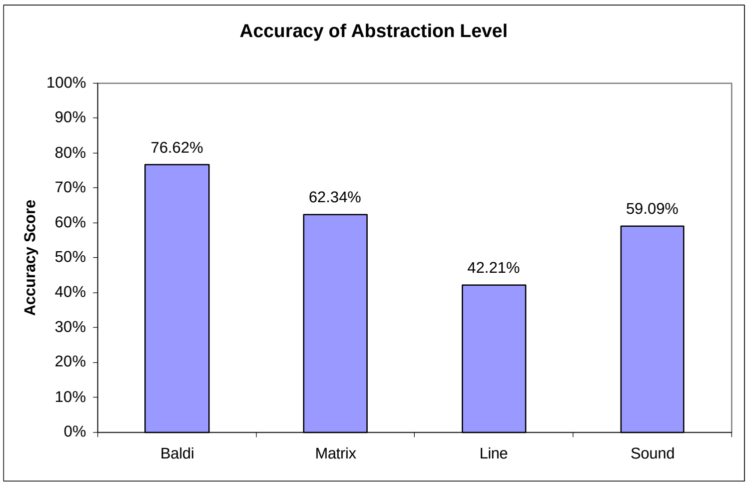


Figure 14: Accuracy scores for each abstraction level

Discussion

The accuracy score significantly depends on the abstraction level ($F(3,27)=14.758$, $p<.001$) and even the highly abstracted line version still scores well above the chance level (42.2%).
Within the visual stimuli the emotions happy and neutral are the most accurately recognized categories and the emotions disgust and fear are the least.
This is in line with the latest research (Schiano, et al., 2000).

The reason for the extreme low score of the disgusted matrix face (0%) is the absence of a lifted nose, which is an essential part of disgusted human faces. The fearful matrix faces were most often confused with surprised and sad faces. The low scores of the disgusted and fearful matrix faces resulted in their redesign (see pretest 2).

Within the audio stimuli the emotions happy and neutral were the most accurately recognized categories and the emotions anger, disgust and fear the least. However, all of them are well above the chance level so that no redesign was necessary.

13.1.2 Pretest 2

To ensure the quality of the stimuli a second pretest was performed. A set of matrix faces were redesigned (see pretest 1) by professional designers. The major change was the addition of a nose. Two students and one professional actor produced facial expressions. The images would be used in the final test and therefore it was necessary to pretest their quality.

Participants

12 members of the NMSA Group of Philips Research (9 male, 3 female) between the age of 21 and 35 participated in the experiment.

Stimuli

Photographs of two students and one professional actor were used in addition to the redesigned matrix face. Each set consisted of one stimuli for each of the 7 emotions including a neutral stimulus.

Procedure

The subjects viewed the stimulus images depicting emotional expression and responded with an emotional label for each image, using the forced choice response format.

The stimuli were presented in groups (matrix and human). The order of the stimuli within the groups were randomized.

In each trial the image and the response labels were shown at the same time. The subjects had 10 seconds to chose the one label the best corresponds to the depicted emotion in each image, indicated by a progress bar. When they failed to do so the image was repeated at the end of the group presentation.

Results

Figure 15 presents the comparison between the accuracy scores of the initial and redesigned matrix faces. Figure 16 displays the accuracy scores for each encoder, grouped by the emotion category. Figure 17 shows the accuracy scores for each encoder. Figure 18 indicates the average accuracy score for all encoders for each emotion.

