# Interpreting Human and Avatar Facial Expressions

Sylvie Noël[1], Sarah Dumoulin[1], and Gitte Lindgaard[2]

[1] Communications Research Centre, 3701 Carling Ave. Ottawa, ON, Canada, K2H 8S2
[2] Carleton University, Department of Psychology, 1125 Colonel By Drive, Ottawa, ON, Canada, K1S 5B6
{sylvie.noel, sarah.dumoulin}@crc.ca, glindgaa@connect.carleton.ca

**Abstract.** This paper investigates the impact of contradictory emotional content on people's ability to identify the emotion expressed on avatar faces as compared to human faces. Participants saw emotional faces (human or avatar) coupled with emotional texts. The face and text could either display the same or different emotions. Participants were asked to identify the emotion on the face and in the text. While they correctly identified the emotion on human faces more often than on avatar faces, this difference was mostly due to the neutral avatar face. People were no better at identifying a facial expression when emotional information coming from two sources was the same than when it was different, regardless of whether the facial expression was displayed on a human face or on an avatar face. Finally, people were more sensitive to context when trying to identify the emotion in the accompanying text.

**Keywords:** Emotions, avatars, virtual reality, collaborative virtual world, facial expression.

## 1 Introduction

The importance of emotional displays in face-to-face communication has led designers of virtual worlds and intelligent avatars to incorporate emotional states into their avatars. There have been various attempts to automate the recognition of a person's emotional state in order to animate that person's avatar with the proper emotional expression in real time. These attempts have included extracting the emotion from text [1], from voice cues [24] or through real-time facial tracking [2,9]. However, errors may potentially occur during these automated extractions – even real-time facial tracking could lead to errors if the person's face is occluded in some way. If such an error occurs, people might be confronted with contradictory emotional information from an avatar. How would they react in such a situation?

Researchers in human emotions have been studying the impact of contradictory emotional information for decades [10]. This early research was aimed at exploring the facial dominance theory. This theory, originally put forward in the 1960s by Tomkins [27], states that there are a small number of prototypical facial expressions intrinsically associated with specific basic emotions (happiness, surprise, fear, anger, disgust, sadness, contempt) [see also 8,13,19]. According to this theory, people will use a person's facial expression to base their decision on what emotion that person is

feeling above any other source of information, as long as the facial expression is one of the basic emotions. Early research on people's reactions to contradictions in emotional information seemed to support this theory [15]. However, Carroll and Russell [10] have suggested that this might at least partially be due to the experimental designs adopted. For example, by forcing participants to select an answer from a small set of prototypical emotions, previous researchers excluded the possibility that participants' responses could have been something other than one of the basic emotions. By modifying several elements of the experimental design, including offering more choices and presenting the context orally, Carroll and Russell were able to produce situations where their participants tended to describe the facial expression by using the emotion contained in the scene description rather than the emotion supposedly displayed by the face. These results supported their theory of limited situational dominance, which states that assigning an emotion to someone is based not only on the facial expression but also on the context, and that there are no prototypical emotional facial expressions.

More recently, De Gelder and Vroomen [12] presented a series of pictures ranging from happy to sad which they combined with a sad voice, a happy voice, or no voice. People responded more slowly when the emotions were mismatched. Voices had an impact on how a face was judged: a sad voice caused a happy face to be judged less happy and a happy voice caused a sad face to be judged less sad. This result appears to support the limited situational dominance theory.

These studies have looked at how people interpret information presented on real human faces. We turn now to studies that used synthetic humans as stimuli.

While research has shown that people can recognize emotions presented on avatar faces [6,21], results suggest that people have more difficulty identifying emotions that are presented on a virtual face as compared to a real human face [17,20,26; although see 3,4,5]. However, by varying the position of various facial elements, Fabri et al. [17] were able to find a virtual face whose recognition rate was similar to that of a human face for all of the emotions they tested. The only exception was that of disgust, which they attributed to the lack of detail in the avatar face, which couldn't display the wrinkled nose normally associated with this emotion.

Recent studies have investigated the impact of contradictory information in the presence of synthetic human faces or voices. Hong et al. [18], for example, paired neutral happy and sad voices with neutral happy and sad synthetic or real faces. Their results suggest that it was harder to identify emotions with contradictory than with congruent information. However, no statistical evidence was provided to support this hypothesis, which weakens the evidence considerably.

