

Is there a dynamic advantage for facial expressions?

Chiara Fiorentini

Institute of Child Health, University College London, UK



Paolo Viviani

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



Some evidence suggests that it is easier to identify facial expressions (FEs) shown as dynamic displays than as photographs (dynamic advantage hypothesis). Previously, this has been tested by using dynamic FEs simulated either by morphing a neutral face into an emotional one or by computer animations. For the first time, we tested the dynamic advantage hypothesis by using high-speed recordings of actors' FEs. In the dynamic condition, stimuli were graded blends of two recordings (duration: 4.18 s), each describing the unfolding of an expression from neutral to apex. In the static condition, stimuli (duration: 3 s) were blends of just the apex of the same recordings. Stimuli for both conditions were generated by linearly morphing one expression into the other. Performance was estimated by a forced-choice task asking participants to identify which prototype the morphed stimulus was more similar to. Identification accuracy was not different between conditions. Response times (RTs) measured from stimulus onset were shorter for static than for dynamic stimuli. Yet, most responses to dynamic stimuli were given before expressions reached their apex. Thus, with a threshold model, we tested whether discriminative information is integrated more effectively in dynamic than in static conditions. We did not find any systematic difference. In short, neither identification accuracy nor RTs supported the dynamic advantage hypothesis.

Keywords: facial expressions, dynamic advantage, emotion, identification

Citation: Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision*, 11(3):17, 1–15, <http://www.journalofvision.org/content/11/3/17>, doi:10.1167/11.3.17.

Introduction

Research on emotion recognition has relied primarily on static images of intense facial expressions (FEs), which—despite being accurately identified (Ekman & Friesen, 1982)—are fairly impoverished representations of real-life FEs. As a motor behavior determined by facial muscle actions, expressions are intrinsically dynamic. Insofar as detecting moment-to-moment changes in others' affective states is fundamental for regulating social interactions (Yoshikawa & Sato, 2008), visual sensitivity to the dynamic properties of FEs might be an important aspect of our emotion recognition abilities.

There is considerable evidence that dynamic information is not redundant and may be beneficial for various aspect of face processing, including age (Berry, 1990), sex (Hill & Johnston, 2001; Mather & Murdoch, 1994), and identity (Hill & Johnston, 2001; Lander, Christie, & Bruce, 1999; see O'Toole, Roark, & Abdi, 2002 for a review) recognition. In real life, static information—such as the invariant geometrical parameters of the facial features—and dynamic information describing the contraction of the expressive muscles are closely intertwined and contribute jointly to the overall perception. The relative contribution of either type of cues, which is likely to depend on the meaning that one is asked to extract from the stimulus, is still poorly understood. Pure motion information is sufficient to recognize a person's identity and sex (Hill &

Johnston, 2001). Other studies have shown that face identity is better recognized from dynamic than static displays when the stimuli are degraded (e.g., shown as negatives, upside down, thresholded, pixilated, or blurred). However, the advantage disappears with unmodified stimuli (Knight & Johnston, 1997; Lander et al., 1999). In short, insofar as recognition of identity from complete static images is already close to perfect, motion appears to be beneficial only when static information is insufficient or has been manipulated (Katsiri, 2006; O'Toole et al., 2002).

In comparison to face identity, fewer studies have investigated the role of dynamic information in FE recognition (see Katsiri, 2006, for a review). Taken together, they seem to suggest that the process of emotion identification is facilitated when expressions are dynamic rather than static. However, because of various methodological issues and conceptual inconsistencies across studies, this suggestion needs to be qualified. We can divide the available studies in three main groups.

First, there are studies showing that dynamic information improves expression recognition in a variety of suboptimal conditions, i.e., when static information is either unavailable or is only partially accessible. As in the case of identity recognition, emotions can be inferred from animated point-light descriptions of the faces that neglect facial features (Bassili, 1978, 1979; see also Bruce & Valentine, 1988). Furthermore, in various neuropsychological and developmental conditions, there is evidence that dynamic presentation improves emotion recognition with

respect to static presentation. For instance, Humphreys, Donnelly, and Riddoch (1993) reported the case of an agnostic patient who was almost unable to categorize FEs from static pictures but instead was quite proficient with dynamic point-light displays. Children with psychopathic tendencies have selective impairments in identifying emotion from morphed FEs presented in slow motion (Blair, Colledge, Murray, & Mitchell, 2001). Conversely, children with autism seem to benefit from slow motion information during FE categorization (Gepner, Deruelle, & Grynfeldt, 2001; Tardif, Lainé, Rodriguez, & Gepner, 2007). In summary, the studies of this first group suggest that dynamic information may be sufficient for emotion recognition and has a compensatory role when static information cannot be adequately accessed.

A second group of studies has demonstrated that humans are sensitive to the temporal properties of FEs, the effect being contingent on the required judgments. Observers can reconstruct the actual progression of an expression from scrambled sets of pictures describing different stages of the expression, even when the original sequence has not been shown and consecutive frames contain only subtle transitions in expressions (Edwards, 1998). Moreover, performance improves under time constraints, suggesting that the extraction of dynamic cues in FEs occurs relatively automatically. Two studies demonstrated that the speed of emotion unfolding is an important factor influencing the perception of FEs (Kamachi et al., 2001; Sato & Yoshikawa, 2004). Kamachi et al. (2001) simulated dynamic FEs through morphed animations transforming a neutral face into the full-blown expression of a basic emotion (i.e., happiness, anger, sadness, surprise). The effect of animation speed on identification was found to depend on the emotion. Sadness was identified more accurately from slow sequences, happiness and, to some extent, surprise from fast sequences, and anger from medium-speed sequences. Sato and Yoshikawa (2004) found that the velocity at which an expression is perceived as unfolding “most naturally” differs according to the emotion. In particular, surprise is judged most natural at fast speed, whereas sadness and fear are judged most natural at slow speed.

All studies summarized so far indicate that dynamic information is a significant component of our representation of FEs. However, they do not provide evidence that bears directly on the dynamic advantage hypothesis. Such evidence emerges instead from a third group of studies comparing emotion identification under static and dynamic conditions. Harwood, Hall, and Shinkfield (1999) tested identification in healthy observers and patients with mental retardation. Dynamic stimuli were twelve FEs of the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) posed with average intensity by one male actor and one female actor and recorded from neutral through the expression’s completion; static stimuli were the same displays frozen at the apex. Both groups identified

sadness and anger significantly better from dynamic than from static displays, but the advantage was limited to just these two emotions. Wehrle, Kaiser, Schmidt, and Scherer (2000) investigated whether the specific ways in which FEs unfold influence emotion identification. Using computer-generated simulations of ten different emotions at low and high intensity level, they contrasted static displays with two types of dynamic displays. In one case, facial actions unfolded sequentially, according to the predictions of the Component Process Model (CPM) developed by Scherer (1984). In the other case, facial actions unfolded simultaneously. Identification accuracy was better with dynamic than with static displays, but there was no difference between the two dynamic conditions. Thus, dynamic advantage could not be attributed to the temporal characteristics of the expression.

Ambadar, Schooler, and Cohn (2005) demonstrated that FEs of slight intensity are recognized better from dynamic displays depicting the emergence of the expression rather than from static displays of the final expression. The original facial displays were the FEs of the six basic emotions, posed by twenty-nine untrained students and recorded at 25 frames/s (Kanade, Cohn, & Tian, 2000). The actual FE stimuli were brief movement sequences (3–6 video frames, 100- to 200-ms duration) obtained by truncating the movement before the apex, so that they ended at the first visible display of the expression. By manipulating motion in several ways, the authors ruled out a number of potential explanations of the observed dynamic enhancement. In particular, sampling density was not likely to be a significant factor because dynamic multi-frame sequences failed to produce benefits comparable to those of a truly dynamic condition. Moreover, displays including only the first and last frames of each sequence were identified as accurately as true dynamic displays. Thus, dynamic advantage could not derive from the unique temporal characteristics of each expression. Based on these analyses, Ambadar et al. concluded that the critical advantage afforded by dynamic displays is to enable participants to perceive the “direction” in which FEs have been changing.