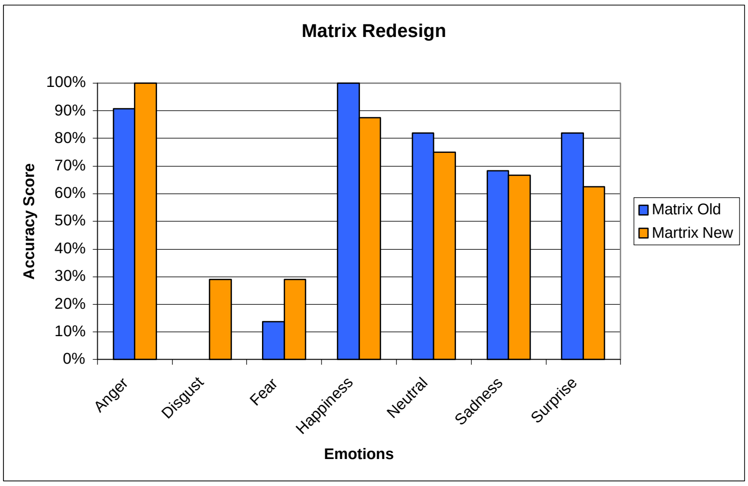


Figure 15: Matrix Redesign

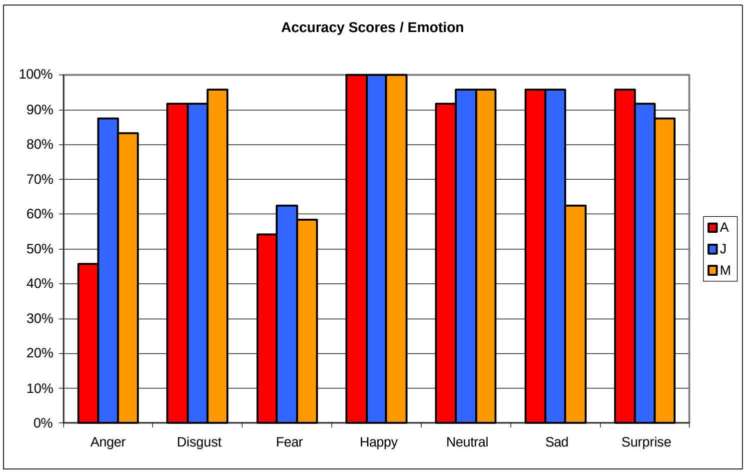


Figure 16: Accuracy Scores per Emotion

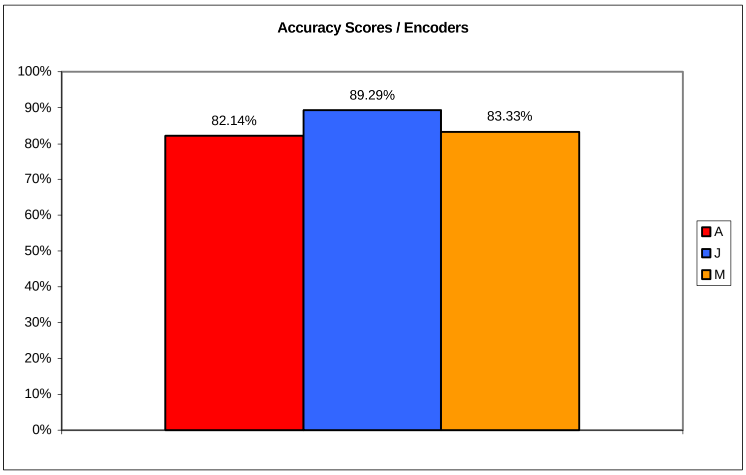


Figure 17: Accuracy Scores per Encoder

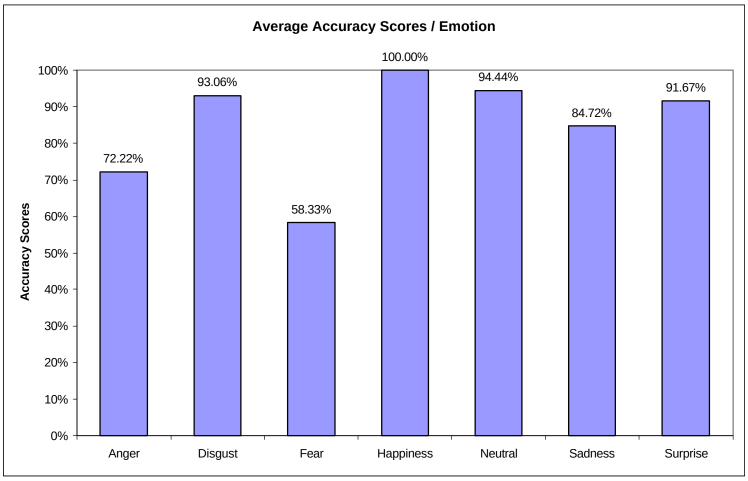


Figure 18: Average Accuracy Scores per emotion

Discussion

The redesign of the matrix faces raised the accuracy score for the categories fear and disgust well above the chance level (both 29%). Overall accuracy was slightly increased from 62% to 64%.

