The identification of unfolding facial expressions

The Identification of Unfolding Facial Expressions

Chiara Fiorentini

Department of Psychology, The Australian National University, Canberra, Australia

Susanna Schmidt

Department of Psychology, University of Turin, Italy

Paolo Viviani

Laboratory of Neuromotor Physiology, Santa Lucia Foundation, Rome, Italy

Corresponding author: Chiara Fiorentini

Department of Psychology, The Australian National University

Canberra, ACT 0200, Australia

Email: chiara.fiorentini@anu.edu.au

Mobile: +61 42 18 69 890

Abstract

We investigated the identification of the facial expressions (FEs) of 6 emotions (anger, surprise, happiness, disgust, fear, sadness) using high-speed (500 f/s) video-recordings of the FE of 5 actors, unfolding from neutral through to apex. Recordings were manually coded frame-by-frame with the Facial Action Coding System (FACS, Ekman, Friesen, & Hager, 2002) in order to identify the occurrence of each facial action contributing to each expression, and its intensity values over time. Recordings were shown in slow-motion (1/20 of recording speed) to 100 observers in a forced choice identification task: participants were asked to identify the emotion during the presentation as one of the six possible alternatives, as soon as they had sufficient information to do so. We analysed the type of response provided by the participants and the associated response time (RT). For each recording and each type of response, we computed the derivative of the cumulative distribution function of the RTs associated to that response, and correlated it with facial activity along each sequence. We found systematic correlations between facial activities, response probabilities and RTs peaks, and significant differences in RTs distribution for right and wrong answers. Overall, the results suggest that facial actions contribute individually and cumulatively to emotion identification.

Keywords: Facial Expressions, Identification, Action Units, Emotions

The Identification of Unfolding Facial Expressions

Since the seminal works by Darwin (1872/1998) and Duchenne (1876/1999), converging evidence (reviewed in Ekman, Friesen, & Ellsworth, 1982) has shown that major emotions such as happiness, fear or anger are associated with consistent facial expressions (FEs), and are universally and reliably inferred from these facial displays. A critical open question concerns what facial information is actually used by observers in order to recognize emotions. From the descriptive point of view, FEs can be characterized in terms of individual facial actions, or *components* (Smith & Scott, 1997). The most widely used technique for doing so is the Facial Action Coding System (FACS, Ekman, Friesen, & Hager, 2002), which identifies the *components* with certain changes in facial appearance produced by specific muscle actions (Action Units, AUs). In principle, any facial expression can be described as a unique combination of AUs.

One of the issues being debated is therefore the relationship between *components* and the overall *configuration* resulting from the activation of all these components. Two main theoretical positions can be distinguished. According to *categorical models,* individual components have no intrinsic meaning, and the information required for accurate emotion identification can be extracted only from the overall configuration (Bimler & Paramei, 2006). Along the same line, some authors made the additional claim (Calder, Young, Perrett, Etcoff, & Rowland, 1996; Campanella, Quinet, Bruyer, Crommelinck, & Guerit, 2002; Etcoff & Magee, 1992) that emotion categorization is a true case of categorical perception.

On the other side, *componential models* hold instead that at least some affective meaning can be ascribed to FE components (Scherer, 1992; Smith & Scott, 1997). In this view the fact that certain AUs are present in different FEs is actually taken to suggest that emotions do have components that can be shared. For instance, raised eyebrows and wide opened eyes, which are usually present in both fear and surprise FEs, would correspond to a shared cognitive state, namely that one is uncertain about his/her situation, and is attending to the environment in an attempt to

reduce the uncertainty. However, the overall configuration which is generally recognized as "fear" conveys information about the expresser's emotional state which is unique to fear, and that is not captured by any of the components of fear taken independently (Smith & Scott, 1997).

Evidence about what facial information contributes to emotion categorization, and about the relative dominance of the configuration over components is mixed (e.g. Bassili, 1979; Bimler & Paramei, 2006; Calder, Young, Keane, & Dean, 2000; Ellison & Massaro, 1997; Martin, Slessor, Allen, Phillips, & Darling, in press).

Cross-cultural research on emotion recognition provides the primary evidence in favour of the view that FEs are evaluated as configurations. Ekman and colleagues (Ekman, 1972, 1989, 1992, 1994, 2004; Ekman & Friesen, 1976) and Matsumoto and colleagues (Biel, Matsumoto, Ekman, Hearn, et al., 1997) showed that photographs of the FEs of anger, fear, happiness, sadness, surprise, and disgust - the so-called "basic" emotions - are recognized very accurately across cultures (mean recognition rate across emotions: 83% in Western cultures; see Russell, 1994 for a review). Such a large consensus suggests that, at least in the case of "basic" emotions, the associated FEs are fairly prototypical. This, in turn, has been taken to imply that these "prototypes" are perceived as "packages", i.e., by evaluating the overall face configuration (Ellison & Massaro, 1997; Smith & Scott, 1997).

Other lines of research have addressed more directly the problem of how FEs are perceived by correlating identification performance to the type and distribution of specific components across the face (e.g. Bassili, 1979; Calder et al., 2000; de Bonis, De Boeck, Pérez-Diaz, & Nahas, 1999; Ellison & Massaro, 1997; Fiorentini & Viviani, 2009). This was generally done by limiting or manipulating the area of the face that observers could see, and then comparing judgments made from one facial area with judgments made from another.

The results, however, do not lend themselves to a unique interpretation. When only part of the face is shown, the salient portion for identifying an expression varies according to the emotion.

In particular, the mouth region is more crucial than the eye region for identifying happiness and disgust, whereas the opposite is true for anger, sadness, and fear (Calder et al., 2000; Hanawalt, 1944; Sullivan, Ruffman, & Hutton, 2007). When the two halves of the face display incongruent information (e.g. fear in the upper half and happiness in the lower half), as in "chimerical portrayals", only some expressions, such as surprise and happiness, can be unambiguously identified from only one area of the face (de Bonis et al., 1999). Instead, the recognition of negative emotions, especially fear, requires the integration of multiple cues across the face (de Bonis et al., 1999; Morris, de Bonis, & Dolan, 2002).

Attempts have also been made to manipulate (Ellison & Massaro, 1997; Fiorentini & Viviani, 2009; Nusseck, Cunningham, Wallraven, & Bülthoff, 2008) or measure (Ekman, Friesen, & Tomkins, 1971; Kohler, Turner, Stolar, Bilker, Bresinger, et al., 2004) selected expressive components. By controlling independently eyebrow frown and smiling mouth in computer-generated stimuli Ellison and Massaro (1997) reported that emotion identification is most efficient when both facial features display congruent emotional meaning. However, the smiling mouth alone is sufficient for a good performance regardless of whether the eyebrows indicate the same emotion. This result was later confirmed by Kohler and colleagues (2004) using photographs of real faces. By measuring with FACS the facial movements of four expressions, they showed that upturned mouth, raised cheeks, and tightened lids were jointly diagnostic features for happiness. However, only the upturned mouth was both necessary and sufficient for emotion identification.

Nusseck, Cunningham, Wallraven, & Bülthoff (2008) obtained similar results by manipulating the motion of selected facial areas. For happiness and surprise, the mouth was sufficient for identification, whereas other expressions (e.g. sadness and disgust) were identified reliably only when consistent information could be gathered from several facial areas. Therefore, at least in the case of certain basic emotions, identification does not require a full coherent complement of facial features (Martin et al., in press).

Further evidence that the diagnostic value of expressive components depends on the emotion was provided by Fiorentini and Viviani (2009) who applied a morphing technique to produce ambiguous stimuli in which the upper and lower part of the face displayed graded blends of different expressions. The results showed that the two most salient facial features (the eyes and the mouth) are taken into account separately, and weighed differently in the identification process, depending on the expression: the mouth was the most important feature for the identification of Happiness and Disgust, whereas the eyes were most relevant for identifying Anger and Fear.

