Perception of Blended Emotions: From Video Corpus to Expressive Agent

Stéphanie Buisine[1], Sarkis Abrilian[2], Radoslaw Niewiadomski[3,4],
Jean-Claude Martin[2], Laurence Devillers[2], and Catherine Pelachaud[3]

[1] LCPI-ENSAM, 151 bd de l'Hôpital,75013 Paris, France
`stephanie.buisine@paris.ensam.fr`
[2] LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France
`{sarkis, martin, devil}@limsi.fr`
[3] LINC, IUT of Montreuil, Univ. Paris 8, 140 rue Nouvelle France,
93100 Montreuil, France
`c.pelachaud@iut.univ-paris8.fr, radek@dipmat.unipg.it`
[4] Department of Mathematics and Computer Science, University of Perugia, Italy

Abstract. Real life emotions are often blended and involve several simultaneous superposed or masked emotions. This paper reports on a study on the perception of multimodal emotional behaviors in Embodied Conversational Agents. This experimental study aims at evaluating if people detect properly the signs of emotions in different modalities (speech, facial expressions, gestures) when they appear to be superposed or masked. We compared the perception of emotional behaviors annotated in a corpus of TV interviews and replayed by an expressive agent at different levels of abstraction. The results provide insights on the use of such protocols for studying the effect of various models and modalities on the perception of complex emotions.

1   Introduction

Affective behaviors in Embodied Conversational Agents (ECAs) can be quite useful for experimental studies on the perception of multimodal emotional behaviors as one can turn on/off a given signal or even a given modality. Real life emotions are often complex and involve several simultaneous emotions [15, 17, 33]. They may occur either as the quick succession of different emotions, the superposition of emotions, the masking of one emotion by another one, the suppression of an emotion or the overacting of an emotion. We refer to blend of emotions to denote these phenomena. These blends produce "multiple simultaneous facial expressions" [30].

Depending on the type of blending, the resulting facial expressions are not identical. A masked emotion may leak over the displayed emotion [17]; while superposition of two emotions will be shown by different facial features (one emotion being shown on the upper face while another one on the lower face) [17]. Distinguishing these various types of blends of emotions in ECA systems is relevant as perceptual studies have shown that people are able to recognize facial expression of felt emotion [14, 37] as well as fake emotion [16] from real life as well as on ECAs [27]. Moreover, in a study on deceiving agent, Rhem and André [29] found that the users were able to

differentiate when the agent was displaying expressions of felt emotion or expression of fake emotion.

Video corpora of TV interviews enable to explore how people behave during such blended emotions not only by their facial expression but also by their gestures or their speech [11]. Yet, these corpora call for means of validating subjective manual annotations of emotion. Few researchers have used ECAs for validating such manual annotations by testing how people perceive the replay of annotated behaviors by an agent. Ten Ham et al. [34] compared the perception of a video of a human guide vs. an agent using the same speech and similar non-verbal behaviors during a route description task but they did not consider emotion. Becker et al. [5] conducted a study to evaluate the affective feedback of an agent in a card game. They found that the absence of negative emotions from the agent was evaluated as stress-inducing whereas the display of empathic feedback supported the acceptance of the agent as a co-equal opponent.

Aiming at understanding if facial features or regions play identical roles in emotion recognition, researchers performed various perceptual tasks or studied psychological facial activity [4, 7, 8, 20]. They found that positive emotions are mainly perceived from the expression of the lower face (e.g. smile) while negative emotion from the upper face (e.g. frown). One can conclude that reliable features for positive emotion, that is features that convey the strongest characteristics of a positive emotion, are in the lower face. On the other hand, the most reliable features for negative emotion are in the upper face.

Based on these findings we have developed a computational model for facial expressions of blend of emotions. It composes facial expressions from those of single emotions using fuzzy logic rules [26]. Very few models of blended emotions have been developed so far for ECAs. The interpolation between facial parameters of given expressions is commonly used to compute the new expression [3, 12, 27, 31].

This paper reports on an experimental study aiming at evaluating if people detect properly the signs of different emotions in multiple modalities (speech, facial expressions, gestures) when they appear to be superposed or masked. It compares the perception of emotional behaviors in videos of TV interviews with similar behaviors replayed by an expressive agent. The facial expressions of the agent are defined using one of two approaches, namely the computational model of blend of emotions (hereafter called "facial blending replay"), or the annotation of the facial expressions from the video ("multiple levels replay"). We are also interested in evaluating possible differences between visual only vs. audio-visual perception as well as possible gender differences. We aim to test if findings reported in [18, 21] can be replicated here, that is if women tend to be better at recognizing facial expressions of emotions.