The human faces were recognized much better. The average accuracy score for all encoders was 84.9% and the differences in the accuracy scores between the encoder (Figure 17) were not significant ($F(2,22)=3.187$, $p<.061$). The amateur students performed just as good as the professional actor.

The recognition was highest for happiness (100%) and lowest for fear (58%). These results are in line with Ekman's classical results, both in terms of relative pattern and absolute levels of performance.

13.1.3 Pretest 3

To write a realistic game script we performed a pretest. Two subjects (both from the NMSA group) played 3 games, each consisting of 30 rounds, with a real dice.

The time needed for each game was recorded to have an indicator for the length of the final experiment. The number of turns per round, the result of each round, the general behavior and the emotional expressions of the players were recorded to have a realistic base for the game script of the final experiment.

Results

- Each game took between 7-8 minutes
- The average number of turns per round was 2.27 with a minimum of 1 and a maximum of 5.
- The players took more to time with their decision if a value equal or higher than 15 was thrown.
- Players tend to pass at throws equal or higher than 16.
- Emotional expression occurred most frequently at the end of each round.

		A			B			total round	
round	throws	m win	m pas	m loose	c win	c pass	c loose	m total	c total
1	4	12						24	0
2	2				4			0	8
3	1	19				19		38	19
4	2	16				16		32	16
5	2	13						26	0
6	4	19				19		38	19
7	3				15			0	30
8	1				20			0	40
9	2		19		19			19	38
10	1		17		17			17	34
11	2	11						22	0
12	2				3			0	6
13	2				17			0	34
14	2	19				19		38	19
15	2	18						36	0
16	2	20				20		40	20
17	3				14			0	28
18	1		17		17			17	34
19	2	12						24	0
20	1		20		20			20	40
21	3				7			0	14
22	2	17				17		34	17
23	2	19				19		38	19
24	3	15						30	0
25	2		17		17			17	34
26	3		19		19			19	38
27	4	6						12	0
28	3	15						30	0
29	2	14						28	0
30	4				11			0	22
Average	2.3						score	599	529
Time	8 min								

Figure 19: Game 1

		A			B			total round	
round	throws	m win	m pas	m loose	c win	c pass	c loose	m total	c total
1	5				14			0	28
2	2	19				19		38	19
3	3	13						26	0
4	3				11			0	22
5	2				3			0	6
6	1		19		19			19	38
7	2				10			0	20
8	3	15						30	0
9	2	13						26	0
10	3		18		18			18	36
11	1	19				19		38	19
12	3	17						34	0
13	1	18				18		36	18
14	2				5			0	10
15	2				12			0	24
16	2				12			0	24
17	3				15			0	30
18	2	17				17		34	17
19	2	9						18	0
20	1		20		20			20	40
21	1	17				17		34	17
22	4				10			0	20
23	1	18				18		36	18
24	2	18				18		36	18
25	1	19				19		38	19
26	2	20				20		40	20
27	2				8			0	16
28	2	19				19		38	19
29	3	11						22	0
30	2	19						38	0
Average	2.166667						score	619	498
Time	7 min								

Figure 20: Game 2

		A			B			total round	
round	throws	m win	m pas	m loose	c win	c pass	c loose	m total	c total
1	1		20		20			20	40
2	3	17				17		34	17
3	2			0	16			0	32
4	2			0	10			0	20
5	3	8					0	16	0
6	2		15		15			15	30
7	2			0	11			0	22
8	2			0	11			0	22
9	1		19		19			19	38
10	3			0	17			0	34
11	3	3					0	6	0
12	2	7					0	14	0
13	2	19				19		38	19
14	4	16					0	32	0
15	2		16		16			16	32
16	4			0	16			0	32
17	3	14					0	28	0
18	2	15					0	30	0
19	3	12					0	24	0
20	2	10					0	20	0
21	1		20		20			20	40
22	1	19				19		38	19
23	2			0	7			0	14
24	3	20				20		40	20
25	2			0	8			0	16
26	2	16					0	32	0
27	3		19		19			19	38
28	3			0	6			0	12
29	3		20		20			20	40
30	3	20		0		20		40	20
Average	2.366667						score	521	557
Time	8 min								