In summary, the above studies - mostly conducted with static stimuli - yielded both evidence that identification may be driven by single facial components, and evidence that certain FEs are identified only by taking into account the entire facial configuration.

As emotional expressions are dynamic events, one might argue that still pictures are not ideally suited as stimuli for investigating emotion perception. Indeed, just as they do for face recognition (e.g. Hill & Johnston, 2001; Lander, Christie, & Bruce, 1999), dynamic cues do afford useful information also for recognizing FEs (Bassili, 1979; Edwards, 1998; Humphreys, Donnelly, & Riddoch, 1993; Wherle, Kaiser, Schmidt, & Scherer, 2000). Bassili (1978, 1979) showed that FEs are recognized reliably even from point-light schematic displays, which suppress most figural cues while preserving movement information. Humphreys et al. (1993) reported the case of an agnosic patient who was unable to identify static FEs, but was relatively accurate with point-light displays. Sensitivity to temporal cues was demonstrated by an experiment (Edwards, 1998) in which participants were able to order correctly scrambled sequences of still frames describing the temporal course of various expressions.

By showing the gradual rise of an expression *pari passu* with the activation of the different facial components, dynamic stimuli can provide information about how expressive cues are integrated into the coherent percept of a specific emotion. Wehrle et al. (2000) tested the assumption that expression identification is driven in a principled way by the time course of the

activated facial components, as predicted by Scherer's componential model (Scherer & Ellgring, 2007). Using computer-generated simulations, their experiment contrasted static FEs with two types of dynamic FEs, one in which AUs were activated according to the sequence postulated by Scherer and colleagues (Scherer & Ellgring, 2007), the other in which all AUs were activated simultaneously as a complete configuration. Identification accuracy was better with dynamic than with static FEs, but there was no difference between the two types of transitions. However, one might argue that the lack of differences between the two conditions in this study was actually due to the use of simulated FEs in which the time course of facial components was determined *a priori*. As far as we know, nobody has investigated the role of individual components in the identification of dynamic portrayals of real FEs.

The aim of this study is to test the hypothesis that emotion identification is achieved by attending selectively to a diagnostic subset of facial features. Our experimental strategy is to correlate identification responses with the actual time course of AUs activation. This strategy is motivated by the demonstration (Fiorentini, Schmidt & Viviani, unpublished results) that, for most basic emotions, the associated AUs converge towards the apex with different time courses. At any one time prior to the apex, the recordings show different combinations of AUs with different intensities. Thus, if indeed some affective meaning can be ascribed to FE components (see above), the distribution of identification responses given during the unfolding of the expression should depend on the time course of the activated facial components. For all basic emotions, we video-recorded at high speed the FEs of five actors and generated a fine-grained description of the facial activity by coding with FACS the frames of the recordings. The recordings were then shown in slow motion to the observers who were asked to identify as soon as possible the unfolding emotion. Finally, response times and response accuracy were correlated with the evolving pattern of AUs activation. From the *categorical* and *componential* views summarized above, one can derive different predictions about response distributions associated with the identification

process.If expressions are dealt with as configurations, and identification involves the evaluation of the complete "package" of AUs, correct response times should cluster near the apex of the expression, with a fairly uniform distribution. Alternatively, identification may be a bottom-up process in which evidence from individually diagnostic components is collected until a decision threshold is crossed. If so, one expects correct responses to occur when these emotion-specific components reach a sufficient intensity, even before the full deployment of all other components. Moreover, if more than one component has diagnostic value, response time distributions may be multimodal.

Method

Generation and coding of the stimuli

Encoders. Five professional actors (age range: 25-40 years; three females), regularly active on stage in Geneva who were paid for their services. Actors are identified with the letters A to E.

Recording procedure. Actors were recorded in separate sessions. They were asked to produce the FE associated with six emotions: Anger, Surprise, Fear, Happiness, Sadness, and Disgust. At the beginning of the session, actors read short scenarios involving situations that typically elicit the corresponding emotion (Bänziger, Pirker, & Scherer, 2006[1]). They were encouraged to perform spontaneously, trying to actually trigger the target emotion. Expressions were performed without vocalization. The six portrayals were produced in a counterbalanced order during a single session. Each portrayal was repeated a few times for subsequent selection. If actors felt that a rendering was not optimal they could repeat it.

A high-frequency digital camera (Nac Hot Shot II, NAC Image Technology) filmed the transition from a neutral face to the fully deployed target expression. At the recording speed of 500 frames/s, all FE were described by less than 800 frames. The face was filmed against a dark

[1] One example of the scenarios used to induce the desired emotion Fear (translations of the original French text).

"It's after midnight and I'm walking back home. For a while I'm alone. Then suddenly I realize that a man is stalking me. Hearing his steps behind me, I accelerate and so he does. I start running and he starts running too. Feeling that he has grabbed my coat, I turn back and see a knife in his other hand."

background, in full frontal view. Figure 1 shows a sample of equally spaced images from the Happiness sequence of actor D.

---------- *Inserire Figura 1* ----------

Figure 1. Example of stimuli. Twelve samples from the Happiness sequence of actor D. In the original recording samples were spaced by 50 ms. At the presentation rate of 25 frames/s adopted for the experiment samples were actually spaced by 1 s. Actual stimuli were in colour.

Facial expression selection

The sequences used for the experiment were selected according to two criteria. First, following Wagner (1997) and Gosselin et al. (1995), after each recording, we asked actors to judge their performance. Performances considered inadequate were immediately discarded. Among the stored recordings, actors were then asked to select the one in which they had best succeeded in conveying only the intended emotion. Thus, we retained one portrayal for each actor for each emotion.

In the second step of the selection procedure, we checked the validity of the selected sample of 30 portrayals by showing to 20 independent observers (age range: 25-38; 12 female) the single frames corresponding to the apex of the expressions. Observers had to score on a 5-point scale the extent to which each basic emotion was expressed in each portrayal (1 = *I completely disagree that emotion x is present in the portrayal*, 5 = *I completely agree that emotion x is present in the portrayal*). The rating was performed in individual sessions. Portrayals were shown one by one in random counterbalanced order, with no repetition, and scores were given after each portrayal. For 23 out of 30 portrayals mean ratings of the target emotion were equal or greater than 4, and only in 4 cases they were lower than 3.5 (i.e., Disgust, Fear, and Sadness of actor A, and Fear of actor E). In order to have all actors express all the emotions under investigation, these 4 expressions were nonetheless included in the final sample.

Post-processing

Each frame (TIFF format) of the recording was processed with Photoshop CS3 to equalize overall luminance, contrast, and chromatic spectrum, and scaled to a standard dimension (864 × 1074 pixels). Across actors and emotions the original sequences showed only small differences in the length of the initial "neutral" phase of the expression (i.e. the time between the beginning of the recording and the onset of an expressive action on the actor's face). Instead, there were substantial differences in the duration of the expression (i.e., from onset to apex). Sequences were cut at the minimum length (770 frames) that accommodated the longest movement (Surprise by actor B) and began with 20 frames showing the neutral face before the first noticeable muscle contraction (excess frames were discarded; range 5-20 frames). The return phase to the neutral face was replaced by copies of the last frame representing the apex of the expression. The length of this final portion varied across participants and emotions (range: 0-270 frames). Finally, the sequences were transformed in digital video with Adobe After Effects, and compressed at a speed of 25 frames/s, yielding a 30.8 s video-clip (wmv format).

FE Coding

FEs were coded using the most updated version of FACS (Ekman, et al., 2002). FACS identifies facial components as individual muscle actions (Action Units, AUs) which determine visually detectable changes in the face appearance. In Table 1, we reported the correspondence between the AUs discussed in this paper and the main facial changes associated to them.