Finally, as mentioned above, the study of Kamachi et al. (2001) focused on the effect of presentation speed rather than the comparison between dynamic and static stimuli. Nevertheless, their results suggest that dynamic stimuli are identified less reliably than static ones, especially sadness at high speed and anger at low speed. The authors argue that the reason for the difference is that dynamic stimuli starting from a neutral face showed a recognizable emotional expression for a shorter time than their static versions (see also Katsiri, 2006).

Reconciling the results of these studies is difficult because of methodological differences and limitations. The main limitation concerns the nature of the dynamic stimuli. Simulating movement by morphing a neutral face into a full expression allows for speed control as in

Kamachi et al. (2001), but there is no guarantee that the time course of the simulated facial actions is anywhere close to the time course of the real ones (Katsiri, 2006; Sato & Yoshikawa, 2004). A significant difference among the studies mentioned above concerns the richness and intensity of the stimuli. Wehrle et al. (2000) used synthetic expressions that, unlike real ones, were hard to identify in the static condition, perhaps because the facial expression model was underspecified. In fact, the possibility to generalize their results to natural faces has been questioned (Katsiri, 2006). In addition, the stimuli tested by Ambadar et al. (2005) were somewhat underspecified insofar as they represented emotional FEs of subtle intensity. The fact that both studies reported a dynamic advantage effect may be taken to suggest that the advantage emerges only when static stimuli are ambiguous and difficult to identify. If so, however, it would be difficult to account for the results of Harwood et al. (1999) and Kamachi et al. (see also Katsiri, 2006), both of which used unambiguous static images. Differences between the tasks in previous studies are also worth mentioning. In the study by Ambadar et al., participants were encouraged to watch repeatedly each recording and to label the expression by choosing among seven options (the six emotions examined plus neutral). Kamachi et al. showed each clip only once for its full duration and asked participants to rate its intensity in all of four emotional categories (happy, sad, surprised, and angry) on a 7-point scale. In Wehrle et al., after watching each stimulus, participants chose the most appropriate emotion label from ten available options and rated its intensity on a 5-point scale. Finally, Harwood et al. asked participants to judge each stimulus by choosing one of the six possible emotions, presented either as verbal labels (i.e., anger, fear, etc.) or as pictorial representations (e.g., a clenched fist for anger, a snake for fear). In conclusion, the above studies differ considerably as far as the stimuli are concerned (number and kind of posers, number of emotions portrayed, stimulus format, and duration).

Overall, the picture emerging from the few studies that addressed directly the dynamic advantage hypothesis is ambivalent. Clear evidence that naturally unfolding dynamic FEs are identified more effectively than static portrayals of the expression apex is still lacking. The present study is an attempt to test the dynamic advantage hypothesis with a different task and a new method for generating dynamic stimuli. Briefly, we recorded with high temporal resolution the unfolding of several posed expressions from a neutral face to the apex and used a custom morphing algorithm to generate graded blends between pairs of dynamic FEs. These blends are then used as stimuli in a classical two-alternative, forced-choice (2AFC) identification task. The key advantage of this technique with respect to previous studies is that the blending process applies simultaneously to the facial features and to their evolution in time, preserving in all cases the realism of the stimuli. Measures of performance with dynamic stimuli are then contrasted with those obtained with static stimuli,

which are again graded blends of the apexes of the same pair of expressions. In addition to comparing response accuracy, we extend the contrast between static and dynamic conditions by considering also the corresponding response times. If the dynamic advantage is real, it should emerge from at least one of these contrasts.

Methods

Participants

Fifteen students from the Faculty of Psychology at the University of Geneva (6 males; 9 females, age range: 20–35) volunteered for the experiment. All participants had normal or corrected-to-normal vision. The experiment was approved by the Ethical Committee of the University.

Stimuli

Facial expression prototypes

Prototype stimuli were 11 dynamic video clips, chosen from a sample validated in a previous study (Fiorentini, Schmidt, & Viviani, [manuscript in preparation](#)), each displaying the unfolding of a facial expression from neutral to apex. The selected video clips displayed the expressions of happiness, anger, fear, sadness, surprise, and disgust, all posed by professional actors (1 male, 2 females) identified as A, B, and C. Prototype selection was made by choosing for each actor the recordings that were best recognized according to previous results (Fiorentini et al., [manuscript in preparation](#)) and/or most suitable to morphing manipulation. Each actor contributed with various expressions (actor A: Anger, Fear, Sadness; actor B: Happiness, Sadness, Disgust; actor C: Anger, Fear, Surprise, Happiness, Disgust), and for all emotions except Surprise, two portrayals of two different actors were used.

Each frame (TIFF format) of the recordings was processed with Photoshop CS3 to equalize overall luminance, contrast, and chromatic spectrum and scaled to a standard dimension (864 × 1074 pixels). Sequences were cut at the minimum length (500 frames) that accommodated the longest movement (Happiness by actor C). Sequences began with 20 frames showing the neutral face before the first noticeable muscle contraction and ended with the expression apex. Copies of the last frame representing the apex of the expression replaced the return phase to the neutral. The length of this final portion varied across actors and emotions.

Expression blends

The actual stimuli used in the experiment were graded blends between two prototype expressions of the same actor, either in the form of morphed pictures (static

condition) or of morphed video clips (dynamic condition). Prototypes were paired by taking into account: (1) the available prototypes for each actor, and (2) how well two prototypes could be morphed one into the other, in terms of the overall image quality of the resulting intermediate expression. In total, we created 8 FE blends, by pairing the original expressions as following: the portrayals of actor A were used in two pairs (Anger–Fear and Fear–Sadness), those of actor B were used in two pairs (Happiness–Disgust and Happiness–Sadness), and those of actor C were used in four pairs (Anger–Disgust, Fear–Surprise, Happiness–Disgust, and Anger–Fear). The pairs Anger–Fear and Happiness–Disgust were replicated in actor C, in order to gain some information about the impact of an actor’s characteristics on observers’ performance.

For each of these 8 pairs, static morphed pictures were created by morphing the apex frame of one prototype expression (e.g., Anger) into the other (e.g., Fear). A custom-designed morphing software (LOKI; Viviani, Binda, & Borsato, 2007) generated a sequence of 51 equally spaced intermediate samples of the linear transformation between prototypes. Essentially, the same procedure was used to create morphed video clips for the dynamic condition. However, in this case, morphing was applied to each pair of corresponding frames along the entire length of the two template sequences (i.e., from neutral to apex), generating, as before, 51 intermediate images. In this way, in both dynamic and static conditions, observers’ performance can be evaluated over an analogous range of values, corresponding to different stages of the transformation of one expression into the other. The crucial difference between the static and dynamic conditions is that in the former motion is absent; moreover, static stimuli depict just the apex of the original expression. The morphing procedure is exemplified in Figure 1.

A crucial step of the morphing technique is the identification of the same set of salient landmarks in both original prototypes (see Fiorentini & Viviani, 2009 for details). When prototypes are two sequences of frames instead of two single pictures, this operation should, therefore, be performed on each pair of corresponding frames (500 in total), which would be exceedingly time-consuming. We circumvented this difficulty by manually positioning the selected landmarks only on a certain number of key frames, which were the same in both prototypes (about one key frame every 30 to 50 frames, depending on the amount of change in facial features present in a given portion of the sequence). Then, LOKI interpolated automatically the corresponding landmarks on the intermediate pictures between successive key frames.

The morphing algorithm produced sequences (each containing 51 pictures) that were graded transformations of one original recording into the other. By compressing the morphed sequences at a speed of 120 fps with VirtualDub (<http://www.virtualdub.org/>), we converted them into digital movies (AVI format) of 4.18-s duration. For both

static and dynamic conditions, the stimuli actually tested during the experiment included only a subset of the morphed expressions generated, i.e., the blends with rank order from 15 to 35, for a total of 21 different stimuli for each pair and each condition. Four examples of dynamic stimuli (corresponding to the endpoints 1 and 50 of the morphing sequences Fear–Sadness of actor A and Anger–Fear of actor C) are available as [Supplementary material](#).