Section 2 summarizes our previous work and describes how to replay multimodal emotional behavior from manual annotations. The replay integrates models of expressive behaviors and blended facial expressions. Section 3 describes the protocol. The results are presented and discussed in sections 4 and 5. We conclude in section 6 on the use of such protocols for studying the effect of various models and modalities on the perception of blends of emotions.

2   Annotating and Replaying Multimodal Emotional Behaviors

In order to study multimodal behaviors during real-life emotions, we have collected a corpus of emotionally rich TV interviews [10]. Several levels of annotation were manually coded using Anvil [23]: some information regard the whole video (called the "global level"); while some other information are related to emotional segments (the "local level"); at the lowest level, there is detailed time-based annotation of multimodal behaviors. Three expert coders defined the borders of the emotionally consistent segments of the clip and labeled each resulting segment with one or two labels. The annotation of multimodal behavior includes gesture expressivity since it was observed to be involved in the perception of emotion [24].

Besides, we have created an ECA system, Greta, that incorporates communicative conversational and emotional qualities [28]. Our model of expressivity is based on studies by researchers such as [19, 35, 36]. We describe expressivity by a set of 6 dimensions: Spatial extent, Temporal extent, Power, Fluidity, Repetition and Overall activity [22]. The Greta system takes as input a text tagged with communicative functions described with APML labels [9] as well as values for the expressivity dimensions that characterize the manner of execution of the agent's behaviors. The system parses the input text and selects which behaviors to perform. Gestures and other non-verbal behaviors (facial expressions and gaze behaviors) are synchronized with speech. The system looks for the emphasis word. It aligns the facial expressions and the stroke of a gesture with this word. Then it computes when the preparation phase of the gesture is as well as if a gesture is hold, co-articulates to the next one, if time between consecutive gestures allows it, or returns to the rest position.

We have defined two corpus-based approaches to design different Greta animations based on the video annotations [26]. The "multiple levels replay" approach involves the level of annotation of emotions, and the low-level annotations of multimodal behaviors (such as the gesture expressivity for assigning values to the expressivity parameters of the ECA, and the manual annotation of facial expressions) [25]. The "facial blending replay" approach is identical to the "multiple levels replay" approach except for facial expressions: it uses a computational model for generating facial expressions of blend of emotions [25]. More details are provided below on how these two approaches have been used in our perceptual study.

3   Experimental Protocol

3.1   Protocol Description

The goals of our experiment are to 1) test if subjects perceive a combination of emotions in the replays as in the original videos, and 2) compare the two approaches for replaying blended emotions. We have selected two different video clips of TV interviews for this study, each featuring a different type of blend.

The 1st clip (video #3, 3rd segment) features a woman reacting to a recent trial in which her father and her brother were kept in jail. As revealed by the manual annotation of this video by 3 expert coders, her behavior is perceived as a superposition of anger and despair. This is confirmed by the annotation by 40 coders with various

levels of expertise [2]. This emotional behavior is perceived in speech and in several visual modalities (gaze, head movements, torso movements and gestures).

The 2nd clip (video #41) features a woman pretending to be positive after having received the negative election results of her political party, thus masking her disappointment by a smile. Such a video has been annotated as a combination of negative labels (disappointment, sadness, anger) and positive labels (pleased, serenity). The annotation of multimodal behaviors reveals that, for this segment, her lips show a smile but a tense smile that is with pressed lips.

With respect to the contextual cues provided by the audio and the visual channels that might influence the subjects' perception of emotion, both channels provide information on the location (outdoor for video #3, indoor room with other people in video #41). Video #3 features both head and hands movements. Video #41 features only the face in close-up (the hands are not visible). The politician seen in video #41 is not a major figure.

40 subjects (23 males, 17 females), age between 19 and 36 (average 24) had to compare the original videos and the different Greta animations. 33 subjects were students in computer science, 7 were researchers, teachers or engineers. The experiment included two conditions: first without audio, and then with audio. In each condition, the subjects played the original video and four different animations. Two animations were specified with data from the literature on basic emotions in facial expressions [14] and body movements [35]. The two other animations were generated with the two approaches mentioned above for replaying annotated behaviors.

Thus, for the superposition example of emotion in clip #3, four animations were designed: 1) Anger, 2) Despair, 3) multiple levels replay, and 4) facial blending replay. For the facial blending replay, the values assigned to the gesture expressivity parameters were computed from the multiple levels replay (e.g. from the manual annotation of perceived expressivity of hand gestures).

Similarly, for the masking of emotion example in clip #41, the four animations were: 1) Joy, 2) Disappointment, 3) multiple levels replay, and 4) facial blending replay.