Nass et al. [22] paired emotional human and digitized voices (happy, sad) with emotional text (happy, sad) such that the two either had congruent or contradictory emotional content. When text and voice matched emotional content, people rated the content higher on emotion than when they did not match: happy stories were judged as happier and sad stories were judged as sadder. People also preferred content with matching stimuli, but found the stories more credible when the voice and text did not match.

Creed and Beale [11] investigated the impact of mismatches between facial emotional content and vocal emotional content, combining an animated avatar face with a female human voice. They used happy, warm, neutral, and 'concern' emotions.

Mismatched stimuli that had either a happy or a warm element, be it in the voice or the face, were judged warmer and happier than those that had no such element.

The results from these three studies suggest that people viewing emotional avatar faces may be influenced by other contextual clues. However, the contradictory results from the studies concerning the impact of context on emotional identification for human faces and the paucity of studies on avatar faces leave us with many unanswered questions. In this paper, we set out to answer two main questions. Do people who are presented with contradictory emotional information identify the expression on a face differently than people who are presented with congruent emotional information? And does emotional information have a different impact depending on whether the displayed face is that of a human or that of an avatar?

## 2 Method

### 2.1 Participants

The 56 participants were recruited from the student population of a local university in return for course credit. Four participants were removed, two because English was not their first language[1] and two because they had made a mistake following instructions. The 52 remaining participants were placed randomly into four groups of nine females and four males each. Age varied between 17 and 51, with an average of 19.8 years.

### 2.2 Stimuli

Stimuli were static emotional images and short texts. The images were of a male avatar face and of a male human face showing five different emotions (anger, happiness, neutral, sadness, surprise). The avatar face (see Figure 1) is a Facial Action Coding System (FACS) [14] compliant face, originally created by Fabri et al. [17]. The facial expressions selected were based on an earlier experiment testing the recognition rate of various emotions [23].

The human face images (see Figure 1) were taken from DaFEx, a database of animated human facial expressions [7]. From the DaFEx's short clips, we selected single frames from movies made by the same actor. The selection criterion was to match as closely as possible the avatar's facial expressions.

---

[1] This was done in order to ensure that participants understood the list of emotion names presented to them.
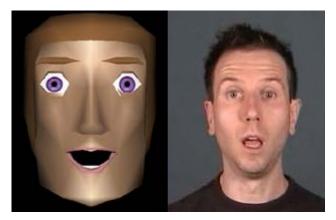
**Fig. 1**. Surprised avatar and human faces

The texts were short two-sentence comments as told from the point of view of the person represented by the avatar or the human face (see Figure 2). For each of the emotions, we created five texts.

> I applied for a really great job last week. They just called me to tell me that I got the job

**Fig. 2**. Happy text #5

We used 10 pictures (five avatar, five human) and 25 texts (5 for each emotion). Each facial emotion was associated with five texts of different emotional content. For example, the happy human face could be presented with a happy text (congruent situation) or with an angry, neutral, sad or surprised text (incongruent situations). The same texts were presented for the avatar and the human face in each condition (i.e., if sad text #4 was associated with a happy human face, it was also associated with a happy avatar face).

### 2.3 Measures

For each picture/text combination, participants were asked to accomplish two tasks. The ratings task consisted of judging each face on trustworthiness, sincerity, appropriateness, intensity, and convincingness. Because of the large amount of data created by this task, these results are not presented here. The identification task consisted of identifying the emotional content of the text and the image separately. This paper presents the results from the identification task only. The ratings and the identification tasks were presented on two different screens.

## 2.4 Procedure

Participants saw only human faces or avatar faces, but they all performed both the identification and the rating tasks. The order in which these tasks were presented was counter-balanced to avoid serial position effects. Thus the four groups of participants were as follows:
1. Avatar faces, identification first
2. Avatar faces, ratings first
3. Human faces, identification first
4. Human faces, ratings first

For the identification-first groups, the identification screen (Figure 3) was always presented first, followed immediately by the ratings screen (Figure 4). The order was reversed for the ratings-first groups. Participants were randomly assigned to one of these four groups.