Procedure

Experiments were run in a dimly illuminated, quiet room. Participants were seated in front of a computer screen (Eizo, FlexScan 2410W 24" monitor; resolution: 1920 × 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). The task (forced-choice identification) was to indicate which prototype of the pair the stimulus was more similar to (e.g., “Is the stimulus more like anger or more like fear?”). Responses were entered by pressing two keys on the computer keyboard. Pairs of emotions were tested in separate sessions, with session order counterbalanced between participants. Each stimulus (either picture or video clip) was presented 20 times for a total of $21 \times 20 = 420$ trials. The stimuli were presented in a random order, with the constraint that there should be at least three trials before repeating the same stimulus. In the static condition, a frame was displayed for a maximum of 3 s, whereas in the dynamic condition the display time corresponded to the duration of the video clips (4.18 s). In both conditions, participants were encouraged to respond as fast and as accurate as possible. No feedback was given. Response times were reckoned from the onset of the display with millisecond accuracy. Aside from a very few cases, responses were entered before the time limit, producing the immediate disappearance of the stimulus. The experiment was self-paced, each trial starting after recording the previous response. Participants could take a short rest by pressing a “Pause” button instead of entering a response. A session lasted between 20 and 30 min, sessions with video clips being generally longer than sessions with static pictures. Immediately before each session, participants were shown the prototypes of the tested pair and performed some practice trials. Dynamic stimuli were tested a few months after static stimuli.

Results

Recognition accuracy

For each face pair and each condition, we estimated the psychometric function relating the rank order of the stimuli

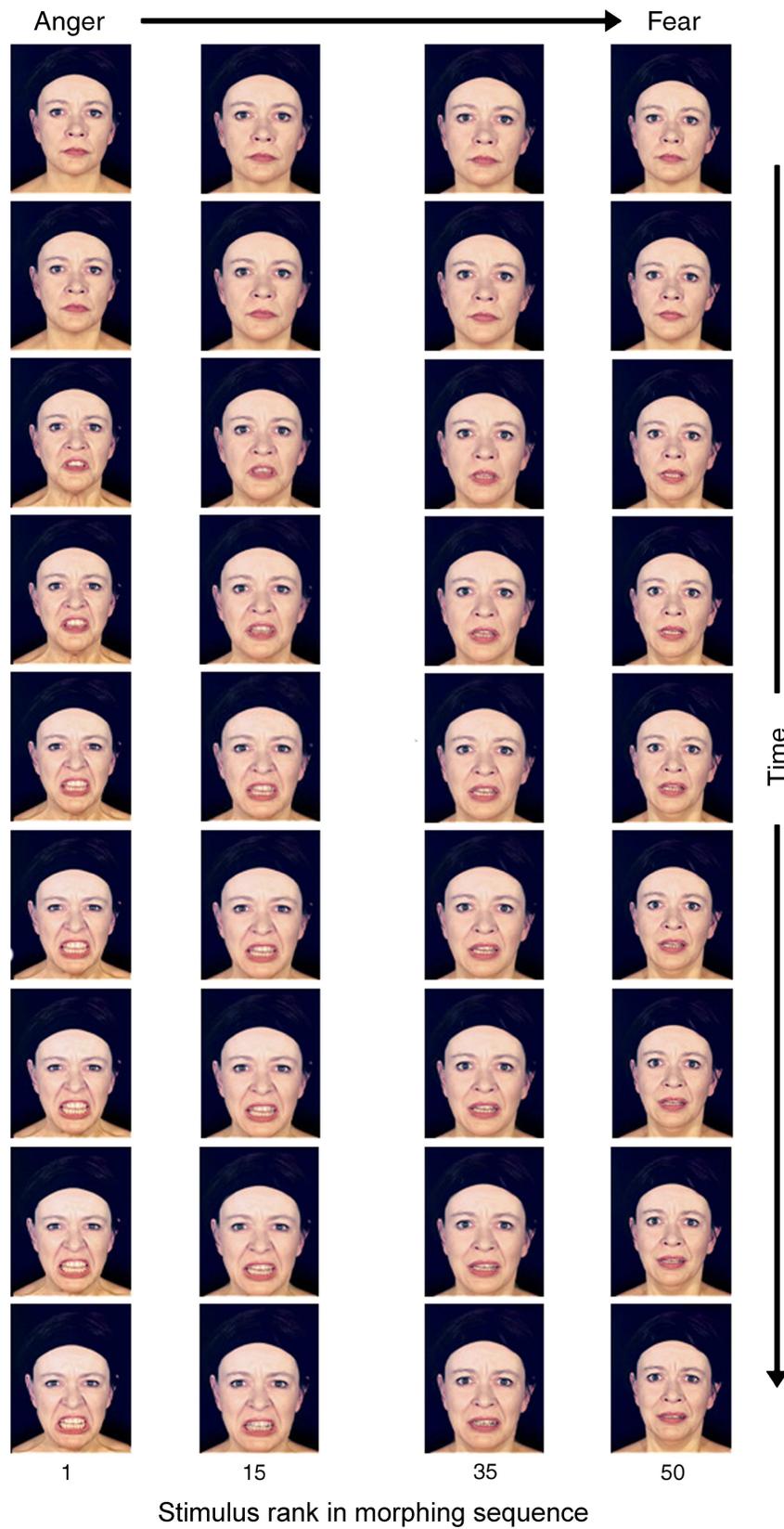


Figure 1. Morphing procedure for generating blends between pairs of FEs. Leftmost and rightmost columns: Samples (evenly spaced in time) of the video clips describing the unfolding of Anger and Fear of actor A, respectively. Middle columns: Beginning (rank 15) and end (rank 35) of the sequence of 21 morphed video clips used as stimuli in the dynamic condition. The last row depicts the final frames of the video clips used as stimuli in the static condition.

Actor	Expression pair	Condition	PSE			JND		
			Lower	Mean	Upper	Lower	Mean	Upper
A	Anger–Fear	Static	25.27	25.45	25.62	6.85	7.56	8.17
		Dynamic	25.24	25.44	25.62	6.93	7.66	8.28
	Fear–Sadness	Static	23.50	23.70	23.94	8.27	9.12	9.72
		Dynamic	23.56	23.79	24.03	8.86	9.77	10.56
B	Happiness–Sadness	Static	23.76	23.96	24.18	6.41	7.08	7.74
		Dynamic	26.04	26.25	26.47	8.04	8.81	9.53
	Happiness–Disgust	Static	22.50	22.66	22.84	5.34	5.86	6.38
		Dynamic	22.60	22.76	22.94	5.86	6.37	6.83
C	Anger–Fear	Static	25.53	26.02	25.77	9.76	10.74	11.93
		Dynamic	25.42	25.62	25.84	7.69	8.49	9.23
	Anger–Disgust	Static	25.86	26.05	26.23	6.79	7.38	7.97
		Dynamic	25.51	25.72	25.91	7.10	7.87	8.60
	Happiness–Disgust	Static	24.61	24.77	24.94	5.68	6.26	6.73
		Dynamic	24.38	24.57	24.77	7.09	7.64	8.14
	Fear–Surprise	Static	24.24	24.41	24.59	6.53	7.27	7.86
		Dynamic	25.14	25.33	25.52	7.26	7.94	8.55

Table 1. Point of Subjective Equality (PSE), Just Noticeable Difference (JND), and associated confidence intervals in dynamic and static conditions. *Note:* Upper/lower: Bounds of 0.95 confidence interval of estimated parameters. Boldface: Parameter means with non-overlapping confidence intervals.

within the morphing sequence to the relative frequency $P(B)$ of identifying the stimulus as template B (e.g., for the pair Anger–Fear, the response frequency of identifying the stimulus as “fear”).

Psychometric functions were characterized by the PSE (i.e., point of subjective equality, corresponding to $P(B = 0.5)$) and the JND (i.e., just noticeable difference, computed as the inverse of the slope of the function at the PSE). These two parameters were estimated as follows. First, we interpolated with a least-square algorithm the data points with a logistic function $f(x) = 1 / (1 + \exp(-(x - c) / a))$. Then, by definition $PSE = c$ and $JND = 4a$.

The relevant issue here is whether dynamic stimuli were recognized more accurately than static ones. As the JND is a reliable indicator of the discriminative power between the two endpoints of the continuum (i.e., the two prototypes), we compared JND values across pairs and conditions. By a bootstrap procedure, we estimated the confidence intervals associated to all JND and PSE mean values. [Table 1](#) summarizes the numerical results. [Figure 2](#) shows the corresponding 16 (8 pairs \times 2 conditions) psychometric functions.

