
ELSEVIER
VISION
RESEARCH
An International Journal for Functional Aspects of Vision
First target
Second target
Third target (Short conditions)
Third target (Long conditions)
Straight condition
Reversed condition
Biochemistry & Cell Biology • Molecular Biology & Genetics
Anatomy, Physiology, Pathology & Pharmacology • Optics, Accommodation & Refractive Error
Circuitry & Pathways • Psychophysics • Perception • Attention & Cognition
Computational Vision • Eye Movements & Visuomotor Control
ISSN 0042-6989 | Volume 51 | Issue 15 | 1 August 2011

Vision Research 51 (2011) 1751–1759

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Time course of discrimination between emotional facial expressions: The role of visual saliency

Manuel G. Calvo [a,*], Lauri Nummenmaa [b,c,d]

[a] Department of Cognitive Psychology, University of La Laguna, Tenerife, Spain
[b] Brain Research Unit, Low Temperature Laboratory, Aalto University School of Science, Espoo, Finland
[c] Department of Biomedical Engineering and Computational Science, Aalto University School of Science, Espoo, Finland
[d] Turku PET Centre, Turku, Finland

ARTICLE INFO

Article history:
Received 15 December 2010
Received in revised form 8 April 2011
Available online 12 June 2011

Keywords:
Facial expression
Discrimination
Visual saliency
Saccade
Emotion

ABSTRACT

Saccadic and manual responses were used to investigate the speed of discrimination between happy and non-happy facial expressions in two-alternative-forced-choice tasks. The minimum latencies of correct saccadic responses indicated that the earliest time point at which discrimination occurred ranged between 200 and 280 ms, depending on type of expression. Corresponding minimum latencies for manual responses ranged between 440 and 500 ms. For both response modalities, visual saliency of the mouth region was a critical factor in facilitating discrimination: The more salient the mouth was in happy face targets in comparison with non-happy distracters, the faster discrimination was. Global image characteristics (e.g., luminance) and semantic factors (i.e., categorical similarity and affective valence of expression) made minor or no contribution to discrimination efficiency. This suggests that visual saliency of distinctive facial features, rather than the significance of expression, is used to make both early and later expression discrimination decisions.

1. Introduction

The aim of the current study was twofold. First, we investigated how quickly happy facial expressions can be discriminated from each of the other five basic emotional expressions (i.e., angry, sad, fearful, disgusted, and surprised) and neutral expressions, when two faces are presented simultaneously. Second, we estimated how much expression discrimination relies on perceptual and semantic attributes of the face stimuli during the early and late stages of the recognition stream.

1.1. Recognition of single expressions vs. discrimination between expressions

There is consistent evidence of a 'happy face advantage' in recognition or categorization tasks where viewers decide which emotion a face conveys. Happy expressions are identified more accurately and/or faster than all the other basic emotional expressions (Calvo & Lundqvist, 2008; Juth, Lundqvist, Karlsson, & Öhman, 2005; Leppänen & Hietanen, 2004; Loughead, Gur, Elliott, & Gur, 2008; Palermo & Coltheart, 2004; Tottenham et al., 2009). This effect has been observed across different facial stimulus sets and response modalities (i.e., manual, verbal, and saccadic). In addition, paradigms manipulating stimulus visibility have shown that happy faces are more resistant than other expressions to pre- and post-masking (Maxwell & Davidson, 2004; Milders, Sahraie, & Logan, 2008) and to reduction of stimulus display times (Calvo & Lundqvist, 2008; Milders et al., 2008), and that happy faces have stronger perceptual dominance in binocular rivalry paradigms (Alpers & Gerdes, 2007; Yoon, Hong, Joorman, & Kang, 2009).

Given such an advantage in the recognition of happy expressions, it is important to examine the mechanisms responsible for their discrimination from other expressions. In most prior research, the face stimuli were presented singly. In contrast, in normal social contexts we often encounter arrays of faces whose expressions we need to discriminate. To our knowledge, only two prior studies have assessed expression discrimination between simultaneously presented faces (Bannerman, Milders, & Sahraie, 2009; Calvo & Nummenmaa, 2009). In these studies, two faces appeared at the same time on a computer screen, and the task required a speeded saccade towards the face predefined as a target. Calvo and Nummenmaa (2009) paired neutral faces with happy, surprised, disgusted, angry, fearful, or sad faces. Happy faces were discriminated from neutral faces faster (as revealed by saccade latencies) than were the other emotional faces. Bannerman et al. (2009) presented neutral faces paired with either happy or fearful faces. The emotional target faces were discriminated faster from neutral distracter faces than vice versa, but no differences between happy and fearful faces appeared.

* Corresponding author.
E-mail addresses: mgcalvo@ull.es (M.G. Calvo), nummenmaa@neuro.hut.fi (L. Nummenmaa).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.06.001

Accordingly, there is solid evidence that happy expressions are recognized faster than other expressions when faces are presented in isolation, but there are inconsistencies regarding discrimination of simultaneously presented faces. In addition, the Bannerman et al. (2009) and Calvo and Nummenmaa (2009) studies assessed discrimination between emotional and neutral faces, but not between emotional faces themselves. Also, the factors responsible for discrimination between expressions, and how they impinge upon the recognition time course, have not been scrutinized previously. In the current study, we extended prior research on facial expression recognition. First, we compared latencies in a saccade vs. a manual response task. Second, we assessed discrimination between happy and each of the other five basic emotional expressions. Third, we examined the role of perceptual and semantic attributes of the face stimuli in the discrimination process. Briefly, a happy (target) and a non-happy (distracter) face were presented briefly (30 ms) and simultaneously side by side. Participants were asked to either press a key on (manual response) or look at (saccadic response) the side where the target appeared. Discrimination was assumed to occur when the proportion of correct responses (to the target) exceeded that of incorrect responses (to the distracter).