Figure 21: Game 3

13.2 Appendix B: The game

The Rules

The goal of the game is to maximize your points. Player A starts the game by rolling a 20 sided dice. In the first round, he has to pass the dice to the player B. Player B must decide either to accept the number and roll again or to pass. If he decides to pass he gets points, equal to the number on the dice and player A gets the double of that. If player B decides to roll the dice again than he must role more than the previous number. If he fails, player A gets the double amount of points of the last throw and player B gets nothing. If he succeeds it is player A's turn, and so on.
If a player rolls a 20 then the opponent cannot roll more. Instead, he has to roll a 20 again. The starting player alternates (ABABABAB…).

Structure:

The subjects observed six games:

- human vs. human* using photos of human faces to express emotions
- human vs. machine* using photos of human faces to express emotions
- human vs. machine* using Baldi to express emotions
- human vs. machine* using a matrix face to express emotions
- human vs. machine* using a matrix face and audio to express emotions
- human vs. machine* using audio to express emotions

* observed by subject

The order of the games were counterbalanced:

Game	Subject					
	1	2	3	4	5	6
human photo	1	3	5	2	4	6
machine photo	5	1	3	6	2	4
machine baldi	3	5	1	4	5	2
machine matrix visual	6	4	2	1	6	3
machine matrix audio/visual	2	6	4	3	1	5
machine matrix audio	4	2	6	5	3	1

Table 14: Game alternation

Each game consisted of 30 rounds. . Each round ends with the winning of one player. The sequence of winning and loosing rounds were randomized for each game, but the end result and the number of won and lost rounds were equal. The subject had to answer one question (dependent variables) about one emotional expression that occurred during each round. (6 emotions x 5 question = 30 rounds). By default, the neutral face was displayed. Each round consisted of 2-4 turns. Every turn offered 3 states in which an emotional expression might have occurred. The following table presents a typical round.

Turn	State	Result	Expression
1	Opponent roles	-	Neutral
1	Opponent result	14	Neutral
2	Player decision	-	Neutral
2	Player roles	-	Neutral
2	Player result	10	Sadness

Table 15: Two turns of one round

The 5 question about each of the 6 emotional expression had to occurred in the same context. The subjects had to observe 5 times the same round. To prevent boring patterns the order of the questions and expressions were randomized. Furthermore, three minor variations off each of the 6 basic round were used in a randomized order across the 6 games. We considered these minimal changes in the context in order to increase the diversity of the game as negligible. Without noticing, the subjects were observing the 6 basic rounds over and over. Asking every time a different question forced the subjects to focus their attention.

13.3 Appendix C: The interface of the software.

First, the software inquired the age and the gender of the subject. Second it would present the training game and last it showed the real experiment.

On the very left and right we placed a score counters. They kept track of the points the players. Next to them, we placed turn indicators. Whenever it was the human's /computer's or the opponents turn it would show a red frame and a label (Human/Machine or Opponent). In the middle of the scoreboard, we placed a small panel, which showed the dice. Next to it on both sides, we placed a history of the last throws. We discovered the need for it during our pretest. The subjects were not able to remember what the last throw was and therefore had difficulties evaluating the situation.