**Fig. 3**. Identification screen

After noting the participants' age and gender, the experimenter asked them to imagine either that they were using a text chat system that could display static images of chatters' faces (for the human face conditions) or that they were in a virtual world where people communicated through text (for the avatar face conditions). They were asked to imagine that the person they were talking to had just typed a short text message and displayed the image to go with it, and that they would be asked to answer a series of questions on that person.

For each of the five emotions tested, the participant saw five combinations of image and text (one congruent and four incongruent), for a total of 25 stimuli. The order of the combined text and image stimuli was randomized for each subject. A computer application was used to present the text-face combinations. There was no limit on the amount of time that participants could take to answer the questions.
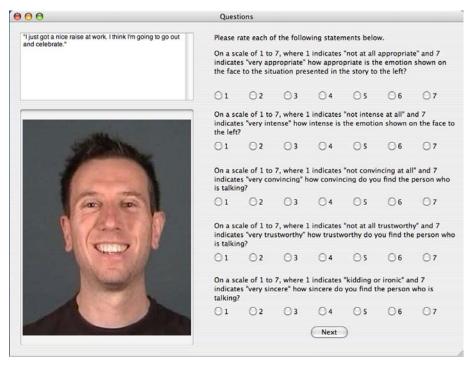


**Fig. 4**. Ratings screen

For the identification task, participants were asked to describe, separately, the emotional content of the text as well as of the face. To do this, they were asked to choose from a list of seven different possible emotions (angry, disgusted, fearful, happy, neutral, sad, surprised). There was also an "other" choice, in which case they were invited to type in the emotion they thought was displayed. In total, people entered an original emotion name that did not appear in the list 134 times for both faces and texts (approximately 5 percent of all answers).

## 2.5 Apparatus

The experiment was run on a Mac Powerbook G4 laptop with a 12-inch screen. A special application was created to present the stimuli and record the participants' answers. The screen size of the image was 8 cm wide by 10.5 cm high.

# 3 Results

In the following text, the "expected emotion" corresponds, for the faces, to the emotion typically associated to the facial expression according to Ekman and Friesen's [14] FACS coding; and, for the texts, to the emotion originally assigned to the text by the researchers.

**Pictures**

In this section, we present the results for the identification of the emotions presented on the faces. We begin by checking whether the order of the two tasks (identification first or ratings first) had an impact on the identification task. We then verify whether people are assigning the expected emotion more often than another emotion for both face types. We next test to see if there is a difference in the identification task depending on the face type. Finally, we explore the impact of the various emotional texts (congruent and incongruent) on the identification of each facial expression.

To test for any effects of presentation order, we compared the number of correct identifications made by participants in the groups who completed the ratings task first to the number made by participants in the groups who completed the identification task first. The ratings-task-first groups selected the expected emotion 559 times and some other emotion 91 times, while the identification-task-first group selected the expected emotion 535 times and some other emotion 115 times. This difference did not reach statistical significance ($\chi^2$ (1) = 3.3, p>0.05). Therefore we have ignored the task order in the following analyses.

In order to ascertain the frequency with which participants assigned the expected emotions to the faces, we examined all participants' responses to all the stimuli, both congruent and incongruent. This showed that participants gave the expected answer 1094 times and gave another answer 206 times. This is significantly different ($\chi^2$ (1) = 606.6, p<0.001). Contrary to Carroll and Russell's [10] results, but in agreement with other studies [15], our participants overwhelmingly described the face as displaying the emotion typically associated with the facial expression.

The data were then examined to determine if the performance of participants who saw a photorealistic face differed from that of participants who saw an avatar face. Participants in the photorealistic face condition selected the expected emotion 574 times and they selected another emotion 76 times, while those in the avatar condition selected the expected emotion 520 times and another emotion 130 times. This difference is significant ($\chi^2$ (1) = 16.8, p<0.001). Participants in the photorealistic condition thus assigned the expected emotion to a human face more often than participants in the avatar condition.

Table 1 presents the percentage of time people gave the expected emotion for the human and the avatar faces for each stimulus combination. For both the human and the avatar faces, participants had a strong tendency to assign the expected emotion to the face, with the exception of the sad human face and the neutral avatar face.

