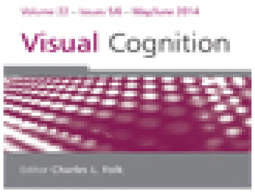

VISUAL COGNITION

ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: https://www.tandfonline.com/loi/pvis20

When a face is (or is not) more than the sum of its features: Configural and analytic processes in facial temporal integration

David Anaki & Morris Moscovitch

To cite this article: David Anaki & Morris Moscovitch (2007) When a face is (or is not) more than the sum of its features: Configural and analytic processes in facial temporal integration, VISUAL COGNITION, 15:6, 741-763, DOI: 10.1080/13506280601130883

To link to this article: https://doi.org/10.1080/13506280601130883

Published online: 20 Jul 2007.

VISUAL COGNITION, 2007, 15 (6), 741–763

Psychology Press
Taylor & Francis Group

When a face is (or is not) more than the sum of its features: Configural and analytic processes in facial temporal integration

David Anaki

Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Canada

Morris Moscovitch

Rotman Research Institute, Baycrest Centre for Geriatric Care, and Department of Psychology, University of Toronto, Toronto, Canada

To investigate temporal integration in face recognition, top and bottom halves of pictures of famous people were presented sequentially, either upright or inverted, with varying temporal intervals between the two halves. The inversion effect, a marker of configural processing, was comparable across 0–400 ms intervals, but decreased at intervals exceeding 400 ms (Exp. 1). When an interfering stimulus appeared during the interval between the two face parts (Exp. 2), it disrupted the integration of the parts but not their perception. This is the first report of such an effect. Thus, performance equalled the combined accuracy of each part when presented alone, which in turn was worse than when they were integrated. Our findings indicate that (a) configural processing of faces depends on integration of face parts that are maintained temporarily in a visual buffer; (b) without integration, identification depends on recognition of individual parts whose contributions are additive; and (c) an interfering visual stimulus can obstruct integration, but leaves perception of individual parts intact. The ability to integrate temporally separated face parts into a unified representation is discussed in light of theories of face perception and temporal integration.

Please address all correspondence to *either* David Anaki, Department of Psychology, Hebrew University, Mount Scopus Jerusalem, Israel, 91905 (e-mail: danaki@mscc.huji.ac.il), *or* Morris Moscovitch, Department of Psychology, University of Toronto, 100 St. George St., Toronto, ON, Canada M5S 3G3 (E-mail: momos@psych.utoronto).

We express our deep appreciation to Philippe Schyns and three anonymous reviewers for their helpful comments and suggestions on earlier versions of this paper. We thank Marilyne Ziegler, Irina Nica, and Aliza Goodman for their help in programming, running, and analysing the experiments.

http://www.psypress.com/viscog DOI: 10.1080/13506280601130883

The rich visual information contained in complex stimuli such as objects, scenes, and faces, often precludes their complete and instantaneous perception. Moreover, many stimuli in natural contexts are partly occluded by opaque surfaces, and different parts of them are exposed to the perceiver gradually over time. Piecemeal registration, in which newly perceived visual elements are incorporated into a semiconstructed representation, therefore, is required. This temporal integration mechanism is comprised of both the transitory maintenance of temporally separated images, as well as their combination into a unified percept. The temporal dynamics involved in the perception and recognition of facial stimuli is the focus of the present study.

Temporal integration has been extensively studied in the domain of pattern, scene and object recognition (for reviews see Brockmole, Wang, & Irwin, 2002; Henderson & Hollingworth, 2003; Hollingworth, 2006; Irwin, 1992; Irwin & Andrews, 1996; Simons & Rensink, 2005). Few studies, however, explored temporal integration in face perception (notable exceptions are Singer & Sheinberg, 2006; Vinette, Gosselin, & Schyns, 2004; Wallis & Bülthoff, 2001). Indeed, a few studies recently used the change detection paradigm, in which participants were asked to detect changes in faces across views, to investigate whether accurate representations of perceived facial stimuli can be retained in memory (Barton, Deepak, & Malik, 2003; Buttle & Raymond, 2003; Davies & Hoffman, 2002; Humphreys, Hodsoll, & Campbell, 2005; Palermo & Rhodes, 2003; Ro, Russell, & Lavie, 2001). However, it is important to dissociate between the *process* of temporal integration and the memory *structure* which allows this process to ensue. The change detection paradigm mainly addresses the latter issue but not the former.

The scarce interest in temporal integration in the face processing domain may be unjustified. Different processes are claimed to underlie face and object recognition, and the nature of temporal integration of faces can differ from that of objects. Various theoretical accounts argue, for example, that object recognition is achieved through analytic or part-based processes, whereas holistic or configural processes trigger face recognition (Farah, 2004; Maurer, Le Grand, & Mondloch, 2002; Moscovitch, Winocour, & Berhmann, 1997; Peterson & Rhodes, 2003). Although the debate concerning the exact definition of these terms is not resolved, there is some agreement that in part-based processes the object is identified on the basis of its elementary parts, while in configural or holistic processing part-decomposition is absent or minimal. Rather, the stimulus is apprehended as a perceptual *gestalt*, or as a representation derived from computation of the spatial-relational information of the facial features (Peterson & Rhodes, 2003). The finding that inversion interferes more with face than object identification (Yin, 1969) is traditionally ascribed to the disruption in extracting configural or holistic information from the inverted stimuli and reverting to analytic processes. Thus, the necessity and nature of temporal integration in face perception and recognition, and its influence on holistic or configural processing, is an issue that is yet to be addressed.

