Comparing a novel model based on the transferable belief model with humans during the recognition of partially occluded facial expressions

Zakia Hammal
Département de Psychologie, Université de Montréal, Canada

Martin Arguin
Département de Psychologie, Université de Montréal, Canada

Frédéric Gosselin
Département de Psychologie, Université de Montréal, Canada

Humans recognize basic facial expressions effortlessly. Yet, despite a considerable amount of research, this task remains elusive for computer vision systems. Here, we compared the behavior of one of the best computer models of facial expression recognition (Z. Hammal, L. Couvreur, A. Caplier, & M. Rombaut, 2007) with the behavior of human observers during the M. Smith, G. Cottrell, F. Gosselin, and P. G. Schyns (2005) facial expression recognition task performed on stimuli randomly sampled using Gaussian apertures. The model—which we had to significantly modify in order to give the ability to deal with partially occluded stimuli—classifies the six basic facial expressions (Happiness, Fear, Sadness, Surprise, Anger, and Disgust) plus Neutral from static images based on the permanent facial feature deformations and the Transferable Belief Model (TBM). Three simulations demonstrated the suitability of the TBM-based model to deal with partially occluded facial parts and revealed the differences between the facial information used by humans and by the model. This opens promising perspectives for the future development of the model.

Keywords: facial features behavior, facial expressions classification, Transferable Belief Model, Bubbles

Citation: Hammal, Z., Arguin, M., & Gosselin, F. (2009). Comparing a novel model based on the transferable belief model with humans during the recognition of partially occluded facial expressions. *Journal of Vision, 9*(2):22, 1–19, http://journalofvision.org/9/2/22/, doi:10.1167/9.2.22.

Introduction

Facial expressions communicate information from which we can quickly infer the state of mind of our peers and adjust our behavior accordingly (Darwin, 1872). To illustrate, take a person like patient SM with complete bilateral damage to the amygdala nuclei that prevents her from recognizing facial expressions of fear. SM would be incapable of interpreting the fearful expression on the face of a bystander, who has encountered a furious Grizzly bear, as a sign of potential threat (Adolphs, Tranel, Damasio, & Damasio, 1994).

Facial expressions are typically arranged into six universally recognized basic categories *Happiness, Surprise, Disgust, Anger, Sadness,* and *Fear* that are similarly expressed across different backgrounds and cultures (Cohn, 2006; Ekman, 1999; Izard, 1971, 1994). Facial expressions result from the precisely choreographed deformation of facial features, which are often described using the 46 Action Units (AUs; Ekman & Friesen, 1978).

Facial expression recognition and computer vision

The study of human facial expressions has an impact in several areas of life such as art, social interaction, cognitive science, medicine, security, affective computing, and human-computer interaction (HCI). An automatic facial expressions classification system may contribute significantly to the development of all these disciplines. However, the development of such a system constitutes a significant challenge because of the many constraints that are imposed by its application in a real-world context (Pantic & Bartlett, 2007; Pantic & Patras, 2006). In particular, such systems need to provide great accuracy and robustness without demanding too many interventions from the user.

There have been major advances in computer vision over the past 15 years for the recognition of the six basic facial expressions (for reviews, see Fasel & Luettin, 2003; Pantic & Rothkrantz, 2000b). The main approaches can be divided in two classes: Model-based and fiducial points approaches. The model-based approach requires the design of a deterministic physical model that can represent

doi: 10.1167/9.2.22 Received January 28, 2008; published February 26, 2009 ISSN 1534-7362 © ARVO

accurately all the geometrical properties of faces, especially muscle activity in faces. This turned out to be extremely difficult to achieve. Moreover, a model-based approach usually involves an intensive training stage. Finally, the trained model is sometimes unable to represent individual differences. These shortcomings led us, like some others, to attack the problem differently using local facial cues—the fiducial points approach—as described in the next section.

Fiducial points approach

The fiducial-based systems compare the deformation of the permanent facial features (eyes, eyebrows, and mouth) with a reference state, i.e., the neutral facial expression. In this approach, the facial movements are quantified by measuring the geometrical displacement of facial landmarks between two frames. They are then classified into AUs or into the six basic facial expressions according to the obtained observations.

For the recognition of three upper Action Units (AUs, which represent a contraction or a relaxation of one or more muscle) corresponding to the deformation of the upper half of the face and six lower AUs corresponding to the deformation of the bottom half of the face, Lien, Kanade, Cohn, and Li (1998), for example, proposed a hybrid method based on manually detected feature points (tracked by Lucas–Kanade tracking algorithm—Lucas & Kanade, 1981—in the remaining of the sequence) and furrows. They used a Hidden Markov Model for each facial state characterized by one AU or combination of AUs. The main drawback of the method is the number of Hidden Markov Models required to detect a great number of AUs or combinations of AUs involved in the classification of facial expressions. Using the same data as Lien et al., Tian, Kanade, and Cohn (2001) used a Neural Network and obtained better classification results than Lien et al.

Pantic et al. (Pantic & Patras, 2005, 2006; Pantic & Rothkrantz, 2000a, 2000b) and then Zhang and Qiang (2005) were the first to take into account the temporal information in the classification process of the AUs and the facial expressions. Pantic et al. (Pantic & Patras, 2005, 2006; Pantic & Rothkrantz, 2000a, 2000b), Hammal et al. (2007) used two face models made of frontal and side views. The eyes, eyebrows, and mouth are automatically segmented and transformed into AUs through the application of a set of rules (Pantic & Rothkrantz, 2000a); and each AU is divided into three time segments: the onset, the apex (peak), and the offset. The classification is performed using Ekman's Facial Action Coding System (FACS), which consists in describing each facial activation as a combination of one or more specific AUs (Ekman & Friesen, 1978).

A limitation associated with the uses of FACS, however, is that the FACS scores are only qualitative and thus provide no categorization rule. Most of the systems making use of the FACS aim at recognizing the different AUs without actually recognizing the displayed facial expression. These systems then bypass the problem of doubt between multiple expressions that can occur. Overcoming this limitation, Zhang and Qiang (2005) proposed a FACS-based model classifying the six basic expressions. It uses a multi-sensory information fusion technique based on a dynamic Bayesian network. Eyes, eyebrows, nose, mouth, and transient features are used for Action Unit (AU) detection. The permanent facial features are automatically detected in the first frame and then tracked in the remaining frames of the sequence. The classification result obtained at time $t - 1$ is added to the characteristic features vector at time t. Contrary to the FACS-based methods described above, the classification results of the AUs obtained by the dynamic Bayesian network are combined using a rules table defined by the authors (2005) to associate to each AU, or combination of AUs, only one of the six basic facial expressions.

Rather than using the FACS modeling process, Tsapatsoulis, Karpouzis, Stamou, Piat, and Kollias (2000) and Pards and Bonafonte (2002) proposed a description of the six basic facial expressions that employs the MPEG-4 coding model, an object-based multimedia compression standard. The MPEG-4 measurement units are the Facial Definition Parameters (FDPs; Tekalp & Ostermann, 2000), a set of tokens that describe minimal perceptible actions in the facial area. The distances between the 6 FDP points allow the modeling of a set of Facial Animation Parameters (FAPs) to describe each one of the six basic facial expressions (Tekalp & Ostermann, 2000). Tsapatsoulis et al. used fuzzy inference for the classification, whereas Pards and Bonafonte used Hidden Markov Models, which offered better results.

The fiducial-based representation requires accurate and reliable facial feature detection and tracking to cope with variations in illumination and the non-rigid motion of facial features. Based on these considerations, the chosen classification approach should allow the modeling of the noise in the segmentation results. Contrary to the classifiers described above, the model proposed here overcomes this problem by using the Transferable Belief Model (TBM) as a classifier, which takes into account the noise and the imprecision in the fiducial points segmentation (see Hammal et al., 2007).

This short review of the state of the art in the computer vision and pattern recognition community shows that great efforts have been made to automatically recognize facial expressions. However, the human visual system remains far ahead of the pack. What is it that makes humans so efficient at classifying facial expressions?

Cues used efficiently by the humans to recognize basic facial expressions

Researchers in psychology have studied the parts of the face that human observers find most useful for the recognition of facial expressions. Boucher and Ekman (1975) claimed that the bottom half of the face is mainly used by human observers for the recognition of Happiness and Disgust and that the whole face is used for the recognition of Anger and Surprise. Bassili (1978, 1979) noted that the whole face leads to a better recognition of the basic facial expressions (74.4%) than the bottom part of the face (64.9%), which leads to a better recognition than the top part of the face (55.1%). Gouta and Miyamoto (2000) concluded that the top half of the face allows a better recognition of Anger, Fear, Surprise, and Sadness, whereas the bottom half is better for Disgust and Happiness.

These experiments and similar others show that different portions of faces vary in their importance for the recognition of facial expressions. One major limitation of this research, however, is the coarseness of the results. Another is the biasness of the search, which is usually limited to specific facial Actions Units thought to be involved in the recognition of the facial expression studied. More recently, Smith et al. (2005) made a finer and less biased analysis of the relative importance of facial features in the discrimination of the basic facial expressions. Their experiment used *Bubbles,* a psychophysical procedure that prunes stimuli in the complex spaces characteristic of visual categorization, in order to reveal the information that effectively determines a given behavioral response in a recognition task (Gosselin & Schyns, 2001).

The Bubbles experiment of Smith et al. (2005)

In the paper by Smith et al. (2005), the *Bubbles* technique was applied to determine the information underlying the recognition of the six basic facial expressions plus Neutral. The stimuli were produced by randomly sampling 70 facial expression images from the California Facial Expressions database[1] at five scales using scale-adjusted Gaussian filters (see Figure 1 and Smith et al., 2005 for details).

Fourteen participants were each submitted to 8,400 sparse stimuli and were instructed to identify which of the seven studied facial expressions was displayed. A model observer was built to benchmark the information available for performing the task. On each trial, the model determined the Pearson correlation between the sparse input and each of the 70 possible original images revealed with the same Gaussian apertures. Its categorization response was the category of the original image with the highest correlation to the sparse input. The experiment revealed the precise location and scale information correlated with accurate categorization of the six basic expressions plus Neutral in the human and model observers (see Figure 2).

The proposed contribution

As already stated, humans remain the most robust facial expression recognizers in ecological conditions and their performance is far better than that of any proposed model.