In an effort to determine if performance differed between the human and the avatar faces on the various stimuli combinations, a series of Mann-Whitney U tests were performed. To take into account the numerous pairwise comparisons undertaken, we

adjusted the threshold to 0.001. There were significant differences between the human and the avatar faces in the cases of the neutral face combined with an angry text (U=169, p<0.000) and the neutral face combined with a sad text (U=143, p<0.000). In both these cases, participants gave the expected emotion more often for the human face than for the avatar face. Perhaps this difference is due to our participants being more sensitive to context when presented with an avatar rather than a photorealistic face. Looking at the answers that people gave for the avatar face, we find that, for the neutral face combined with an angry text, the answers most often given were "neutral" (11 times) and "sad" (11 times), while for the neutral face combined with a sad text, the answers most often given were "sad" (14 times) and "neutral" (8 times). These results show that many people confused the avatar neutral face with a sad expression, particularly when the context was one of negative affect.

**Table 1**. Percentage of people assigning the expected emotion to a facial expression, according to facial type, facial emotion, and text emotion (congruent stimuli are in bold).

| | Text | | | | | |
|---|---|---|---|---|---|---|
| Face | Angry | Happy | Neutral | Sad | Surprised | |
| | *Human Face* | | | | | *Average* |
| Angry | **84.6%** | 84.6% | 92.3% | 96.2% | 92.3% | *90.0%* |
| Happy | 100% | **100%** | 100% | 100% | 100% | *100%* |
| Neutral | 92.3% | 96.2% | **96.2%** | 88.5% | 96.2% | *93.9%* |
| Sad | 69.2% | 57.7% | 57.7% | **80.7%** | 65.4% | *66.2%* |
| Surprised | 88.5% | 92.3% | 88.5% | 100% | **88.5%** | *91.5%* |
| | *Avatar Face* | | | | | |
| Angry | **84.6%** | 69.2% | 65.4% | 84.6% | 80.8% | *76.9%* |
| Happy | 88.5% | **100%** | 96.2% | 96.2% | 84.6% | *93.1%* |
| Neutral | 42.3% | 69.2% | **69.2%** | 30.8% | 69.2% | *56.2%* |
| Sad | 80.8% | 92.3% | 76.9% | **88.5%** | 76.9% | *83.1%* |
| Surprised | 92.3% | 92.3% | 88.5% | 84.6% | **96.2%** | *90.8%* |

The faces that showed the most variance in percentage of expected responses as the accompanying text differed are the human sad face and the avatar neutral face. A Cochrane Q was used to assess any differences between the various stimuli combinations. It revealed marginally significant differences between the various texts combined with the human sad face (Q(4)=9.9, p=0.042). For the sad human face, the congruent condition increased somewhat people's tendency to assign the expected emotion to that face as compared to incongruent conditions. In the case of the avatar neutral face, these differences were significant (Q(4) = 16.6, p=0.002). This appears to be due to the low recognition score for the sad text (30.8%). Removing this particular condition resulted in a non-significant Q (Q(3)=6.8, p=0.079). As noted earlier, when the neutral face was combined with a sad text, participants tended to assign a sad emotion as often as a neutral emotion to the avatar face.

In summary, our participants chose the expected emotion more than any other emotion, and they did this more often for the human than for the avatar face. The difference between the two face types was greatest with the neutral expression – people tended to confuse the neutral avatar face with a sad face, especially when the accompanying text contained a negative affect. Finally, the only face whose

identification was different for congruent vs. incongruent emotion combinations was that of the human sad face, although this effect was marginal.


**Texts**

In this section, we look at how people perceived the emotions contained in the texts. As with the pictures, we begin by checking for any impact of task order on the text identification task. We then verify whether people were assigning the expected emotion to the texts and whether the facial type (human or avatar) had an impact on this identification task. Finally, we explore the impact of the accompanying facial expressions on the identification of each emotional text.

Does task presentation order have an impact on the identification of the emotions contained in each text? The two groups who completed the ratings task first selected the expected emotion 465 times and some other emotion 185 times, while the two groups who completed the identification task first selected the expected emotion 488 times and some other emotion 162 times. This difference was not significant ($\chi^2$ (1) = 2.1, p>0.10). Therefore we ignored task order in the following analyses.