In a recent study (Anaki, Boyd, & Moscovitch, 2007), we investigated the characteristics and time boundaries of facial temporal integration by presenting faces of famous and nonfamous people, segmented horizontally into three parts and presented in sequence. Each part contained one salient facial feature: The top part consisted of the upper part of the head, and was sliced just below the eyes. The middle part included the nose, and was sliced just above the lips. The lower part contained the mouth, and was cut below the chin. The interstimulus interval (ISI) between the parts varied between 0 and 700 ms, and the faces were presented either upright or inverted. The findings revealed that even when face parts were separated by a 200 ms interval between each facial slice, an inversion effect was obtained, which was comparable to that obtained with faces presented as a whole. A significant decrease in the inversion effect, and at times even its elimination, was observed only when the temporal interval between the face parts surpassed 200 ms.

On the basis of these findings we concluded that temporal integration is a pervasive process in face perception. Facial parts can be stored temporarily in a short-term visual buffer and be merged with incoming additional face parts to form a whole face. These integration processes can occur only within a limited temporal framework of approximately 450 ms, which is the time between the onset of the first face part to the offset of the third face part in the 200 ms condition. More importantly, the emergence of an inversion effect, comparable to that obtained in whole face presentation, indicates that the integration of the face segments is performed in a manner sensitive to the holistic or configural aspects of the face. Thus, in addition to the encoding of facial features from the three facial parts, supplementary visual information is extracted from the unified representation at short intervals. The nature of this additional information could be either the computation of relational information between the features of the face across the parts, or the creation of a template. However, when the interval between the parts is long, identification relies only on the information embedded in the separate parts.

Although the study described above has shown that temporal integration exists in the face perception domain, as in the object perception realm, the results from the two areas of research are not completely convergent. Specifically, previous research had identified four different memory stores in which integration could occur: Visible and informational persistence (known also as iconic memory), visual short- term memory (VSTM), and visual long-term memory (VLTM). Visible persistence is a sensory trace that is phenomenologically perceivable 80–100 ms after stimulus onset. Informational persistence is a more durable visible trace (150–300 ms poststimulus offset), which contains rich visual information, yet lacks the vivid visual experience that characterizes visible persistence (Coltheart, 1980). VSTM abstracts the visual information and maintains it for several seconds (Phillips, 1974). It is less vulnerable to visual interference than visible and informational persistence, but its capacity is limited to three–four objects (Luck & Vogel, 1997) and it lacks detailed spatial information (Irwin, 1991). Finally, the qualities of the visual representations in VLTM are similar to those of VSTM, but the capacity of VLTM is greater and extends over long periods of time (Hollingworth, 2004).

Studies have shown that both visible persistence (Coltheart, 1980; Irwin & Yeomans, 1986) and VSTM (Brockmole et al., 2002; Hollingworth, Hyun, & Zhang, 2005), as well as VLTM, could support temporal integration, but not informational persistence (e.g., Di Lollo, 1980; Irwin, 1992, 1996). However, in our previous study (see above), face integration was apparent even when the interval between the first and last parts was 450 ms long, which is within the timeframe of informational persistence. In addition, the face segments were not integrated when the interval between each part was 700 ms although previous research had documented integration in VSTM.

One major difference between our previous study with faces and studies investigating temporal integration in pattern completion is that in the former the faces were divided into three parts and not two, as in the latter. As a result, the value of 450 ms during which temporal integration was apparent for face parts was not an ISI value but the total time required to present all the face parts, while the actual ISI between each segment was 200 ms. This precludes a full comparison between faces and objects, and also prevents one from accurately delineating the time course of facial temporal integration. One could argue, for example, that the time course of facial temporal integration is similar to that of pattern temporal integration, and that the performance in the previous study resulted from the integration of only two adjacent parts (separated by an interval of 200 ms). These two parts consisted of sufficient detailed information to allow identification.

In addition, the maximal ISI used in our previous study was only 700 ms. Although no integration was observed in this condition, the possibility that integration could be supported by VSTM cannot be ruled out without exploring longer ISIs. To that end, faces were partitioned in the present study into two parts, and were presented with a variable interval between them, ranging between 0 and 1600 ms. In Experiment 1, we explored the influence of the ISI on the inversion effect in an attempt to determine what are the time limits of temporal integration in faces. In Experiment 2, the nature of the short-term memory buffer was examined by introducing visual interfering stimuli after each segment and exploring the influence of this distractor on the integration process.

EXPERIMENT 1

In Experiment 1 we attempted to identify possible factors which might account for the discrepancy found between face and pattern integration. First, to explore the reason why informational persistence supported facial integration while it failed to support pattern integration, the facial stimuli were designed in a similar mode as pattern stimuli, namely, two segments instead of three. Second, to investigate the possibility that VSTM plays a role in temporal integration of faces, but at a later stage than that of pattern and object integration, the ISI condition was lengthened to 1600 ms.

Method

Participants. Eighty-four undergraduate students at the University of Toronto participated in the experiment for course credit or monetary reimbursement. All participants had normal or corrected-to-normal vision.

Materials. The critical stimuli consisted of 120 frontal-view images of famous people from different fields (e.g., actresses, politicians, etc.) downloaded from public accessible sites in the web. Pictures were converted into a 256 grey-level format (74 dpi) and subtended 5.6° in width and 7.2° in height.

Each face was divided into a top part, consisting of the upper part of the head sliced below the eyes, and a lower part containing the nose and the mouth and cut below the chin (Figure 1). Each part was presented for 17 ms with an interval of 0, 50, 100, 200, 400, 800, or 1600 ms between each part. The parts were always presented in their correct location (i.e., top part above lower part in the upright condition, but below the lower part in the inverted condition), with the top part always presented first. In our previous study (Anaki et al., 2007) we found that the order of part presentation had no effect across ISI and orientation conditions.

Procedure. Twelve participants were randomly assigned to each of the seven ISI conditions. Although interleaving the ISI conditions in a within-subject design would have prevented possible observer's biases, pilot studies have shown that it would have required reduction of the number of trials per condition, resulting in decreased power. Stimuli were displayed on an IBM colour monitor controlled by E-Prime software (Psychological Software Tools, Inc., 2000), implemented in an IBM PC-compatible computer.