One of us (S.S.) - a certified FACS coder - performed the coding using the uncompressed frame sequences. The coding was performed every 10 frames (20 ms of real time). Thus, each sequence was described by 78 records reporting identity and intensity of all active AUs.

Table 1

Action Units (AUs) and their Changes in Facial Appearance

AU number	Changes in Facial Appearance[a]	Face Portion Involved
1	Raises the inner part of the eyebrow	Eyes, Eyebrows and surrounding region
2	Raises the outer part of the eyebrow	
4	Lowers the eyebrows and pulls them together	
5	Raises the upper lid, exposing more of the upper portion of the eyeball	
6	Raises the cheek, produces "crow's feet" and wrinkles below the eye	
7	Raises and tightens the lower eyelid	
9	Wrinkles and pulls the skin upwards along the sides of the nose	Nose
10	Raises the upper lip and pushes the infraorbital triangle upwards(produces a bend in its shape)	Lips and mouth
11	Deepens the middle portion of the nasolabial furrow	
12	Pulls the lip corner up diagonally toward the cheekbone	
14	Tightens the mouth corners and produces a dimple-like wrinkle beyond the lip corners	
15	Pulls the corner of the lips down and produces some pouching, bagging, or wrinkling of the skin below the lip corners	
16	Pulls the lower lip down, flattens the chin boss	
17	Pushes the chin boss and lower lip upwards	
20	Stretches the lips horizontally, elongates the mouth	
23	Tightens the lips, making the lips appear more narrow	
24	Presses the lips together, narrows the lips	
25	Parts lips	
26	Lowers mandible	
27	Pushes mandible downwards	
38	Dilates the nostrils	Nose
43	Closes the eyes	Eyes

Note: [a]Changes in Facial Appearance*:* main facial changes produced by the corresponding AU (leftmost column). For a complete description of all associated changes, see Ekman, Friesen, & Hager (2002).

Identification task

Participants. One-hundred undergraduate students from the Faculty of Psychology and Educational Sciences at the University of Geneva (age-range: 20-35; 68 females) participated in exchange of course credits. All participants had normal or corrected-to-normal vision. The experiment was approved by the Ethics Committee of the University. Informed consent was obtained from the participants.

Procedure. The experiment was run in a dimly illuminated, quiet room. Participants were seated in front of a computer screen (Eizo, FlexScan 2410W 24'' monitor; resolution: 1920 x 1200 pixels; sampling rate: 60 Hz) at a distance of about 57 cm (at this distance 1 cm on the screen corresponds to 1 deg of visual angle). Before the experiment, we informed participants that stimuli were slow-motion video-recordings of the FE of six emotions identified by their standard name. On each trial, we showed one video-clip and asked participants to identify (forced choice) the unfolding emotion as soon as they felt confident to do so. At 25 frames/s the presentation was 20 time slower than the real FE. Responses were entered through a response box (DirectIn Custom response Box, by EMPIRISOFT Research Software) with one RT button and six category buttons labelled with the tested emotions. Two sequential responses were requested. Hitting the first button stopped the presentation of the stimulus and recorded the response time (RT) with 1 ms accuracy. Then, the chosen emotion was entered with the category buttons. In all but a few exceptional cases, responses were given well before the end of the presentation. Otherwise, response time was set to the maximum duration (30.8 s).

Each video-clip was presented 4 times for a total of $30 \times 4 = 120$ trials (a pilot study showed that increasing the number of repetitions beyond this limit might produce perceptual learning effects). The stimuli were presented in a random order, with the constraint that there should be at

least three trials before repeating either the same emotion or the same actor. A session lasted about 20 minutes.

Results

FE analysis

For each actor and emotion Table 2 reports the AUs involved in the FEs in the order of activation, the number of active AUs (N), the average duration of the expressions (D), the average intensity of the AUs (I), and an estimate of the synchronization among AUs (S). The number of active AUs is an index of the *complexity* of the expression. Across actors, it ranged from a minimum of 3 (Happiness) to a maximum of 10 (Fear) AUs. The intensity with which facial actions are produced is an index of *expressivity* computed by averaging the peak intensity of all active AUs. Number and intensity of the AUs differed across emotions (Number: $F(5, 24) = 8.543$, $p < .001$; Intensity: $F(5, 24) = 4.676$, $p < .01$). There was no significant difference among actors (Number: $F(4, 25) = 0.613$, $p = .657$; Intensity: $F(4, 25) = 1.052$, $p = .401$). With respect to complexity, two groups emerged, one comprising more complex FEs (i.e. Anger, Fear, and Disgust), the other including simpler ones (i.e. Surprise Happiness, and Sadness). Pair-wise comparisons showed that Surprise expressions were more intense than all other emotion portrayals.

Duration was defined as the interval between the first noticeable muscle contraction (the onset point) and the last intensity change in any AU. Averaged over actors, durations across emotions were fairly similar ($F(5, 24) = .400$, $p = .844$). Instead, duration did differ across actors ($F(4, 25) = 4.215$, $p < .05$). The mean intensity level of an expression did not correlate significantly with its duration, as very short emotions (e.g. Surprise) could be quite intense as well.

Synchronization was estimated as the mean over all pairs of AUs of the coefficient of linear correlation between their intensity. The mean synchronization over expressions and actors was .71, with a minimum of .48 (Anger by actor C) and a maximum of .97 (Surprise by actor C). Overall, synchronization did not vary significantly either across actors ($F(4,25) = 0.492$, $p = .742$) or across

emotions ($F(5,24) = .2.582$, $p = .053$). Synchronization was negatively correlated with duration ($r = -.538$, $p < .01$) and positively correlated with intensity ($r = .471$, $p < .01$) indicating that rapid, intense expressions tended to unfold in a more coordinate fashion.

Table 2

Analysis of FE recordings

FE	Actor	AUs[a]	N[b]	D[c] (ms)	I[d]	S[e]
	A	4,7,10,15,25,26,17	7	1000	2.86	0.66
	B	4,10,6,15,25,9,17,26,16	9	960	3.78	0.74
Disgust	C	1,26,4,2,7,9,15,25,17	9	640	3.56	0.81
	D	2,10,15,4,5,6,7,1	8	920	2.88	0.65
	E	7,4,15,25,1,10,6,17	8	640	3.50	0.79
Mean			8.2	832	3.31	0.73
	A	1,2,5,26,25,38,4,20	8	1180	2.63	0.29
	B	5,26,38,4,25,20,10,6,1	9	920	2.33	0.58
Fear	C	1,2,26,25,5,38,4,20	8	740	3.25	0.57
	D	2,27,1,25,4,5,20,6,7,17	10	540	3.30	0.83
	E	1,2,27,5,25,20,7	7	1000	3.57	0.76
Mean			8.4	876	3.02	0.61
	A	4,7,12,1,15	5	820	2.60	0.68
	B	1,4,11,6,7,17,15	7	900	3.43	0.73
Sadness	C	1,15,4,17	4	220	3.00	0.90
	D	7,1,4,11,43,6	6	900	3.00	0.80
	E	15,1,4,12,7,6	6	820	3.00	0.65
Mean			5.6	732	3.01	0.75
	A	25,5,1,2,27,6,12	7	780	3.00	0.75
	B	5,12,6,25,14,26	6	1380	2.83	0.54
Happiness	C	6,12,25	3	360	4.00	0.91
	D	2,25,12,5,6,1	7	960	4.00	0.51
	E	12,1,2,25,6	5	680	2.80	0.66

Mean			5.6	832	3.33	0.67

(table continues)

FE	Actor	AU[a]	N[b]	D[c] (ms)	I[d]	S[e]
	A	1,2,5,38,26	5	360	4.40	0.95
	B	1,2,5,26,38,25	6	1540	4.00	0.83
Surprise	C	1,2,27,5,25	5	300	4.60	0.97
	D	2,27,25,5,1	5	780	4.20	0.90
	E	1,2,27,25,5,17	6	560	4.00	0.84
Mean			5.4	656	4.24	0.90
	A	24,4,5,7,17	5	760	2.40	0.79
	B	10,25,26,4,16,7,9,20,5	9	760	3.89	0.67
Anger	C	5,23,25,26,4, 1,16	8	840	3.50	0.48
	D	27,25,2,4,9,20,23,5	8	960	3.63	0.79
	E	26,25,9,23,7,17,5,4	8	680	3.13	0.28
Mean			7.6	800	3.31	0.60

Note. [a]AUs: Sequence of active action units from onset to apex. [b]N: Total number of active AUs. [c]D: Duration of the sequence from the first noticeable action to the last intensity change in any AU. [d]I: Intensity. Average of peak values over all active AUs. [e]S: Synchronization. Average over all pairs of AUs of the linear correlation coefficient between intensities.