1.2. Time course and perceptual/semantic factors in expression discrimination

To explore the time course of facial expression discrimination, we used a 2AFC (two-alternative-forced-choice) paradigm with saccadic and manual responses. This combination allows us to investigate whether perceptual and semantic factors support expression recognition at different stages. Saccades are executed much faster than manual responses, and thus more detailed information of the stimuli could be acquired by the time of manual responding than by the time of saccadic responding (see Bannerman et al., 2009; Crouzet, Kirchner, & Thorpe, 2010). Saccades might thus be particularly sensitive to perceptual attributes of the stimuli (such as visual saliency), while manual responses would allow for the accrual of semantic information (such as perceived expression category). It is possible that, while a perceptual criterion is used at early stages of facial expression discrimination, a semantic criterion could guide discrimination at later stages. If so, saccade latencies should be influenced by perceptual differences between the to-be-discriminated faces, whereas manual response latencies should be more influenced by semantic differences.[1]

To determine the role of perceptual factors in facial expression discrimination, we computed global image statistics such as luminance and energy, as well as image similarity between a target and a distracter face.[2] In addition, local visual saliency of the eye, nose, and mouth regions was modeled by using the iNVT+ algorithm (Itti & Koch, 2000; see also Itti, 2006; Walther & Koch, 2006). The resulting saliency map represents the relative visual conspicuity of an image and its parts as a function of a combination of intensity, color, and spatial orientation. Various models have proposed that image statistics and visual saliency influence initial shifts of covert and overt attention (see Torralba, Oliva, Castelhano, & Henderson, 2006). Supporting evidence has shown that the initial distribution of eye fixations is determined by the saliency weights of the different parts of the image (Calvo & Nummenmaa, 2008; Parkhurst, Law, & Niebur, 2002; Underwood & Foulsham, 2006), and that visual discriminability arises from the physical saliency of the target in relation to the distracters (Nothdurft, 2006). Accordingly, if discrimination of facial expressions is initially driven bottom-up by perceptual factors, discrimination will be easier when differences in image statistics and visual saliency increase. The more perceptually dissimilar the target and the distracter face are, the more accurate and faster discrimination will be.

To examine how semantic and affective factors contribute to expression discrimination, we obtained measures of expression category membership (i.e., the probabilities and latencies of classifying faces as happy vs. non-happy), and emotional valence (i.e., the probabilities and latencies of classifying the faces as pleasant vs. not pleasant) for each face stimulus. Such a distinction is theoretically important: While categorization involves classifying faces into discrete expressions, affective evaluation involves a more general differentiation in a pleasantness–unpleasantness dimension (see Calder, Burton, Miller, Young, & Akamatsu, 2001; Fiorentini & Viviani, 2009). Both expression categorization and affective evaluation can be performed early in the visual processing stream (Calvo & Nummenmaa, 2009; Calvo, Nummenmaa, & Avero, 2010; Carroll & Young, 2005; Lipp, Price, & Tellegen, 2009). Accordingly, if facial expression discrimination is driven by category information, categorical dissimilarity between the target and the distracter face should lead to faster discrimination. If recognition is driven by affective content, performance should be facilitated by affective dissimilarity between the target and the distracter face.

2. Method

In the main experiment, we used a 2AFC protocol with manual and saccadic responses to investigate facial expression discrimination. In two additional studies, we obtained indices of expression category membership, and assessed affective valence of each face stimulus. We also computed global image properties of the face stimuli and the local visual saliency of face regions that are critical for emotional expression.

2.1. Participants

Thirty-two psychology undergraduates (aged from 19 to 23 years old) participated for course credit. Of them, 16 (12 female) were randomly assigned to the saccade task and another 16 (12 female) to the manual response task.

2.2. Stimuli

We selected 210 digitized color photographs from the KDEF (Karolinska Directed Emotional Faces; Lundqvist, Flykt, & Öhman, 1998) stimulus set. The face stimuli portrayed 30 individuals (15 females: KDEF no. 01, 02, 03, 05, 07, 09, 11, 13, 14, 19, 20, 26, 29, 31, 33; and 15 males: KDEF no. 03, 05, 06, 08, 10, 11, 12, 13, 14, 17, 22, 23, 29, 31, 34), each posing seven expressions (neutral, happiness, anger, sadness, disgust, surprise, and fear). Non-facial areas (e.g., hair, etc.) were removed by applying an ellipsoidal mask. Each face subtended a visual angle of 8.4° (height) × 6.4° (width), and was presented against a black background.

2.3. Apparatus

The stimuli were presented on a 21″ monitor with a 120-Hz refresh rate. In the saccade task, eye movements were recorded with an EyeLink II tracker (SR Research Ltd., Mississauga, Ontario,

[1] Alternatively, it is possible that the decision occurs at about the same time and is based on the same information in both the saccade and the manual task, but that the post-decision motor movement has different durations. If so, saccade and manual response accuracy would be equivalent in both tasks, although latencies would be shorter in the saccade task.

[2] By perceptual factors we refer to physical stimulus properties that are accessible to the visual system but are devoid of any meaning themselves. We distinguish between simple image statistics based on image intensity distribution (luminance and energy) and complex visual properties (image similarity and saliency) that combine various simple properties.

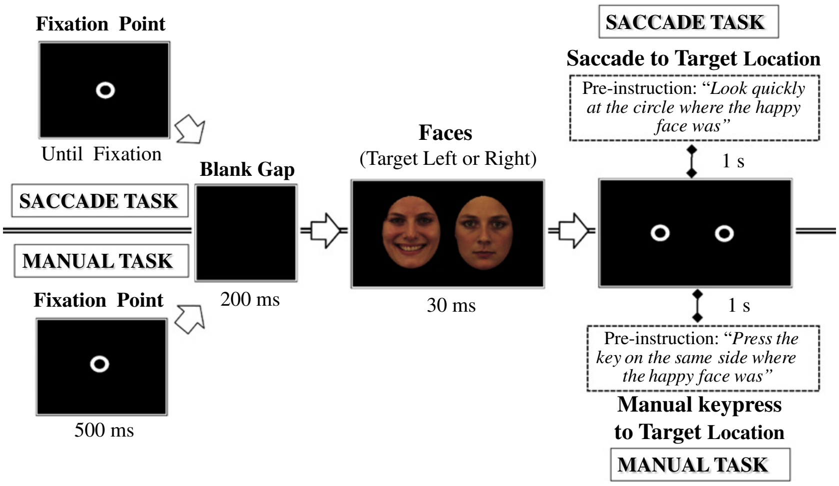

Fig. 1. Sequence of events and overview of basic characteristics of a trial.

Canada) at a 500-Hz sampling rate and <0.5° spatial resolution in pupil tracking mode. For the manual task, no eyetracker was used and responses were obtained through finger keypresses on a computer keyboard. Participants had their head positioned on a chin and forehead rest at a 60-cm viewing distance.

2.4. Procedure

In the saccade task, we performed a 9-point calibration routine followed by validation. Each trial (see Fig. 1) began with a central drift correction circle (0.5°). When the participant fixated this circle (in the saccade task; or following a 500-ms display of the circle in the manual task), there was a 200-ms gap period with a black screen. This gap allows attention to disengage from the fixation point before the face stimuli appear, and accelerates subsequent attention shifts and eye movements (see Fischer & Weber, 1993). Following the gap, two lateralized faces were presented simultaneously for 30 ms,[3] one to the left and the other to the right of fixation, with the inner edges 2.5° away from the central fixation point. Finally, two fixation circles appeared for 1 s, each placed at the center of the location where the faces had been displayed. The participants were to make a speeded saccade to the circle that replaced the target happy face (saccade task) or a speeded keypress with the index finger of the corresponding side (manual task). Following 12 practice trials, each participant was presented with 180 experimental trials randomly.