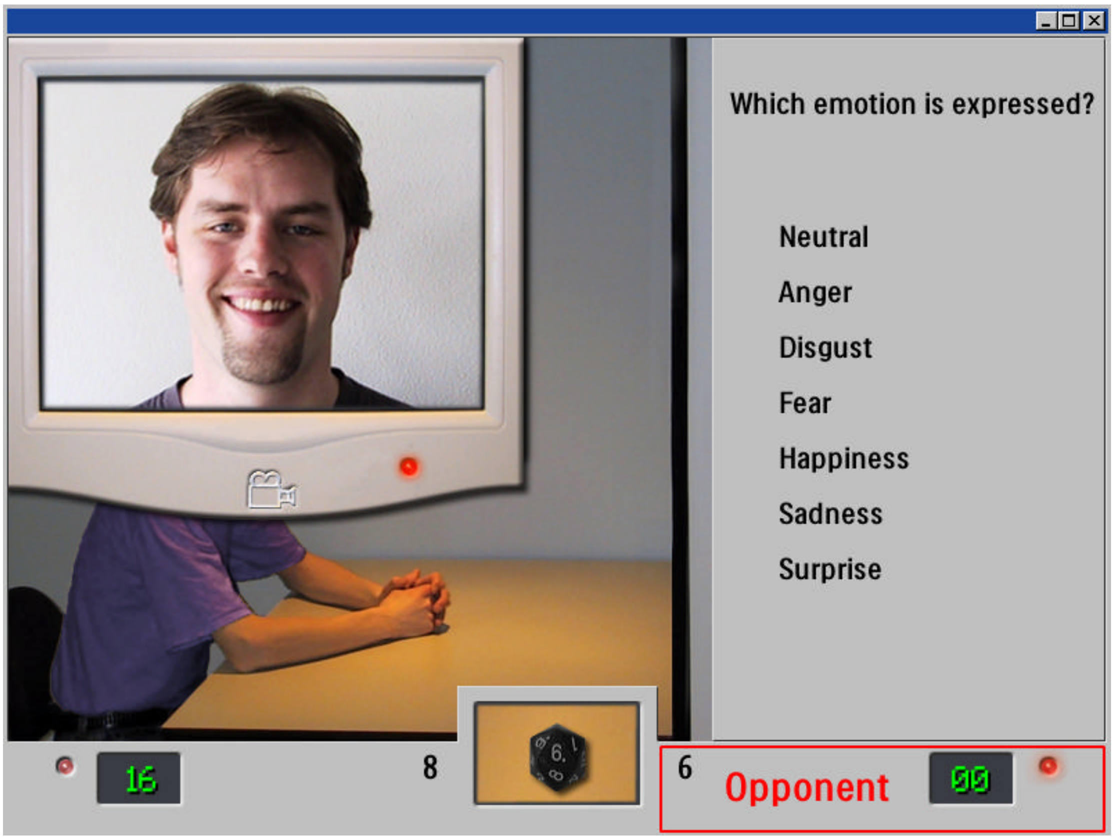
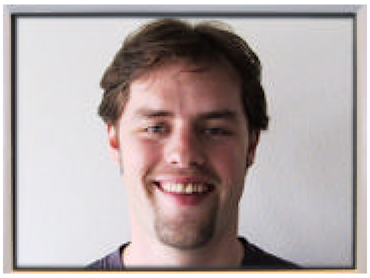
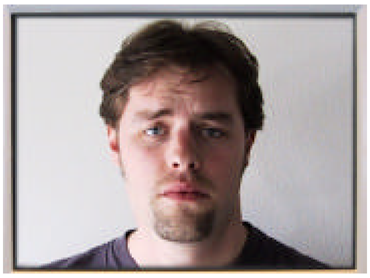


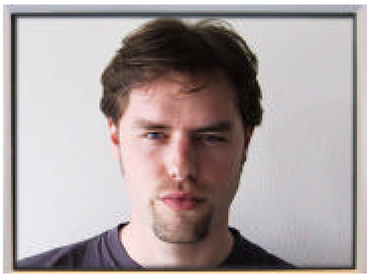
Figure 22: The interface. The human player started with an 8. The opponent rolled a 6 and therefore lost the round. The human player gets 2x8=16 points.

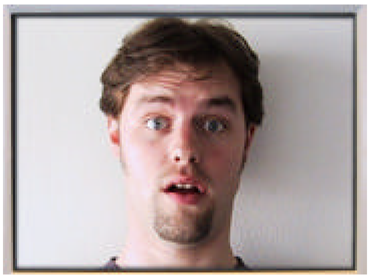
The software stored all the answers of the subjects in a log file. We used it as the base for out statistical analyses.