Identification rates

Table 3 reports mean and range across actors of the response probabilities for all stimulus/response pairs. A two-way ANOVA (6 [emotion] × 5 [actor], repeated measures, arcsin transformation) on correct responses detected significant differences for emotion ($F(5, 495) = 107.846$, $p < .001$) and actor ($F(4, 396) = 22.394$, $p < .001$), and a significant emotion × actor interaction ($F(20, 1980) = 52.674$, $p < .001$). Collapsing the probability of correct identification over emotions, expressive ability varied somewhat among actors (A: .68; B: .75; C: .77; D: .81; E:

.77). Pair-wise LSD comparisons showed that mean accuracy for actor A was significantly smaller than all other means, and mean accuracy for actor D was significantly higher. Collapsing over actors showed that differences among emotions were more marked (Disgust: .57; Fear: .61; Sadness: .89; Happiness: .94; Surprise: .73; Anger: .80). Pair-wise differences between emotions were all significant, except the one between Disgust and Fear.

Table 3

Response probabilities

FE	Response[a]					
	Disgust	Fear	Sadness	Happiness	Surprise	Anger
Disgust	**.57**	.02	.33	.00	.00	.07
	[.31-.89]	[0-.06]	[.06-.64]	--	--	[0-.28]
Fear	.07	**.61**	.03	.00	.27	.02
	[0-.24]	[.42-.86]	[0-.10]	--	[.04-.38]	[0-.05]
Sadness	.03	.01	**.89**	.06	.01	.00
	[.08-.02]	[0-.04]	[.55-.99]	[0-.27]	[0-.05]	--
Happiness	.00	.00	.02	**.94**	.04	.00
	--	--	[0-.06]	[.90-.98]	[.01-.09]	--
Surprise	.01	.24	.02	.00	**.73**	.00
	[0-.01]	[.09-.41]	[0-.09]	--	[.49-.88]	--
Anger	.09	.07	.01	.00	.02	**.80**
	[0-.30]	[0-.28]	[0-.07]	--	[0-.02]	[.63-.95]

Note. [a]Mean probabilities and range (brackets) for each stimulus-response pair. Boldface: Probability of correct responses.

Response times

If specific combinations of AUs contributed sequentially and cumulatively to the identification of an expression, correct responses should cluster around the points along the presentation where the discriminating AUs reach a critical activation level. To test this hypothesis we computed the RT probability density functions for both correct and incorrect responses. To do so we adopted the following procedure. First we computed the cumulative distribution functions

by pooling the responses of all participants. The distributions with relative frequency ≥ .07 were interpolated by a generalized logistic function, f(t) = 1/(1+exp(P(t))), where P(t) is a 6th degree polynomial. Finally, the RT density functions were estimated by the analytical derivative of the interpolation. For all actor/expression combinations Figure 2 shows the cumulative RT distributions for correct responses. We identified the centres of the response clusters by the relative maxima (modes) of the density distributions. Figure 3 and 4 show examples in which a single-peaked RT distribution is aligned with the FACS coding of facial activity (AUs intensity as a function of time).

---------- Inserire Figura 2 ----------

Figure 2. RT distributions. Each panel summarizes the results for the indicated expression. Heavy lines are the cumulative RT distribution for correct responses (all participants). Thin lines are least-square interpolations using a generalized logistic model (see Method). Letters identify the distributions for the five actors. Shown inset the probability of a correct identification for each actor and the average across actors.

The three main findings were: 1) In all cases, RT modes for correct responses preceded the apexes of the expressions. 2) Most RT probability density distributions showed clear evidence of one or more response clusters. Moreover, for several expressions RT modes corresponded to a change in the intensity of a group of AUs. 3) There were systematic shifts between the RT distributions for correct and incorrect responses to a given portrayal. The correspondence between facial activity and RT distributions for correct and incorrect identifications is described in detail in the next two sections.

Correct responses

If expressions are processed as configurations, RTs for correct responses should cluster near their apexes. On the other hand, if identification relies on individual diagnostic components,

correct responses should occur when these emotion-specific components reach a sufficient intensity, even before the expression is fully deployed. For all expressions, RT modes preceded the apexes of the expressions (Table 4), suggesting that expression identification relies on subsets of AUs, or at least, on expressions where all units are active, but have not yet reached full intensity.

Table 4

Apexes and RT modes for correct responses

FE	Actors										
	A		B			C		D		E	
	Apex[a]	RT Mode 1	Apex	RT Mode 1	RT Mode 2	Apex	RT Mode 1	Apex	RT Mode 1	Apex	RT Mode 1
Disgust	19.2	8.3	19.8	7.0	**12.1**	13.8	6.3	18.8	10.9	12.4	7.7
Fear	24.1	8.6	17.6	9.9		15.0	10.0	11.2	6.7	24.8	10.0
Sadness	16.8	14.8	18.2	12.5		5.1	4.5	17.7	7.5	15.4	7.3
Happiness	14.5	6.5	27.6	13.0		6.7	4.0	19.1	6.9	13.9	2.0
Surprise	7.0	5.9	30.8	6.9	**18.5**	6.0	5.0	16.4	6.3	10.3	7.0
Anger	16.5	10.6	18.3	4.2		16	7.6	19.1	7.8	13.4	4.1

Note. [a]Time of the apex and RT modes are expressed in seconds. Boldface: the two cases in which the majority of correct responses is represented by the second RT mode.

Furthermore, if responses depended on critical AUs changes, these changes must precede the mode of the RT distributions. Assuming that it takes approximately 100 ms to process novel visual stimuli and at least 100 ms to respond (ref?), the critical AUs for triggering a response should occur at least 200 ms before the corresponding modes. For each actor and each emotion, we recorded the intensity of the AUs active during the 5 frames (i.e. 200 ms of playback time) preceding the RTs modes (Table 5). The relationship between identification accuracy, RTs distributions, and facial activity for each of the six emotions is detailed below.

Disgust. Differences in accuracy among actors (see insets in Figure 2) were related to differences among the corresponding RT density distributions. The distribution had a single, well-defined peak for actors C and E, two distinct peaks for actor B, and was relatively shallow for actors A and D. In those last two cases, the cues which may be salient for perceptual identification (AU4: furrowing the brows, 10: raising the upper lip, and 15: pulling the lip corners down), appeared to be spread along the expressive sequence.