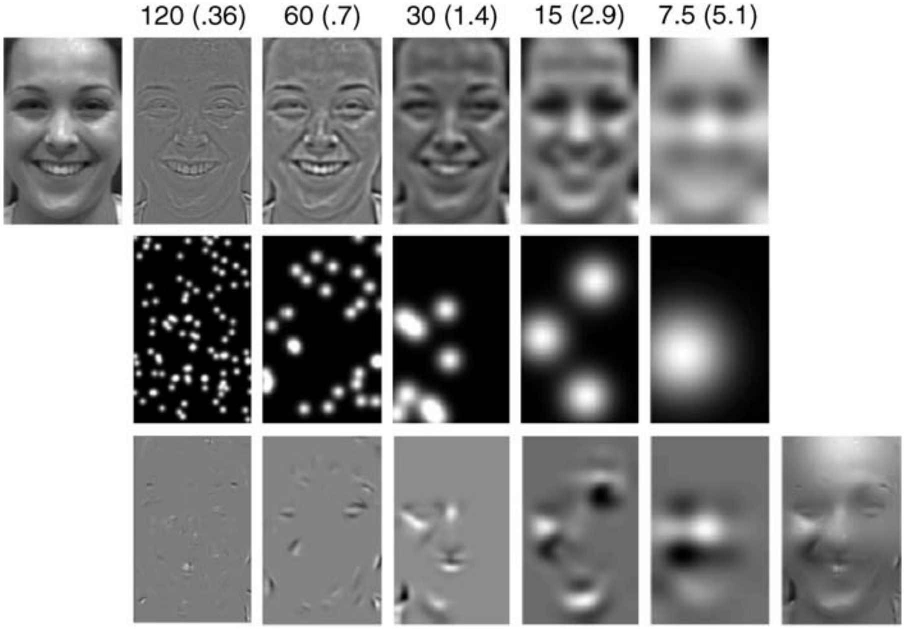


Figure 1. Illustration of the stimulus-generation process used in Smith et al. (2005). First, as shown in the top row, each original face was decomposed into five spatial frequency bandwidths of one octave each. Each bandwidth was then independently sampled with randomly positioned Gaussian apertures (second row). The third row shows the resulting sampling of facial information. The sum of information samples across scales produced an experimental stimulus, e.g., the rightmost picture in the third row.

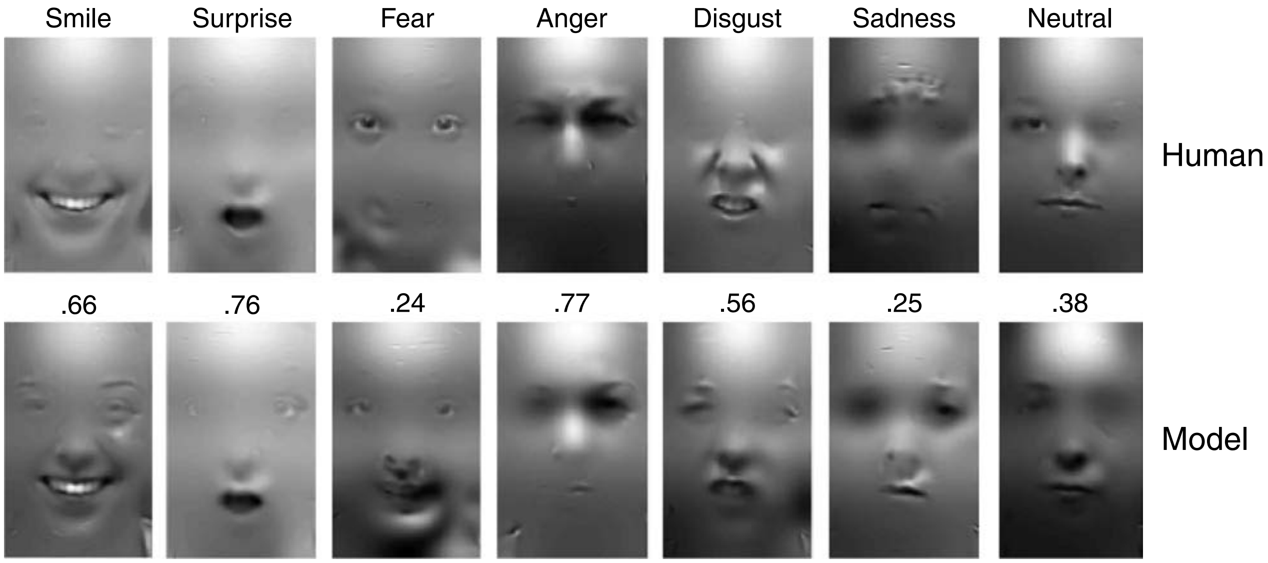


Figure 2. Location and scale information used by the human and model observers for the categorization of the six basic expressions plus Neutral (from Smith et al., 2005).

The ultimate goal of this work is to set up an automatic model for facial expressions classification inspired by humans. Here, more modestly, we compared the behavior of a novel implementation of the TBM-based model proposed by Hammal et al. (2007) to the behavior of humans in the same experimental conditions to determine where they differ and, hence, where we should focus our attention in future implementations of the model. For this comparison, we chose the most complete data set on human facial expression recognition presently available, i.e., the Smith et al. (2005) *Bubbles* experiment described above. We had to significantly modify the model proposed by Hammal et al. (2007) for the classification of stimuli displaying the six basic facial expressions plus *Neutral* to give it the ability of dealing with sparse stimuli like the ones encountered in a *Bubbles* experiment as well as in real life (Zeng, Pantic, Roisman, & Huang, 2009).

The remainder of the paper is organized as follows: First, we present a brief description of the Transferable Belief Model (TBM)-based model proposed by Hammal et al. Then we adapt the Hammal et al. model to the Smith et al. *Bubbles* experiment. Finally, we compare the behaviors of the model and humans in three simulations and draw conclusions regarding future implementations of the model.

The Transferable Belief Model for basic facial expression recognition of Hammal et al. (2007)

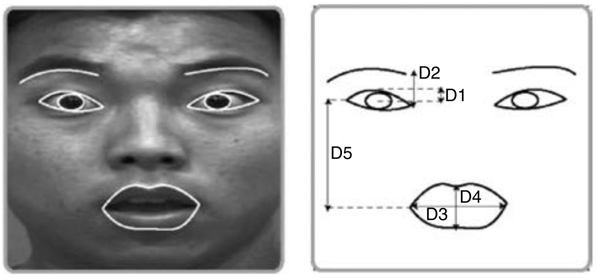
To deal with the sparse stimuli used in the Smith et al. experiment, we adapted the TBM-based model proposed by Hammal et al. (2007). In this section, we sketch the basic facial expression recognition model described by Hammal et al. (2007) before turning to its adaptation. The model mainly consists of three processing stages: data extraction, data analysis, and classification. In the data extraction stage, frontal face images are presented to the system and the contours of the eyes, eyebrows, and mouth are extracted, which leads to a skeleton of the facial features. From this skeleton, several distances characterizing the facial feature deformations are computed.

In the data analysis stage, the numerical values of the characteristic distances are mapped to symbolic states that qualitatively encode how much a given distance differs from its corresponding value in the Neutral state. Then each facial expression is characterized by a combination of characteristic distance states.

In the classification stage, the Transferable Belief Model (TBM) is applied to recognize the six basic facial expressions plus the Neutral expression. Further details regarding these three stages are provided in the Data extraction section, Data analysis section, and Classification process section.

Data extraction

The first step in the Hammal et al. (2007) facial expression model is the automatic extraction of the contours of the permanent facial features (eyes, eyebrows, and mouth—see Hammal, Eveno, Caplier, & Coulon, 2006). However, the automatic segmentation of the mouth requires chromatic information and is not possible on the sole basis of the achromatic information of the face set used by Smith et al. (Figure 3). Thus, for the simulations



D1 Eye opening, distance between: the center of the iris and upper eyelids or upper and lower eyelids

D2 Distance between the interior corner of the eye and the interior corner of the eyebrow

D3 Mouth opening width, distance between left and right mouth corners

D4 Mouth opening height, distance between upper and lower lips

D5 Distance between a corner of the mouth and the corresponding external eye corner

Figure 3. (Top) Facial features segmentation. (Bottom) Characteristic distances description.

reported in this paper, the segmentation of the facial feature contours was performed manually in the original frame (i.e., before the application of the Bubbles mask). The characteristic points corresponding to the contours of each permanent facial feature were manually detected and then the corresponding curves were automatically traced. The permanent facial feature deformations occurring during facial expressions were then measured by five characteristic distances D_1 to D_5 (Figure 3) extracted from the characteristic points corresponding to the contours of the permanent facial features. Each distance is normalized with respect to the distance between the centers of both irises in the analyzed face. This makes the analysis independent of the variability of face dimensions and of the position of the face with respect to the camera. In addition to distance normalization, only the deformations with respect to the Neutral expression are considered.

Data analysis

The aim of the data analysis stage is to characterize each facial expression using the characteristic distances that have been measured in the previous stage. A two-step procedure is then proposed: first, a symbolic description, named state, is associated to each distance. Second, rules are defined, which establish how the symbolic states relate to particular facial expressions.

For the purpose of the first step, the numerical values of the characteristic distances are mapped to symbolic states that encode how much a given distance differs from its corresponding value in the Neutral state. A numerical to symbolic conversion is carried out using a fuzzy-like model for each characteristic distance D_i (Hammal et al., 2007 for more details). It allows the conversion of each numerical value to a belief in five symbolic states reflecting the magnitude of the deformation. S_i if the current distance is roughly equal to its corresponding value in the Neutral expression, C_i^+ vs. C_i^- if the current distance is significantly higher vs. lower than its corresponding value in the Neutral expression, and S_i or C_i^+ noted $S_i \cup C_i^+$ vs. S_i or C_i^- noted $S_i \cup C_i^-$ (the sign $\cup$ means logical or) if the current distance is neither sufficiently higher vs. lower to be in C_i^+ vs. C_i^-, nor sufficiently stable to be in S_i.

Figure 4 shows an example of this mapping for the characteristic distance D_2 (distance between eye corner and eyebrow corner) for several video sequences going from Neutral to Surprise expression and coming back to Neutral, which have been obtained from different individuals. We observe similar evolutions of the characteristic distance associated with the same facial expression. The characteristic distance D_2 always increases in the case of Surprise because people raise their eyebrows. Thus, D_2 evolves from the equal state (S_2) to the significantly higher state (C_2^+) via an undetermined region ($S_2 \cup C_2^+$) corresponding to a doubt between the two considered states.

The conversion from the numerical D_i values to symbolic states is carried out using the function depicted in Figure 5. The threshold values defining the transition from one state to another $\{a_i, b_i, c_i, d_i, e_i, f_i, g_i, h_i\}$ have been derived through a statistical analysis of the Hammal–Caplier database (2003)[2] for each characteristic distance.

For each distance D_i, the minimum threshold a_i is averaged across the minimum values of the characteristic distance D_i for all the facial expressions and all the subjects. Similarly, the maximum threshold h_i is obtained from the averaging of the maximum values of the characteristic distance D_i for all the facial expressions and all the subjects. The middle thresholds d_i and e_i are defined as the mean of minimum and maximum, respectively, of the characteristic distances D_i on Neutral facial images for all the subjects (Hammal et al., 2007).