Of the two faces on each trial, one was always happy (target) and the other (distracter) was sad, angry, fearful, disgusted, surprised, or neutral. The target and the distracter faces presented on each trial always corresponded to different individuals. This made the task more ecologically valid, given that in real life the people we see at a given time are different rather than clones. Also, by using two different identities we ensured that discrimination could not be made on the basis of trivial physical differences between the target and the distracter (see Calvo & Nummenmaa, 2009).

2.5. Measures

Saccade latencies were recorded from the onset of the face pair until the first eye movement with an amplitude over 2° was initiated towards one of the circles that replaced the faces. We only analyzed eye movements that occurred after the gap period and that the Eyelink system classified as a saccade (saccade velocity threshold = 30°/s; saccade acceleration threshold = 4000°/s). Manual response latencies were recorded from the onset of the face pair until a keypress was made by the participant. In addition to the mean latency for correct responses, we computed the *minimum* latency, i.e., the *earliest* time point where the proportion of saccadic or manual responses correctly directed towards the target face exceeded that of erroneous responses towards the distracter. This would reveal when sufficient information is available to perform the task with above chance accuracy. To estimate the minimum time required for encoding, we first divided the expression-wise response latency distributions into 20-ms bins. Next, we computed the proportion of correct and erroneous responses in each bin and searched for the first bin that contained significantly more correct than erroneous responses, and was followed by at least five successive bins with more correct than erroneous responses (see Kirchner & Thorpe, 2006).

2.6. Design

One between-subjects factor (type of task: saccade vs. manual response) was combined with one within-subjects factor (expression of the non-happy face distracter paired with the happy face target: neutral, angry, sad, disgusted, surprised, and fearful). Each participant was presented with each target happy face six times (three on the left and three on the right visual field), once with a different distracter face category (neutral, angry, etc.). On each of four counterbalancings (with four participants each on each task), a target face was randomly paired with a distracter face of a different identity. The side (left or right) of the distracter for a given identity and category varied across counterbalancings.

2.7. Assessment of perceptual attributes of face stimuli: global low-level image properties and local visual saliency

Visual saliency and low-level image properties influence initial orienting of covert attention and eye movements (see Section 1.2). It is thus possible that discrimination performance could be affected by target–distracter differences in such perceptual characteristics. To examine this issue, we first obtained the *luminance* and *energy*[4] statistics of each stimulus face as a whole, with Matlab

[3] Our choice of the 30-ms stimulus display was based on prior relevant studies using the 2AFC task (20-ms display; Bannerman et al., 2009; Kirchner & Thorpe, 2006). We slightly increased the stimulus exposure because the threshold for expression recognition above chance of singly presented faces is around 25 ms (Calvo & Lundqvist, 2008; Milders et al., 2008).

[4] Energy is the amount of signal conveyed by the different spatial frequencies of the image (e.g., Näsänen, 1999).

7.0 (The Mathworks, Natick, MA). These values were based on the image pixel intensities. We also measured *image similarity* by computing Pearson correlations of the pixel intensities (i.e., grayscale luminance) for each happy face and the corresponding distracter. Second, we computed the *visual saliency* of three main regions of each face (eye, nose, and mouth), by means of the iLab Neuromorphic Vision C+ Toolkit (Itti & Koch, 2000). In the neuromorphic model, the visual input is first decomposed and processed by feature (contrast, orientation, and color) detectors mimicking the response properties of retinal neurons, lateral geniculate nucleus, thalamus, and V1. These features are then integrated for a neural saliency map that is a graded representation of the visual conspicuity of each pixel in the image. In our face stimuli (8.4° high × 6.4° wide), the vertical visual angles covered by each face region were as follows: eyes (1.6°), nose (1.8°), and mouth (1.6°). Finally, we calculated *difference scores* of luminance, energy, and saliency for each face target–distracter pair. Difference scores would thus indicate how different the target and distracter faces are.

2.8. Assessment of semantic attributes of face stimuli: category membership and emotional valence

Facial expressions can be assigned to discrete categories as a function of specific features (e.g., smile, frown, wide-open eyes, etc.; see Calvo & Marrero, 2009; Kohler et al., 2004). Expressions also convey affect along a more global unpleasant vs. pleasant valence dimension. Both categorical and affective information can be relevant for discriminating between faces. To assess *category membership* and *emotional valence* of the face stimuli, we conducted an additional study. Twenty undergraduates (15 females) not participating in the main experiment performed an expression categorization task, and another 20 (15 females), a valence evaluation task. Each face stimulus was presented singly and centrally for 30 ms. By pressing one of two keys, participants classified each face as happy or not (categorization task) or decided whether it conveyed a pleasant expression or not (evaluation task). To obtain an index of *category membership*, we computed the probability that participants (wrongly) classified non-happy faces as "happy", and the reaction times for (correctly) classifying them as "not happy". Fewer errors and shorter reaction times were assumed to reflect more categorical distance between happy targets and non-happy distracters. To get an index of *affective similarity*, we computed the probability that non-happy faces were classified as "pleasant", and the speed at which a non-happy face was classified as "not pleasant". A lower probability score and a faster response would reveal more affective differences from the genuine happy expressions.

3. Results

3.1. Discrimination performance in the saccade and manual response tasks

A 2 (type of task) × 6 (expression of distracter) repeated-measures (expression) ANOVA was conducted on the dependent measures, with Greenhouse–Geisser correction for unequal variances. Bonferroni corrections ($p < .05$) were used for post hoc multiple comparisons in these and all the following analyses. Mean scores and significant multiple contrasts—as indicated by superscripts—are shown in Table 1.

For the proportion of *correct responses*, the main effects of task, $F(1, 30) = 23.96$, $p < .0001$, $\eta_p^2 = .444$, revealed more accurate discrimination in the manual ($M = .941$) than in the saccade ($M = .808$) task. The weak effect of expression, $F(4.43, 132.80) = 2.73$, $p < .05$, $\eta_p^2 = .083$, showed no significant differences after Bonferroni corrections (all $ps < .15$). Correct responses to the target happy face were thus equivalent when paired with neutral, sad, angry, fearful, disgusted, and surprised face distracters. The expression by task interaction was not significant ($F < 1$).

Table 1
Mean accuracy (in proportion of correct responses) and reaction times (in ms) in the saccade task (SRT) and the manual task (MRT), for each distracter expression category.