In order to determine the degree to which participants assigned the expected emotion to the texts, we examined all participants' responses to all of the congruent and incongruent stimuli. The expected emotion was assigned to the text 953 times and another emotion was provided 347 times. This difference was significant ($\chi^2$ (1) = 282.5, p<0.001).

Participants in the human face group chose the expected emotion 476 times (and another emotion 174 times) while the avatar group chose the expected emotion 477 times (and another emotion 173 times). There is no significant difference between the two groups ($\chi^2$ (1) = 0.004).

Since there was no difference between the two face-type conditions, we combined the data from the two conditions for the next analysis. Table 2 presents the percentage of participants who assigned the expected emotion to the texts for each stimulus combination. Was there a difference between the various emotion combinations for each emotional text? Assigning a threshold of 0.001, we found a significant difference between the various text and emotional face combinations for the angry text (Q(4)=18.9, p=0.001), the neutral text (Q(4) = 29.5, p<0.001), the sad text (Q(4)=61.8, p<0.001), and the surprised text (Q(4) =28.0, p<0.001). The only one that does not achieve significance is the happy text (Q(4)=10.9, p=0.027). Could these differences be due to people giving the expected emotion more often for the congruent stimuli as compared to the incongruent stimuli? In the case of the angry text, this appears to be the case. When the angry text is combined with an angry face, 75% of the people gave the expected answer, while the number drops below 54% for the surprised, happy, and sad faces. Although 73% of people assigned the expected emotion to the neutral face combined with the angry text, this is not a true counter-example, since the neutral face is de facto without an emotional expression. For that stimulus combination, the only source of emotional information should be the text, and so it is not surprising to see that the text emotion dominates. This pattern (high response rate for the congruent stimuli and the neutral face combined with the emotion text; low response rates for all other incongruent combinations) is also found

for the sad text. However, this is not the case for the surprised text, where not only does the congruent combination (surprised text and surprised face) not receive the highest percentage of identification (this goes rather to the combination of surprised text and sad face), but the neutral face combination shows the lowest rate of identification of the expected emotion. Finally, in the case of the neutral text, the percentage of people who chose the expected emotion was similar when it was accompanied by either the angry, neutral or sad face.

**Table 2.** Percentage of people assigning the expected emotion to the text according to emotion in text and on face, for all groups (congruent stimuli are in bold)

| Face | Text | | | | |
| | Angry | Happy | Neutral | Sad | Surprised |
| --- | --- | --- | --- | --- | --- |
| Angry | **75.0%** | 73.1% | 100% | 42.3% | 75.0% |
| Happy | 46.2% | **94.2%** | 80.8% | 44.2% | 67.3% |
| Neutral | 73.1% | 80.8% | **98.1%** | 88.5% | 46.2% |
| Sad | 46.2% | 78.8% | 96.2% | **88.5%** | 88.5% |
| Surprised | 53.9% | 76.9% | 75.0% | 63.5% | **80.8%** |
| *Average* | *58.9%* | *80.8%* | *90.0%* | *65.4%* | *71.6%* |

In summary, our participants chose the expected emotion for the text more than any other emotion. There was no difference between the groups that saw a human face and those that saw an avatar face. People were sensitive to context: the percentage of people who gave the expected emotion varied significantly according to the accompanying face, except in the case of the happy texts. Congruent stimuli combinations were better recognized than incongruent stimuli combinations in the case of the angry and the sad text (although when the facial expression was neutral, people also gave the expected emotion at a very high rate), but not in the case of the neutral or the surprised texts.

## 4. Discussion

The main goal of this experiment was to investigate the impact contradictory emotional information might have on people's ability to identify an emotion on a virtual or a human face.

Our participants overwhelmingly assigned the expected emotion to the face, regardless of whether that face was a photorealistic human face or an avatar face, although people in the avatar condition were somewhat less likely to do so. However, the difference between the human and the avatar face was only significant for two stimuli combinations: a neutral face with a sad text and a neutral face with an angry text. In these conditions, many people tended to interpret the neutral avatar face as sad. Finally, while we might have expected that people would assign the expected emotion more often in the congruent condition, this was not the case. The only emotional face that showed some difference between the various stimuli combinations was the neutral avatar face and this result was due solely to the very low recognition rate when that face was paired with a sad text.