Each trial began with a 1000 ms centrally presented fixation cross. Following its offset, the two parts of the face appeared either in an upright or inverted orientation with varying ISIs between the parts. The orientation of each face was determined randomly for each participant. After the

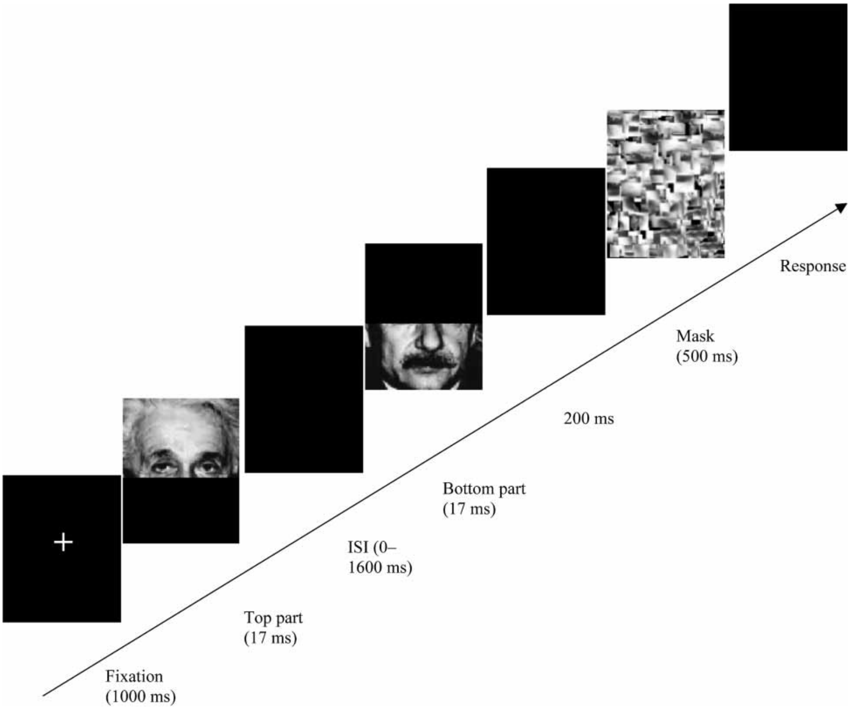

Figure 1. Sequence of events in a typical trial in Experiment 1. A target face was presented centrally either as a whole or segmented into parts, with varying interval lengths between the parts, followed by a mask.

presentation of the second part of the face a black screen appeared for 200 ms followed by a 500 ms distractor, created from minute pieces of facial features. Participants responded by keypress whether the face was familiar or not. Participants were instructed to respond that a face was familiar only if they had seen this face in the past *and* could provide specific information about the person. Specific information was defined as the name of the person (e.g., Prince Charles) or unique details about him or her (e.g., *Princess Diana's husband and heir to the English throne*). Participants were told that general information (e.g., *an actor*, *a politician*) was not sufficient to render a familiar response.

The response initiated the appearance of a unique number, designated for each face, at the lower left side of the screen. Participants then wrote their response (the name of the person or any identifying information about him or her) on an answer sheet, which was supplied to them before the experiment. They then proceeded to the next trial by pressing the "z" key. Sixty faces were presented upright and sixty faces were shown inverted. No face was repeated twice and the same 120 faces were shown to all participants.

After seeing the faces in parts, participants were presented with all the faces again and responded in the same manner as before. The faces were presented as a whole, in an upright orientation and shown for unlimited time. Participants were asked to refrain from any modifications of the responses to the first part of the session if their response to the second part were different. This latter part was administered to differentiate between faces that were not identified due to the experimental manipulation and faces that were unknown to the participant.

Results

The ability of each participant to identify the faces was influenced both by the condition in which he or she was tested *and* his or her familiarity with the faces. In order to control for these latter effects, the second part of the experiment was introduced, in which each face appeared without any constraints. Using these scores we computed the accuracy of each participant in the first part of the experiment relative to the total number of faces identified in the second part of the experiment (i.e., P[correct in the first part in condition ISI $\times$ | correct in the second part]). Thus, for example, if a participant identified 40 upright faces (out of 60) in the first part of the experiment and 48 faces in the second part, his or her accuracy was computed as .83 and not .67. Prior to that, an ANOVA was conducted to ensure that the accuracy level in the second part of the experiment was equivalent across the different ISIs, $F(6, 77) = 1.09$, $MSE = 113.79$, $p > .37$. The participants' responses were scored by two experimenters who determined that an answer was correct if a specific name was supplied (e.g., *Clint Eastwood*), or detailed information, attesting to the familiarity with this person, was provided (e.g., *the good guy in the movie The Good, the Bad and the Ugly*). Both experimenters were blind to the ISI condition from which a specific response sheet was taken. In the majority of the responses (89%), a specific name was given. The interjudge agreement in cases where information, but no names, was supplied was .94.

As can be seen in Figure 2, upright faces (65%) were recognized better than inverted faces (17%). More importantly, the difference in accuracy between upright and inverted faces (i.e., inversion effect) was equivalent in short and intermediate intervals (0–400 ms), and was diminished considerably only in the long ISIs (i.e., 800 and 1600 ms). This observation was confirmed in a two-way ANOVA, which yielded main effects of orientation and ISI, $F(6, 77) = 1301.35$, $MSE = 0.007$, $p < .0001$, and $F(6, 77) = 5.71$, $MSE = 0.019$, $p < .0001$, respectively, and also an interaction between