Subjects had to assign a value between 1 (high similarity with the video) and 4 (low similarity) to each animation (Fig. 1). The order of presentation of the superposition and masking example, and the location on the graphical interface of the corresponding animations in the audio and no audio conditions were counterbalanced across subjects. Subjects could assign the same similarity value to several animations.

After each condition, subjects had to answer a questionnaire. They had to report on their confidence when assigning similarity values. They could select between 5 confidence scores: 1) I clearly perceived differences between the 4 animations and I easily compared them to the video (4-point confidence score), 2) I perceived some differences that enabled me to do my evaluation (3-point), 3) I perceived some differences but had difficulties to compare the animations with the video (2-point), 4) I perceived few differences between the animations and had a lot of difficulties to evaluate them (1-point), 5) I did not perceive any differences between the animations (0-point).

In the questionnaire, subjects also had to annotate the emotions that they perceived in the animation they ranked as the most similar to the original video. They could select one or several emotion labels from the same list of 18 emotional labels that had been used for the annotation of the videos in a previous experiment [2].
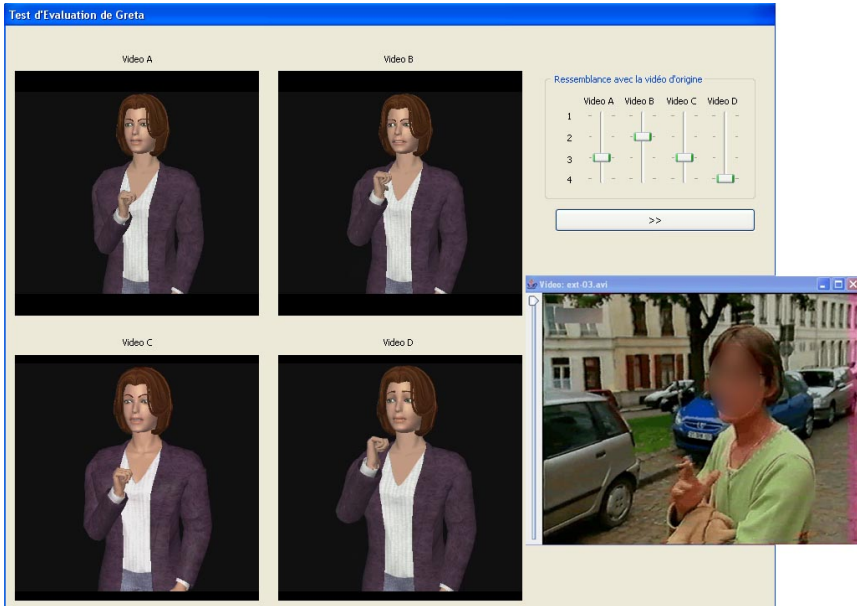
Fig. 1. Screen dump of the superposition example ; 4 different animations and 4 sliders for selecting a similarity value for each animation ; the original video #3 (non blurred during the test) is displayed separately ; the video and the animations feature the facial expressions and the hand gestures ; the masking example is similar to this display but focuses on the face in the video #41 and in the corresponding ECA animations

3.2 Using "Multiple Levels" and "Facial Blending" Replays in the Study

As we explained above, the "multiple levels replay" and the "facial blending replay" differ only by the computation of facial expressions [26]. In this section, we explain how they were used for the perception study.

Our computational model of facial expressions arising from blends of emotions is used in the "facial blending replay". It is based on a face partition approach. Any facial expression is divided into *n* areas. Each *area* represents a unique facial part like brows or lips. The model computes the complex facial expressions of emotions and distinguishes between different types of blending (e.g., superposition and masking). The complex facial expressions are created by composing the face areas of the two source expressions. Different types of blending are implemented with different sets of fuzzy rules for the computation of the complex facial expression. The fuzzy rules are based on Ekman's research on blends of emotions [17].

Figure 2 shows the agent displaying the masked expression of disappointment (computed as similar to sadness) and fake joy. The images a) and b) display the expressions of disappointment and joy, respectively. Image d) shows the masking expression computed by the "facial blending replay". We can notice that the absence of orbicularis oculi activity as indicator of unfelt joy[13] is visible on both images (c) and (d), the annotated video and the corresponding Greta simulation.