The intermediate threshold b_i is computed as the threshold a_i of the distance D_i assigned to the lower state C_i^- augmented by the median of the minimum values of the distance D_i over all the image sequences and for all the subjects. Likewise, the intermediate threshold g_i is computed as the threshold h_i of the distance D_i assigned to the higher state C_i^+ reduced by the median of the maximum values over all the image sequences and for all the subjects. The thresholds f_i and c_i are obtained similarly (Hammal et al., 2007).

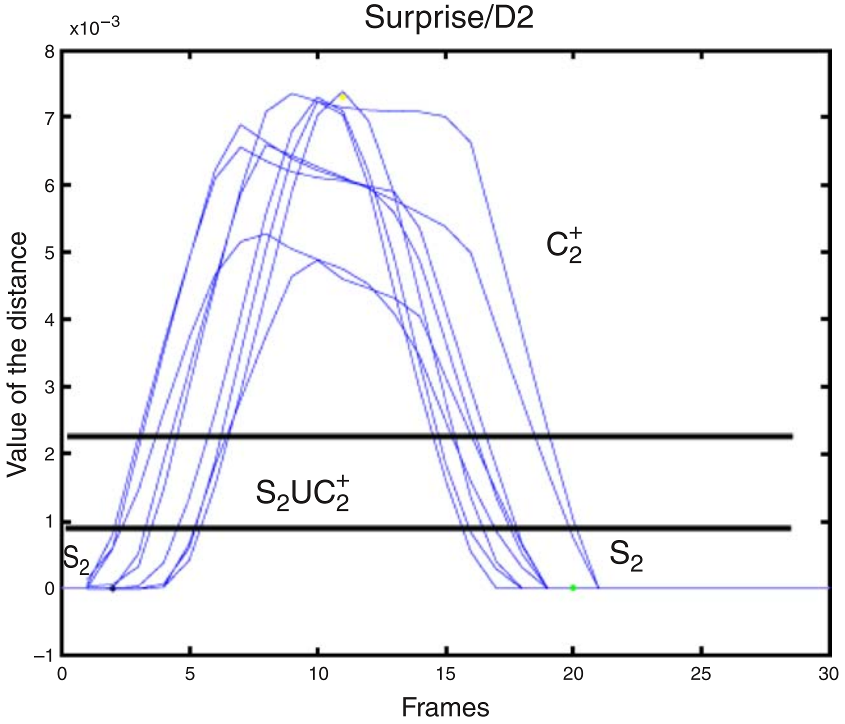


Figure 4. Time course of the characteristic distance D_2 (distance between the interior corner of the eye and the interior corner of the eyebrow) and its corresponding state values for the *Surprise* expression for 9 subjects with no formal acting training from the Hammal–Caplier database (one curve per subject).

After the numerical to symbolic conversion, each facial expression is characterized by a combination of the resulting characteristic distance states according to the rules displayed in Table 1. This mapping of facial expressions to characteristic distance states has been obtained by heuristic analysis and has been validated by MPEG-4 (Malciu & Preteux, 2001), a widely used description of the deformations undergone by the facial features for the six basic facial expressions plus Neutral (Hammal et al., 2007). For instance, a Surprise expression is characterized by the fact that the eyebrows are raised (D_2 is in C^+ state), the upper eyelids are open (D_1 is in C^+ state), and the mouth is open (D_3 is in C^- state and D_4 is in C^+ state).

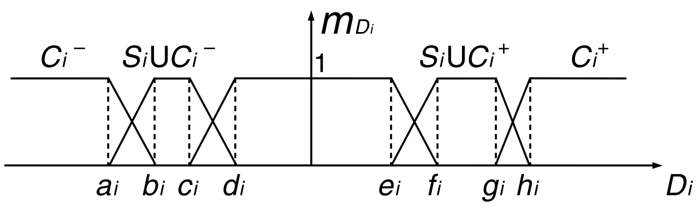


Figure 5. Model of basic belief assignment based on characteristic distance D_i. $S_i \cup C_i^+$ (vs. $S_i \cup C_i^-$) means S_i or C_i^+ (vs. S_i or C_i^-) and corresponds to the doubt between S_i and C_i^+ (vs. S_i and C_i^-). For each value of D_i, the sum of the pieces of evidence of the states of D_i is equal to 1.

It must be underlined, however, that humans are not all or none, be it for the production or the recognition of facial expressions of emotion. Facial expressions may include a blend of expressions, which makes human observers often hesitating between several expressions. Moreover, automatic facial features segmentation can lead to measurement errors on the characteristic distance states. This means that a facial expression analyzer should be capable of dealing with noisy data. Such a system should model the doubt on the characteristic distance states and generate conclusions such that the associated certainty varies with the certainty of facial points localization and tracking. For these reasons, an all-or-none system based only on the logical rules of Table 1 is not sufficient to reliably recognize facial expressions. These issues can be

	D_1	D_2	D_3	D_4	D_5
Happiness (E_1)	C^-	$S \cup C^-$	C^+	C^+	C^-
Surprise (E_2)	C^+	C^+	C^-	C^+	C^+
Disgust (E_3)	C^-	C^-	$S \cup C^+$	C^+	S
Anger (E_4)	C^-	C^-	S	$S \cup C^-$	S
Sadness (E_5)	C^-	C^+	S	C^+	S
Fear (E_6)	C^+	$S \cup C^+$	$S \cup C^-$	$S \cup C^+$	$S \cup C^+$
Neutral (E_7)	S	S	S	S	S

Table 1. *Di* states corresponding to each expression.

directly tackled by the use of the Transferable Belief Model (TBM, Smets, 2000).

Based on the facial features segmentation, the TBM is first used to model the characteristic distance states. In order to determine the current expression, a fusion process of the states of the five characteristic distances is then performed based on the TBM combination process (see Fusion process section). The TBM has already demonstrated its suitability for the classification of facial expressions (Hammal, Caplier, & Rombaut, 2005; Hammal et al., 2007). Indeed, it has been validated against two well-benchmarked databases, the Cohn–Kanade database (2000) and CAFE database (Dailey, Cottrell & Reilly, 2001), as well as that of Hammal-Caplier (2003).

The TBM is well adapted for the design of a fusion approach where various independent sensors or sources of information collaborate together to provide a more reliable decision. The facial expression classification in our model is based on the TBM fusion process of all the information resulting from the states of the characteristic distance states.

Classification process

The Transferable Belief Model (TBM) is a model of representation of partial knowledge (Smets, 1998; Smets & Kruse, 1994). It can be understood as a generalization of probability theory. It can deal with imprecise and uncertain information and provides a number of tools for the integration of this information (Denoeux, 2008; Smets, 2000). It considers the definition of the frame of discernment $\Omega' = \{H_1, \ldots, H_N\}$ of N exclusive and exhaustive hypotheses characterizing some situations. This means that the solution to the problem is unique and is one of the hypotheses of Ω'.

Using the TBM approach requires the definition of the Basic Belief Assignment (BBA) associated with each independent source of information. The BBA assigns an elementary piece of evidence $m^{\Omega}(A)$ to every proposition A of the power set $2^{\Omega'} = \{\{H_1\}, \{H_2\}, \ldots, \{H_N\}, \{H_1, H_2\}, \ldots, \Omega\}$. In the current application the independent sensors correspond to the different characteristic distances and the hypotheses H_i correspond to the six basic facial expressions plus *Neutral*. The first step in the classification process then is to perform an intermediate modeling between the numerical values of the characteristic distances D_i and the required expressions. More precisely, the Basic Belief Assignment related to the characteristic distance states is defined (see Equation 1 below). Then using the rules (see Table 1) between the symbolic states and the facial expressions, the BBAs of the facial expressions according to each characteristic distance are deduced. Finally, the combination process of the BBAs of all the distance states (and then the corresponding expressions) leads to the definition of the BBAs of the facial expressions using all the available information (see Fusion process section).

The BBA $m_{Di}^{\Omega Di}$ of each characteristic distance state D_i is defined as

$$\begin{aligned} & m_{Di}^{\Omega_{D_i}} : 2^{\Omega_{D_i}} \rightarrow [0, 1] \\ & A^{\Omega_{D_i}} \rightarrow m_{Di}^{\Omega_{Di}}(A), \sum_{A \in 2^{\Omega_{Di}}} m_{Di}^{\Omega_{Di}} = 1, \end{aligned} \tag{1}$$

where $\Omega_{D_i} = \{C_i^+, C_i^-, S_i\}$, the power set $2^{\Omega_{Di}} = \{\{C_i^+\}, \{C_i^-\}, \{S_i\}, \{S_i, C_i^+\}, \{S_i, C_i^-\}, \{S_i, C_i^+, C_i^-\}\}$ is the frame of discernment (the set of possible propositions and subset of propositions), $\{S_i, C_i^+\}$ vs. $\{S_i, C_i^-\}$ is the doubt (or hesitation) state between the state C_i^+ vs. C_i^- and the state S_i. The piece of evidence $m_{Di}^{\Omega_{Di}}$ associated with each symbolic state given that the value of the characteristic distance D_i is obtained by the function depicted in Figure 5. $m_{Di}^{\Omega_{Di}}(A)$ is the belief in the proposition $A \in 2^{\Omega_{Di}}$ without favoring any of propositions of A in case of doubt proposition. This is the main difference when compared with the Bayesian model, which implies equiprobability of the propositions of A. A is called the focal element of $m_{Di}^{\Omega_{Di}}(A)$ whenever the belief on A $m_{Di}^{\Omega_{Di}}(A) > 0$. Total ignorance is represented by $m_{Di}^{\Omega_{Di}}(\Omega_{D_i}) = 1$. To simplify, the proposition $\{C_i^+\}$ is noted C^+ and the subset of propositions $\{S_i, C_i^+\}$ is noted $S \cup C^+$ (i.e., S or C^+ that corresponds to the doubt state between S and C^+).

The Transferable Belief Model for partially revealed basic facial expressions

Bubbles modeling process

The originality of the current work consists in making a fusion architecture based on the TBM that is capable of modeling partially occluded facial expressions as encountered in the Smith et al. *Bubbles* experiment and, more generally, in real life. Thus, instead of using all the characteristic distances (Hammal et al., 2007), only those revealed by the Gaussian apertures are used. The TBM is well suited for this: It facilitates the integration of a priori knowledge and it can deal with uncertain and imprecise data, which is the case with *Bubbles* stimuli. Moreover, it is able to model the doubt between several facial expressions in the recognition process. This property is important considering that "binary" or "pure" facial expressions are rarely perceived (people usually display mixtures of facial expressions (Young et al., 1997). Also, the proposed method allows Unknown expressions, which correspond to all facial deformations that cannot be categorized into one of the predefined facial expressions.