For most pairs of FEs, the JND was lower in the static condition than in the dynamic condition. However, only in three cases the associated confidence intervals did not overlap (Happiness–Sadness, actor B: JND static = 7.08, JND dynamic = 8.81; Happiness–Disgust, actor C: JND static = 6.26, JND dynamic = 7.64). The opposite difference was significant in only one case (Anger–Fear, actor C: JND static = 10.74, JND dynamic = 8.49). Across face pairs, the static/dynamic difference was not significant (t -test for

paired samples, $t(7) = 0.970$, $p = 0.364$). The PSE also varied between pairs and between conditions; but only for two cases, Happiness–Sadness (actor B) and Fear–Surprise (actor C), we detected significant differences between static and dynamic conditions (Happiness–Sadness: PSE static = 23.96, PSE dynamic = 26.25; Fear–Surprise: PSE static = 24.41, PSE dynamic = 25.33), suggesting that in the static condition Sadness and Surprise were more salient than Happiness and Fear, respectively, and that the opposite was true in the dynamic condition. In no case the opposite was true. Across face pairs, the static/dynamic difference was not significant (t -test for paired samples, $t(7) = 0.967$, $p = 0.366$). In short, no definite trend emerged from the comparison between static and dynamic conditions.

Statistical analysis (ANOVA, 8 [face pair] \times 21 [rank] \times 2 [condition], repeated measures, Greenhouse–Geisser correction, arcsin transformation) detected significant differences for face pair ($F(7, 98) = 13.342$, $p < 0.001$) and rank ($F(20, 280) = 1171.357$, $p < 0.001$) and a significant face pair \times rank interaction ($F(140, 1960) = 5.664$, $p < 0.001$). As suggested by the analysis of PSE and JND values across pairs and conditions, there was neither a main effect of condition ($F(1, 14) = 0.661$, $p = 0.430$) nor an interaction between condition and face pair. However, the interaction face pair \times rank \times condition was significant ($F(140, 1960) = 1.691$, $p < 0.001$), which resulted from the presence of a static advantage for two face pairs (Happiness–Sadness of actor B and Happiness–Disgust of actor C) and a dynamic advantage for one face pair (Anger–Fear of actor C). This pattern of results does not confirm the hypothesis that recognition accuracy per se is

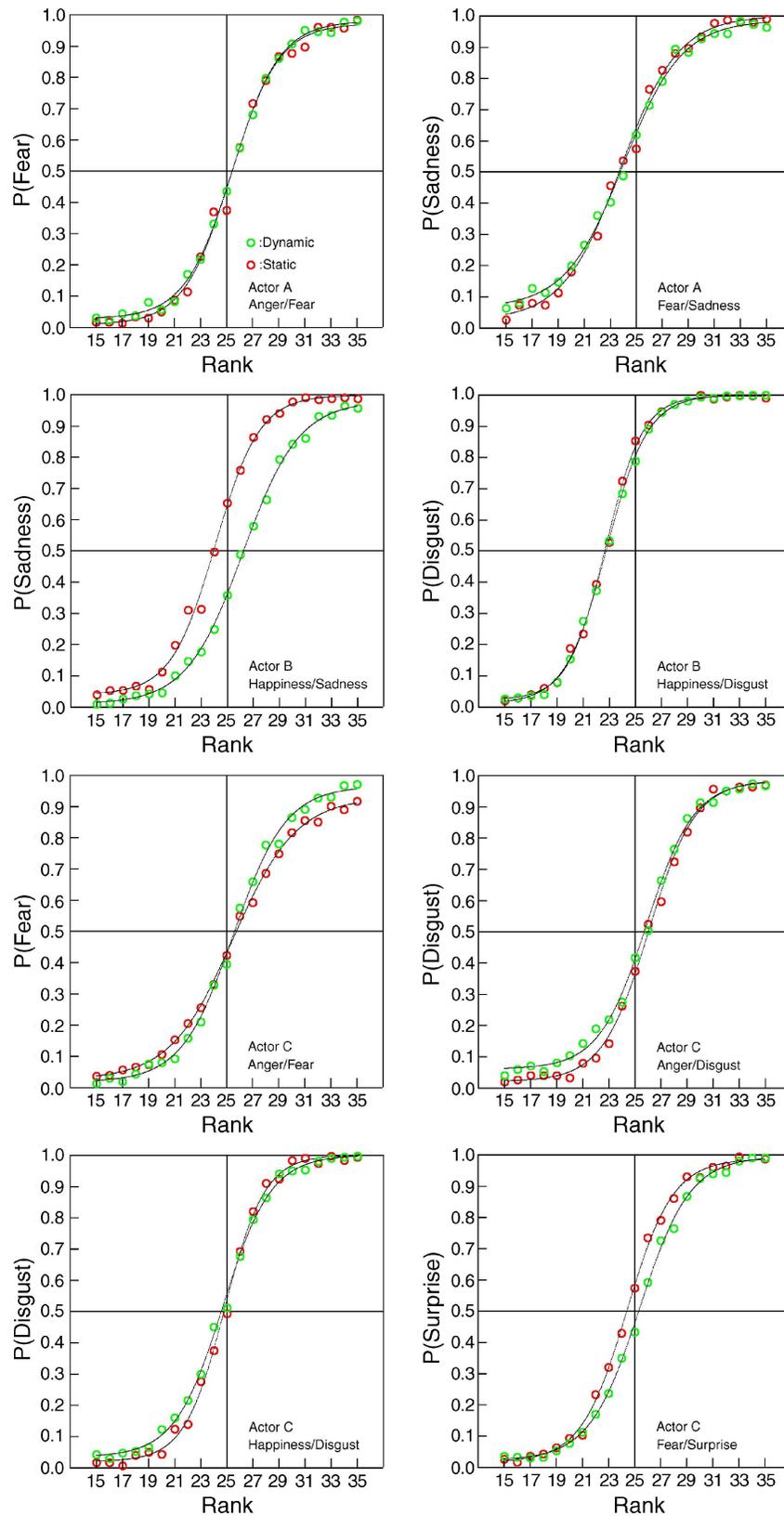


Figure 2. Psychometric functions. Ordinate: Probability of indicating the stimulus as closer to the second FE in the indicated pair. Abscissa: Rank order of the stimulus along the morphing sequence from the first to the second FE in the indicated pair. Data points (red: static condition; green: dynamic condition) were calculated by pooling the results for all participants. Continuous lines: Interpolation of the data points by a logistic function. Psychometric functions are summarized by the PSE and the JND.

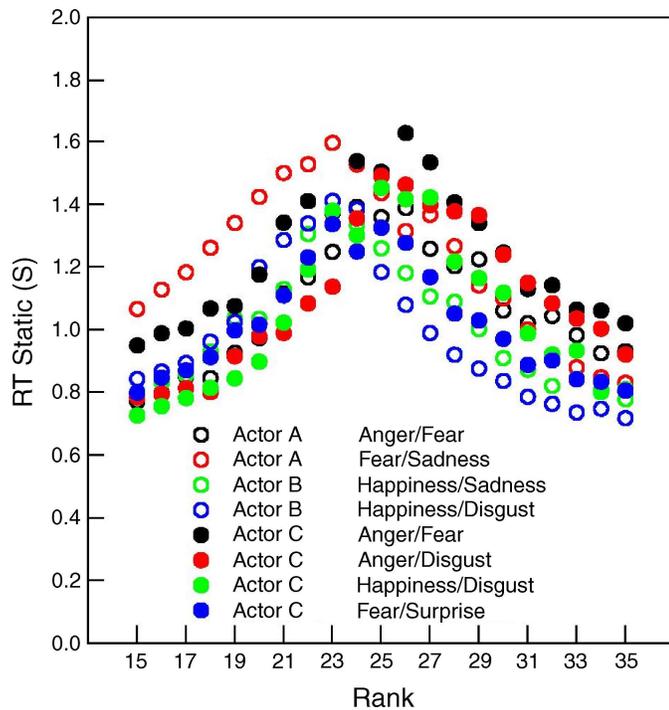


Figure 3. Average response times (RTs) for each face pair in the static condition as a function of stimulus rank.

superior when the actual sequence of muscle contractions is available. Instead, performance seems to depend more strongly on the specific face pair examined.