These results are different from those obtained by Carroll and Russell [10]. One possible explanation for this is the differences in the design of the two experiments. Carroll and Russell limited their stimuli to a few combinations of text and face emotions that could be mistaken for each other, while our stimuli contained a variety of emotion combinations, including some that could not be confused (e.g., happy and angry). In addition, our text stimuli were very short and presented visually, while Carroll and Russell's were long and were read to the participants.

Our results also differ from others that studied the impact of contradictory emotions using synthetic human faces or voices [11,18,22]. While Hong et al. [18] thought that there was a difference between congruent and incongruent data, we found no such difference. Nass et al. [22] found that people judged a specific emotion as more intense when presented with two congruent sources of emotional information as opposed to two incongruent sources. Creed and Beale [11] found that people judged a specific emotion as more intense in incongruent stimuli when at least one of the two sources of emotion contained that emotion than when neither source contained the emotion. However, our research differed from these last two in that they asked their participants to judge the intensity of each emotion on a scale, whereas we asked participants to name the emotion being presented. It is possible that we might have obtained similar results if we had asked people to judge each individual emotion on a scale[2].

Our results agree with the predictions made by the facial dominance theory [27]. Apart from the neutral avatar face, the majority of our participants assigned the expected emotion to the facial expression. In the case of the neutral face, combining it with a text containing a negative emotion caused a large number of our participants to interpret the face as sad. It is possible that some of the neutral avatar traits are somewhat similar to those on the sad face such that the inclusion of text with negative affect was enough to push some people into interpreting the neutral face as sad. However, these results do not necessarily support Carroll and Russell's [10] limited situational dominance. If that had been the case, then the neutral face should have been interpreted as angry with the angry text and as sad with the sad text. This was not the case.

In the case of assigning emotions to the texts, although our participants generally selected the expected emotion, the various combinations of facial expressions and text had more impact on people's choices than they did when assigning emotions to the faces, except for the happy text. There were two cases where congruent stimuli (and a neutral face combined with an emotional text) were assigned the expected emotion more often than for incongruent stimuli: the angry text and the sad text. However, this was not the case for the surprised text or the neutral text. There are several possible explanations for these results. Our text stimuli were very short, using only two sentences to establish an emotion, compared to Carroll and Russell's [10] use of a full paragraph to set the mood in their experiment. This may have resulted in ambiguous text stimuli, which would explain why people were more sensitive to the various combinations of facial expressions and text emotions. Another potential explanation is that people interpret others' words based on several different input sources. Thus, if a

---

[2] While we did have a question in the ratings task on intensity, this was overall intensity of the facial expression, not the intensity of each of the possible emotions.

person says they are angry and at the same time they are crying, the viewer may rightly conclude that the speaker is hurt and sad. This may have had an impact on how people interpreted the text stimuli. However, it is important to note that, for most of the stimuli combinations, a majority of people did select the expected emotion for the texts.

In conclusion, people appear to interpret facial emotions in a similar way regardless of whether the face expressing the emotion is that of a human or of a medium fidelity avatar. The only exception was the avatar with a neutral face, which was more difficult to interpret in the presence of text with an unpleasant emotion. People are more sensitive to other sources of emotional information when trying to interpret the emotion in short texts, though people's reactions differ depending on the emotion presented in the text.

It is generally safe to use avatars for conveying emotions in an application, but care should be taken when attempting to convey neutrality. We cannot be sure if, in this case, the difficulty with the neutral avatar face was due to the particular avatar used. Independent testing of avatar facial expressions should therefore be performed prior to implementation.