Fear. With the exception of actor D (P(Correct)= .86), the portrayal of Fear was somewhat ambiguous. All RT distributions showed a single, well-defined peak. In all but one case (actor E), the peak occurred when AU4 (furrowing the brows) and /or AU20 (stretching the lips) had reached a considerable intensity. The absence of one of these critical AUs, or an atypical relative phase among them seemed to affect recognition. For instance, because in actor E AU20 (stretching the lips) was very subtle and A4 (furrowing the brows) was absent, Fear was very often confused with Surprise (Table 3). Likewise, in actor B both AU4 (furrowing the brows) and AU20 (stretching the lips) were present and well detectable. However, the first part of activation sequence was misleading because of the presence of AU10 (raising the upper lip). AU5 (raising the upper lids), with (actor A and C) or without (actor B, D, and E) AU1 and 2 (raising the brows) is also

consistently associated with correct responses, suggesting that wide opened eyes are judged as a very prototypical component of the fear expression.

Table 5

AUs associated with correct responses

FE	Actor	AUs[a]	Common AUs[b]
Disgust	A	4,7,10,15	**4**[5] **10**[4] **15**[4] **7**[3]
	B	4,10	
	C	1,4,2,7,9,15	
	D	10,15,4,	
	E	7,4,15,1,10,17	
Fear	A	1,2,5, 4	**5**[5] **4**[4] **20**[3]
	B	5,4,20	
	C	1,2,26,25,5,4	
	D	4,5,20	
	E	5,20	
Sadness	A	4,7,1,15	**1**[5] **4**[5] **7**[3] **15**[3]
	B	1,4,7	
	C	1,15,4,17	
	D	1,11,43,	
	E	15,1,4,7,	
Happiness	A	25,6,12	**12**[5]
	B	12	
	C	6,12	
	D	12	
	E	12	
Surprise	A	1,2,5,38,26,	**1**[5] **2**[5] **5**[5] **25/26**[5]
	B	1,2,5,26	
	C	1,2,27,5,25	
	D	2,2725,5,1,	
	E	1,2,27,25,5	
Anger	A	24,4,5,7,17,	**4**[4]
	B	10,25,26,4,	
	C	23,25,4	
	D	25,4,9	
	E	25,26,9,23,7,17	

Note. [a]AUs: Action Units most involved in correct identification and active at the first peak of RT distributions. AUs are reported in order of activation. [b]Common AUs: Action Units present in the majority of portrayals (boldface). In brackets the number of portrayals

Sadness. The joint activation of AU4 (furrowing the brows), 7 (tightening the lower eyelids), and 15 (pulling the lip corners down) were relevant for correct identification. The slow

onset of the movement in four portrayals also proved helpful for identifying Sadness. The RTs peak in actor C was both sharp and early. By contrast, correct responses for actors A, B, D, and E were considerably spread along the movement sequence. In two cases the RTs peak occurred in correspondence of a complete (actor C) or almost complete (actor E) pattern. The presence in these portrayals of AU1 (raising the inner brow), 4 (furrowing the brows) and 15 (pulling the lip corners down) resulted in almost perfect recognition. In the other cases the peak occurred in correspondence of different cues (actor B: AU1, raising the inner brow, 4, furrowing the brows, and 7, tightening the lower eyelids; actor D: AU1, raising the inner brow, 11, which deepens the nasolabial furrow, and 43, which closes the eyes).

Happiness. For all actors the probability of correct recognition exceeded .90 and the RT distributions showed a single, well-defined peak. For actors A, B, and C most responses occurred after the full deployment of all AUs involved. For actors A, B, and C most responses occurred after the onset of all constituting AUs, although their intensity was not fully deployed. RTs differences among actors correlated with the onset of a small number of AUs that appear to be critical for correct identification. In all cases AU12 (pulling the lip corners up, producing the "smile") closely preceded the RTs mode. For actors A and C AU12 was activated almost simultaneously with AU6 (raising the cheeks, making wrinkles below and around the eyes). In the remaining cases AU6 was delayed and might have been responsible for later responses.

Surprise. Surprise was also generally well recognized (P(Correct)=.73). For actors A, C, D, and E, where AUs activation was fairly rapid and synchronous, most correct responses occurred in correspondence of AUs 1 (raising the inner brow), 2 (raising the outer brow), 5 (upper lid raise) and 26 (mandible lowered). Instead, in the portrayal of actor B the onset of AU1 and AU2 (raising the brow) preceded that of the other AUs and responses were delayed until AU5 (upper lid raise) reached a sufficient intensity.

Anger. The shape of RT distributions did not correlate with recognition accuracy, which was generally high (P(correct) = .80). In two cases there was a single, well-defined peak, which was associated either with AU4 (furrowing the brows) and AU9 (making wrinkles along the sides of the nose; actor D) or with AU23 (lips tightening), and 25 (parting the lips; actor C). For two other actors (B and E) a first peak correlated with early facial activity (actor B: AU4 (furrowing the brows), 10 (raising the upper lip), 25 (parting the lips) , 26 (jaw drop); actor E: AU7 (tightening the lower eyelids), 9, 23, 25), while a second peak corresponded to the subsequent activation of additional AUs (actor B: AU5, raising the upper lids, 9, making wrinkles around the nose, 16, which pulls the lower lip down, and 20, stretching the lips; actor E: AU4, furrowing the brows, 5, raising the upper lids, and 17, which pushes the chin boss and lower lip upwards).

Incorrect responses

Additional evidence that responses were triggered by the local structure of the sequences of AUs emerged from the analysis of incorrect responses. In many cases, when emotion A was mistakenly identified as emotion B the response occurred in correspondence of AUs clusters that are actually shared by A and B (not necessarily at the same place in the activation sequence). In the following, we consider two of the most common misidentifications[2].

Surprise/Fear. The FE of Surprise and Fear tended to be confused more often than any other pair of emotions (Table 3), because of the similarity between the corresponding AUs profiles (Table 2). The timing of the misidentifications revealed a regular pattern across actors (Table 6). The upper and lower parts of the table correspond to Surprise and Fear portrayals, respectively. For each expression, and for both correct and wrong responses we indicated the AUs active before the peak of the RT distribution, the timing of peak, and the response frequency.

[2] Another major group of confusions involved Disgust, Anger, and Sadness. Disgust was the most ambiguous emotion (across actors: P(correct)=.57), the most frequent confusion being with Sadness (P(Sadness/Disgust)=.33). Also in this case, cues shared by different emotions were mostly responsible for these confusions. Due to space limitations, these confusions will not be detailed here. Additional material is available online.

Table 6

Analysis of Surprise/Fear misidentifications

Portrayal/Response	Actor	AUs intensity differences [a]	RT mode (s) [b]		P(Response) [c]		K-S [d]
			Surprise	Fear	Surprise	Fear	
Surprise/Fear	A	1[+1] 2[+1] **5**[+1]	6.0	9.3	.80	.18	.096
	B	**5**[+1]	19.2	22.2	.49	.41	<.001
	C	**5**[+1]	6.3	7.6	.60	.39	.001
	D	2[+1] **5**[+1]	6.2	9.0	.88	.11	.002
	E	1[0] 2[0] 5[0]	7.6	8.5	.88	.09	.082
			Surprise	Fear	Surprise	Fear	
Fear/Surprise	A	1[-1] 2[-1] **5**[-1] **4**[-1]	6.1	9.9	.32	.65	<.001
	B	5[0] **4**[-1] **20**[-1]	10.8	11.5	.08	.53	.246
	C	**5**[-2] **4**[-2]	8.0	11.2	.38	.58	<.001
	D	**4**[-1] **5**[-1] **20**[-1]	5.1	8.3	.04	.86	<.001
	E	5[0] 20[0]	11.1	11.3	.56	.42	0.138

Note. [a] For each AU involved in misidentifications is indicated (square brackets) the difference in intensity between the level of activation at the time of wrong and correct responses. Boldface: critical AUs for response choice. [b] RT modes for the indicated responses. [c]Response probabilities. [d]K-S: Significance of the difference between RT distributions for correct and wrong responses (Kolmogorov-Smirnoff test).