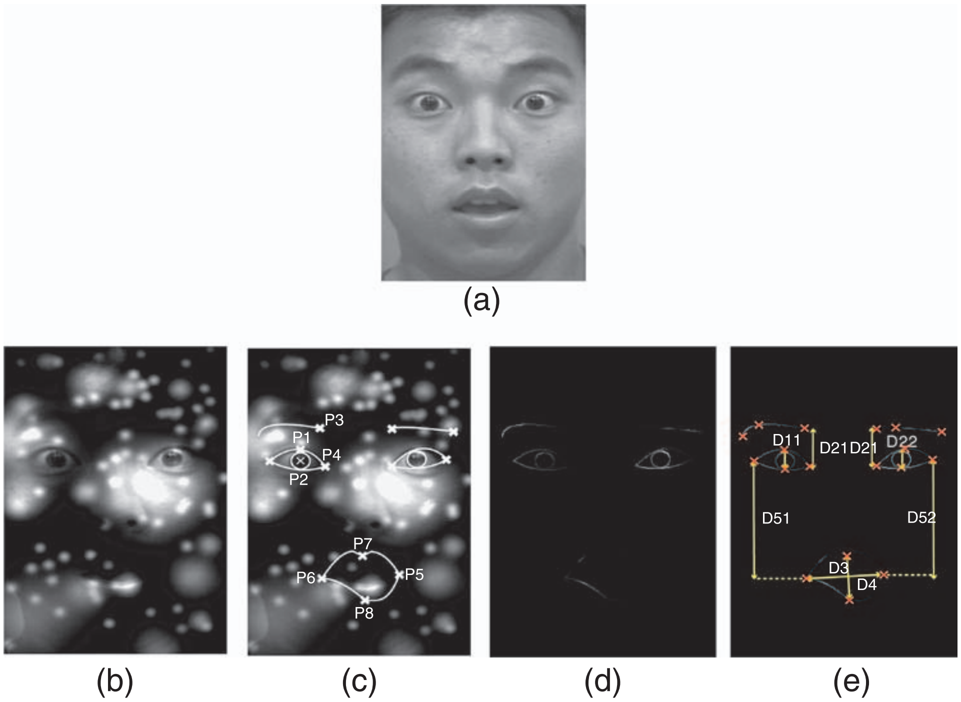


Figure 6. (a) Original frame. (b) Result of the intersection (multiplication) between the original frame and the *Bubbles* mask—some facial features are not visible even for human observers. (c) Superposition of the segmentation results of the facial feature contours made manually on the original frame (i.e., before the application of the Bubbles mask as shown in (a)) on the frame (b). (d) Revealed facial feature contours after the intersection process (multiplication), the appearance intensity of the contours varies according to the size, the position, and the number of the Gaussian apertures. (e) Based on the contours in (d), the characteristic points (used in computing the 5 characteristic distances) for which the pixel intensities are different from 0 are identified (red crosses) and the distances computed from these points. D_{i1} and D_{i2} correspond, respectively, to the left and right sides of the characteristic distance D_i, except for D_3 and D_4, which are associated with the mouth.

Distance measurements

Among the five characteristic distances only those retrievable from the facial features revealed by the Gaussian apertures on each trial are identified. Figure 6 summarizes these different processing steps. First, Gaussian apertures are applied to the face image (see Figure 6b). It can be seen that some facial parts are not revealed. The permanent facial features need then to be segmented from the sparse facial expressions. However, due to the difficulty of applying the permanent facial features segmentation after the application of the *Bubbles* mask (i.e., the collection of Gaussian apertures that sample a face on a particular trial), the intersection between the *Bubbles* mask and the contours of the facial features is performed in two steps. First, the segmentation of the permanent facial features is made manually on the original frame (i.e., before the application of the Bubbles mask, see Data extraction section). The characteristic points corresponding to each contour are manually detected. Figure 6c shows an example of the corresponding contours. However, it should be noted that even human experts do not obtain perfect segmentation results and a weak dispersion of the detected points appears, which leads to (sometimes large) imperfections in the corresponding contours. Most importantly, however, the characteristic distances are measured based only on the characteristic points and not on the corresponding contours. Thus, the small dispersion errors of the characteristic points do not significantly affect the classification process. This claim is based on the results of a quantitative evaluation using a ground truth corresponding to the results of the manual detection of the characteristic points by human experts (see Hammal et al., 2006).

Second, the intersection between the contours and the results of the application of the *Bubbles* mask is done revealing a subset of the contours of the permanent facial features and then of the corresponding characteristic points (see Figures 6c and 6d). The appearance intensity of the contours and of the characteristic points varies according to the size, the position, and the number of the Gaussian apertures (see The Bubbles experiment of Smith et al. (2005) section and Figure 6d). However, as reported below only the characteristic points are used for the computation of the characteristic distances. The characteristic points for which the pixel intensities are different from 0 are identified (red crosses in Figure 6e). Finally, all distances computed from contour points different from 0 are identified and taken into account in the classification process (see Figure 6e).

There exists some degree of asymmetry in human facial expressions. Indeed, work has recently been carried out on the analysis of asymmetrical facial feature deformations, especially in the case of spontaneous facial expressions (Cohn & Schmidt, 2004; Schmidt, Ambadar, Cohn, & Reed, 2006). However, in the present work, we assume that facial expressions and then the corresponding facial features behavior are symmetrical and then each characteristic distance value D_i is considered as the mean between its corresponding left and right side values as

$$\begin{cases} (D_{i1}+D_{i2})/2 \text{ if } D_{i1} \text{ and } D_{i2} \text{ are revealed} \\ D_{i1} \text{ if } D_{i2} \text{ is not revealed} \qquad 1 \leq i \leq 5, i \neq 3, i \neq 4\,, \\ D_{i2} \text{ if } D_{i1} \text{ is not revealed} \end{cases} \tag{2}$$

where D_{i1} and D_{i2} correspond, respectively, to the left and right sides of the characteristic distance D_i, except for D_3 and D_4, which concern the mouth (see Figure 6e).

Once the intersection step is completed the Basic Belief Assignments (BBAs) of the identified characteristic distances are computed (see Data analysis and Classification process sections). In order to model the stimuli used by human observers, the appearance intensities of the corresponding characteristic distances are taken into account in the fusion process. How this was done is explained in the following section.

Discounting

In the TBM, discounting (Smets, 2000) is used to take into account the reliability of the sensors by altering their corresponding Basic Belief Assignments. If we know that a sensor is reliable, the belief function it provides is accepted without any modification. If we know that it is entirely unreliable, the information coming from the source is considered as irrelevant. In the current work discounting was used to weaken or inhibit the characteristic distances used for the classification process by weighting their corresponding BBAs as a function of the degree to which they are sampled by the Gaussian apertures (see Figure 6d).

More specifically, the discounting operation is controlled by a discount rate α, $0 \leq \alpha \leq 1$, which allows computing the new piece of evidence noted ${}^{\alpha}m$ (see Equation 3 and Smets, 2000) for each proposition A according to its current piece of evidence m and the discounting rate α as

$$\begin{aligned} {}^{\alpha}m(A) &= \alpha * m(A) \\ {}^{\alpha}m(A \cup \bar{A}) &= 1 - \alpha * (1 - m(A \cup \bar{A})), \end{aligned} \tag{3}$$

where $\bar{A}$ corresponds to the complement of the proposition A. If the distance is fully reliable $\alpha = 1$, then m is left unchanged (i.e., ${}^{\alpha}m(A) = m(A)$). If the distance is not reliable at all $\alpha = 0$, m is transformed into the vacuous BBA (i.e., ${}^{\alpha}m(A) = 0$).

In the Smith et al. *Bubbles* experiment, the revealed facial parts used for the classification process appear with different intensities. This can be understood as differences in reliability for the corresponding distances. *Discounting* was used to weight the contribution of each characteristic distance D_i according to its intensity represented by *inten* (D_i). This leads to five *discounting* parameters α_i ($1 \leq i \leq 5$), one for each characteristic distance D_i.

The values α_i can be computed by learning (Elouadi, Mellouli, & Smets, 2004) or by optimizing a criterion (Mercier, Denoeux, & Masson, 2006) when the reliability of the sensors is uncertain or unknown. In the current work, the reliability of the sensors (the characteristic distances) is known and is equal to their appearance intensity after the application of the Bubbles mask. Then the corresponding reliability parameters α_i are equal to *inten*(D_i).

Each characteristic distance D_i is computed by measuring the distance between two points relative to their distance in the neutral state. As reported above, each distance is considered only if the intensities of its two associated points are both different from 0. Then its intensity is taken as the mean of the intensities of its associated points. For example α_1 the discounting parameter of D_1 was computed as

$$\alpha_1 = \text{inten}(D_1) = (\text{inten}(P1) + \text{inten}(P2))/2, \tag{4}$$

where inten($P1$) and inten($P2$) correspond, respectively, to the intensities of pixels P1 and P2, which are different from 0 (see Figure 6c).

Other choices could have been made for the computation of the discounting parameters. For example, we could have opted for multiplying the intensities of the associated points. No evidence is currently available to indicate which of several possible options is the best. However, given the large number of trials used with changing size, position, and number of the Gaussian apertures, a large range of values for each α_i is tested allowing us to reach our goal of analyzing the response of the system to the inhibition or the discounting of the required information.

To evaluate the influence of the characteristic distance intensities, we also considered the special case where the discounting parameters of all used distances ($D_i \neq 0$) were fully reliable (i.e., $\alpha_i = 1$ for $1 \leq i \leq 5$; see the second simulation in Simulations section).

Once the discounting parameters α_i of all the characteristic distances used were set, the corresponding BBAs were redefined according to Equation 3.

Fusion process

The main feature of the TBM is the powerful combination operator that integrates information from different sensors. Based on the rules listed in Table 1 and in order to take into account all the available information, the facial expression classification is based on the TBM fusion process of all the D_i states.

The BBAs $m_{D_i}^{\Omega_{D_i}}$ of the states of the characteristic distances are defined on different frames of discernment (see Classification process section). For the fusion process, it is necessary to redefine the BBAs on the same frame of discernment 2^Ω, where $\Omega = \{Happiness\ (E_1), Surprise\ (E_2), Disgust\ (E_3), Fear\ (E_4), Anger\ (E_5), Sadness\ (E_6), Neutral\ (E_7)\}$ is the set of expressions.