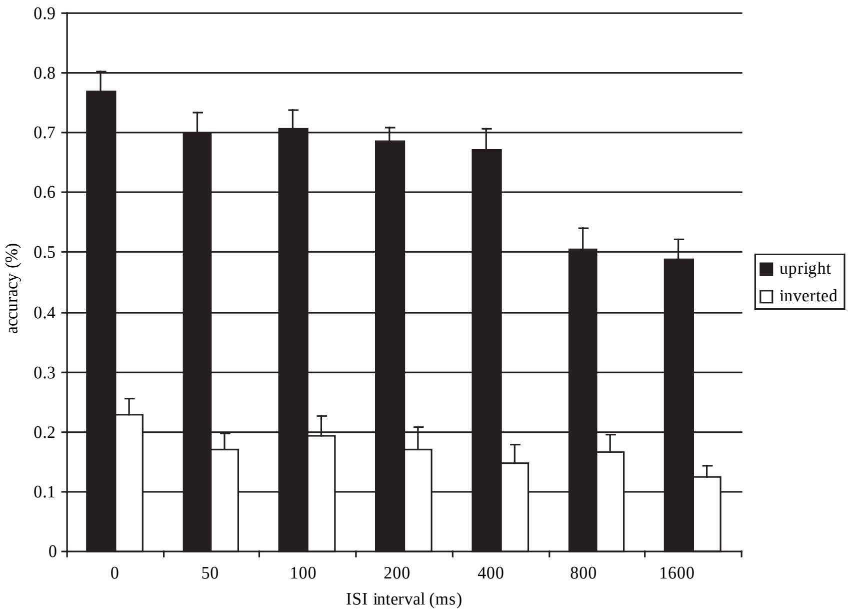

Figure 2. Percentage of faces identified correctly in Experiment 1 as a function of interval between face parts (ISI) and orientation (upright, inverted).

orientation and ISI, $F(6, 77) = 6.02$, $MSE = 0.007$, $p < .0001$. Post hoc tests (Bonferroni) revealed that the comparable inversion effect in the 800 and 1600 ms ISIs differed from the inversion effects in all the other conditions, the latter conditions not differing among themselves. Probing the source of the diminution of the inversion effect showed that while for the inverted faces a similar level of accuracy was obtained across all the ISI conditions, $F(6, 77) = 1.14$, $MSE = 0.011$, $p > .35$, a different pattern was found for upright faces, $F(6, 77) = 9.48$, $MSE = 0.015$, $p < .0001$: While the accuracy level of all short and intermediate intervals was similar, the performance in the long intervals was significantly lower. Bonferroni post hoc tests comparing all conditions to one another did not reveal any differences between 0 and 400 ms interval conditions, which in turn differed from the 800 and 1600 ms interval conditions.

We also tested the discontinuity between the short and intermediate intervals (0–400) ms and the long intervals (800 and 1600 ms) by modelling the relationship between the inversion effects size and the interval length. We fitted a linear equation to the observed data in the 0, 50, 100, 200, and 400 ms conditions (Figure 3). Based on the regression line obtained we extrapolated the predicted inversion effect in the 800 ms (50.3%) and 1600 ms (47.8%) conditions and tested whether the obtained values differed from

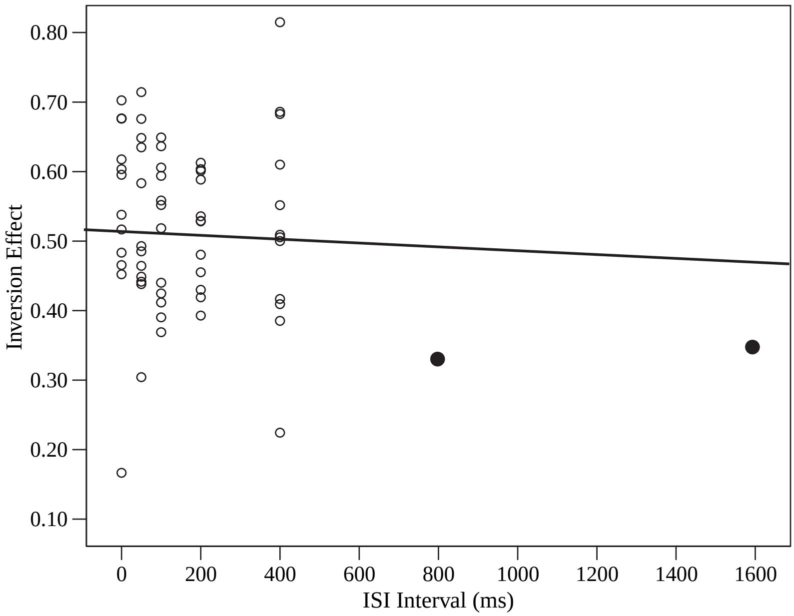

Figure 3. Scatterplot of individual values of the inversion effects in the different ISIs (0–400 ms). The regression line is plotted in black and the black circles depict the mean inversion effects obtained in the 800 and 1600 ms ISIs.

the ones predicted. The significant effects, $t(11) = 4.87$, $p < .0001$, and $t(11) = 3.94$, $p < .005$, for the 800 ms and 1600 ms, respectively, indicate that there is a discontinuity between the conditions.

Discussion

The results of Experiment 1 replicate and extend those of our previous study (Anaki et al., 2007). The findings show a similar magnitude of inversion effect across all ISI conditions up to 400 ms, a timeframe that encompasses both visible and informational persistence. A significant decrease in the inversion effect was observed only when the ISI exceeded 400 ms, with the reduction due mainly to the decrease in the identification of upright faces. These results mirror our previous findings in the three-segment paradigm, where a significant reduction in the inversion effect was observed when the interval between the parts exceeded 200 ms (totalling 450 ms between the first and last face segments). Thus, temporal integration does not seem to be supported by VSTM.

These findings suggest that when the temporal interval between the two parts is short, participants are able to integrate them and form a whole representation, on which configural processes can operate. For this to occur, a time-limited buffer must exist, which maintains the first facial part until it could be incorporated with the incoming second part. This buffer is operative at 400 ms but not at intervals of 800 ms or more, where the formation of a complete representation is not possible. As a result, the identification of the face, whose parts are separated by long intervals, was based mainly on the individual face parts. In the inverted condition, where holistic or configural processes do not occur even when whole faces are presented (although see Murray, 2004), the ISI manipulation has little influence. Identification in the inverted conditions was based on the individual features, mainly the eyes, which are presented first and are more salient than other facial parts.