	Facial expression of distracter					
	Neutral	Sad	Angry	Fearful	Disgusted	Surprised
Saccade task						
Accuracy						
M	.827	.833	.804	.777	.798	.810
SD	.112	.096	.109	.116	.148	.132
RTs						
M	315^{a}	318^{a}	326^{ab}	330^{ab}	334^{b}	340^{b}
SD	64	69	62	68	66	59
Manual task						
Accuracy						
M	.961	.959	.940	.933	.922	.932
SD	.054	.044	.055	.069	.053	.057
RTs						
M	522^{a}	523^{a}	549^{ab}	561^{ab}	573^{b}	575^{b}
SD	60	77	88	82	73	85

Note. Means with different superscripts (horizontally) are significantly different; means with the same superscript or no superscript are equivalent. Superscripts with two letters indicate that the mean scores are equivalent to those having either one or the other letter, or both (e.g., scores with an 'ab' superscript indicate that they are equivalent to those with either 'a', 'b', or 'ab' superscripts).

In the analysis of mean *response latencies*, main effects of task, $F(1, 30) = 89.92$, $p < .0001$, $\eta_p^2 = .750$, and expression, $F(3.42, 102.72) = 9.77$, $p < .0001$, $\eta_p^2 = .246$, emerged, with no interaction, $F(3.42, 102.72) = 1.97$, $p = .12$, *ns*. Mean saccade latencies ($M = 327$ ms) were shorter than manual response latencies ($M = 550$ ms). Post hoc contrasts revealed that responses to the target happy face were faster when it had to be discriminated from neutral or sad faces than when paired with surprised or disgusted faces, with angry and fearful faces not differing significantly from the other four categories, both for the saccade task, $F(5, 75) = 6.00$, $p < .0001$, $\eta_p^2 = .286$, and the manual task, $F(5, 75) = 5.85$, $p < .0001$, $\eta_p^2 = .281$.

To determine the *minimum* saccade latency, the proportions of correct and incorrect responses for each 20-ms time bin were compared by means of pairwise *t* tests for dependent samples. Regarding *saccades*, the earliest time window where correct responses exceeded incorrect responses was the 200-ms bin for neutral faces, $t(15) = 2.41$, $p < .05$, the 220-ms bin for sad faces, $t(15) = 3.00$, $p < .01$, the 240-ms bin for angry faces, $t(15) = 2.52$, $p < .025$, the 260-ms bin for fearful faces, $t(15) = 2.24$, $p < .05$, and the 280-ms bin for disgusted, $t(15) = 2.71$, $p < .025$, and surprised, $t(15) = 2.63$, $p < .025$, faces. With respect to *manual* responses, the earliest time window was the 440-ms bin for neutral, $t(15) = 2.28$, $p < .05$, and sad, $t(15) = 2.42$, $p < .05$, faces, the 460-ms bin for angry faces, $t(15) = 2.09$, $p = .05$, the 480-ms bin for fearful, $t(15) = 2.71$, $p < .025$, and disgusted, $t(15) = 2.93$, $p < .025$, faces, and the 500-ms bin for surprised faces, $t(15) = 2.48$, $p < .025$. Latency distributions are shown in Fig. 2a–f.

3.2. Analysis of global image statistics and local visual saliency

Table 2 shows the absolute values of global luminance and energy of the faces, as well as local saliency of the eye, nose, and mouth regions, for each expression. For these variables, we analyzed the *difference scores* between happy faces and those in each of the non-happy categories (i.e., non-happy distracter–happy target scores), by means of one-way ANOVAs (6: non-happy expressions). The pixel-by-pixel image similarity correlation between the target and the distracter faces was also analyzed.

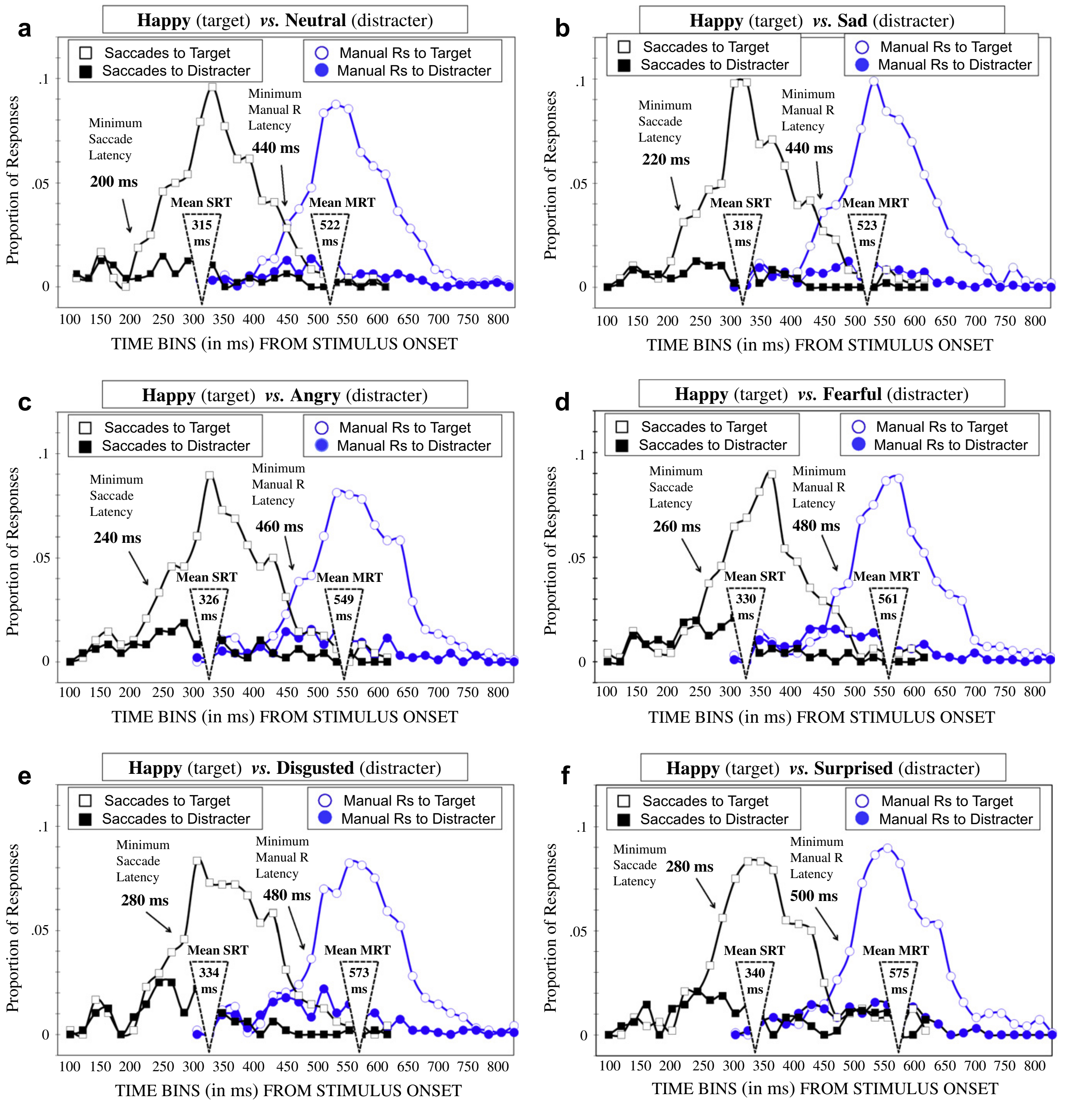

Fig. 2. (a–f) Saccadic and manual reaction time distribution of correct and erroneous responses (in proportion of responses) across 20-ms time bins for each combination of emotional expressions, in the saccade task and the manual response task. Arrows indicate the earliest time point when correct responses to the target (happy) face significantly exceeded incorrect responses to the distracter (non-happy) face. Triangles show the mean saccade (SRT) and manual (MRT) latencies of correct responses.