From the rules table and the BBAs of the states of the characteristic distances $m_{D_i}^{\Omega_{D_i}}$ (see The transferable Belief Model for basic facial expressions recognition of Hammal et al. (2007) section), a set of BBAs on facial expressions $m_{D_i}^{\Omega}$ is derived for each characteristic distance D_i. In order to combine all this information, a fusion process of the BBAs $m_{D_i}^{\Omega}$ of all the states of the characteristic distances is performed using the conjunctive rule of combination noted $\oplus$ (see Equation 5 (Denoeux, 2008; Smets, 2000); Equation 6 shows the mathematical definition of the corresponding symbol) and results in m^Ω the BBA of the corresponding expressions

$$m^\Omega = \oplus m_{D_i}^{\Omega}. \tag{5}$$

For example, if we consider two characteristic distances D_i and D_j with two BBAs $m_{D_i}^{\Omega}$ and m_{Dj}^{Ω} derived on the same frame of discernment, the joint BBA $m_{D_{ij}}$ is given using the conjunctive combination (orthogonal sum) as

$$m_{D_{ij}}^{\Omega}(A) = (m_{D_i}^{\Omega} \oplus m_{D_j}^{\Omega})(A) = \sum_{B\cap C=A} m_{D_i}^{\Omega}(B)*m_{D_j}^{\Omega}(C), \tag{6}$$

where A, B, and C denote propositions, the sign $\cap$ means logical AND, and $B \cap C$ denotes the conjunction (intersection) between the propositions B and C. This leads to propositions with a lower number of elements and with more accurate pieces of evidence.

Decision process

The decision is the ultimate step in the classification process. It consists in making a choice between various hypotheses E_e and their possible combinations. Making a decision is associated with a risk except if the result is sure ($m(E_e) = 1$). As it is not always the case (more than one expression can be recognized at once), several decision criteria can be used (Denoeux, 2008; Smets, 2000).

To compare directly our classification results with those obtained by the human participants of Smith et al., the decision was made using the maximum of pignistic probability $BetP$ (see Equation 7 and Smets, 2005), which only deals with singleton expressions:

$$BetP : \Omega \to [0,1]$$

$$C \to BetP(C) = \sum_{A\subseteq\Omega, C\in A} \frac{m^\Omega(A)}{(1-m^\Omega(\phi))*Card(A)}, \forall C\in\Omega, \tag{7}$$

where $BetP(C)$ corresponds to the pignistic probability of each one of the hypothesis C of A, ϕ corresponds to the conflict between the sensors, and $Card(A)$ corresponds to the number of elements (hypothesis) of A.

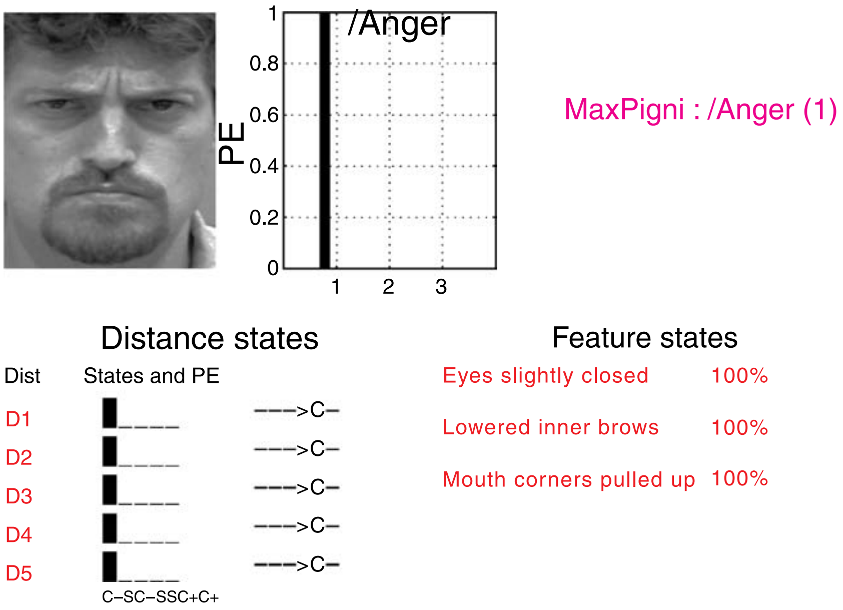
Simulations

The simulations were performed on all basic facial expression stimuli employed by Smith et al. That is, 100,800 stimuli (7,200 stimuli per subject * 14 subjects). Three simulations were carried out: first, using all the characteristic distances (see The transferable Belief Model for basic facial expressions recognition of Hammal et al. (2007)); second, using discounting by all-or-none inhibition of the characteristic distances (see Discounting section); and, third, using discounting by graded inhibition of the characteristic distances (see Discounting section).

User interface

The complete interface defined by Hammal (2006a, 2006b for a more detailed description) is used and described below. It summarizes all the information extracted and analyzed for each frame.

The proposed approach deals with the facial expression classification as well as the description of the characteristic distance states used and their corresponding facial feature deformations. To do this, each characteristic distance state is translated into deformations of the corresponding facial features (with its corresponding piece of evidence). In the case of the current application, on each trial, the characteristic distances apparent after the application of the *Bubbles* mask are identified. Figure 7 presents an example of the information displayed during the analysis of Anger expression. In this example all the characteristic distances are identified and used, but it is not always the case that all the characteristic distances are identified as explained above. The interface is divided into five different regions: in the upper left region, the frame to be analyzed; in the upper middle region, the result of the BBAs of the expressions (in this case only the Anger expression appears with a piece of evidence equal to 1); in the upper right region, the decision result based on the



MaxPigni : /Anger (1)

Distance states

Dist	States and PE	
D1		--->C−
D2		--->C−
D3		--->C−
D4		--->C−
D5		--->C−

Feature states

Eyes slightly closed	100%
Lowered inner brows	100%
Mouth corners pulled up	100%

Figure 7. User interface displaying: current frame; the BBAs of the expressions (in this case only anger is identified with a piece of evidence equal to 1); the decision classification (maximum pignistic probability: MaxPigni) in this case is Anger; distance states estimation and the corresponding facial feature deformations (according to their Neutral state) with their corresponding pieces of evidence.

pignistic probability with its value; in the lower left region, the states of the characteristic distances and their pieces of evidence; in the lower right region, the corresponding facial feature deformations.

Results and discussion with all characteristic distances

For this simulation all characteristic distances were used. Three kinds of results are reported in Figure 8: first, the singleton results that are associated with a high level of confidence and during which only one expression was recognized; second, the double results during which two expressions were recognized at the same time (this occurs when the model hesitates between two expressions); third, the Total results, which is the sum of the singleton results and of the corresponding double results.

The classification results show that two double results occur frequently. The doubt between the Happiness and Disgust facial expressions appears to be due to the doubt states in Table 1. Based only on the five characteristic distances used, the system is sure that the current expression is one of the two and not any of the others. More information is required, however, to dissociate between them. Another pair of expressions the system has difficulty in discriminating is Surprise and Fear. Interestingly, these expressions are also notoriously difficult to discriminate for human observers (e.g., Roy et al., 2007). Figure 9 shows an example for which the system remains in doubt rather than taking the risk of making a wrong decision. Based on the pignistic probability decision, the doubt yields the same probability (see Equation 7) for the two considered expressions.

Considering the doubt states as a correct classification decision, we obtain the final correct response rates by summing the singleton and the corresponding double results, which gives the Total results (green histograms in Figure 8). The best classification rates are obtained for Happiness (100%), Anger (100%), Surprise and Disgust (about 75% each). As found in other studies (Black & Yacoob, 1997; Rosenblum, Yacoob & Davis, 1996; Yacoob & Davis, 1996) poor performance is obtained for the Sadness expression (25%). In the present case, this may reflect the fact that the classification rule used for Sadness lacks important pieces of information.

The last histogram in each graph corresponds to the total ignorance of the system (see Figure 8, Ign). These cases correspond to the facial feature deformation, which do not correspond to any of the seven expressions considered (conflict) and are thus recognized as an Unknown expression (see Hammal et al., 2007). The piece of evidence of Unknown expression is equal to 1 in these cases. However the pignistic probability is based on normalized BBAs (see Decision process section) where the piece of evidence for the conflict (Unknown expression) is redistributed across the whole set of seven expressions.

The bar "exp + noise" corresponds to the cases where the system recognizes the considered expression, but it cannot dissociate it from other expressions. This did not occur in the present simulation, but it did in the following.