Response times

Response times (RTs) computed from stimulus onset showed a complex pattern of results. Both in static (Figure 3) and dynamic conditions (Figure 4, left panel) RTs for all pairs of FEs were inversely related to the distance of the stimulus from prototypes. However, mean RTs were quite different between conditions (Table 2). In the static condition, where stimuli showed the expression apex, the mean time to reach identification was of the order of 1000–1300 ms (mean RT over FEs and ranks: 1088 ms).

Static RTs averaged over ranks were significantly correlated with JND values for corresponding FE pairs ($r = 0.863$, $p < 0.01$), indicating a strong relationship between how accurately an expression is identified and how fast observers can take a decision. Statistical analysis (two-way ANOVA, 8 [face pair] \times 21 [rank], repeated measures, Greenhouse–Geisser correction) detected a main effect of both face pair ($F(7, 98) = 6.493$, $p < 0.001$) and rank ($F(20, 280) = 36.486$, $p < 0.001$), as well as a face pair \times rank interaction ($F(140, 1960) = 3.918$, $p < 0.05$). The large effect of morphing step was further qualified by the presence of a quadratic trend ($F(1, 14) = 43.649$, $p < 0.001$) showing that RTs were significantly longer for the

most ambiguous stimuli near the middle of the morphing continuum than for prototypes.

RTs in the dynamic condition (Figure 4, left panel) were considerably longer than under static conditions (mean RT over FEs and ranks: 2007 ms). As in the static condition, statistical analysis (two-way ANOVA, 8 [face pair] \times 21 [rank], repeated measures, Greenhouse–Geisser correction) detected a main effect of face pair ($F(7, 98) = 30.460$, $p < 0.001$) and rank ($F(20, 280) = 28.898$, $p < 0.001$), as well as a face pair \times rank interaction ($F(140, 1960) = 3.886$, $p < 0.005$). Likewise, in this condition, there was a significant quadratic regression term in the relationship between morphing step and RT ($F(1, 14) = 50.542$, $p < 0.001$). Most responses to dynamic stimuli were given before the expression reached its maximal intensity, the only systematic exception being the pair Happiness–Sadness of actor B. To describe the temporal relationship between RTs and the time for an expression to reach its apex, we adopted the following procedure. First, we coded with the Facial Action Coding System¹ (FACS; Ekman, Friesen, & Hager, 2002) a subset of the video clip frames (1 frame every 20 ms) and described the deployment of each prototype expression by the time course of the intensity of each intervening action unit (AU) according to the FACS [1–5] scale (see Figure 5). Then, we identified the time (hereafter, Peak Time: PT) when the summed intensity of all AUs reaches its maximum value. PT values for all prototypical FEs are reported in Table 3.

Finally, because all stimuli were blends between two FEs, we estimated the PT values for each of the 21 dynamic morphed FEs (steps from 15 to 35 of the morphing sequence) by linear interpolation between the PTs for the corresponding prototypes. The right panel of Figure 4 shows the temporal relationship between RT and PT by plotting their difference as a function of stimulus rank. The average difference RT – PT (relative RT) is negative for all but one pair of FEs, which confirms that, in the dynamic condition, reliable emotion identification is possible before the full deployment of the corresponding expression. It should be stressed that across pairs of FEs, RT varied *pari passu* with PT. In particular, the values of RT and PT measured at the PSE showed a significant positive correlation ($r = 0.874$, $p < 0.01$). However, the order of relative RTs was markedly different from that of absolute RTs. For instance, the Happiness–Sadness pair of actor B, which produced the fastest responses (Figure 4, left panel), was also the only pair for which responses were given systematically after the full deployment of the expression.

The fact that, in most cases, observers responded while facial action was still in progress may suggest that, unlike response accuracy, RT data substantiate the dynamic advantage claim. In other words, dynamic stimuli may be more effective than static ones in that they allow a decision to be taken before discriminative information is fully available. To test this hypothesis, one has to estimate the

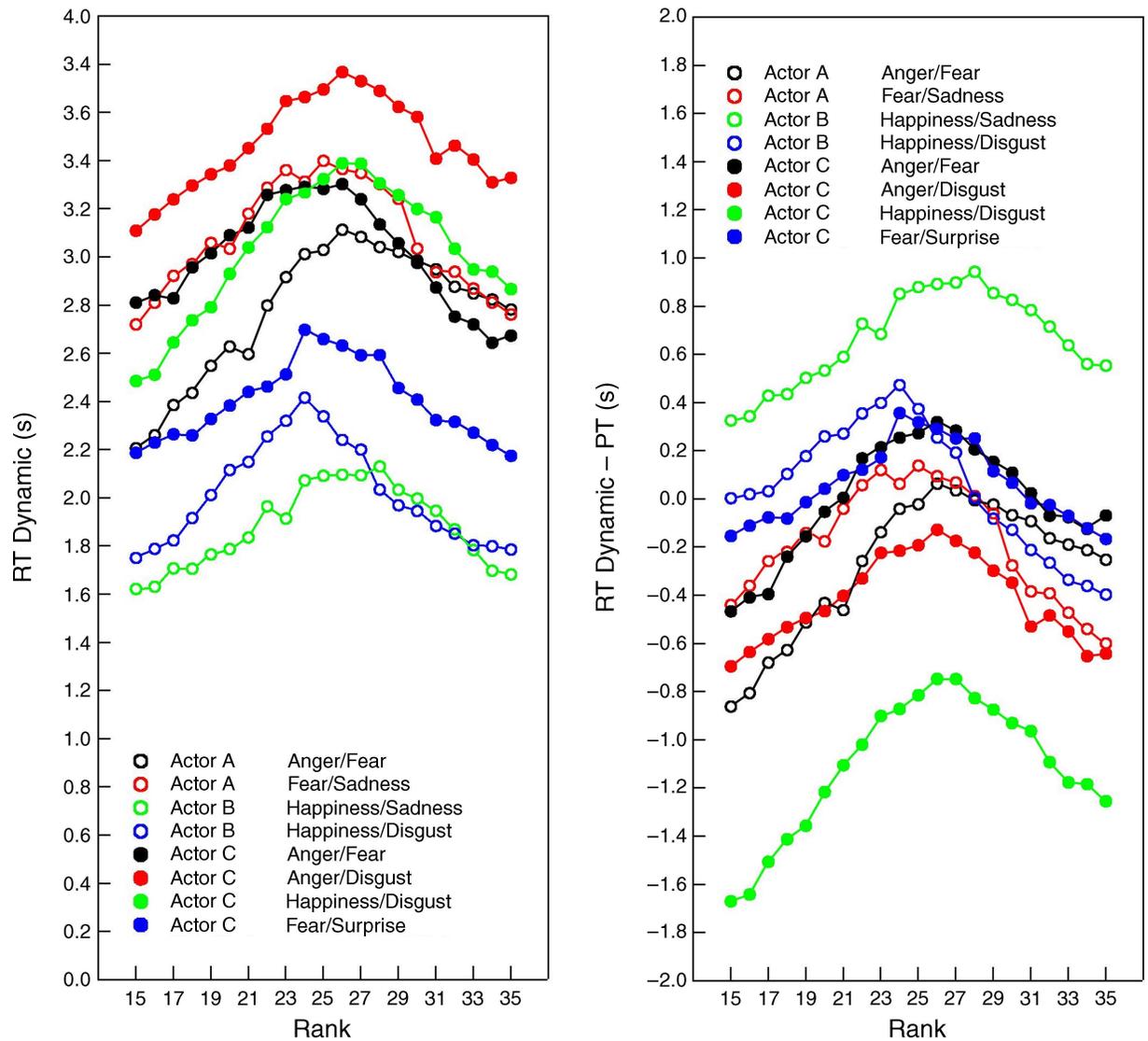


Figure 4. (Left) Average response times (RTs) for each face pair in the dynamic condition as a function of stimulus rank. (Right) Relative distance of average RT from peak time (PT). PT for blends of FEs was estimated by linear interpolation of the PTs for the two prototypes in each pair of FEs.