Luminance and energy difference (distracter–target) scores, and pixel-by-pixel image similarity correlation. Mean luminance and energy difference scores varied as a function of expression, $F(5, 174) = 20.83$, $p < .0001$, $\eta_p^2 = .374$ (luminance); $F(5, 174) = 8.85$, $p < .0001$, $\eta_p^2 = .203$ (energy). Luminance differences were greater for neutral faces than for all the other faces. Energy differences were greater for angry, sad, disgusted, and fearful faces than for surprised and neutral faces. Neutral faces had *higher* (instead of lower) luminance, $t(29) = 7.84$, $p < .0001$, and energy, $t(29) = 2.29$, $p < .05$, than happy faces. Image similarity between the target and the distracter faces was also affected by expression, $F(5, 174) = 3.27$, $p < .01$, $\eta_p^2 = .086$, with the only significant contrast indicating that similarity was higher for neutral than for surprised faces. These findings are contrary to a merely low-level image account of discrimination between expressions (see Section 4).

Visual saliency difference (distracter–target) scores. No significant differences appeared as a function of expression in saliency values for the eye, $F(5, 174) = 1.39$, $p = .23$, *ns*, or the nose ($F < 1$) region. However, a reliable effect emerged for the mouth region, $F(5, 174) = 7.38$, $p < .0001$, $\eta_p^2 = .175$. Importantly, post hoc comparisons revealed that the neutral and the sad face differences from happy faces were greater than those of the surprised and the disgusted faces; the angry and the fearful faces were somewhere in

Table 2
Mean values for perceptual and semantic characteristics of faces in each expression category.

	Facial expression						
	Happy	Neutral	Sad	Angry	Fearful	Disgusted	Surprised
Perceptual							
Global properties							
Luminance	71.3	76.0^{a}	69.9^{b}	70.3^{b}	69.0^{b}	68.4^{b}	70.2^{b}
Energy ($\times 10^{-7}$)	2,310	2,349^{a}	2,187^{b}	2,202^{b}	2,220^{b}	2,185^{b}	2,300^{a}
Correlation	–	.915^{a}	.905^{ab}	.900^{ab}	.897^{ab}	.909^{ab}	.885^{b}
Local saliency							
Eye region	0.35	1.48	2.41	2.22	2.16	2.28	3.26
Nose region	0.15	1.20	1.35	1.25	0.81	0.79	1.21
Mouth region	9.10	2.62^{bc}	2.05^{bc}	4.20^{b}	4.41^{b}	5.85^{ab}	6.28^{ab}
Semantic							
Category membership							
Prob. "Happy"	98.00	2.67^{b}	1.03^{c}	0.87^{c}	2.03^{b}	2.83^{b}	4.87^{a}
RT "No-happy"	546	620^{ab}	597^{b}	587^{b}	654^{a}	643^{a}	656^{a}
Emotional valence							
Prob. "Pleasant"	96.33	55.77^{a}	3.40^{cd}	1.53^{d}	4.73^{c}	6.13^{c}	39.60^{b}
RT "No-pleasant"	577	724^{ab}	644^{cd}	618^{d}	708^{ab}	692^{bc}	742^{a}

Note. Prob. "Happy": probability (in percentage) of responding "happy". RT "No-happy": latencies (in ms) in correctly responding that a non-happy face was "not" happy (or that the happy face was happy). Prob. "Pleasant": probability (in percentage) of responding "pleasant". RT "No-pleasant": latencies (in ms) in responding that a non-happy face was "not" pleasant (or that the happy face was pleasant). Means with different superscripts (horizontally) are significantly different; means with the same superscript or no superscript are equivalent. Superscripts with two letters indicate that the mean scores are equivalent to those having either one or the other letter, or both (e.g., scores with an 'ab' superscript indicate that they are equivalent to those with 'a', 'b', or 'ab' superscripts).

between, with no significant differences from the other expression categories. Furthermore, the mouth of happy faces was more salient than that of all the other expressions, $F(6, 209) = 21.57$, $p < .0001$, $\eta_p^2 = .389$.

3.3. Analysis of semantic attributes of the faces: category membership and emotional valence

Table 2 shows the mean accuracy scores, as well as the speed-based category membership and emotional valence indices, which were analyzed by means of one-way ANOVAs (6: non-happy expressions).

Category membership. The probability that non-happy faces were judged to be happy varied as a function of expression, $F(5, 145) = 25.19$, $p < .0001$, $\eta_p^2 = .465$. Angry and sad faces were the most different from happy faces (i.e., the least errors), followed by fearful, neutral, and disgusted faces, with surprised faces being most likely classified as happy. Response times were also affected by expression, $F(5, 145) = 4.65$, $p < .01$, $\eta_p^2 = .138$. Angry and sad faces were judged as not happy faster than the other face categories, followed by neutral faces, with disgusted, fearful, and surprised faces being rejected as happy more slowly. Compared with all the non-happy faces, the truly happy faces were more likely to be considered as happy, $F(6, 174) = 8629.95$, $p < .0001$, $\eta_p^2 = .997$.

Affective valence. Expression influenced the probability that non-happy faces were evaluated as pleasant, $F(5, 145) = 154.88$, $p < .0001$, $\eta_p^2 = .842$, and the time to respond that they were not pleasant, $F(5, 145) = 16.33$, $p < .0001$, $\eta_p^2 = .360$. Angry faces were the least likely to be assessed as pleasant, followed by sad, fearful, and disgusted faces; surprised and neutral faces were considered as pleasant to a significant extent. Regarding response times, angry faces were particularly fast to be evaluated as not being pleasant, followed by sad, disgusted, fearful, and neutral faces, with responses being slowest for surprised faces. The truly happy faces were more likely to be judged as pleasant than all the non-happy expressions, $F(6, 174) = 425.27$, $p < .0001$, $\eta_p^2 = .936$.