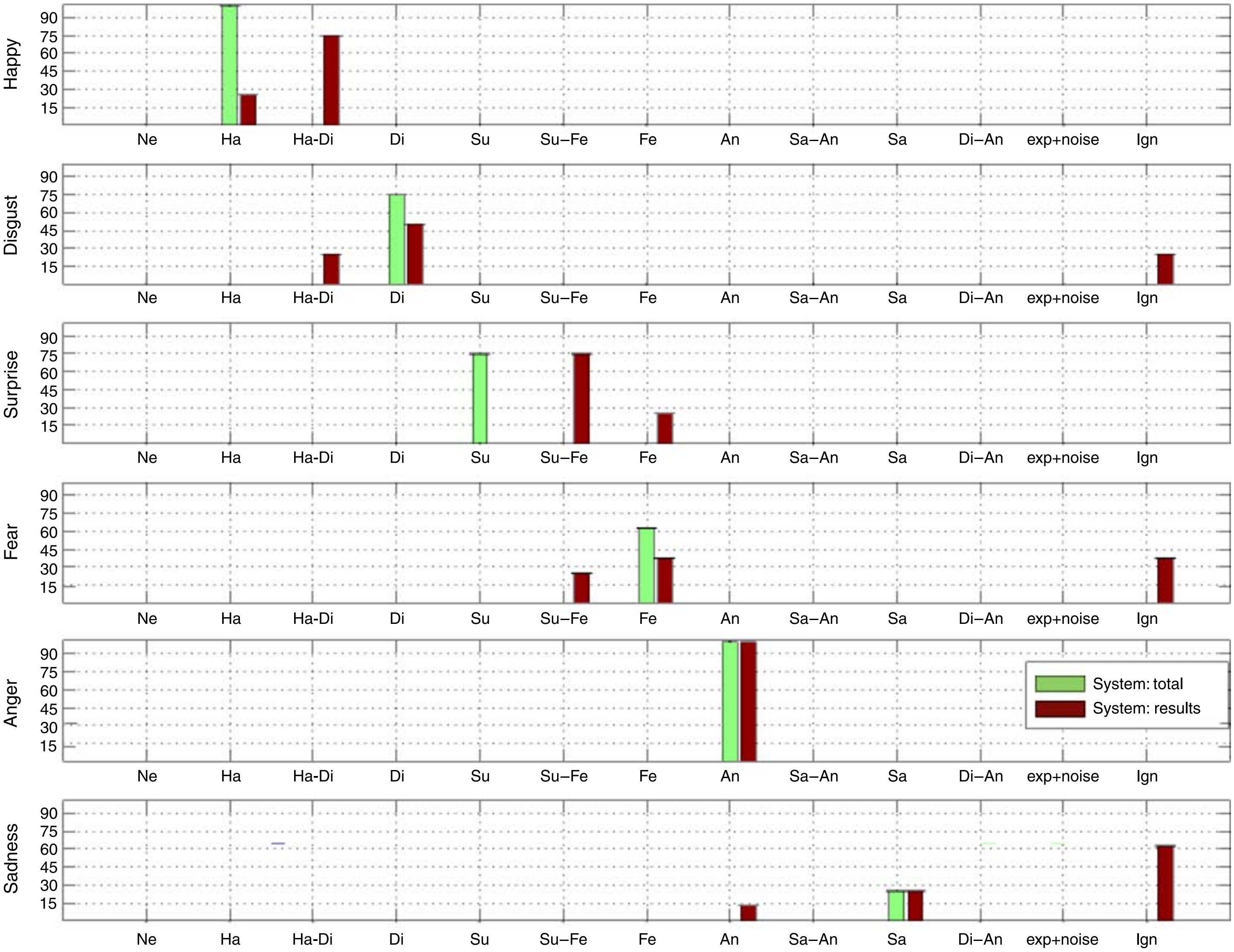


Figure 8. The means of the classification results with all characteristic distances. Ne: *Neutral,* Ha: *Happiness,* Di: *Disgust,* Su: *Surprise,* Fe: *Fear,* An: *Anger,* Ign: *Total ignorance*. Each graph corresponds to the presented facial expression. The horizontal axis corresponds to the system's responses. Three kinds of results are presented: the singleton results given by the system associated with a high level of confidence (red); the double results occurring when the system hesitates between two expressions (also red); the Total results, which is the sum of the singleton results and of the corresponding double results (green). This sum corresponds to the total correct responses of the model. "exp + noise" corresponds to the cases where the system recognizes the expression, but it cannot dissociate it from other expressions (triple or quadruple equiprobable expressions).

Results and discussion with all-or-none inhibition of the distances

For the simulations reported in this section, all the characteristic distances revealed by the Gaussian apertures were used (see Distance measurements section). More precisely, the distances that are completely hidden (the corresponding appearance intensity is equal to zero) by the mask are completely inhibited and those that are revealed (the corresponding appearance intensity is different from zero) are used completely. The results are illustrated in Figure 10. The classification performance of the human observers in the Smith et al. experiment is also reported.

The best performances are obtained for Anger (84%), Surprise (76%), and Fear (60%). This is consistent with human observers. Compared with human observers, the classification rates for Anger, Surprise, and Fear are not significantly different (two-way ANOVA, $P > 0.01$). The classification rate for Happiness (45%) is lower than that in the first simulation and lower than that for humans. Similar to the performance obtained using all characteristic distances, the worst classification rate was obtained with Sadness (28%).

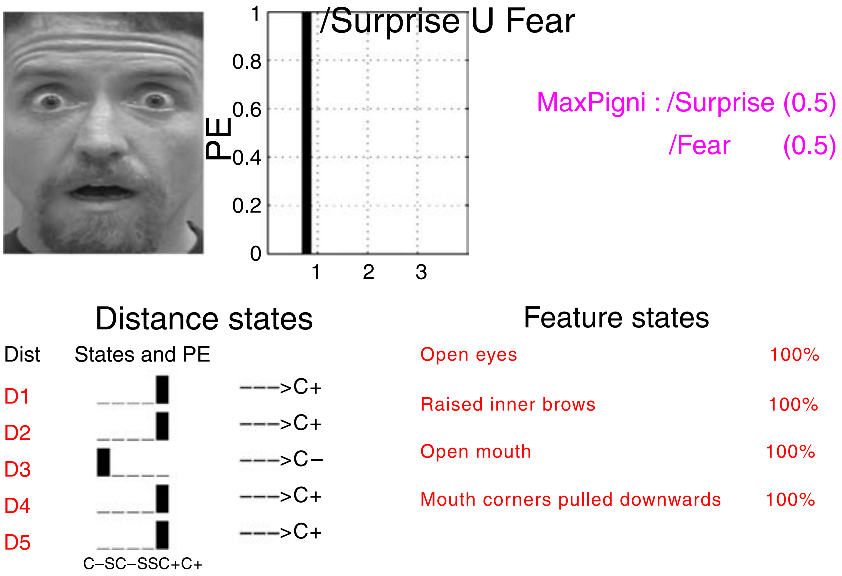


Figure 9. Example of classification of Surprise expression. Result interface displays Surprise classification where the system remains in doubt between Surprise and Fear rather than taking the chance of making a wrong decision; based on the pignistic probability decision, the doubt yields the same probability for the two expressions.

The inhibition process seems to have affected the model's behavior in three ways. First, the appearance of new double results (Sadness and Anger, Disgust and Anger) is due to the inhibition of the specific characteristic distances, which previously prevented their occurrence. Second, compared with the previous simulation for which all characteristic distances were fully taken into consideration, this new simulation resulted in similar but noisier classification. For example, Happiness performance decreases as doubt between Happiness and other expressions becomes apparent (see Figure 10, bar "exp + noise"). The system recognizes that the current expression is Happiness, but due to the inhibition of specific characteristic distances, it cannot dissociate it from other expressions as reliably as in the previous simulation. Figure 11 shows the system hesitating between Happiness, Disgust, Surprise, and Fear based on the state of distance D_4, the only one available for this particular combination of facial expression image and Gaussian apertures. Here, the characteristic distances required to distinguish between these expressions are inhibited. Third, the inhibition reduced the Ignorance rates for Disgust, Fear, and Sadness. These results mean that some characteristic distances are necessary for the recognition of some expressions while others increase the doubt and then their inhibition increases the recognition.

Results and discussion with graded inhibition of the distances

In this final simulation, the classification rates are based on the characteristic distances revealed by the Gaussian apertures and weighted by how much they were revealed by these apertures. Compared to simulation 2, the distances hidden by the Bubbles mask are still completely inhibited but the visible distances are weighted according to their level of visibility (i.e., appearance intensity).

The best classification rates are obtained for Anger (84%), Surprise (77%), and Fear (75%; see Figure 12). The classification rates for Sadness (61%) and Disgust (59%) are greater than in the second simulation, where the characteristic distances revealed by the Gaussian apertures were fully used. The results for Happiness in this new simulation are the same as in the previous. Thus, the correct classification rates for all expressions increased compared with the second simulation, except for Happiness. Conversely, Ignorance rates decreased for all expressions. For example the Ignorance rates for the Disgust, Fear, and Sadness expressions are now null while their classification rates have increased relatively to the previous simulation. Finally, compared with the performances of human observers (white bars in Figure 12), the model-based classification rates (Total results, green bars in Figure 12) of Anger are better and those of Surprise, Fear, and Sadness are not significantly different (two-way ANOVA, $P > 0.05$).

At this point, it is clear that the inhibition of some characteristic distances (simulation 2) leads to a decrease in classification performance and an increase in uncertainty relatively to when all the relevant information is available (simulation 1). In contrast, weighting the characteristic distances with the mean intensity of their end points (simulation 3) leads to a better fit between the model and human observers. At first glance, there seems to be a good fit between the full TBM model and human observers. In particular, their average performances are comparable. However, the Pearson correlations computed on a trial-by-trial basis tell a different story.

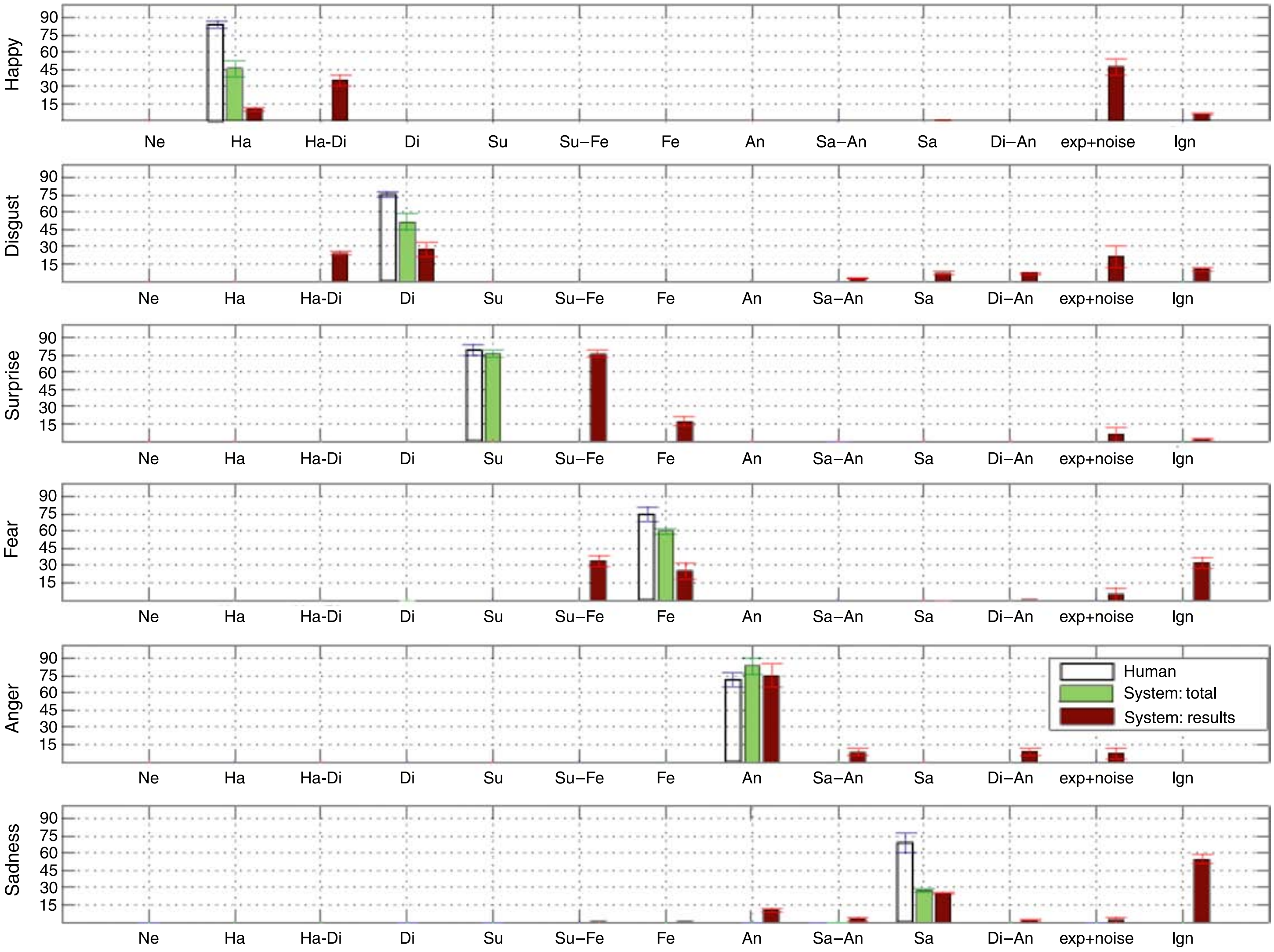


Figure 10. The means and standard deviations of the classification results with the inhibition process of the characteristic distances. Ne: *Neutral,* Ha: *Happiness,* Di: *Disgust,* Su: *Surprise,* Fe: *Fear,* An: *Anger,* errors: *Total ignorance*. Red and green bars correspond to the system results, with the same conventions as in Figure 8, and white bars show the human performances obtained in the Smith et al. study.