Actor	Expression pair	RT (ms)	
		Static	Dynamic
A	Anger–Fear	1071	2278
	Fear–Sadness	1221	3079
B	Happiness–Sadness	1014	1878
	Happiness–Disgust	993	2019
C	Anger–Fear	1238	3008
	Anger–Disgust	1104	3469
	Happiness–Disgust	1046	3028
	Fear–Surprise	1022	2401

Table 2. Mean response times (RT) in static and dynamic conditions.

FE	Actor	PT (ms)
Anger	A	3080
	C	3670
Fear	A	3000
	C	2330
Sadness	A	3500
	B	1000
Happiness	B	1420
	C	4170
Disgust	B	2500
	C	4080
Surprise	C	2330

Table 3. Mean peak time (PT) of prototype FE.

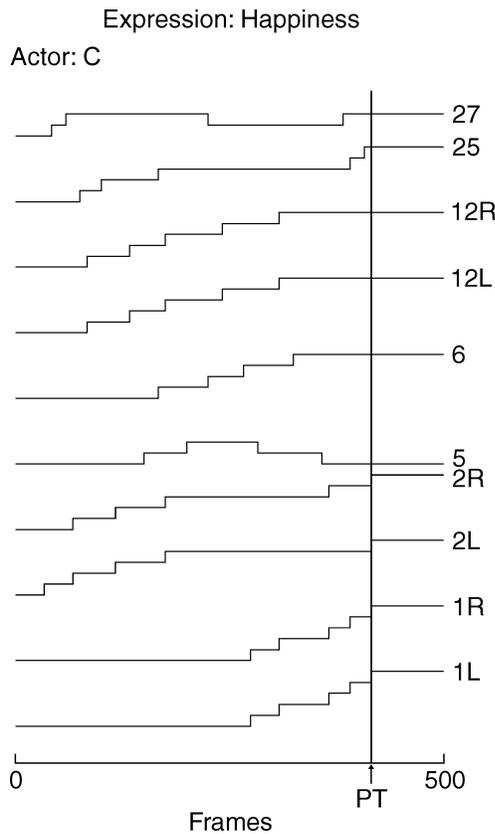


Figure 5. Example of FACS coding of the unfolding of an FE from neutral to apex. Coding was performed on individual frames of video recording (frames spaced by 20 ms). Traces describe the time course of the intensity (1–5 scale) of the AUs activated by the expression. Numerical codes to the right identify the AUs according to FACS notation. “L”: the AU occurs on the left side of the face, “R”: the AU occurs on the right side. Changes in facial appearance produced by the listed AUs are given as follows: 1: raises the inner part of the brow; 2: raises the outer part of the brow; 5: raises the upper lid; 6: raises the cheeks, narrows the eye’s aperture, causes wrinkles below and around the eyes; 12: pulls the corners of the mouth, producing the “smile”; 25: the lips part; 27: the jaw drops. Peak time (PT) marks the point in time after which no further AU changes are discernible (apex of the FE).

time course of the information during the presentation of the stimuli. This was done by using again the time course of the intensity of the AUs. First, for each prototype expression, we normalized the global curve of activity (sum of the intensity of all AUs) to the value reached at PT so that the apex corresponded to a value of 1. As shown in Figure 6, for many pairs of FEs the surge of activity for the two expressions is noticeably different. Then, the activity curve for each blend was computed again by a linear interpolation between steps 15 and 35 of the morphing sequence.

The procedure for estimating the effectiveness of dynamic stimuli relative to that of static stimuli is based on the assumption that RTs in both static and dynamic conditions signal the moment when the information acquired from the beginning of the presentation has reached a decision threshold. The assumption was specified further by positing that the amount of discriminative information acquired up to time T is the integral of the activity level from $t = 0$ to $t = T$. In the dynamic case, let $a_D(k, t)$ be the time-varying information acquired from the k th stimulus in the morphing sequence ($15 \leq k \leq 35$).

By definition, the corresponding information in the static case is just $a_S(k, t) = 1$. Finally, let us define α as the effectiveness of dynamic information relative to that of static information. Then, our assumption can be formalized as

$$\alpha(k) \int_0^{RT_D} a_D(k, t) dt = \int_0^{RT_S} a_S(k, t) dt = RT_S. \quad (1)$$

Therefore,

$$\alpha(k) = \frac{RT_S}{\int_0^{RT_D} a_D(k, t) dt}. \quad (2)$$

Effectiveness $\alpha(k)$ was finally estimated by inserting in the above formula the result of numerical integration of

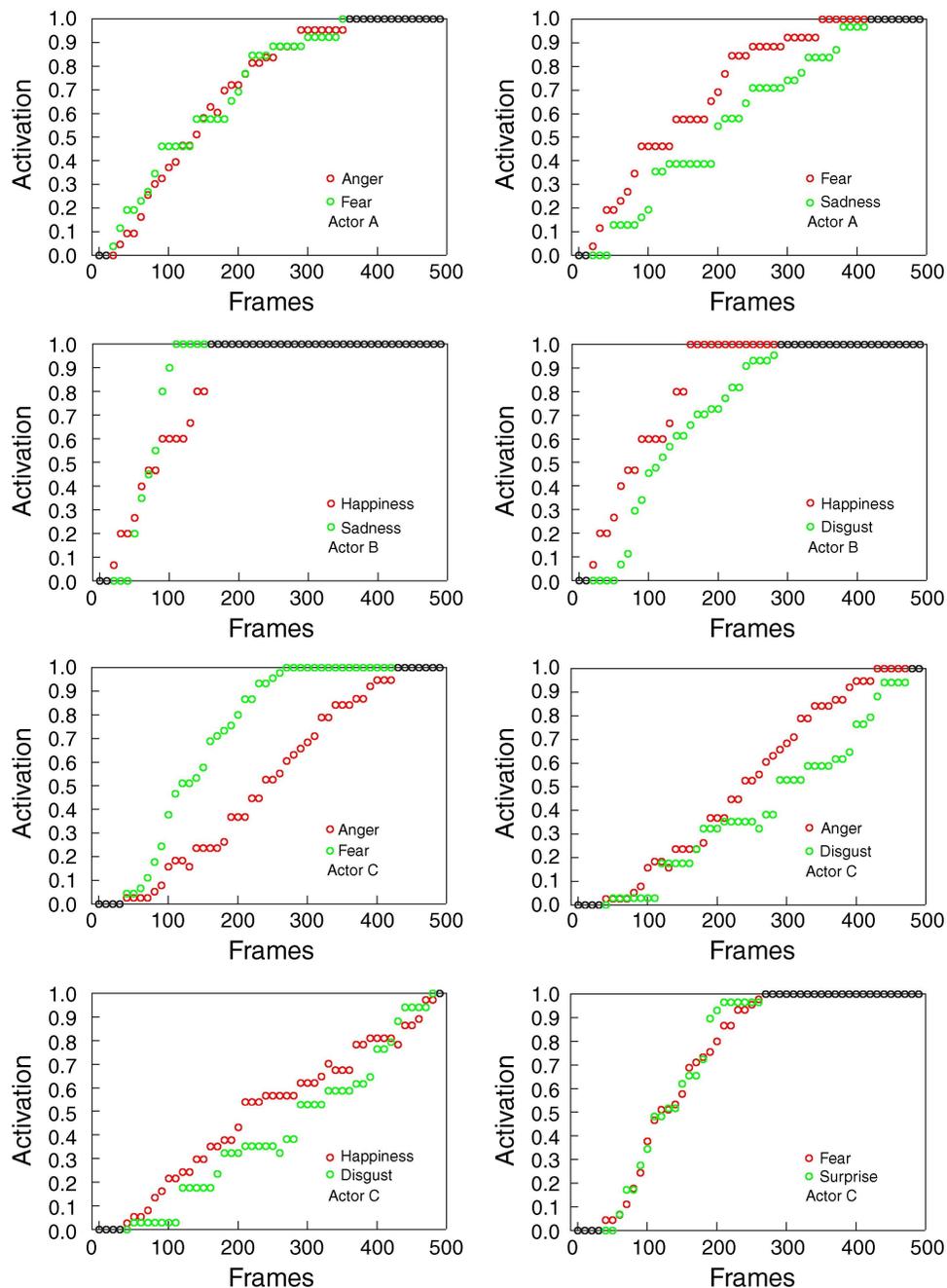


Figure 6. Activation profiles computed by averaging the intensities of all active AUs. Activation is normalized to the value reached at PT and varied between 0 (neutral) and 1 (apex).

the empirical $a_D(k, t)$ activation functions. Figure 7 plots $\alpha(k)$ for each morphing step k and each indicated face pair. The fact that effectiveness does not depend in a systematic way on the morphing step (as did response times) suggests that our assumption is consistent and that the definition above captures an aspect of the perceptual performance that is virtually independent of the stimuli. If so, the results shown in Figure 7 do not confirm the dynamic advantage hypothesis. Indeed, in all but one case (Happiness–Disgust of actor C) the dominant tendency is for $\alpha(k)$ to be

somewhat smaller than 1, indicating that static pictures are actually processed in a slightly more effective way than dynamic ones.