3.4. Relationship between actual discrimination performance and perceptual and semantic factors

To examine the contribution of perceptual and semantic factors to expression discrimination, we computed non-parametric Spearman's ρ correlations between these variables and response accuracies and latencies in the 2AFC tasks (see Table 3). *Visual saliency* was the best predictor of discrimination performance. Greater differences in saliency of the mouth region of happy targets, in relation to the paired distracters, were associated with higher recognition accuracy and shorter mean and minimum latencies across both the saccade and the manual response tasks. Furthermore, the correlations between visual saliency and discrimination latencies remained significant after the contribution of all the other predictors was controlled by means of partial correlations (all *rs* $\geqslant$ .88, *ps* < .05, two-tailed). There was a non-significant tendency for greater differences in *luminance*, and for lower *image correlation*, to be related to higher accuracy and faster correct responses.[5] The trend approached significance for *category membership*, with faster responding that non-happy faces were "not" happy (hence greater category differences) being related to faster discrimination. In contrast, no index of *emotional valence* was related to discrimination performance. In Fig. 3, these relationships are shown graphically for a selected and representative group of variables.

4. Discussion

We investigated the time course of discrimination between happy and other facial expressions by measuring saccadic and manual response latencies in a forced-choice task with two faces presented simultaneously. A consistent pattern emerged for both response modalities. Happy faces were discriminated most efficiently (i.e., most accurately *and* rapidly) from neutral and sad faces, and least efficiently from disgusted and surprised faces, with discrimination from angry and fearful faces being of intermediate difficulty. The minimum latencies of correctly directed saccades indicated that accurate discrimination was possible as early as 200 ms from stimulus onset. Visual saliency of the mouth region was a critical factor in facilitating discrimination: The more salient the mouth was in happy face targets vs. paired non-happy distracters, the faster discrimination was.

[5] Analysis of *local* luminance, energy, and image similarity for the eye, nose, and mouth regions did not yield any significant correlations with discrimination performance. This shows that such image statistics do not account for the significant effects of saliency.

Table 3
Spearman rho correlations (two-tailed) between measures of discrimination performance and measures of perceptual and semantic stimulus attributes.

	Saccade response task (SRT)			Manual response task (MRT)		
	Accuracy	Mean RT	Min. RT	Accuracy	Mean RT	Min. RT
Perceptual						
Global properties						
Luminance	.60	−.37	−.49	.60	−.60	−.49
Energy	.20	−.20	−.32	.43	−.43	.15
Correlation	.31	−.66	−.55	.43	−.43	−.59
Local saliency						
Eye region	−.26	−.60	−.55	−.49	−.49	−.28
Nose region	−.61	.37	.46	−.54	.54	.62
Mouth region	.60	−.94**	−.93**	.89*	−.89*	−.93**
Semantic						
Category membership						
Prob. "Happy"	.09	.60	.58	−.54	.54	.52
RT "No-happy"	−.37	.71#	.67	−.60	.60	.74#
Emotional valence						
Prob. "Pleasant"	.14	.09	.06	−.03	.03	.09
RT "No-pleasant"	.03	.31	.23	−.14	.14	.31

Note. Min. RT: minimum latencies. Prob. "Happy": probability of responding "happy". RT "No-happy": latencies (in ms) in correctly responding that a non-happy face was "not" happy (or that the happy face was happy). Prob. "Pleasant": probability of responding "pleasant". RT "No-pleasant": latencies (in ms) in responding that a non-happy face was "not" pleasant (or that the happy face was pleasant).
$p \leqslant .10$ (two-tailed; $p \leqslant .05$, one-tailed).
* $p < .025$ (two-tailed).
** $p < .01$ (two-tailed).

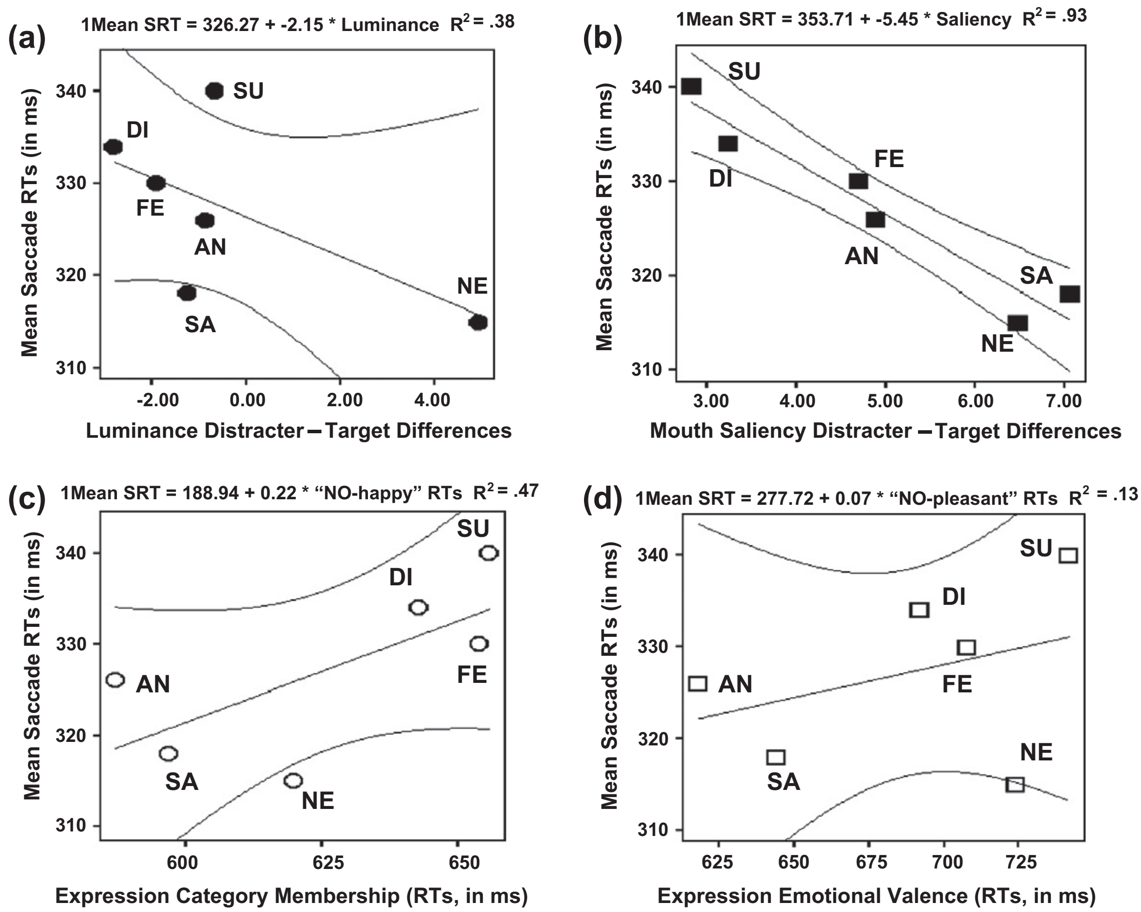

Fig. 3. Linear regressions between (1) distracter (non-happy)–target (happy) face differences in (a) global face luminance and (b) mouth region saliency, as well as (c) category membership (RTs for "No-happy" responses), and (d) emotional valence (RTs for "No-pleasant" responses), and (2) mean saccade latencies to the target, with prediction of the mean at 95% confidence interval. NE: neutral; SA: sad; AN: angry; FE: fearful; DI: disgusted; SU: surprised, distracters.