It is noteworthy that although the inversion effect was reduced significantly in the long interval conditions, it was not eliminated entirely. This inversion effect probably results from the upside-down presentation of the individual facial parts themselves. This influence of inverting the facial segments is discernible even when temporal integration fails, since rich configural and holistic information is still embedded in the facial parts (Leder & Bruce, 2000). Moreover, inversion also disrupts processing of local cues (Rhodes, Brake, & Atkinson, 1993). Thus, even when temporal integration across parts is absent, the inversion effect is not expected to disappear but to be reduced disproportionately, primarily because of a drop in performance in the upright condition when integration is prevented.

It should also be noted that despite the widespread use of the face inversion effect in the literature to investigate configural aspects of face perception, it is considered only an indirect measure since the configuration of the face is not directly manipulated. However, the relationship between the processing of relational information between facial parts (i.e., configural processing) and the inversion effect probably is based on solid empirical findings (for review see Rossion & Gauthier, 2002). In addition, differential sensitivity to short and long temporal intervals between face parts, similar to those found in the current study, has been reported in recent studies. In those studies, face configuration was manipulated directly, either by misalignment of its top and bottom half (Anaki et al., 2007, Exp. 1) or by using incongruent chimeras comprised of face halves taken from different people (*composite face effect*; Singer & Sheinberg, 2006). Thus, the decrease in the inversion effect observed in the present study as a function of ISI could be attributed with a high degree of certainty to disruption of configural processes.

Before fully endorsing the proposed interpretation that the decrease in the inversion effect in Experiment 1 is related to the failure of integration, resulting in a reduced ability to process the faces configurally, an alternative account should be considered. According to this account, participants may have relied only on the top part, eye segment, in their attempts to identify the face. They were more successful in identifying the upright faces in the short intervals since the response delay was also short (i.e., the bottom part appeared after 0–400 ms). In the longer ISIs, where response was delayed, the memory trace of the top part began to decay, affecting the response's accuracy. Hence, both interpretations incorporate in their accounts the temporary nature of the short-term buffer. But while the *integration account* claims that this limitation impedes integration and consequently configural processing, the alternative *decay account* asserts that no integration occurs at all, and the decrease in accuracy as the ISI lengthens is the result of memory decay. Experiment 2 will attempt to decide which account is more tenable.

EXPERIMENT 2

According to the integration interpretation, the failure to integrate facial parts across long intervals was attributed to the limited temporal capacities of the visual buffer. Since the persistence of the representation in the buffer is short-lived, incoming visual information cannot be integrated with the fading stored representation. If this buffer is an integral component of the integration process, its operation could also be disrupted by the presentation of a non-related visual stimulus while the buffer is retaining the first facial part for integration with the second facial part. As a result, the incoming visual information will replace the existing facial representation and prevent the integration of the facial parts. This interference will be more detrimental when the nonrelated visual stimulus appears after the top first part of the face than after the bottom second part since integration in the latter condition most likely will be completed before the interfering stimulus appears. Such interference should have little effect on perception if the decay account is correct, since the same amount of time and interference ensues regardless of whether the irrelevant stimulus appears before or after the bottom half of the face.

In Experiment 2, we presented upright faces with intervals of 200 or 800 ms between the top and bottom parts. These two intervals represent conditions in which temporal integration is either present or absent. Only upright faces were used because the results of Experiment 1 showed that the lack of temporal integration is expressed mainly in the upright condition. At each ISI, the two facial parts appeared in one of three conditions: (a) Without any visual interference between the parts—no interference, similar to Experiment 1; (b) an interfering visual stimulus appearing 150 ms after the presentation of the top part—top interference; and (c) an interfering stimulus appearing 150 ms after the appearance of the lower part—bottom interference (see Figure 4). Note that the relatively long interval between the facial part and the interfering stimulus precludes the operation of the latter as a perceptual mask