(a) disappointment   (b) joy   (c) original video   (d) masking of disappointment by joy computed in the "facial blending" replay
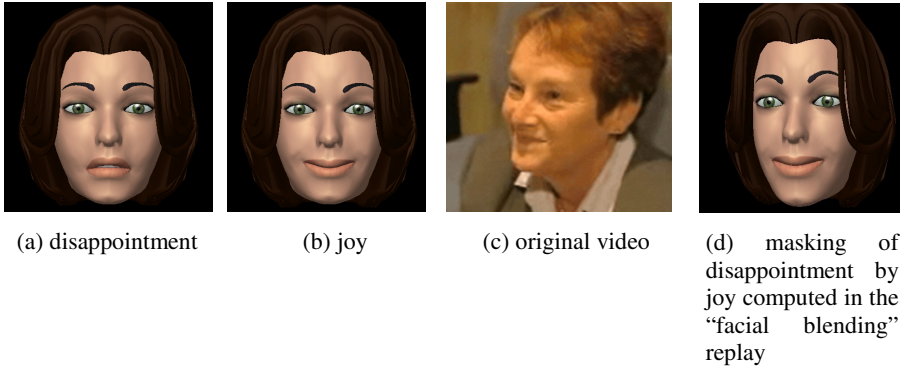
Fig. 2. Disappointment masked by joy

In this example, single emotions (disapointment (a) and joy (b)) are defined in the system using Ekman's research. In the "facial blending replay", the facial expression is computed using the blending model for masking. In the "multiple levels replay", the facial expressions are not generated from system predefined information. Instead, facial parameters such as brows movements, gaze direction, or mouth tension, have been specified out of the manual annotations of the original video. A correspondance table between the manual annotations and MPEG-4 Facial Animation specifications has been defined in this purpose.

With respect to the audio channel, in order to avoid a bias due to speech synthesis quality in the evaluation of similarity between the ECA animations and the original video, we used in the animations the real speech from the original video.

4   Results

4.1 Superposition of Emotions

We computed the number of times each animation was ranked as the closest to the video. In the no audio condition, Anger is perceived as the closest animation by 61% of the subjects (multiple levels replay 20%, facial blending 9%, Despair 9%). In the audio condition, Anger is perceived as the closest animation by 33% of the subjects (multiple levels replay 26%, facial blending 24%, Despair 17%). The perception of superposed emotions in the 1st clip was also examined using an analysis of variance with Audio output (no audio, audio) and Animation (multiple levels replay, facial blending replay, anger, despair) as within-subjects factors. Gender of subjects (male, female) was included as between-subjects factor. Rankings of animations were converted into similarity scores (the first rank became a 3-point score of similarity; the fourth rank became a 0-point score). The main effect of Animation proved to be significant ($F(1/114)=15.86$; $p<0.001$, see Fig. 3). The similarity score for the Anger animation is significantly higher than the multiple levels replay ($t(39)=3.05$; $p=0.004$); the multiple levels and facial blending replays are not significantly different from one another ($t(39)=1.65$; NS); and the similarity score of the Despair animation

tends to be lower than the facial blending replay (t(39)=1.83; p=0.076). The Audio * Animation interaction appeared significant (F(1/114)=5.98; p=0.001, see Fig. 4): this interaction shows that the Animation effect is highly significant in the no-audio condition (F(3/114)=24.11; p<0.001) whereas it is only marginal in the audio condition (F(2/114)=2.42; p=0.087).

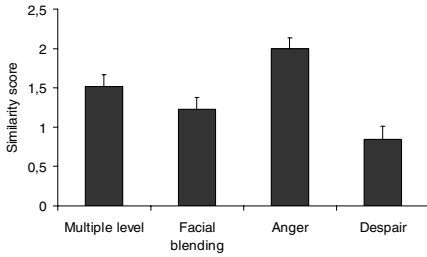**Fig. 3.** Similarity scores of the 4 animations with the superposition example
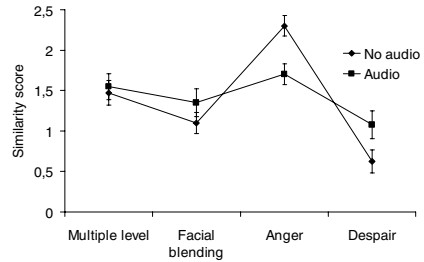
Fig. 4. Similarity scores of the 4 animations as a function of the condition

The Gender * Animation interaction is also significant (F(3/114)=3.61; p=0.016, see Fig. 5). Female subjects gave significantly lower similarity scores to the facial blending replay than male subjects (t(38)=2.70; p=0.010). The similarity scores of the 3 other animations are not significantly different between male and female subjects.
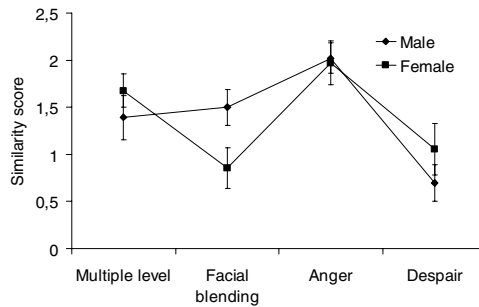
Fig. 5. Similarity scores of the 4 animations as a function of subjects' gender