4.1. Time course of discrimination between happy and non-happy faces

Saccade latencies revealed that the earliest time point at which happy faces could be reliably discriminated from other expressions varied between 200 and 280 ms. In saccadic response tasks faces are distinguished from non-face objects as early as 100–110 ms (with mean saccade latencies around 140 ms), and such discrimination of faces is faster than that of other categories like animals or vehicles (Crouzet et al., 2010). Faces may thus have a privileged access to the cognitive system due to the adaptive importance of recognition of other people's identities, motivational states and intentions. The minimum saccade latencies that we found for conscious expression categorization are close to the timing of the electrophysiological N170 component of brain activation, as assessed by event-related-potentials (ERPs). The N170 potential is considered as the earliest neural signature of category-specific face processing (Bentin, Allison, Puce, Perez, & McCarthy, 1996). Although N170 is thought to reflect structural encoding of faces, it is also modulated by affect, with emotional faces triggering larger N170 amplitudes than neutral faces (see Eimer & Holmes, 2007). If we assume a delay of about 20–25 ms for the target-guided saccades to be programmed (Schiller & Kendall, 2004), this implies that happy faces can be recognized as different from other expressions within a 175–255-ms post-stimulus range, which is just a little above the N170 latency.

Responses were on average 220 ms faster for saccadic than for manual responses. However, discrimination accuracy was higher for manual (94% accuracy) than for saccadic (81%) responses. The higher accuracy in the manual task is consistent with the hypothesis that the decision is based on additional information not yet available when the decision is made in the saccadic task (see Bannerman et al., 2009; Crouzet et al., 2010), or that the same visual input receives additional processing in the manual task. The data involving faster, yet less accurate, saccadic responses suggest that the decision is made earlier in the saccade than in the manual task, rather than the faster (saccade) or slower (manual) responses reflecting mere speed of the post-decision motor mechanisms. Nevertheless, virtually the same response latency distribution—although shifted on the temporal axis—emerged for both response systems. The expressions that were discriminated faster in the saccade task were also discriminated faster in the manual task. In addition, the relationship between the perceptual stimulus properties and discrimination performance was equivalent for the saccade and the manual task. This suggests that some critical facial information that is accessed very early in the saccade task is also re-used and makes a significant contribution to later discrimination decisions in the manual task.

4.2. Role of perceptual and semantic factors in facial expression discrimination

What information is accessed early and drives expression discrimination? To answer this question, we compared the pattern of discrimination performance with the pattern of differences in perceptual and semantic face characteristics. For both the saccadic and the manual task, there was a discrimination advantage for happy vs. neutral and sad faces, and a disadvantage for happy vs. disgusted and surprised faces, with an intermediate difficulty for angry and fearful faces. Accordingly, if discrimination depends on bottom-up processes relying on perceptual information, then differences in global luminance, energy, image similarity, and/or local visual saliency between happy and non-happy faces should be greater for neutral and sad faces, and lower for disgusted and surprised faces, with angry and fearful faces in between. In contrast, if discrimination involves top-down processes relying on semantic information, then differences in perceived category membership and/or emotional valence between the happy and the non-happy faces would be greater for neutral and sad faces, and lower for disgusted and surprised faces, with angry and fearful faces in between.

Regarding the role of perceptual factors, luminance, energy, and image similarity did not make a significant contribution to discrimination efficiency. First, happy faces did not generally have higher luminance or energy than other faces. Second, neutral faces (whose discrimination was fastest) actually had higher rather than lower luminance and energy than happy faces. And, third, image similarity was higher between neutral and happy faces than between surprised (whose discrimination was slowest) and happy faces. The opposite pattern would have been predicted if differences in such low-level image factors accounted for discrimination efficiency. Presumably, luminance, energy, and image similarity made no contributions to discrimination because this involved *identifying* the target face *expression*, of which the mere global image statistics are not informative. In contrast, visual saliency of the mouth region was highly related to discrimination efficiency. The typical smile in the mouth region is a *distinctive* feature of happy expressions (Calvo & Marrero, 2009; Kohler et al., 2004), and therefore provides reliable information for expression discrimination (see below). It is thus reasonable that the more salient such a distinctive feature is in a target face in relation to that of a distracter (as was the case for neutral and sad faces), the faster discrimination will be, and vice versa (for disgusted and surprised faces).

Regarding semantic factors, the relationship between category membership and discrimination performance only approached significance, and correlations between affective valence and discrimination indices were close to zero. This raises the issue of whether categorical or affective information was processed. One argument and two types of data are relevant to address this issue. First, a discrimination task—in which a target face belonging to an *expression category* has to be *identified* against a distracter face of another category—must presumably rely on information that defines such categories. Second, differences in ERP waveforms between neutral and emotional faces occur around 120–180 ms post-stimulus (see Eimer & Holmes, 2007), which suggest that some type of semantic or affective (not merely physical) information of facial expressions is obtained early. And, third, affective priming has been found following happy, or liked, faces at prime-probe SOAs of 300–750 ms (Calvo et al., 2010; Lipp et al., 2009; Nummenmaa, Peets, & Salmivalli, 2008). This suggests that affective information is indeed extracted from emotional faces, albeit with some delay. In the current study, it is possible that categorical and emotional information were *obtained* at some point, but were only minimally *used* due to their being overshadowed by the earlier extracted, and simpler to be managed, visual saliency information, which then was retained also for later discrimination stages.