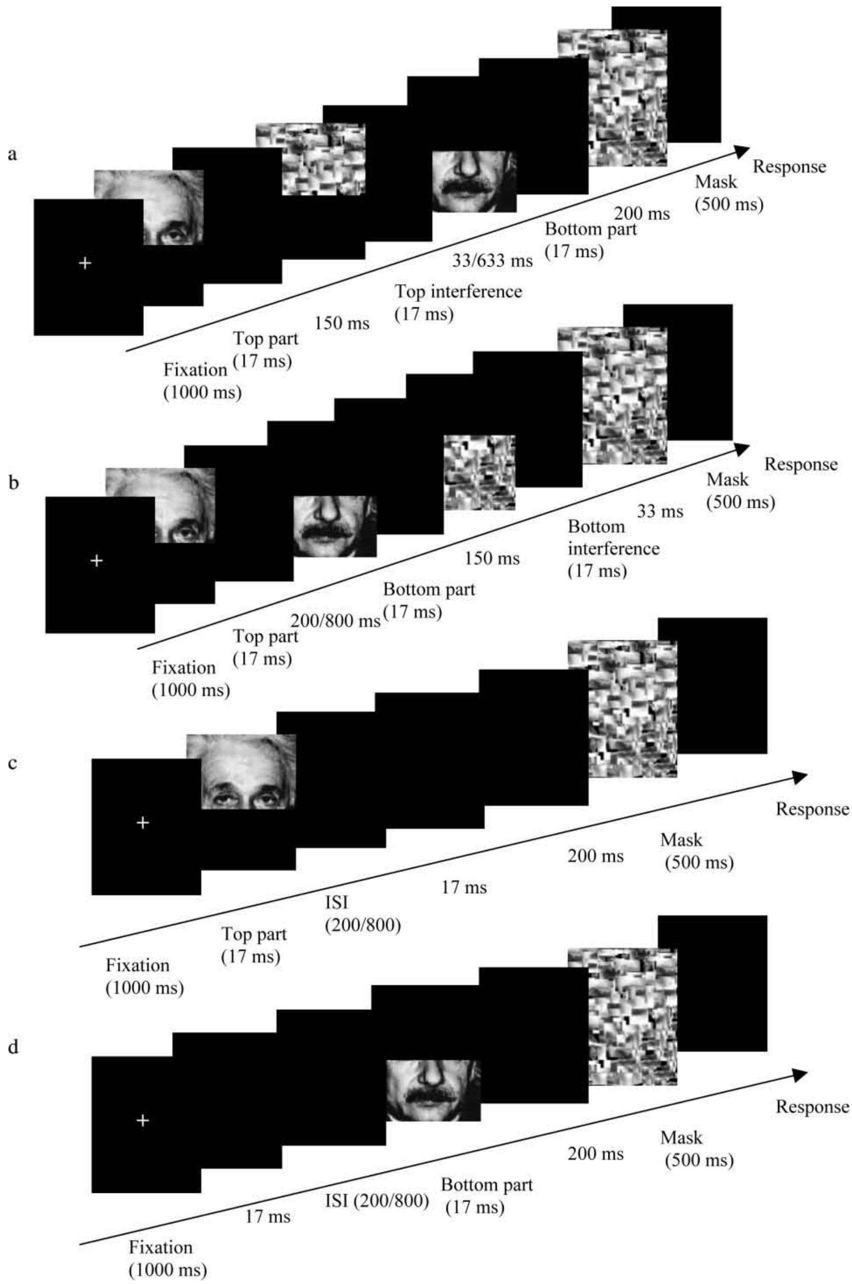

Figure 4. Examples of conditions in Experiment 2: Top interference (a), bottom interference (b), top only (c), bottom only (d).

(Loffler, Gordon, Wilkinson, Goren, & Wilson, 2005; Rolls & Tovee, 1994). Nevertheless, the interfering stimulus is expected to impede the integration process by replacing the facial part in the short-term buffer.

Performance in the no interference condition was expected to be superior in the 200 ms ISI condition compared to the 800 ms ISI condition, because of the additional integration processes that occur in the former but not in the latter condition. Different predictions were made for the two interference conditions in each ISI; at 200 ms ISI, interference was expected to be more detrimental for integration when the interfering stimulus followed the top part than the bottom part: Since the top part has to be maintained in the visual buffer until its integration with the bottom part, the interfering stimulus was expected to replace the top part in the buffer. Indeed, similar consequences were predicted to occur when the interfering stimulus appeared after the bottom part as well, but the integration process was estimated to end by the time the visual stimulus appeared. In the 800 ms ISI condition, where integration is not possible, comparable accuracy levels were predicted for the two interference conditions.

The decay interpretation would concur that identification will be greater in the 200 ms ISI no interference condition than in the longer 800 ms ISI condition, due to decay of the trace. Yet, it will predict no differences in accuracy between the three conditions of the 200 ms ISI interval. Since the identification of the face is based, according to this account, on the top part only, the manipulation which is intended to disrupt integration will be ineffective. The different hypotheses are depicted in Figure 5.

Two more conditions, top only and bottom only part presentations, were investigated in Experiment 2. They were added to test the claim, advanced above by the integration account, that performance in the conditions in which temporal integration is prevented, is based on additive information from each of the parts separately. If so, the hypothesis is that the overall accuracy in the conditions where no temporal integration occurs will be similar to the *additive* accuracy in the top only and bottom only conditions. According to decay account, however, accuracy of the top only condition in the 200 ms ISI interval would be greater than in the 800 ms ISI, where the decay of the top part's memory will be operating (see Figure 5).

Method

Participants. Thirty-two undergraduate students at the University of Toronto participated in the experiment.

Materials. The 120 critical stimuli were identical to the famous faces used in Experiment 1. In each ISI condition the parts were presented randomly in five conditions (24 faces or face parts per condition): No interference, top interference, bottom interference, top only, and bottom only.