The overall confidence score reported by subjects in the questionnaire as associated to the ranking of the animations amounts to 2.8/4. There is no main effect of Audio on this variable (F(1/38)=0.45; NS) but the main effect of Gender tends to be significant (F(1/38)=3.82; p=0.058). Female subjects had higher confidence scores (3.1/4) than male subjects (2.6/4).

In the questionnaire, the subjects also had to characterize the animation that they had ranked as the closest to the original video, each subject being allowed to choose more than one label. In the no-audio condition, 82.5% of subjects perceived anger,

whereas only 12.5% perceived despair. In the audio condition, the main emotion perceived by subjects was sadness (85% of subjects). Anger was perceived by 67.5% of subjects and despair by 42.5% of subjects. For the no audio condition, only 6 subjects (15%) selected a single label (4 of them for the Anger animation). For the audio condition, only 3 subjects (7,5%) did select a single label (each time the "Despair" label for the animation was selected). Table 1 summarizes the perception of superposed vs. single emotions using macro-classes categories (Anger, Sadness). These results show that, even though the animation perceived as the closest to the original video was the "Anger" animation, subjects nevertheless perceived it as a combination of several emotions.

Table 1. Percentages of subjects who perceived each macro-class of emotion in the animation that they had ranked as the closest to the original video of superposed emotions

	Anger + Sadness	Anger but not Sadness	Sadness but not Anger	Neither Anger nor Sadness	Total
No audio	52,5 %	40 %	0 %	7,5 %	100 %
Audio	85 %	0 %	12,5 %	2,5 %	100 %

4.2   Masking of Emotions

Regarding the masking emotion example, in the no audio condition, Joy was perceived as the closest animation by 40% of the subjects (facial blending 33%, multiple levels replay 20%, Disappointment 7%). In the audio condition, facial blending was perceived as the closest animation by 38% of the subjects (multiple levels replay 27%, Joy 24%, Disappointment 11%).

The perception of masked emotion was also studied by means of an analysis of variance with Audio output (no audio, audio), Animation (multiple levels replay, facial blending replay, Joy, Disappointment) as within-subjects factors and Gender (male, female) as between-subjects factor. The main effect of Animation was the only one that proved significant ($F(1/114)=18.07$; $p<0.001$, see Fig. 6). The similarity score of the facial blending replay is significantly higher than the multiple levels replay ($t(39)=2.05$; $p=0.047$) but not significantly different from Joy ($t(39)=1.36$; NS). The difference between Joy and the multiple levels replay is not significant either ($t(39)=0.83$; NS). However, the similarity score of the Disappointment animation is significantly lower than the multiple levels replay ($t(39)=4.77$; $p<0.001$).

The overall confidence score of subjects for the ranking of these animations is 2.5/4. The analysis showed no main effect of Audio ($F(1/38)=0.21$; NS) nor Gender ($F(1/38)=0.12$; NS) on this variable. In the no audio condition, 35% of subjects rated Disappointment with the label Embarrassment. Joy and Pleased were perceived by 15% and 32.5% of subjects respectively. Disappointment was mentioned by only 5% of subjects. In the audio condition, Worry and Sadness were the most perceived emotions (47.5% and 35% respectively). Joy and Pleased represented only 5% and 7.5%. Positive emotions (e.g. Pleased, Serenity, Joy) represented 32% of labels in the no
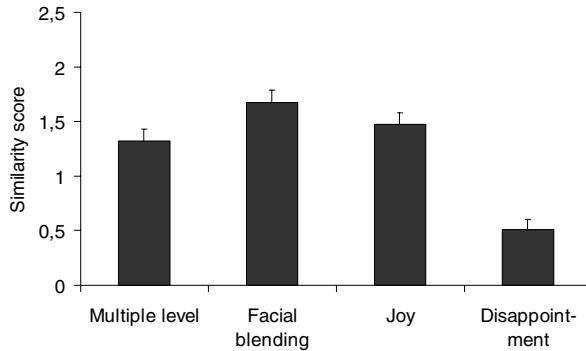
Fig. 6. Similarity scores of the four animations with the original video in the masking example

audio condition and 13% in the audio condition. Negative emotions (e.g. Embarrassment, Doubt, Sadness) represented 43% of labels in the no audio condition and 81% in the audio condition. Only 4 subjects selected a single label in the no audio condition and only 3 subjects in the audio condition. Moreover, some subjects (11 in the no audio condition, 9 in the audio condition) used both positive and negative labels to describe the animations. Table 2 summarizes this perception of mixed vs. single emotions using macro-classes categories.