# References

1. Albrecht, I., Haber, J., Kähler, K., Schröder, M. Seider, H.-P.: May I talk to you? :-) – Facial animations from text. In Proc. Pacific Graphics 2002, pp. 77-86. IEEE Computer Society (2002)
2. Bailenson, J.N., Yee, N., Merget, D., Schroeder, R. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. Presence: Teleoperators and Virtual Environments, 15 (4), 359-372, (2006)
3. Bartneck, C. Affective Expressions in Machines. Master's Thesis, Stan Ackerman Institute, Eindhoven (2000)
4. Bartneck, C. Affective expressions in machines. Ext. Abstracts CHI 2001, 189-190, ACM Press (2001)
5. Bartneck, C. How convincing is Mr. Data's smile: Affective expressions of machines. User Modeling and User-Adapted Interaction, 11, 279-295 (2001)
6. Bartneck, C. and Reichenbach, J. Subtle emotional expressions of synthetic characters. Int. J. Human-Computer Studies, 62, 179-192 (2005)
7. Battocchi, A., Pianesi, F. Goren-Bar, D. The properties of DaFEx, a database of kinetic facial expressions. J. Tao, T. Tan, and R.W. Picard (eds.) ACII 2005, LNCS 3784, pp. 558-565 (2005)
8. Buck, R. The Communication of Emotion. New York: Guilford Press (1984)
9. Burford, D., Blake, E. Real-time facial animation for avatars in collaborative virtual environments. Proc. South African Telecommunications Networks and Applications Conference, 178-183 (1999)
10. Carroll, J.M., Russell, J.A. Do facial expressions signal specific emotions? Judging emotion from the face in context. J. Person. Soc. Psych, 70 (2), 205-218 (1996)

11. Creed, C., Beale R. Psychological responses to simulated displays of mismatched emotion expressions. Interacting with Computers, 20 (2), 225-239 (2008)
12. De Gelder, B., Vroomen, J. The perception of emotions by ear and by eye. Cognition and Emotion, 14 (3), 289-311 (2000)
13. Ekman, P. Universals and cultural differences in facial expressions of emotion. In J.K. Cole (Ed.), Nebraska Symp. On Motivation, vol. 19, pp. 207-283. Lincoln, University of Nebraska Press (1972)
14. Ekman, P., Friesen, W. Facial Action Coding System, Consulting Psychologists Press (1978)
15. Ekman, P., Friesen, W., Ellsworth, P. What are the relative contributions of facial behavior and contextual information to the judgment of emotion? In P. Ekman (Ed.) Emotion in the Human Face, (2$^{nd}$ Ed., pp. 111-127). New York, Cambridge University Press (1982)
16. Ehrlich, S.M., Shciano, D.J. and Sheridan, K. Communicating facial affect: It's not the realism, it's the motion. Proc. CHI 2000, ACM Press, 252-253 (2000)
17. Fabri, M., Moore, D. and Hobbs, D. Mediating the expression of emotion in educational collaborative virtual environments: An experimental study. Int. J. of Virtual Reality, 7, 66-81 (2004)
18. Hong, P. Wen, Z. Huang, T.S. Real-time speech-driven face animation with expressions using neural networks. IEEE Transactions on Neural Networks, 13 (1), 100-111 (2002)
19. Izard, C.E. The Face of Emotion. New York, Appleton-Century-Crofts (1971)
20. Kätsyri, J.Klucharev, V., Frydrych, M. and Sams, M. Identification of synthetic and natural emotional facial expressions. Proc. ITRW Audio Visual Speech Proc. Conference, (2003)
21. Ku, J. et al. Experimental results of affective valence and arousal to avatar's facial expressions. Cyberpsychology & Behavior,8 (5), 493-503 (2005)
22. Nass, C. Foehr, U.G., Somoza, M. The effects of emotion of voice in synthesized and recorded speech. Proc. AAAI Symp Emotional and Intelligent II: The Tangled Knot of Social Cognition (2001)
23. Noël, S., Dumoulin, S., Whalen, T., Stewart, J. Recognizing emotions on static and animated avatar faces. Proc. HAVE 2006 (2006)
24. Nogueiras, A., Moreno, A., Bonafonte, A., Marino, J.B. Speech emotion recognition using hidden markov models. Proc. Eurospeech 2001 (2001)
25. Russell, J.A., Bacharowski, J-A., Fernandez-Dols, J.M. Facial and vocal expressions of emotion. Annu. Rev. Psychol. 54, 329-349 (2003)
26. Spencer-Smith, J. et al. Making faces: Creating three-dimensional parameterized models of facial expression. Behavior Research Methods, Instruments & Computers, 33 (2), 115-123 (2001)
27. Tomkins, S.S. Affect, Imagery, Consciousness (vols. 1 and 2). New York, Springer. (1962-1963).