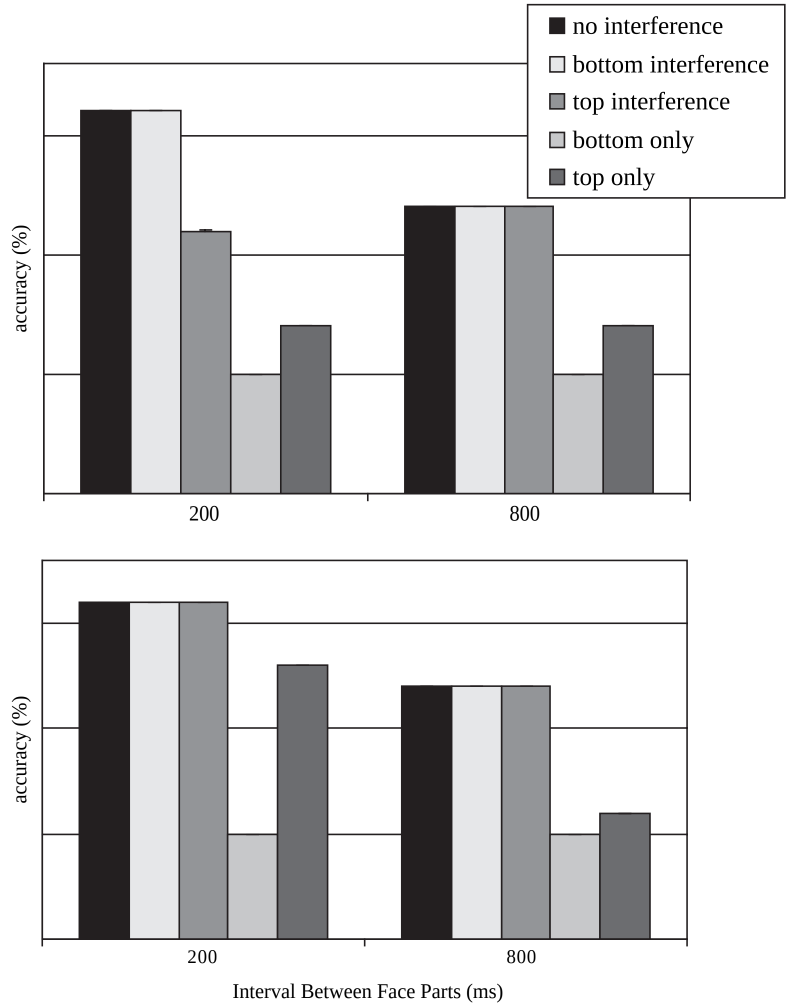

Figure 5. Hypothesized results for Experiment 2. According to the integration hypothesis (top panel) accuracy should be lower in the top interference condition in the 200 ms compared to other conditions, while similar to the no interference and bottom interference in the 800 ms ISI. In addition, the bottom and top only conditions should be comparable in the two ISIs and their sum should equal conditions in which no integration occurs (e.g., top interference). According to the decay hypothesis (bottom panel) same accuracy should be obtained across all conditions within each ISI, and performance in the top only condition should be higher in the 200 ms ISI than in the 800 ms ISI. To view this figure in colour, please see the online issue of the Journal.

Procedure. Sixteen participants were assigned randomly to the two ISI conditions. The apparatus and display format were identical to those of Experiment 1. In the no interference condition, following the offset of a

1000 ms fixation, the two parts of the upright face appeared for 17 ms each, separated by an ISI of either 200 or 800 ms. After the presentation of the second part of the face, a black screen appeared for 200 ms followed by a 500 ms mask. Participants responded in the same manner as in Experiment 1.

In the top interference condition, an interfering stimulus, its size equivalent to the top part, was presented 150 ms after the offset of the top part (to prevent masking effects) for 17 ms (Figure 4a). In the bottom interference condition, the interfering stimulus appeared under the same conditions but following the bottom part of the face (Figure 4b). The interfering stimulus was prepared in a similar fashion to the mask but its size was different. In the top only and bottom only conditions, the bottom part or the top part, respectively, were deleted, and replaced by a black screen (Figure 4c–d).

As in Experiment 1, following the presentation of the 120 faces, participants were shown the faces again but presented in a whole upright format for unlimited time, and were asked to identify them.

Results and discussion

The accuracy of each participant was computed in the same manner as reported in Experiment 1 after ensuring a comparable level of performance in the second part of the experiment. The mean accuracy of the participants as a function of condition is presented in Figure 6.

A two-way ANOVA with one between-subjects factor, ISI (200 and 800 ms), and one within-subject factor, presentation type (no interference, top interference, bottom interference, top only, bottom only) was performed on the proportion of correct responses in the first part of the experiment. This analysis yielded a main effect of presentation type, $F(4, 27) = 170.14$, $MSE = 0.009$, $p < .0001$, and an interaction between ISI and presentation type, $F(4, 27) = 3.97$, $MSE = 0.009$, $p < .01$. We then performed the relevant comparisons pertaining to the specific hypotheses.

Replicating the results observed in Experiment 1, we found greater accuracy in the short ISI (79%) compared to the long ISI (66%) in the no interference condition, $t(30) = 2.65$, $SE = 0.05$, $p < .01$. In addition, the resemblance of the two interference conditions (top/bottom interference) to the no interference condition differed at the two ISIs. In the short interval condition, the accuracy of the bottom interference condition was similar to the no interference condition, $t(15) < 1$, and both conditions were significantly more accurate than the top interference condition, $t(15) = 4.42$, $SE = 0.03$, $p < .0001$, $t(15) = 4.26$, $SE = 0.03$, $p < .001$, for the no interference and bottom interference, respectively. In contrast, in the long ISI condition, the two interference conditions did not differ from the no interference condition,

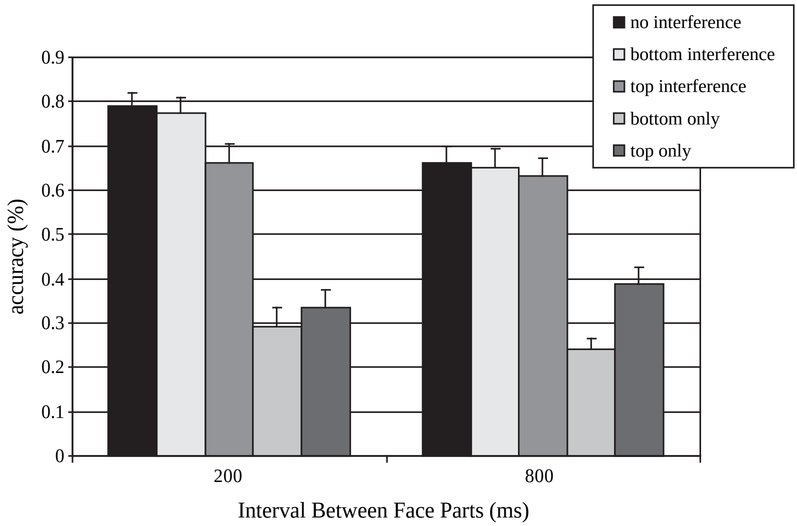

Figure 6. Percentage of faces identified correctly in Experiment 2 as a function of interval between face parts (200 ms, 800 ms) and condition of presentation. To view this figure in colour, please see the online issue of the Journal.