Surprise → “Fear”. When the stimulus was Surprise, the RTs peak for the wrong response “Fear” occurred after the peak for the correct response (an example in Figure 3). In four actors (actors A, B, C, and D) the facial activity related to the later (Fear) peak was similar to that related to the earlier (Surprise) peak, the main difference being a higher intensity of some AUs. In particular, it appears that, at a high intensity, AU5 (raising the upper eyelids) elicits the perception of Fear. Thus, the reason of confusion seems to be that observers occasionally delayed the response because evidence for Surprise was not deemed sufficient. When the response was finally given, the intensity of the active units had reached a higher level, which elicited the confusion with Fear.

---------- *Inserire Figura 3* ----------

Figure 3. RT and AUs profiles. Results for one representative portrayal. Upper part: Time course of the intensity of the indicated AUs activated by the portrayal. Sampling interval: 20 ms. Lower part: RT density distribution for correct responses estimated by differentiating the analytical interpolation of the cumulative distribution (see Figure 2). Numbers indicate the AUs intensities corresponding to the single well-defined RT mode. The two lower panels contrast the RT density distributions for correct (upper) and wrong (lower) responses to a “Surprise” portrayal. On average, wrong identifications of the portrayal as “Fear” occurred later than correct responses, when active AUs had reached top intensity.

Fear → “Surprise”. The results for the opposite confusion emphasize further the role of features in triggering the identification process. On average, correct responses occurred later than wrong ones. The delay varied across actors, but correct RTs always depended on facial activity reaching a higher level of activation (Fear-related AUs were 1-level more intense when the response was “Fear”). In most cases the RTs peak for the wrong response “Surprise” (Figure 4) was correlated with the same cluster of AUs, i.e., 1 and 2 (raising the brows), 5 (raising the upper lid), 25, 26/27 (opening the mouth at various degrees) that triggered correct identification of

Surprise (Figure 3). Apparently, confusion occurred when facial actions shared by both FE were wrongly interpreted as sufficient activity for a full-blown Surprise expression rather than the early unfolding of the Fear expression. The fact that RT peaks associated with correct responses were delayed suggests that identifying Fear relies crucially on the detection of later cues associated only to that emotion (AU4, furrowing the brows, and 20, stretching the lips).

Figure 4. RT and AUs. Same format as Figure 3. The two panels contrast the RT density distributions for correct (upper) and wrong (lower) responses to a “Fear” portrayal (same actor as in

Figure 4). On average, wrong identifications of the portrayal as "Surprise" occurred before than correct responses, when active AUs had not reached yet top intensity.

Discussion

We investigated the relative weight of configural and component-based cues for the identification of emotional expressions. To address the issue we used a set of high-speed video-recordings of the transition between a neutral face and the posed, full-blown expressions corresponding to six basic emotions. The recordings were coded frame by frame with FACS to describe in detail time course and intensity of the facial components of the expression. Slow-motion versions of the recordings were then presented as stimuli in a perceptual task asking observers to identify the unfolding emotion as soon as they felt confident to do so. The nature of the cues relevant for identification was investigated by correlating response time distributions for correct and incorrect categorizations with facial activity. With regard to the components/configuration debate, the results suggest that emotion identification implies both an independent evaluation of expressive components and the cumulative integration of the components into an overall configuration.

Response times and identification accuracy

As in previous studies (Biehl et al., 1997; Ekman, 1994; Galati et al., 1997; Wallbott & Scherer, 1986) identification accuracy depended on both the actor and the portrayed emotion. Mean recognition rates over our large sample of observers (76%, see Table 3) were somewhat lower than those reported by Ekman (1994; 83%), and Biehl et al. (1997; 83%), but definitely higher than those reported by Galati et al. (1997; 51%) and by Wallbott and Scherer (1986; 63%). There may be several reasons for these discrepancies, including the choice of the encoders and the portrayal selection process. For instance, the very high correct recognition rates reported by Biehl et al. (1997) were obtained by selecting carefully among a set of intense portrayals whose

ecological validity has been questioned (Carroll & Russell, 1997). On the other side, the low recognition rates (e.g. 17% for Fear) in Galati et al. (1997) study may be accounted for by the use of untrained encoders instead of professional actors.

Distributions of correct responses tended to cluster around one or two modes, well before the expression reached the apex (see Table 4). In some cases only a subset of the action units was deemed sufficient to reach a decision; in other cases all units were active, but had not yet reached full intensity. Thus, the main question is the nature of the partial information permitting identification. In keeping with previous findings (e.g. Kohler et al., 2004, Nusseck et al., 2008, Schyns et al., 2007), AU12 (pulling the corners of the lips up) appears as the necessary and sufficient feature for identifying Happiness. This exception aside, RTs modes were never preceded by the activation of a single AU, suggesting that more than just one facial component is required to trigger a response. Thus, in general, the identification process is best described as an on-line integration of several increasingly active AUs converging towards a perceptual solution. As also suggested by the recent findings of Schyns et al. (2007), we assume that the temporal integration of information stops as soon as the diagnostic features for judging a particular expression have been integrated. If so, the AUs active shortly before RTs modes may be considered as the minimum amount of information sufficient for perceptual identification.

This minimum amount cannot be identified uniquely because actors did not activate the same AUs for communicating a given emotion. However, certain consistent patterns emerged from the analysis of correct answers. For instance, Sadness expressions including AU1 (raising the inner brows), 4 (furrowing the brows), 15 (pulling the corners of the lips down) yielded more accurate (actor C) or faster (actor E) identifications than expressions where AU15 was activated later (actor A and B), or expressions containing contradictory cues (e.g. AU12 in actor A). Likewise, in Anger AU4 (furrowing the brows) and AU9 (making wrinkles around the nose) were consistently

associated either with one (actor B) or two (actor D and E) RT peaks, depending on their phase difference.

Responses were generally delayed until the crucial AUs for excluding potential alternatives reached a sufficient intensity, suggesting that AUs maintain their identity during the perceptual integration process. This was the case of AU5 (raising the upper eyelids), which is particularly salient in Surprise. Correct responses were often associated with a broad combination of AUs that, in addition to AU5, also included AU1 (raising the inner brow) and AU2 (raising the outer brow). Yet, in actor B the RTs mode was clearly delayed until AU5 reached a sufficiently high level of activation, as if the intensity of this unit was the crucial discriminating cue.

AUs onset hardly ever correlated with a RT mode. Rather the major determinants of the response were the intensity profiles of all relevant AUs and their relative phase. For example, the most common Fear expression (actors B, C, and D) was characterized by an early activation of AU1, 2 (raising the brows), and 5 (raising the upper lid) followed by the critical emergence of AU4, (furrowing the brows) and/or AU20 (stretching the lips). Conversely, a delayed onset of AU20 with respect to AU4 (as in Actor C) prevented a joint evaluation of the two AUs, and might have affected identification accuracy. Also, an excessive delay between AU4, 20 and all the preceding activity (Actor A) seems detrimental for recognition, as observers tended to respond before the onset of these units.

Overall, the AUs that correlated most consistently with RT peaks for correct responses were also those that are often associated with the selected emotions, namely AU4 with Anger, Disgust, Fear and Sadness (Smith & Scott, 1997), AU 1, 2, 5 with Surprise, and Fear, AU9 and 10 with Disgust, AU15 with Sadness, AU 20 with Fear (Smith & Scott, 1997; Scherer & Ellgring, 2007; Ekman et al., 2002).