Discussion

We tested the suggestion emerging from previous studies (e.g., Ambadar et al., 2005; Harwood et al., 1999; Wehrle

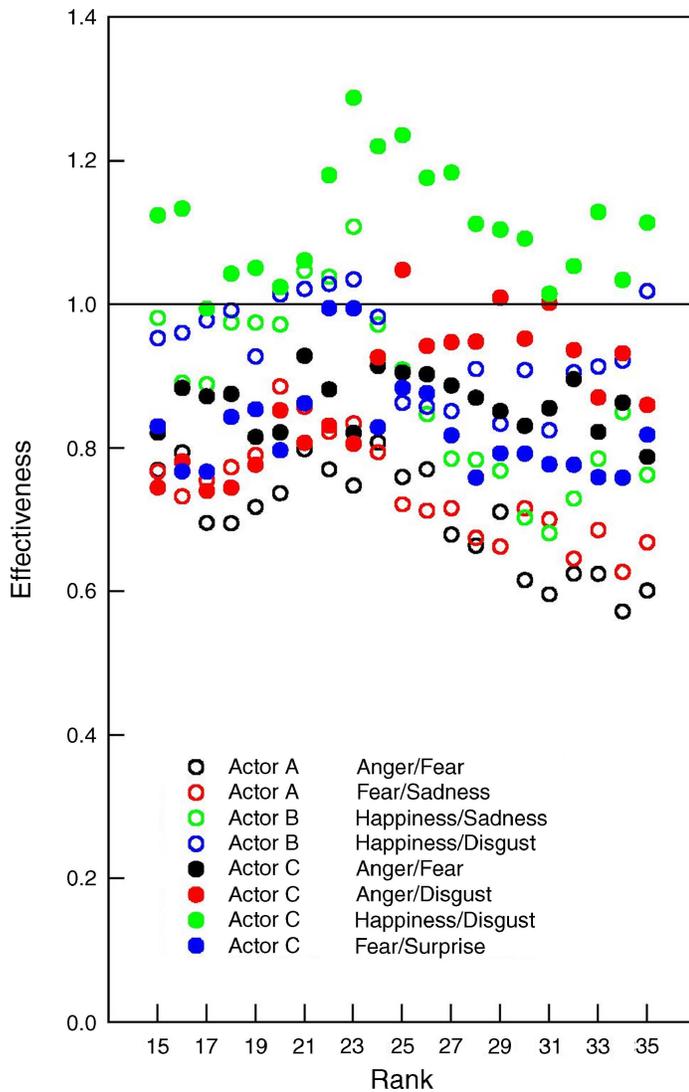


Figure 7. Estimate of the effectiveness with which dynamic information is acquired relative to static information (see Results section). Values greater than 1.0 indicate that information is acquired more effectively under dynamic condition than under static condition.

et al., 2000) that dynamic FEs are recognized more accurately than static FEs. With respect to all previous studies, the test involved a novel approach to the generation of the stimuli that define static and dynamic conditions. In the static condition, stimuli were graded blends of the apex of two FEs, and discriminability was assessed through the corresponding psychometric function. Stimuli for the dynamic condition were graded blends between high temporal resolution recordings of the actual unfolding of the same FEs. Thus, unlike Kamachi et al. (2001) and Wehrle et al. (2000), the time course of the stimuli is a veridical description of the true facial actions. Moreover, with respect to Ambadar et al. (2005) and Harwood et al. (1999), the higher sampling frequency (500 versus 25 frames/s) afforded

an accurate FACS description of the dynamic unfolding of the FEs. Finally, our test of the dynamic advantage hypothesis involved both a quantitative estimation of the discriminative power based on morphing technique for generating FE blends and response time distributions.

If a dynamic advantage exists, one would expect a more effective integration of the perceptual information during the presentation of the dynamic stimuli. If so, the speed–accuracy trade-off should be superior for dynamic stimuli than for static stimuli. This was not borne out by the analysis of the psychometric functions because the JNDs were not systematically lower in the dynamic condition than in the static condition (Table 1). Actually, with one exception (Anger–Fear of actor C) the opposite tendency emerged from the comparison. Thus, taking into account the time course of the expression from a neutral face to the apex does not seem to improve identification accuracy per se. Nevertheless, it was still possible that a dynamic advantage would emerge by taking into consideration also response times.

The fact that most responses to dynamic FEs were given while facial action was still in progress (Figure 4, right panel) does indeed indicate that the dynamic nature of the facial features somehow compensates for their incompleteness. This, however, is insufficient to substantiate the stronger claim that dynamic information is more effective than the static one. To test the validity of this claim, we made a number of assumptions. First, we equated the somewhat vague notion of information with the global level of facial activity reached at any one time during the unfolding of the expression. Summing the activation level of each intervening AU disregards the well-documented fact that not all AUs contribute equally to identify an expression. However, insofar as we have analyzed separately each pair of FEs, our choice may have reduced discriminability but did not introduce biases. Second, we adopted the basic idea underlying the so-called “counting models” (see Luce, 1986) that identification in a forced-choice, two-alternative task is achieved by cumulating discriminative evidence sampled from the stimulus until a criterion threshold is reached. Although these models were originally developed for dealing with classical Choice Reaction Time tasks, in which the time scale is much shorter than in our experiments, they proved adequate to predict the relationship between RTs and discriminability also for time scales of the order of the seconds (Viviani, 1979a, 1979b). Therefore, it may not be unreasonable to assume that their validity extends further, to even longer time scales. Finally, effectiveness was estimated through the single, multiplicative parameter $\alpha(k)$, which amounts to assume that effectiveness is independent of the level of activation reached at any one time during stimulus presentation. While there is no principled basis for this simplifying assumption, the available number of degrees of freedom did not warrant using a more flexible parameterization. In fact, although allowance was made

for effectiveness to depend on the rank order k of the stimulus along the morphing sequence, we found no consistent evidence of such dependence. Thus, it seems safe to assume that the adopted definition of dynamical effectiveness is consistent and captured an aspect of the information acquisition process that, as desired, is independent of the actual expression being shown. If so, the fact that in all but one case (Happiness–Disgust of actor C) effectiveness tended to be smaller than 1 indicates that static pictures were actually processed more effectively than dynamic ones. Thus, the analysis of response times confirmed the conclusion drawn from the analysis of psychometric functions, namely, that the “dynamic advantage” hypothesis for FEs is not supported by empirical evidence.

A ceiling effect may explain why dynamic sequences showing the unfolding of an expression are not recognized more accurately than portrayals of just the apex. There is evidence that pictures of fully developed, intense FEs yield almost perfect identification (Carroll & Russell, 1997). Close-to-ceiling identification rates with static stimuli have been cited by Kamachi et al. (2001) as the reason why accuracy for dynamic FEs was found to be no different—in some cases even poorer—than accuracy for static FEs. Moreover, in our study, two of the three significant differences in accuracy in favor of static stimuli occurred for the pairs Happiness–Disgust (actor C) and Happiness–Sadness (actor B), which are characterized by strongly contrasting facial configurations along the positive–negative valence axis and are discriminated more easily than most other pairs (Fiorentini & Viviani, 2009). This again suggests that when two configurations are most clearly differentiated, the strength of discriminative information at their apex overwhelmed whatever additional gain may accrue from dynamic information gathered during the unfolding of the expressions.