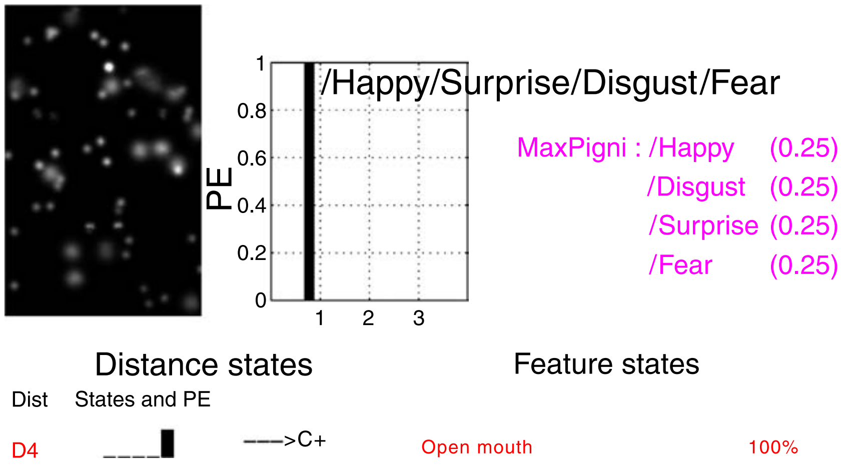


Figure 11. Classification results of the *Happiness* expression with the inhibition of the characteristic distances D_1, D_2, D_3, and D_5. Based only on the state of D_4 the system hesitates between Happiness, Disgust, Surprise, and Fear.

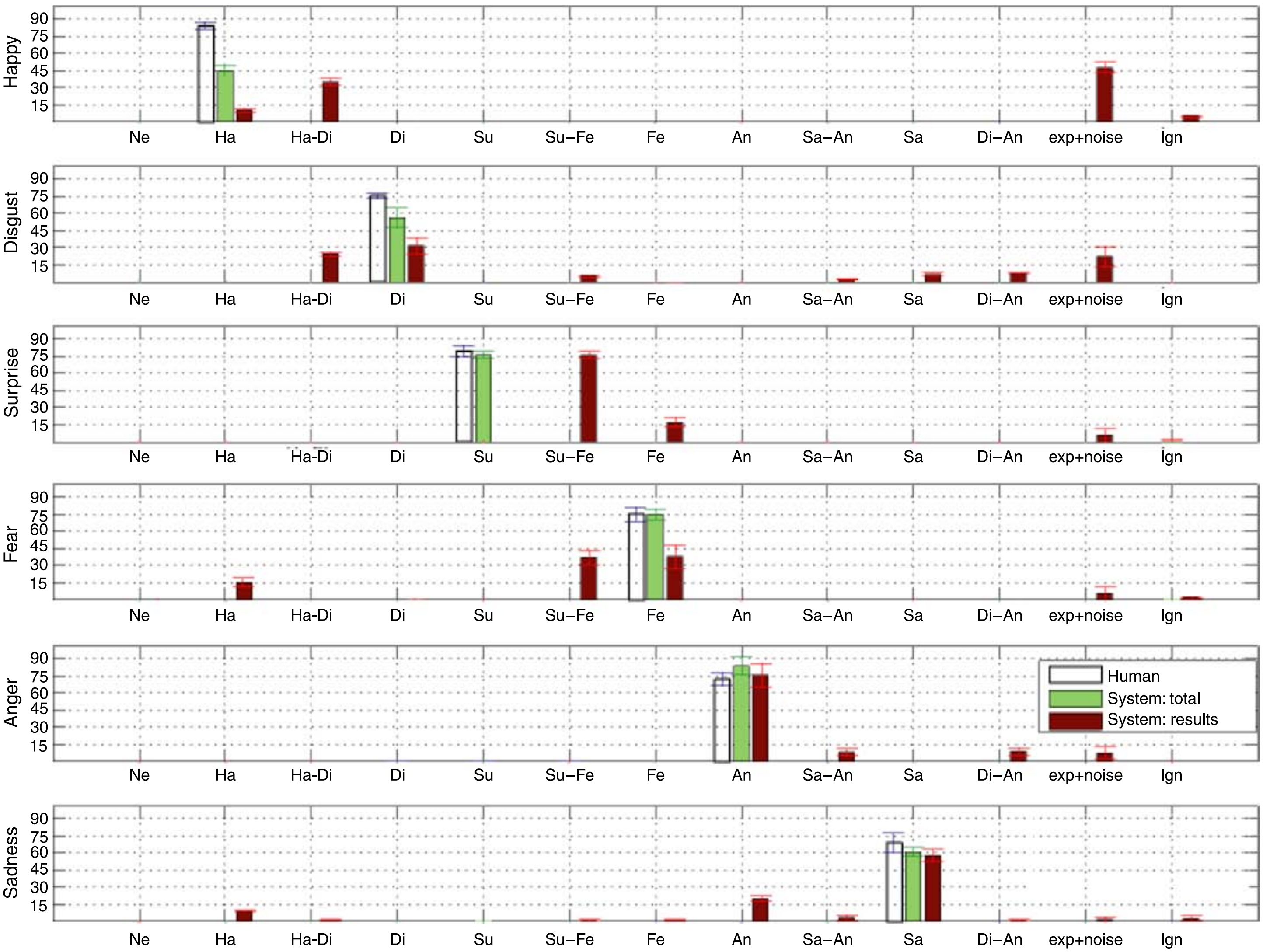


Figure 12. The means and standard deviations of the classification results with a discounting process according to the appearance intensities of the characteristic distances. Ne: *Neutral,* Ha: *Happiness,* Di: *Disgust,* Su: *Surprise,* Fe: *Fear,* An: *Anger,* Ign: *Total ignorance*. Red and green bars correspond to the system results and white bars to the human performances.

To better quantify the fit between humans and model observers, we computed the Pearson correlation coefficients between the model classification results of the third simulation and those of human observers for each of the six facial expressions. The obtained coefficients are: $R_{\text{Happiness}} = 0.02$, $R_{\text{Disgust}} = 0.035$, $R_{\text{Surprise}} = 0.04$, $R_{\text{Anger}} = -0.02$, $R_{\text{Fear}} = 0.003$, $R_{\text{Sadness}} = -0.02$. Based on these results it is clear that even if the classification rates of the model and human classifier are comparable, they do not have the same behavior on a trial-by-trial basis. The difference must pertain to the information used for the recognition. The next section assesses this possibility.

Relative importance of the characteristic distances

To better understand the results of the last simulation, we examined the relative importance of the five characteristic distances for the recognition of each of the basic facial expressions. To do this we have performed a multiple linear regression. The general purpose of multiple linear regression analyses is to learn more about the relationship between several independent variables and a dependent variable. In the present case the dependent variable corresponds to each of the six facial expressions E_e and the independent variables correspond to the five characteristic distances D_i. For example, Happiness (E_1), on a given trial, could be defined as

$$(E_1)_t = x_{1t}*D_1 + x_{2t}*D_2 + x_{3t}*D_3 + x_{4t}*D_4 + x_{5t}*D_5, \tag{8}$$

where x_{1t}, …, x_{it} correspond to the appearance intensities of the characteristic distances D_1, …, D_i.

The aim is to compute the contribution of each characteristic distance for the recognition of E_1, then based on all the available data, we obtain

$$\begin{pmatrix} (E_1)_1 \\ (E_1)_2 \\ \\ (E_1)_n \end{pmatrix} = \begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{pmatrix} * \begin{pmatrix} x_{11}\ldots x_{15} \\ x_{21}\ldots x_{25} \\ \\ x_{n1}\ldots x_{n5} \end{pmatrix}, \tag{9}$$

$$E_1 = d_{E_1} * X_{E_1}, \tag{10}$$

where n corresponds to the number of times E_1 is presented and recognized and x_{ni} corresponds to the appearance intensity (see Discounting section) of the characteristic distance D_i during the recognition of the expression E_1 at time n.

The same modeling is made for the other expressions leading to five equations to be solved as

$$\begin{aligned} E_e &= d_{E_e} * X_{E_e} \\ d_{E_e} &= (X'_{E_e} * X_{E_e})^{-1} * X'_{E_e} * E_e, \end{aligned} \tag{11}$$

where d_{E_e} corresponds to the coefficients of the characteristic distances reflecting their importance for the recognition of each facial expression E_e, $1 \le e \le 6$.

The solution regression coefficients are given in Figure 13. Each histogram comprises 5 regression coefficients, each corresponding to the importance of a characteristic distance for the recognition of the current expression.

To measure the quality of the results obtained for each expression, the corresponding percentages of variance explained R^2 are measured and reported in Figure 13.[3] Except for *Sadness*, the values of R^2 are positive and very high, which reflects a good fit of the data and thus a high confidence in the coefficients obtained.