Misidentifications

We take up the question of why confusion arises by focussing on systematic errors, or "common confusions" (Tomkins & McCarter, 1964). Several causes, including response biases, cognitive biases, and intrinsic similarities in the portrayals may be responsible for these errors. Based on judgment studies of posed photographs, Tomkins and McCarter (1964) concluded that some confusions (e.g. Fear with Surprise and Anger with Disgust) are significantly more stable than others. Moreover, they reported that certain confusions are not reversible (e.g. Fear is much more frequently mistaken for Surprise than the opposite). To explain this pattern Tomkins and McCarter (1964) emphasized response biases as a source of errors by arguing that some affects tend to be denied because they imply danger and are misidentified as affects that are similar but "safer". By the same token, "safe" affects would not be confused with the tabooed ones. An opposite argument led de Bonis (2002) to emphasize instead cognitive biases. Arguably, when one is confronted with ambiguous stimuli, a false alarm (mistaking a "safe" affect for a threatening one) is more adaptive in terms of evolutionary advantage than overlooking a potential danger (Öhman, 1986). Over and above response biases and cognitive factors, a significant role of intrinsic similarity is suggested by the fact that the highest rate of confusion was observed for the pair Surprise/Fear. Indeed, as already noted by Darwin (1872-1998, p. 305-306), Surprise and Fear may be construed as laying on a common continuum, Fear being often preceded by (and sometimes mixed with) Surprise or astonishment, with which it shares the common element of startle and physiologic arousal.

The question then is whether similarity between two expressions can be analyzed in terms of shared identifiable components. Evidence that sets of AUs are preferentially involved in the perception of FE emerged from experiments with chimerical faces. Several studies (Bassili, 1979; Calder, Young, Keane, & Dean, 2000; Katsikitis, 1997) demonstrated that the features that are most relevant for identification are located in the upper part of the face for some expressions and in the bottom part for others. Specifically, in the case of Fear the upper face is most relevant (Calder

et al. 2000). An intense AU5 (Wide open eyes) often elicits the Fear response, even when the lower face is morphed into a different expression (Fiorentini & Viviani, 2009). Actually, merely displaying against a uniform black background the white of the eyes of either fearful or happy faces is sufficient to modulate amygdala responsiveness (Whalen et al., 2004). In keeping with this, Adolphs et al. (2005) reported the case of a patient with bilateral amygdala damage who, because of the inability to make normal use of information from the eye region, showed a selective impairment in recognizing the fear expression.

The analysis of RTs supported the hypothesis that the timing with which identifiable AUs subsets are activated is also a significant source of confusion. We found that errors tended to occur at different times with respect to correct responses, and also at different times for opposite confusions. In particular, responses "Surprise" to a Fear expression occurred on average before correct responses (Figure 3). We suggest that errors were triggered by the critical AUs for Surprise (i.e., AU1,2, 5, ±25, 26) occurring early on during the unfolding of the expression, before the activation of the telltale cues for Fear (AU4 and/or AU20). If so, it follows that these early components of the Fear expression can indeed be isolated perceptually. The relative timing of AUs subsets within the sequence leading to a full-blown Fear expression also explains why wrong responses "Fear" to a Surprise portrayal tended to be delayed with respect to correct responses (Figure 2). The results for individual actors indicate in fact that responses were delayed when observers failed to perceive the activation of the critical group AU1,2, 5 as sufficiently to signal Surprise, and took that group into account only later in the sequence, when its activation was strong enough to induce confusion with Fear. Insofar as individual AU are assumed to carry functional meaning (Scherer & Ellgring, 2007), perceptual confusions might reflect a more fundamental similarity between the underlying experiences.

Methodological concerns

The experimental strategy adopted by our experiment may rise at least two concerns. First, one might question the ecological validity of posed expressions in emotion research. Even when actors make an effort to self-induce emotions, it has been argued (e.g. Carroll & Russell, 1997) that the expression is still significantly different from that elicited by a true emotional antecedent. Moreover, natural expressive variability among individuals (how quickly the expression is produced, how many AUs are activated, their degree of synchronization) may be enhanced in the case of actors using different strategies to appear emotional. Both points have merits. However, several authors (e.g. Banse & Scherer, 1996; Galati, Scherer, & Ricci Bitti, 1997; Gosselin, Kirouac, & Doré, 1995; Nusseck et al., 2008; Scherer & Ellgring, 2007) have expressed confidence that, with adequate training, actors manage to produce instances of emotional expressions that are sufficiently stable and realistic. Consequently, in our experiment we adopted a technique of emotion self-induction that has been successfully used in previous studies (Bänziger, Pirker, & Scherer, 2006). More importantly, also at normal speed, video-recording spontaneous expressions of emotions in a laboratory setting has proven very difficult, leading some authors (e.g. O'Sullivan, 1982) to argue that "the use of only spontaneous expressions of emotions is neither desirable nor likely" p.?. The difficulty is even greater when a fine-grained temporal description of facial actions is desired because recording at high-speed involve lighting and synchronization constraints that cannot be satisfied if one attempts to record spontaneous expressions. In conclusion, while we acknowledge the limits of the use of actors, it appears that, at least when quantitative psychophysical measures are involved, the available recording technology leaves no alternative to this method. At the same time, we also believe that our main findings are not significantly biased by the use of posed rather than natural expressions.

The second concern is the temporal scaling of the stimuli. Because the scaling was uniform, the relative timing of the dynamic events was the same as in the original expression. However, the absolute timing was quite different, and might have affected the performance. It has

been claimed (cf Introduction) that dynamic displays of FE are better identified than still pictures because motion cues help disambiguate the expression (e.g. Ambadar, 2005). If so, one could also suspect that such "dynamic advantage" occurs only if movement velocity falls within the natural range. Conversely, velocities much lower than normal might be detrimental for identification. Specifically, some motion cues available in real-life may fail to be detected in our experiment because the kinematics of the stimuli was severely slowed down. Two observations speak against this possibility. First, a recent psychophysical study (Fiorentini & Viviani, 2010) has questioned the validity of the dynamic advantage hypothesis. The study failed to detect any significant difference in identification performance between recordings of actual expressive movement and static pictures in which, so to speak, the entire temporal evolution of the expression is collapsed onto the final (apex) configuration. Second, the identification rates measured in our experiment are in the upper range of values reported so far using static pictures as stimuli. Of course, we do not claim that the relationship between AUs activation and RT distributions in our experiment is a faithful, scaled down description of the decision process that takes place in real time. However, the fact that the pattern of correct responses and confusions across emotions is in keeping with previous results, suggests that the observed order relationships and the associated causal link between AU and responses captures some significant aspect of that process.

Conclusions and speculations

To our knowledge, we have presented the first detailed description of the time course with which the components of the facial expression of the basic emotions are activated and converge towards the apex configuration. The crucial point emerging from this analysis is that, contrary to some views (Ekman, 2003), facial expressions are definitely not generated by the synchronous release of pre-packaged array of motor commands. Instead, many facial components follow a distinct, fairly independent and asynchronous temporal evolution. Therefore, the necessary conditions are satisfied for testing experimentally the main question addressed in this study,

namely whether emotional expressions are perceived as a global configuration or with a bottom-up process in which individual cues are integrated progressively as they become available until sufficient diagnostic information is accumulated for a reliable identification. Generally speaking, the results favoured the second hypothesis. The observation that, in most cases, responses tended to cluster around distinct and identifiable phases of the unfolding facial action rather than near the end of the movement suggests that perceptual identifications (both correct and incorrect) are triggered by the relative time course of a diagnostic set of AUs. It should be stressed, however, that not all AUs identified by FACS evolved independently. It is then possible that certain synergies are indeed activated as packages, and that their contribution to identification cannot be factored out.