Conversely, the ceiling effect hypothesis is consistent with the results by Ambadar et al. (2005) who reported a dynamic advantage in the case of subtle expressions where facial actions are not as well marked as in full-blown FEs. Likewise, the schematic faces used by Wehrle et al. (2000) may have resulted in a suboptimal performance in the static condition because they too lack the full complement of detailed facial information present in full-blown FEs.

An important methodological issue needs to be emphasized. As noted in the Introduction section, expressive movements simulated by morphing a neutral face into a full expression differ from biological ones insofar as they unfold linearly, even when temporal realism and biological constraints are respected (Katsiri, 2006). Kamachi et al. (2001) used this morphing technique to generate their dynamic stimuli. Therefore, their failure to detect a dynamic advantage in that study may be due to the artificialness of the dynamic stimuli. This confounding factor was not present in our study. Although we did apply a morphing technique to generate ambiguous stimuli, the templates being blended were faithful descriptions of actual facial movements. Thus, unlike those tested by Kamachi et al., our dynamic FEs

preserved the crucial temporal information contained in the original FEs used as prototypes.

Insofar as they show that information is not integrated more effectively from dynamic FEs than from static FEs, the results of the analysis of RTs are generally in line with the results of the analysis of the psychometric functions. The pair Happiness–Disgust of actor C deviates from this general trend. On the one side, this pair is discriminated more precisely in the static condition (JND = 6.26) than in the dynamic condition (JND = 7.64). On the other side, it is also the only pair for which there is evidence of a dynamic advantage ($\alpha(k) > 1$ for all ks , see Figure 7). To conclude, we consider this exceptional case in greater detail.

In general, the expression of disgust is fairly ambiguous, yielding somewhat intermediate or low recognition scores in classical judgment studies (Russell, 1994). In particular, Disgust of actor C was characterized by a relatively small amount of facial activity, distributed over the entire temporal unfolding of the expression. However, when Disgust was pitted against a very different expression, such as Happiness, it was relatively easy to identify blended stimuli as one or the other prototype. In fact, also the other pairs that included Happiness (Happiness–Sadness and Happiness–Disgust of actor B) were well discriminated (Table 1). The difference in response accuracy in static and dynamic conditions can be well accounted by noting that in the former condition observers could take full benefit from the high distinctiveness of apices. Instead, in the latter condition, the apices were reached only at the end of the presentation (PT = 4.17 s and PT = 4.08 s for Happiness and Disgust, respectively) and responses were given long before PT (mean RT = 3.028 s). Thus, the dynamic advantage estimated by the index $\alpha(k)$ derives entirely from information picked up during the unfolding of the morphed sequence. A comparison among pairs permits one to make hypotheses on the nature of this information.

At any time, the contribution of each template to an intermediate blend depends jointly on the blend rank along the morphing sequence and on the activation functions of the templates. The activation functions of Happiness and Disgust increase slowly, that for Happiness being slightly higher than that for Disgust during most of the unfolding process (Figure 6). If the difference between the levels of activation were the relevant discriminative information, one would expect the dynamic advantage to be strong for the pair Anger–Fear of actor C where the divergence between activation levels is greatest. Alternatively, if activation rate were the key factor, the dynamic advantage should be maximum for the pair Happiness–Sadness of actor B. In fact, the average α values for both these pairs are much lower than that for Happiness–Disgust of actor C. In short, contrary to the hypothesis by Ambadar et al. (2005), the available evidence suggests that FE identification in dynamic conditions is not facilitated by the presence of motion per se. Rather, the crucial factor seems to be the timing with which the AUs sufficient to recognize a facial configuration become available before the expression

reaches its apex. This suggestion is in keeping with the results of a recent experiment (Fiorentini et al., [manuscript in preparation](#)). We showed that when observers are asked to stop the slow motion rendering of an expression as soon as they identify the target expression, RT distributions peak in correspondence with clusters of AU activations that are typical of the target expression. A more direct test of the above hypothesis could be done either by comparing naturally unfolding dynamic FEs with static snapshots of the same expressions sampled at different points in time before the apex or by directly manipulating the relative timing with which telltale AUs are activated before reaching the apex.

Acknowledgments

This work was supported by Swiss National Foundation (SNF) grant 100011-112252 to Paolo Viviani.

Commercial relationships: none.

Corresponding author: Chiara Fiorentini.

Email: fiorentinichiara@gmail.com.

Address: Behavioral and Brain Sciences Unit, UCL Institute of Child Health, 30, Guilford Street, London WC1N 1EH, UK.

Footnote

¹The Facial Action Coding System (FACS) is a comprehensive method of objectively coding facial activity. Using FACS and viewing videotaped facial behavior, coders can manually code all possible facial displays. These are decomposed into 30 elementary facial actions, called Action Units (AUs), which have a specified anatomical basis.

References

- 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*, 403–410.
- 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.
- 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.
- Berry, D. S. (1990). What can a moving face tell us? *Journal of Personality and Social Psychology*, *58*, 1004–1014.
- Blair, R. J. R., Colledge, E., Murray, L., & Mitchell, D. G. V. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology*, *29*, 491–498.
- Bruce, V., & Valentine, T. (1988). When a nod's as good as a wink: The role of dynamic information in facial recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 169–174). New York: John Wiley & Sons.
- Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology*, *72*, 164–176.
- Edwards, K. (1998). The face of time: Temporal cues in the facial expressions of emotion. *Psychological Science*, *9*, 270–276.
- Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*, 238–252.
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *The facial action coding system* (2nd ed.). Salt Lake City, UT: Research Nexus eBook.
- Fiorentini, C., Schmidt, S., & Viviani, P. (2010). The unfolding of facial expressions. *Journal of Nonverbal Behavior* (manuscript in preparation).
- Fiorentini, C., & Viviani, P. (2009). Perceiving facial expressions. *Visual Cognition*, *17*, 373–411.
- Gepner, B., Deruelle, C., & Grynfeldt, S. (2001). Motion and emotion: A novel approach to the study of face processing by young autistic children. *Journal of Autism and Developmental Disorders*, *31*, 37–45.
- Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *American Journal on mental retardation*, *104*, 270–278.
- Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*, 880–885.
- Humphreys, G. W., Donnelly, N., & Riddoch, J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: Neuropsychological evidence. *Neuropsychologia*, *31*, 173–181.
- Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, *30*, 875–887.

- Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on automatic face and gesture recognition* (pp. 46–53). Los Alamitos, CA: IEEE Computer Society Conference Publishing Services.
- Katsiri, J. (2006). *Human recognition of basic emotions from posed and animated dynamic facial expressions*. Unpublished doctoral dissertation. Online in pdf format: <http://lib.tkk.fi/Diss/2006/isbn951228538X/isbn951228538X.pdf>.
- Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition*, 4, 265–273.
- Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory and Cognition*, 27, 974–985.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, Oxford: Oxford University Press.
- Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London*, 258, 273–279.
- O’Toole, A. J., Roark, D., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6, 261–266.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141.
- Sato, W., & Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cognition and Emotion*, 18, 701–710.
- Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale, NJ: Erlbaum.
- Tardif, C., Lainé, F., Rodriguez, M., & Gepner, B. (2007). Slowing down presentation of facial movements and vocal sounds enhances facial expression recognition and induces facial–vocal imitation in children with autism. *Journal of Autism and Developmental Disorders*, 37, 1469–1484.
- Viviani, P. (1979a). A diffusion model for discrimination of temporal numerosity. *Journal of Mathematical Psychology*, 19, 108–136.
- Viviani, P. (1979b). Choice reaction times for temporal numerosity. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 157–167.
- Viviani, P., Binda, P., & Borsato, T. (2007). Categorical perception of newly learned faces. *Visual Cognition*, 15, 420–467.
- Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78, 105–119.
- Yoshikawa, S., & Sato, W. (2008). Dynamic facial expressions of emotion induce representational momentum. *Cognitive, Affective, & Behavioral Neuroscience*, 8, 25–31.