Table 2. Percentages of subjects who perceived each macro-class of emotion in the animation that they had ranked as the closest to the original video of masked emotions

	Joy + Sadness	Joy but not Sadness	Sadness but not Joy	Neither Joy nor Sadness	Total
No audio	7,5 %	10 %	57,5 %	25 %	100 %
Audio	12,5 %	40 %	30 %	17,5 %	100 %

4.3 Effects of the Complex Emotion Models

We conducted an analysis of variance on the 2 examples of blend of emotions (superposition and masking), the 2 conditions (no audio, audio), our 2 approaches of generation (multiple levels replay and facial blending replay), and subjects' gender. The results show no main effect of approach ($F(1/38)=0.01$; NS), i.e. no overall significant difference between the multiple levels replay and the facial blending replay.

5 Discussion

The 1st goal of this study was to test whether subjects perceive a combination of emotions in our ECA animations. Our results show that subjects tend to perceive one emotion as being predominant: in the superposition example, subjects ordered what we have called "basic Anger" first; in the disappointment masked by joy, basic Joy and facial blending replay were rated similarly. However, this result is partially con-

tradicted when analyzing the result from the labeling task where most subjects associated several emotion labels to the animation they considered as being the most similar to the video. The macro-classes of emotion that they selected are consistent with previous annotations of the videos with 3 expert coders (and 40 coders for the superposition example [2]). In the animations of single emotions that we designed for the superposition example, subjects perceived a secondary emotion, possibly through non-verbal or verbal cues brought by our method of copying the original video. This reveals that these animations, designed to display single emotions, feature instead a relative complexity of emotions.

Besides, our experimental protocol enabled to compare two conditions, audio and no audio. The 4 animations corresponding to the superposition example were better discriminated in the no audio condition; the addition of verbal cues had the effect to lower down the differences between the animations. Moreover, users chose a larger set of emotions to label the animations in the audio condition. In the superposition animations, subjects better perceived the sadness / despair dimension in the audio condition. We may thus infer that, in this particular sequence, anger was primarily displayed by nonverbal behaviors whereas despair was mostly expressed by the verbal stream.

In the disappointment-masked-by-joy example, subjects perceived negative emotions better in the audio condition. This suggests that the person in this video sequence controlled her nonverbal behavior (positive emotions perceived in the mute condition) better than her verbal behaviors (negative cues perceived in the audio condition). This strong effect of verbal behavior may be due to the fact that we used the original voice and not a synthesized voice. This perception of valence of emotion is reported in Table 3. It is compatible with previous studies on the whole corpus in which 2 coders annotated the videos in 3 conditions (Audio only, Video only, Audio and video) [1]. We had observed a high level of quantitative agreement over the whole corpus for the 3 conditions for "Anger". The coders agreed on the "Pain" label for the "Audio only" and "Audio and video" conditions but not for the "Video only" condition, which could be explained by the fact that acoustic cues show well this emotion.

Table 3. Valence of labels assigned to the animations perceived as being closest to the videos

	Positive only	Negative only	Pos. / Neg. conflict	Neutral
Superposition no audio	2,5%	95%	2,5%	0%
Superposition with audio	0%	100%	0%	0%
Masking no audio	32,5%	30%	35%	2,5%
Masking with audio	7,5%	52,5%	37,5%	2,5%

The second goal of this study was to evaluate our two approaches for replaying blends of emotions (multiple levels replay and facial blending replay). Our global analysis (Section 4.3) showed that none of the two approaches was consistently preferred. We may first mention that the two replay approaches share some common features which may partly explain why it was not so easy to discriminate between

them: they were elaborated on the basis of the same annotations and they both included some behaviors automatically generated by the Greta system. However, we can notice that the facial blending replay was significantly better ranked than the multiple levels replay in the masking example. Conversely in the superposition example, the two types of replay did not differ significantly except for female subjects who gave higher similarity scores to the multiple levels replay.

These results suggest that the facial blending replay was more appropriate in the masking example and the multiple levels replay in the superposition example. These results alone are not sufficient to understand this interaction. It could be due to the nature of complex emotions (superposition, masking), or to the particular emotions tested here (anger / despair, disappointment / joy), or to the modalities available (with or without hand gesture). For example, the size of the face is not the same in our superposition and in the masking examples.