The neural mechanisms that implement the integration of the cues and lead to a response are largely unknown. However, the concept of implicit motor competence affords a promising line of speculation. It is well established that observing a specific biological action activates the neural systems in motor and parietal areas that would subserve the same action (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fogassi, & Gallese, 2001). The fact that the visual representation of a gesture, so to speak, "resonates" in the motor system has been evoked to explain why biological movements are perceived with greater accuracy than non-biological ones (Viviani, 2002). More specifically, the observation that lesions of the sensory-motor areas impair the ability to rate and name facial emotions (Adolphs, 2002; Adolphs, Tranel, & Damasio, 2003) points to the involvement of the sensory-motor system in emotion recognition. If indeed experiencing an emotion and observing the FE of the same emotion both activate a shared neural substrate, the concept of resonance may also provide the basis for a theory of emotion identification (Gallese, 2005, 2007; Goldman & Sripada, 2005). The theory holds that we interpret the overt manifestation of an emotion by an unconscious mental attribution, i.e. by instantiating, undergoing, or experiencing in the mind the emotional state that, according to our own competence, is responsible for that manifestation.

According to the above line of speculation, the pattern of correlations between RT distributions and response accuracy would reflect the attuning between actors and observers. In particular, 1) RT clusters would indicate the phases in the internal simulation when motor resonance attains a threshold; 2) high rates of misidentification (as in the case of Fear → "Surprise") would occur because the ambiguity of the visual configuration maps into an ambiguity of the simulation; 3) differences in RT modes for correct and wrong identifications would suggest that the same input may give rise to different resonances with different time courses.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1,* 21-61.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433,* 68-72.

Adolphs, R., Tranel, D., & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain and Cognition, 52,* 61-69.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. Psychological Science,16(5), 403-410.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology. 70,* 614-636.

Bänziger, T., Pirker, H., & Scherer, K. (2006). Gemep - geneva multimodal emotion portrayals: a corpus for the study of multimodal emotional expressions. In L. Deviller et al. (Ed.), *Proceedings of LREC'06 Workshop on Corpora for Research on Emotion and Affect* (pp. 15-019). Genoa. Italy.

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.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21*, 3-21.

Bimler, D. L., & Paramei, G. V. (2006). Facial-expression affective attributes and their configural correlates: Components and categories. *Spanish Journal of Psychology, 9,* 19-31.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews, Neuroscience, 6*, 641–651.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 551.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3,* 81-117.

Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guerit, J. M. (2002). Categorical perception of happiness and fear facial expressions: An ERP study. *Journal of Cognitive Neuroscience, 14,* 210-227.

Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology, 72,* 164-176.

Darwin, C. (1998*). The expression of the emotions in man and animals*. (3rd ed.). New York*:* Oxford University Press. (Original work published 1872).

de Bonis, M. (2002). Causes and reasons in failures to perceive fearful faces. In M. Katsikitis (Ed.), *The Human Face: Measurement and Meaning* (pp. 149-167). Norwell, Massachusetts: Kluwer Academic Publishers.

de Bonis, M., De Boeck, P., Pérez-Diaz, F., & Nahas, M. (1999). A two-process theory of facial perception of emotions. *C.R. Acad. Sci. III, 322,* 669-675.

Duchenne, B. (1999). *The mechanism of human facial expression*. Cambridge, UK: Cambridge University press. (Original work published 1876).

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation 1971*, (Vol. 19, pp. 207-283). Lincoln, NE: University of Nebraska Press.

Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In A.S.R.Manstead & H. L. Wagner (Eds.), *Handbook of social psychophysiology* (pp. 143-164). Chichester, England: Wiley.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169-200.

Ekman, P. (1994). Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. *Psychological Bulletin, 115,* 268-287.

Ekman, P. (2003). *Emotions revealed.* New York: Times Books.

Ekman, P. (2004). Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. NY: Owl Books.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V., & Hager, J. C. (2002) The Facial Action Coding System (2nd ed.). Salt Lake City, UT: Research Nexus eBook.

Ellison, J. W. & Massaro, D. W. (1997). Featural evaluation, integration, and judgment of facial affect. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 213-226.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44,* 227-240.

Fiorentini, C., & Viviani, P. (2009). Perceiving facial expressions. *Visual Cognition, 17,* 373-411.

Galati, D., Scherer, K. R., & Ricci-Bitti, P. E. (1997). Voluntary facial expression of emotion: comparing congenitally blind to normal sighted encoders. *Journal of Personality and Social Psychology, 73,* 1363-1380.

Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences, 4,* 23-48.

Gallese, V. (2007). Before and below "theory of mind": embodied simulation and the neural correlates of social cognition. *Phil. Trans. R. Soc. B, 362,* 659-669.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119,* 593-609.

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition, 94,* 193-213.

Gosselin, P., Kirouac, G., & Doré, F. (1995). Components and recognition of facial expression in the communication of emotion by actors. *Journal of Personality and Social Psychology, 68,* 83-96.

Hanawalt, N. G. (1944). The role of upper and lower parts of the face as the basis for judging facial expressions: II. In posed expressions and "candid camera". *Journal of General Psychology, 31,* 23-36.

Katsikitis, M. (1997). The classification of facial expression of emotion: a multidimensional-scaling approach. *Perception, 26,* 613-626.

Kohler, C. G., Turner, T., Stolar, N., Bilker, W., Brensinger, C. M., Gur, R. E., et al. (2004). Differences in facial expressions of four universal emotions. *Psychiatry Research*, *128*, 235–244.

Matsumoto, D., & Ekman, P. (1988). Japanese and Caucasian facial expressions of emotion (JACFEE) and neutral faces (JACNeuF). San Francisco, CA: Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University.

Morris, J. S., de Bonis, M., & Dolan, R. (2002). Human amygdala responses to fearful eyes. *Neuroimage, 17,* 214-222.

Nusseck, M., Cunningham, D. W., Wallraven, C., & Bülthoff, H. H. (2008). The contribution of different facial regions to the recognition of conversational expressions. Journal of Vision, 8, 1-23.

Öhman, A. (1986). Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analysis of emotion. *Psychophysiology, 23,* 123-145.

Ortony, A., & Turner, T. (1990). What's basic about basic emotions? *Psychological Review, 97,* 315-331.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Neuroscience Reviews, 2,* 661-670.

Russell J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin. 115,* 102-141.

Russell, J., & Fernandez-Dols, J. M. (1997). The psychology of facial expression. New York: Cambridge University Press.

Scherer, K. R. (1992). What does facial expression express? In K. T. Strongman (Ed.), *International Review of Studies on Emotion* (Vol. 2, pp. 139-165). Chichester, England: Wiley & Sons Ltd.

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion, 7,* 113-130.

Schwaninger, A., Wallraven, C., Cunningham, D. W., & Chiller-Glaus, S. D. (2006). Processing of facial identity and expression: A psychophysical, physiological, and computational perspective. *Progress in Brain Research, 156*, 321–343.

Schyns, P., Petro, L. S., & Smith, M. L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Current Biology. 17,* 1580-1585.

Schyns, P. G., Petro, L. S., & Smith, M. L. (2009). Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: behavioral and brain evidence. *Plos One. 4,* 1-16.

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

Smith, C. A., & Scott, H. S. (1997). A componential approach to the meaning of facial expressions. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 229-254). New York: Cambridge University Press.

Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills, 18,* 119-158.

Viviani, P. (2002). Motor competence in the perception of dynamic events: a tutorial. *Attention and Performance, 19,* 406-442.

Wagner, H. L. (1997). Methods for the study of facial behaviour. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 31-56). New York: Cambridge University Press.

Wallbott, H.G., & Ricci-Bitti, P. (1993). Decoders' processing of emotional facial expression—a topdown or bottom-up mechanism? *European Journal of Social Psychology*, 23, 427–443.

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51,* 690-699.

Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S. et al. (2004). Human amygdala responsivity to masked fearful eye whites. *Science, 306,* 2061.

Author note

This research was supported by FNS Grant #100014-112252 to P.V. We are grateful to Dr. Susanne Kaiser for her helpful advice in all phases of the study.