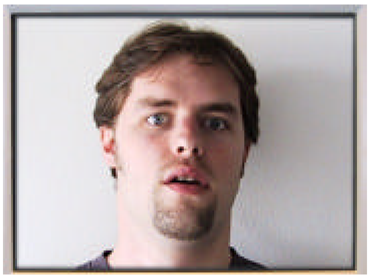
13.4 Appendix D: The stimuli

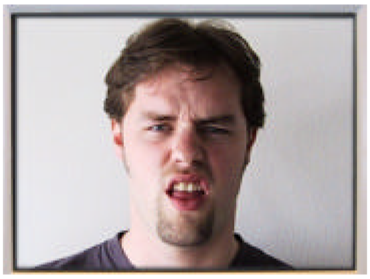

Happiness

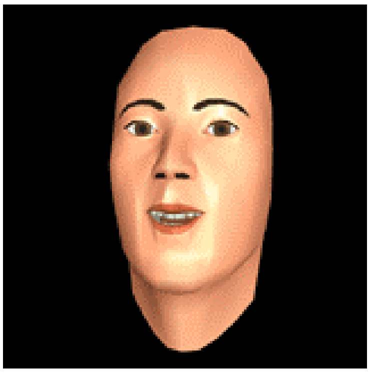

Sadness

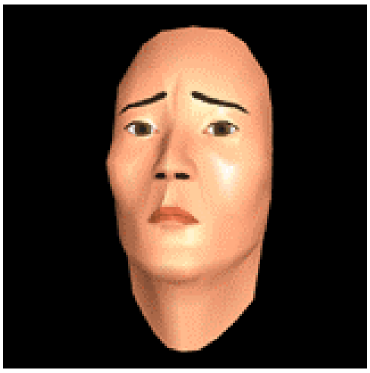

Anger

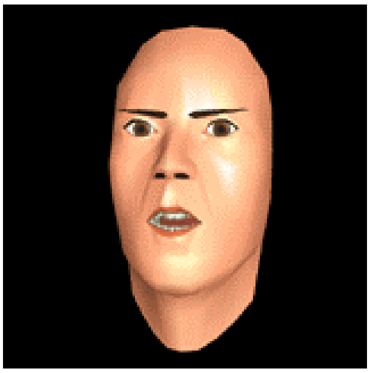

Surprise

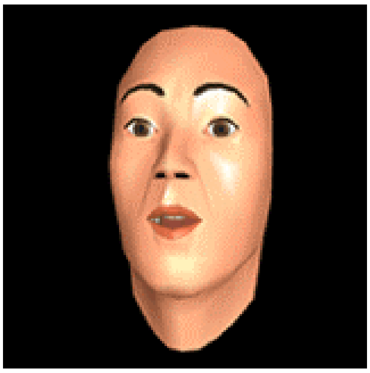

Fear

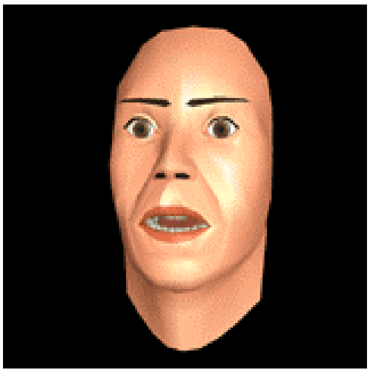

Disgust

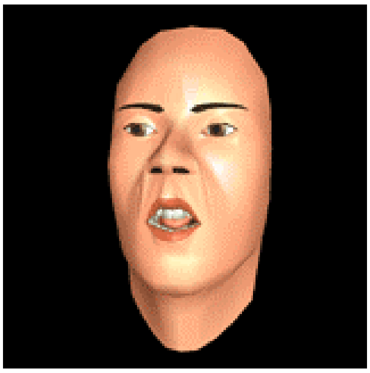

Happiness