4.3. From visual saliency to processing of significance in discrimination of facial expressions

How can visual saliency of the mouth region of happy faces account for their efficient discrimination from non-happy faces? We propose a two-stage processing mechanism. A first stage would involve quick featural detection and analysis of visually salient regions, by relying only on the physical attributes of the faces, with no processing of their significance. The minimal impairment of detection in visual search tasks (Calvo & Nummenmaa, 2008) and recognition in categorization tasks (Calvo et al., 2010; Leppänen & Hietanen, 2007; McKelvie, 1995) when happy faces are presented upside-down—relative to when presented upright—is consistent with such a featural conceptualization. Furthermore, saliency makes the happy mouth easily accessible to extrafoveal vision (Goren & Wilson, 2006) and automatically attracts overt attention (Calvo & Nummenmaa, 2008). Extrafoveal vision is important for discrimination tasks where the to-be discriminated, simultaneously presented, faces appear slightly away from a central fixation point. In such conditions saliency ensures that a prominent feature will be perceived before it is fixated or in the absence of fixations.

In a second stage, the earlier detection of a salient happy mouth would be used for expression recognition and affect retrieval, thus involving semantic processes. Expression information can be reliably extracted from the happy mouth shape (i.e., upturned lip corners, generally surrounding teeth) because this shape is uniquely and systematically associated with happy expressions (Calvo & Marrero, 2009; Kohler et al., 2004). This makes the mouth a highly distinctive feature. The mouth has indeed been found to be the most important, necessary and sufficient, element in the recognition of happy faces (Kontsevich & Tyler, 2004; Leppänen and Hietanen, 2007). The salient smiling mouth would thus be used as a diagnostic cue to retrieve the categorical and affective associations of the happy expression. This explains why visual saliency is crucial not only for the first stage involving perceptual analysis of physical attributes of the faces, but also for the second one involving processing of their significance: The most salient feature is also highly informative.

4.4. Conclusions

Happy faces can be discriminated from other expressions between 200 and 280 ms (minus 20–25 ms for saccade programming), as assessed by saccade latencies, with discrimination efficiency being higher for some—notably neutral and sad—expressions than others—especially disgusted and surprised. Visual saliency of the mouth region is a critical factor in facial expression discrimination. The more salient the mouth is in a target happy face relative to a distracter non-happy face, the more efficient is discrimination. Although semantic—both categorical and affective—information is probably obtained from the faces early, such information does not make a direct contribution to the discrimination process. Rather, semantic information probably becomes dependent on visual saliency, with saliency keeping accessible a highly diagnostic facial feature such as the mouth region, and significance being extracted from this salient feature. At the time of making a discrimination decision, saliency rather than significance is the dominant criterion.

Acknowledgments

This research was supported by Grant PSI2009-07245 from the Spanish Ministry of Science and Innovation, and the Canary Agency for Research, Innovation and Information Society (NEUROCOG Project), and the European Regional Development Fund, to MGC, and by the Academy of Finland Grant #217995 to LN, and the AivoAALTO grant from the Aalto University.

References

Alpers, G. W., & Gerdes, A. B. M. (2007). Here is looking at you: Emotional faces predominate in binocular rivalry. *Emotion, 7*, 495–506.

Bannerman, R. L., Milders, M., & Sahraie, A. (2009). Processing emotional stimuli: Comparison of saccadic and manual choice-reaction times. *Cognition and Emotion, 23*, 930–954.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*, 551–565.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expression. *Vision Research, 41*, 1179–1208.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods, 40*, 109–115.

Calvo, M. G., & Marrero, H. (2009). Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cognition and Emotion, 23*, 782–806.

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. *Journal of Experimental Psychology: General, 137*, 471–494.

Calvo, M. G., & Nummenmaa, L. (2009). Eye-movement assessment of the time course in facial expression recognition: Neurophysiological implications. *Cognitive, Affective, and Behavioral Neuroscience, 9*, 398–411.

Calvo, M. G., Nummenmaa, L., & Avero, P. (2010). Recognition advantage of happy faces in extrafoveal vision: Featural and affective processing. *Visual Cognition, 18*, 1274–1297.

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology, 58A*, 1173–1197.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades towards faces: Face detection in just 100 ms. *Journal of Vision, 16*, 1–17.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*, 15–31.

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

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences, 16*, 553–610.

Goren, D., & Wilson, H. R. (2006). Quantifying facial expression recognition across viewing conditions. *Vision Research, 46*, 1253–1262.

Itti, L. (2006). Quantitative modeling of perceptual salience at human eye position. *Visual Cognition, 14*, 959–984.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

Juth, P., Lundqvist, D., Karlsson, A., & Öhman, A. (2005). Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd. *Emotion, 5*, 379–395.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*, 1762–1776.

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

Kontsevich, L. L., & Tyler, C. (2004). What makes Mona Lisa smile? *Vision Research, 44*, 1493–1498.

Leppänen, J., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research, 69*, 22–29.

Leppänen, J., & Hietanen, J. K. (2007). Is there more in a happy face than just a big smile? *Visual Cognition, 15*, 468–490.

Lipp, O., Price, S. M., & Tellegen, C. L. (2009). No effect of inversion on attentional and affective processing of facial expressions. *Emotion, 9*, 248–259.

Loughead, J., Gur, R. C., Elliott, M., & Gur, R. E. (2008). Neural circuitry for accurate identification of facial emotions. *Brain Research, 1194*, 37–44.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces – KDEF*. CD-ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, Stockholm, Sweden. ISBN 91-630-7164-9.

Maxwell, J. S., & Davidson, R. J. (2004). Unequally masked: Indexing differences in the perceptual salience of "unseen" facial expressions. *Cognition and Emotion, 18*, 1009–1026.

McKelvie, S. J. (1995). Emotional expression in upside-down faces: Evidence for configurational and componential processing. *British Journal of Social Psychology, 34*, 325–334.

Milders, M., Sahraie, A., & Logan, S. (2008). Minimum presentation time for masked facial expression discrimination. *Cognition and Emotion, 22*, 63–82.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research, 39*, 3824–3833.

Nothdurft, H. C. (2006). Salience and target selection in visual search. *Visual Cognition, 14*, 514–542.

Nummenmaa, L., Peets, K., & Salmivalli, C. (2008). Automatic activation of adolescents' peer-relational schemas: Evidence from priming with facial identity. *Child Development, 79*, 1659–1675.

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, 36*, 634–638.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modelling the role of salience in the allocation of overt visual attention. *Vision Research, 42*, 107–123.

Schiller, P. H., & Kendall, J. (2004). Temporal factors in target selection with saccadic eye movements. *Experimental Brain Research, 154*, 154–159.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. (2006). Contextual guidance of eye movements in real-world scenes: The role of global features in object search. *Psychological Review, 113*, 766–786.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*, 242–249.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *The Quarterly Journal of Experimental Psychology, 59*, 1931–1949.

Walther, D., & Koch, C. (2006). Modelling attention to salient proto-objects. *Neural Networks, 19*, 1395–1407.

Yoon, K. L., Hong, S. W., Joorman, J., & Kang, P. (2009). Neural circuitry for accurate identification of facial emotions. *Emotion, 9*, 172–182.