Finally, in the superposition example, male and female subjects judged differently the animation generated by facial blending replay. Further data would be necessary to interpret this effect: for example we could ask subjects to annotate precisely these animations (not only label them, but also annotate the animation of different parts of the face), and examine whether there are gender differences when performing this task. Male and female subjects also gave different confidence scores to their ranking in the superposition example: female subjects were more confident in their answers. Such a result is consistent with classical view that female subjects have better abilities to decode nonverbal cues [18, 21].

6   Conclusions and Future Work

We compared the perception of emotional behaviors annotated in a corpus of TV interviews replayed by an expressive agent at different levels of abstraction. One drawback of our corpus approach based on spontaneous behaviors collected during TV interviews is that it is quite difficult to collect enough data to train statistical models. That is the reason why we use instead an exploratory approach with selected illustrative cases of different complex emotions (e.g. superposition and masking) for validating our representations.

Thus, future work is needed to validate our copy-synthesis approach (annotation of emotion and expressivity, computation of signs of blended emotions, animation by an ECA). Anyhow, the results we reported show that both replay approaches might be of interest for different types of blend of emotions. We also intend to improve the design of our "basic" emotion animations and the method of comparing them with our replay approaches. We will involve complementary videos with respect to the combination of types of blending and modalities. We plan to adapt our copy-synthesis approach so that two different individual emotions that are combined can be assigned different weights. The computational model of the agent expressivity also needs to be improved to better simulate expressive arm movements as well as to better match behaviors observed in such videos (e.g. movement of the torso, separate specifications of expressivity for different body parts). We will use information that we have collected on subjects' personality using the EPI (Eysenck Personality Inventory) questionnaire since introversion / extraversion has been observed to have an impact on the perception of multimodal behaviors displayed by ECAs [6]. Further testing of the influence

of visual and audio channels should also be investigated (e.g. use of synthetic speech, or filtered audio rendering the speech content unintelligible with minimal effect on prosody and voice quality [32]).

We believe that such experimental studies will enable to identify the parts of expressions that are most critical to the perception and display of real-life multimodal emotions.

Acknowledgements

References

1. Abrilian, S., Devillers, L., Buisine, S., Martin, J.-C.: EmoTV1: Annotation of Real-life Emotions for the Specification of Multimodal Affective Interfaces. 11th Int. Conf. Human-Computer Interaction (HCII'2005) (2005) Las Vegas, Nevada, USA

2. Abrilian, S., Devillers, L., Martin, J.-C.: Annotation of Emotions in Real-Life Video Interviews: Variability between Coders. 5th Int. Conf. Language Resources and Evaluation (LREC'2006) (2006) Genoa, Italy

3. Albrecht, I., Schröder, M., Haber, J., Seidel, H.-P.: Mixed feelings: Expression of non-basic emotions in a muscle-based talking head. Special issue of Journal of Virtual Reality on "Language, Speech & Gesture" 8 4 (2005)

4. Bassili, J. N.: Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. Jour. Pers. Soc. Psychol. 37 11 (1979)

5. Becker, C., Prendinger, H., Ishizuka, M., Wachsmuth, I.: Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game. 1st International Conference on Affective Computing & Intelligent Interaction (ACII'2005) (2005) Beijing, China 466-473

6. Buisine, S. Conception et Évaluation d'Agents Conversationnels Multimodaux Bidirectionnels. PhD Thesis. Doctorat de Psychologie Cognitive - Ergonomie, Paris V. 8 avril 2005. Direction J.-C. Martin & J.-C. Sperandio. 2005. http://stephanie.buisine.free.fr/

7. Cacioppo, J. T., Petty, R. P., Losch, M. E., Kim, H. S.: Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. Journal of Personality and Social Psychology 50 (1986)

8. Constantini, E., Pianesi, F., Prete, M.: Recognizing Emotions in Human and Synthetic Faces: The Role of the Upper and Lower Parts of the Face. Intelligent User Interfaces (IUI'05) (2005) San Diego, CA, USA 20-27

9. De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APML, a Markup Language for Believable Behavior Generation. Life-like characters. Tools, affective functions and applications. Springer (2004)

10. Devillers, L., Abrilian, S., Martin, J.-C.: Representing real life emotions in audiovisual data with non basic emotional patterns and context features. 1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005) (2005) Beijing, China 519-526

11. Douglas-Cowie, E., Devillers, L., Martin, J.-C., Cowie, R., Savvidou, S., Abrilian, S., Cox, C.: Multimodal Databases of Everyday Emotion: Facing up to Complexity. 9th European Conf. Speech Communication and Technology (Interspeech'2005) (2005) Lisbon, Portugal 813-816