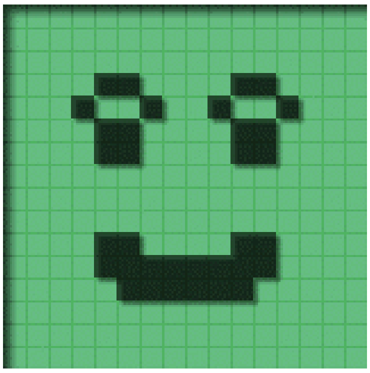

Sadness

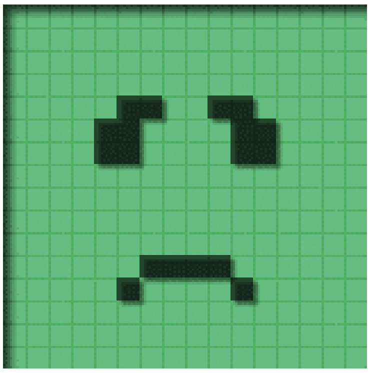

Anger

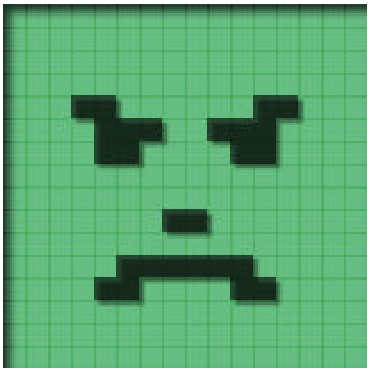

Surprise

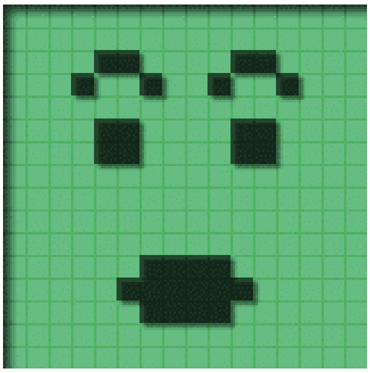

Fear

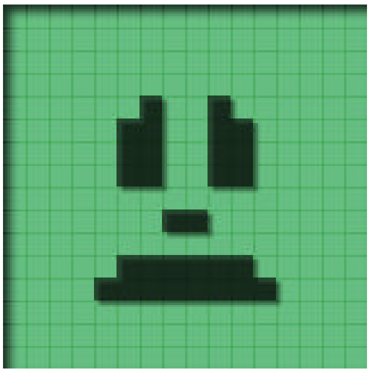

Disgust

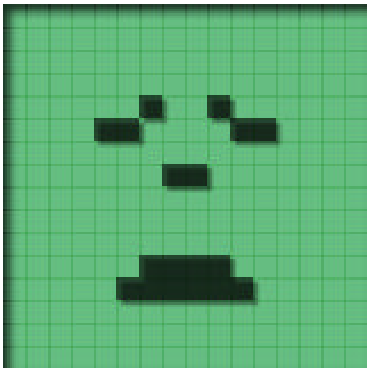

Happiness


Sadness

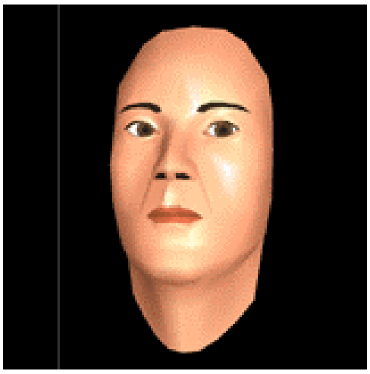

Anger

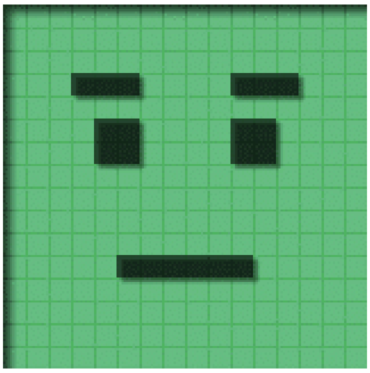

Surprise


Fear


Disgust


Neutral


Neutral


Neutral