We will focus on three aspects of the results displayed in Figure 13. The importance of each characteristic distance for the recognition of each expression is compared with the Smith et al. results. Except for *Anger*, there is an excellent correspondence between the most important characteristic distances for the proposed model and the facial cues used by the ideal observer (or model) of Smith et al. This model uses all the information available to perform the task optimally. These results allow the conclusion that the characteristic distances used summarize the most important information necessary for the classification of the facial expressions in the CAFE database and that the rules (i.e., Table 1) we used reflect ideal but not human information usage. However, the visual cues used by human observer are different from those used by the Smith et al. model observer and the model proposed here. In some cases, human observers show a partial use of the optimal information available for the classification of facial expressions (Smith et al., 2005). For example, humans use the mouth but not eyes for *Happiness* and they use the eyes but not the mouth for

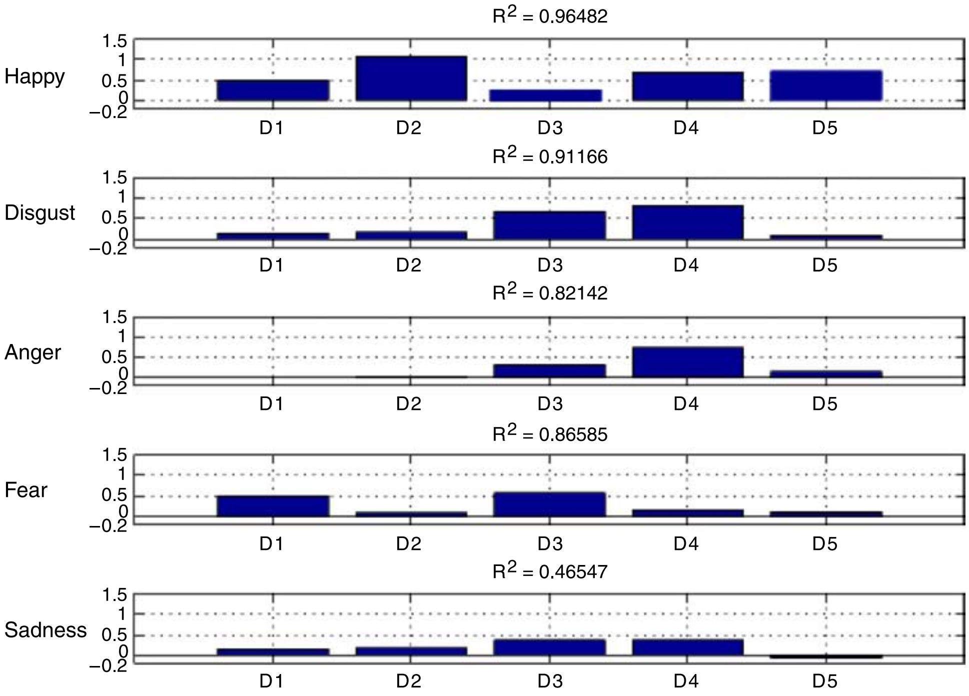


Figure 13. Importance of each characteristic distance for the studied expressions: positive values mean that the corresponding distance and the related expression are positively correlated; 0 means that there is no relation between them; negative values means that the corresponding distance and the related expression are negatively correlated. 1 means the strongest positive possible value, −0.2 is the lowest negative value.

Fear. In other cases, humans use information, which is not optimal: for example the nasolabial furrow in the case of *Disgust* and the wrinkles on the forehead in the case of *Sadness*. Given that humans easily outperform machines at recognizing facial expressions in everyday situations, it appears likely that their alleged "suboptimalities," in fact, reflect robust everyday facial expression statistics, not present in the CAFE face image set. Thus it seems promising for a future implementation of our model to use these "suboptimal" features for the facial features classification (e.g., nasolabial furrow in the case of *Disgust*) and to take into account their relative importance in the classification process.

Conclusion

We have modified the TBM-based model for recognizing facial expressions proposed by Hammal et al. (2007) to allow it to process partially occluded facial stimuli. Next, we have compared the behavior of this model with that of humans in a recent experiment, in which human participants had to classify stimuli expressing the six basic facial expressions plus Neutral that were randomly sampled using Gaussian apertures (Smith et al., 2005). The modified TBM-based model fits the human data better than the original TBM-based model. However, further analyses revealed important differences between the behavior of the modified TBM-based model and human observers. Given that humans are extremely well adapted to real-life facial expression recognition, future work will focus on weighting each visual cue during the classification process according to its importance for the expected expression and on adding other visual cues used by human observers such as wrinkles. The fusion architecture based on the TBM will greatly facilitate this future work.

Acknowledgments

This research was supported by the Canadian Institutes of Health Research and the Le Fonds Québécois de la Recherche sur la Nature et les Technologies.

Commercial relationships: none.

Corresponding author: Zakia Hammal.
Email: zakia_hammal@yahoo.fr.
Address: C.P. 6128, succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada.

Footnotes

[1] See http://www.cs.ucsd.edu/users/gary/CAFE.

[2] Hammal–Caplier database is composed of 19 subjects that displayed 4 expressions (Smile, Surprise, Disgust, and Neutral). Eleven subjects were used for the training and 8 subjects for the test. Each video recording starts with neutral state, reaches the apex of the expression, and goes back to the neutral state. The sequences were acquired in 5-second segments at 25 images/second.

[3] The importance of the distances for the Surprise expression is not reported because this expression is always recognized as a mixture of *Surprise* and *Fear* and never as Surprise only.

References

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature, 372,* 669–672. [PubMed]

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 373–379. [PubMed]

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37,* 2049–2058. [PubMed]

Black, M. J., & Yacoob, Y. (1997). Recognizing facial expression in image sequences using local parametrized models of image motion. *Transactions on Computer Vision, 25,* 23–48.

Boucher, J. D., & Ekman, P. (1975). Facial areas and emotional information. *Journal of Communication, 25,* 21–29. [PubMed]

Cohn, J. F. (2006). Foundations of human computing: Facial expression and emotion. *Proceedings of ACM Multimodal Interfaces,* 233–238.

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *Journal of Wavelets, Multi-Resolution Information Processing, 2,* 1–12.

Cohn–Kanade database (2000). http://vasc.ri.cmu.edu/sidb/html/face/facial%20expression.

Dailey, M., Cottrell, G. W., & Reilly, J. (2001). California facial expressions (cafe). Unpublished digital images. San Diego, CA: University of California.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: Murray.

Denoeux, T. (2008). Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence. *Artificial Intelligence, 172,* 234–264.

Ekman, P. (1999). *The handbook of cognition and emotion: Facial expression*. John Wiley and Sons.

Ekman, P., & Friesen, W. V. (1978). *The facial action coding system (facs): A technique for the measurement of facial action*. Palo Alto, CA: Consulting Psychologists Press.

Elouadi, Z., Mellouli, K., & Smets, P. (2004). Assessing sensor reliability for multisensor data fusion within the transferable belief model. *IEEE Transactions on Systems, Man, and Cybernetics B, 34,* 782–787. [PubMed]

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition, 1,* 259–275.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41,* 2261–2271. [PubMed]

Gouta, K., & Miyamoto, M. (2000). Facial areas and emotional information. *Japanese Journal of Psychology, 71,* 211–218.

Hammal–Caplier database (2003). http://viscog.psy.umontreal.ca/zakia/index.htm.

Hammal, Z. (2006a). *Facial features segmentation, analysis and recognition of facial expressions using the transferable belief model*. PhD thesis, LIS Laboratory, Grenoble, France.

Hammal, Z. (2006b). Dynamic facial expression understanding based on temporal modeling of transferable belief model. *Proceedings of the International Conference on Computer Vision Theory and Application,* Setubal, Portugal, February 25–28.

Hammal, Z., Caplier, A., & Rombaut, M. (2005). A fusion process based on belief theory for classification of facial basic emotions. *Proceedings of the 8th International Conference on Information Fusion (ICIF),* Philadelphia, PA, USA.

Hammal, Z., Couvreur, L., Caplier, A., & Rombaut, M. (2007). Facial expressions classification: A new approach based on Transferable Belief Model. *International Journal of Approximate Reasoning, 46,* 542–567.

Hammal, Z., Eveno, N., Caplier, A., & Coulon, P. Y. (2006). Parametric models for facial features segmentation. *Signal processing, 86,* 399–413.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115,* 288–299. [PubMed]

Lien, J. J., Kanade, T., Cohn, J. F., & Li, C. (1998). Subtly different facial expression recognition and expression intensity estimation. *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (pp. 853–859). Santa Barbara, CA: IEEE.

Lucas, B. D., & Kanade, T. (1981). An iterative image-registration technique with an application to stereo vision. In *Image Understanding Workshop,* (pp. 121–130). US Defense Advanced Research Projects Agency.

Malciu, M., & Preteux, F. (2001). Mpeg-4 compliant tracking of facial features in video sequences. *Proceedings of the International Conference on Augmented, Virtual Environments and 3D Imaging* (pp. 108–111), Mykonos, Greece, May.

Mercier, D., Deunoeux, T., & Masson, M. H. (2006). Refined sensor tuning in the belief function framework using contextual discounting. *Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems* (vol. 2, pp. 1443–1450). Paris, France.

Pantic, M., & Bartlett, M. S. (2007). Machine analysis of facial expressions. In K. Delac & M. Grgic (Eds.), *Face recognition* (pp. 377–416). Vienna, Austria: I-Tech Education and Publishing.

Pantic, M., & Patras, I. (2005). Detecting facial actions and their temporal segmentation in nearly frontal-view face image sequences. *Proceedings of the IEEE International Conference on Systems Man and Cybernetics,* Waikoloa, Hawaii, October.

Pantic, M., & Patras, I. (2006). Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on System Systems, Man, and Cybernetics, Part B: Cybernetics, 36,* 433–449.

Pantic, M., & Rothkrantz, L. J. M. (2000a). Expert system for automatic analysis of facial expressions. *Image and Vision Computing Journal, 18,* 881–905.

Pantic, M., & Rothkrantz, L. J. M. (2000b). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22,* 1424–1445.

Pards, M., & Bonafonte, A. (2002). Facial animation parameters extraction and expression detection using hmm. *Signal Processing, Image Communication, 17,* 675–688.

Rosenblum, M., Yacoob, Y., & Davis, L. S. (1996). Human expression recognition from motion using a radial basis function network architecture. *IEEE Transactions on Neural Networks, 7,* 1121–1137.

Roy, S., Roy, C., Fortin, I., Either-Majcher, C., Belin, P., & Gosselin, F. (2007). A dynamic facial expression database. *Proceedings of the Vision Sciences Society,* Sarasota, Florida.

Schmidt, K. L., Ambadar, Z., Cohn, J. F., & Reed, I., (2006, March). Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. *Journal of Nonverbal Behavior, 30,* 37–52.

Smets, P. (1998). The transferable belief model for quantified belief representation. *Handbook of defeasible reasoning and uncertainty management system* (vol. 1, pp. 267–301). Dordrecht: Kluwer Academic.

Smets, P. (2000). Data fusion in the transferable belief model. *Proceedings of the International Conference on Information Fusion* (pp. 21–33), Paris, France: IEEE.

Smets, P., & Kruse, R. (1994). The transferable belief model. *Artificial Intelligence, 66,* 191–234.

Smets, Ph. (2005). Decision making in the TBM: The necessity of the pignistic transformation. *International Journal of Approximate Reasoning, 38,* 133–147.

Smith, M., Cottrell, G., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions of emotions. *Psychological Science, 16,* 184–189.

Tekalp, M., & Ostermann, J. (2000). Face and 2-D mesh animation in MPEG-4. *Image Communication Journal, 15,* 387–421.

Tian, Y., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23,* 97–115.

Tsapatsoulis, N., Karpouzis, K., Stamou, G., Piat, F., & Kollias, S. (2000). A fuzzy system for emotion classification based on the mpeg-4 facial definition parameter set. *Proceedings of the 10th European Signal Processing Conference,* Tampere, Finland, September 5–8.

Yacoob, Y., & Davis, L. S. (1996, June). Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18,* 636–642.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63,* 271–313. [PubMed]

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, 31,* 39–58. [PubMed]

Zhang, Y., & Qiang, J. (2005, May). Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27,* 699–714.