12. Duy Bui, T. Creating Emotions And Facial Expressions For Embodied Agents. PhD Thesis. University of Twente. 2004.

13. Ekman, P.: Darwin, Deception, and Facial Expression. Annals of the New York Academy of Sciences 1000 (2003)

14. Ekman, P.: Emotion in the human face. Cambridge University Press (1982)

15. Ekman, P.: The Face Revealed. Weidenfeld & Nicolson London (2003)

16. Ekman, P., Friesen, W.: Felt, false, miserable smiles. Journal of Nonverbal Behavior 6 4 (1982)

17. Ekman, P., Friesen, W. V.: Unmasking the face. A guide to recognizing emotions from facial clues. Prentice-Hall Inc., Englewood Cliffs, N.J. (1975)

18. Feldman, R. S., Philippot, P., Custrini, R. J.: Social competence and nonverbal behavior. Fundamentals of Nonverbal Behavior. Cambridge University Press (1991)

19. Gallaher, P.: Individual differences in nonverbal behavior: Dimensions of style. Journal of Personality and Social Psychology 63 (1992)

20. Gouta, K., Miyamoto, M.: Emotion recognition, facial components associated with various emotions. Shinrigaku Kenkyu 71 3 (2000)

21. Hall, J. A., Matsumoto, D.: Gender differences in judgments of multiple emotions from facial expressions. Emotion 4 2 (2004)

22. Hartmann, B., Mancini, M., Pelachaud, C.: Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. Gesture Workshop (GW'2005) (2005) Vannes, France

23. Kipp, M.: Gesture Generation by Imitation. From Human Behavior to Computer Character Animation. Boca Raton, Dissertation.com Florida (2004)

24. Martin, J.-C., Abrilian, S., Devillers, L.: Annotating Multimodal Behaviors Occurring during Non Basic Emotions. 1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005) (2005) Beijing, China 550-557

25. Martin, J.-C., Abrilian, S., Devillers, L., Lamolle, M., Mancini, M., Pelachaud, C.: Levels of Representation in the Annotation of Emotion for the Specification of Expressivity in ECAs. 5th International Working Conference On Intelligent Virtual Agents (IVA'2005) (2005) Kos, Greece 405-417

26. Martin, J.-C., Niewiadomski, R., Devillers, L., Buisine, S., Pelachaud, C.: Multimodal Complex Emotions: Gesture Expressivity And Blended Facial Expressions. Special issue of the Journal of Humanoid Robotics. Eds: C. Pelachaud, L. Canamero. (to appear)

27. Pandzic, I. S., Forchheimer, R.: MPEG-4 Facial Animation. The Standard, Implementation and Applications. John Wiley & Sons, LTD (2002)

28. Pelachaud, C.: Multimodal expressive embodied conversational agent. ACM Multimedia, Brave New Topics session (2005) Singapore 683 - 689

29. Rehm, M., André, E.: Catch Me If You Can - Exploring Lying Agents in Social Settings. Int. Conf. Autonomous Agents and Multiagent Systems (AAMAS'2005) (2005) 937-944

30. Richmond, V. P., Croskey, J. C.: Non Verbal Behavior in Interpersonal relations. Allyn & Bacon Inc. (1999)

31. Ruttkay, Z., Noot, H., ten Hagen, P.: Emotion Disc and Emotion Squares: tools to explore the facial expression face. Computer Graphics Forum 22 1 (2003)

32. Savvidou, S., Cowie, R., Douglas-Cowie, E.: Contributions of Visual and Auditory Channels to Detection of Emotion. British Psychological Society Annual Conference (NI Branch) (2001) Cavan, Republic of Ireland

33. Scherer, K. R.: Analyzing Emotion Blends. Proceedings of the Xth Conference of the International Society for Research on Emotions (1998) Würzburg, Germany 142-148

34. ten Ham, R., Theune, M., Heuvelman, A., Verleur, R.: Judging Laura: Perceived Qualities of a Mediated Human Versus an Embodied Agent. 5th International Working Conference On Intelligent Virtual Agents (IVA'2005) (2005) Kos, Greece 381-393

35. Wallbott, H. G.: Bodily expression of emotion. European Journal of Social Psychology 28 (1998)

36. Wallbott, H. G., Scherer, K. R.: Cues and Channels in Emotion Recognition. Journal of Personality and Social Psychology 51 4 (1986)

37. Wiggers, M.: Jugments of facial expressions of emotion predicted from facial behavior. Journal of Nonverbal Behavior 7 2 (1982)