$F(2, 30) < 1$. Thus, as predicted by the integration account, temporal integration was observed at the short ISI only when the interference appeared *after* the bottom part, but not when it appeared following the top part. In the long ISI condition, where no temporal integration of the parts exists, as demonstrated by the decrease of the inversion effect in Experiment 1, the three conditions were similar in their accuracy level. Since no integration is possible even when no visual interference is presented, the identification of the face stems mainly from perception of its individual parts, perception which is not impeded by the visual interference. Interestingly, performance in all the three conditions at the long ISI condition (i.e., no/top/bottom interference) was similar to the accuracy of the top interference condition at the short ISI condition. Thus, temporal integration was prevented in the short 200 ms ISI condition by the introduction of visual interference which expunged the top part of the face from the short-term buffer. This outcome is similar to that obtained when temporal integration is eliminated by introducing a long temporal interval between the face parts that causes decay of the top part's representation (as in the 800 ms ISI no interference condition).

It is noteworthy that these effects cannot be explained by perceptual masking of the facial parts by the visual interference stimulus. First, the temporal interval between the face and the interfering stimulus is too long for masking to be effective. Second, the pattern of results precludes such an interpretation: If perceptual masking is responsible for the difference between the top and bottom interference at 200 ISI, a similar pattern also should have been observed at 800 ISI. However, the accuracy of the top and bottom interference conditions in the long interval was comparable. In addition, accuracy should have been higher in the no interference condition at 800 ISI compared to the other two interference conditions, yet performance was equivalent across conditions. The absence of these effects indicates that the interfering stimulus allowed perceptual processing of the preceding facial part but obliterated its representation in the visual buffer, which is necessary for temporal integration. To our knowledge no-one else has reported yet such a visual interference effect.

Finally, accuracy was lower when only one part of the face was presented (bottom/top only conditions), $p\text{s} < .0001$, with higher accuracy in the top only than in the bottom only, $p < .05$, due probably to the rich information embedded in the eyes' region (e.g., Fraser, Craig, & Parker, 1990; Haig, 1986; Leder, Candrian, Hubber, & Bruce, 2001; Vinette et al., 2004).[1] When a combined score of these two conditions was computed (which was identical in the two ISI conditions; 63%), it was found to be essentially equal to the four conditions lacking temporal integration, while being significantly lower than the two conditions at 200 ms ISI, in which integration occurred (all $p\text{s} > .01$). This finding supports the claim that in the absence of integration, the processing of the face is based on the additive information obtained from the two facial parts, with no supplementary benefits resulting from the fact that the two parts complement each other to form a whole face.

The results of Experiment 2 support the integration account much more than the decay account. In fact, no prediction based on the latter account was confirmed. The difference found between the conditions in the 200 ms

[1] One might claim that higher accuracy was obtained in the top only condition than in the bottom only condition because of masking effects, as the former condition was presented 417 or 1017 ms before the mask, while the mask preceded the latter condition by only 200 ms. However, as noted earlier, our results clearly show that masking effects are no longer apparent 200 ms (and even earlier) after the stimulus presentation. For example, in Exp. 2 in the 200 ISI condition an interfering mask appeared 150 ms after the bottom part of the face. Nevertheless, the accuracy in this condition was equal to the no mask condition where the whole-face mask appeared 200 ms after the bottom part. The same claim could be demonstrated in the 800 ISI condition, where no differences were found between the three condition. Of course it could be claimed that at ISIs of 150–200 ms masking effects still exist and the small difference between the two ISIs precluded us from seeing any differences. However, since perceptual masking occurs at short ISIs this temporal difference should be significant. Furthermore, in Exp. 1 we have ISIs that are similar to the long ISIs in the top only condition. Specifically, the ISI between the top part and the mask in Exp. 1 varies between 217 (in the 0 ISI) and 1817 ms (in the 1600 ISI). If an active masking is operating at ISIs longer than 150–200 ms we should have seen an *increase* in accuracy as the ISI lengthens, since the ISI between the mask and the top part varies. This was not found.

ISI interval was not predicted by that interpretation. In addition, the expected greater accuracy for the top only condition in the 200 ms ISI condition (compared to the 800 ms ISI interval) was not found.

GENERAL DISCUSSION

The focus of the present study was to explore temporal integration and its fundamental properties in face perception. Three major finding were observed: (1) Face segments, separated by intervals up to 400 ms, can be combined together into a unified representation, which is processed configurally. At longer intervals, however, integration fails, and consequently, so does configural perception. This failure is expressed by a reduction in the inversion effect, stemming mainly from a decline in the identification accuracy of upright faces. (2) When the segments are not integrated, the face's identification is based on information extracted from its individual parts. This was demonstrated in Experiment 2 which showed that when both parts were presented, but not integrated, accuracy equalled the sum of the accuracy of the parts when each was presented alone. (3) Loss of integration could result from the decay of the representation at a long interval but also from an intervening, interfering stimulus, which disrupts the maintenance of face parts in the visual buffer, but not their perception.

We propose that a short-term visual buffer, which maintains visual input temporarily, is the vehicle of the integration process of time-segregated stimuli. The results of Experiment 2 show that the main contribution of this buffer, in the present context, is not the processing of the input per se, but rather in providing a platform for integration: The interfering stimulus inserted after the second bottom segment did not impair perception, attesting that by the time it appeared, higher level information already was extracted from the image. Thus, the detrimental effects of the interfering stimulus, observed when inserted after the first segment, result from preventing the sustenance of the facial part for later integration. Consequently, accuracy in this condition was equivalent to the cumulative accuracy of the parts when presented alone. To our knowledge, this is the first demonstration of a visual interference effect that affects integration of stimulus parts, but not their perception.

These results could be easily accommodated by views claiming that during face perception relational information between features is computed (e.g., Diamond & Carey, 1986; Searcy & Bartlett, 1996). As shown in this study this computation is not dependent on simultaneous presentation of facial features but could be performed even when the different facial components are separated by short intervals. The interim maintenance of facial parts in the visual buffer allows for extraction of the relevant spatial information from which a configural representation of the face is established. The emergence of a comparable inversion effect, which is considered a hallmark of configural processing, under ISI conditions shorter than 400 ms, demonstrates that indeed the product of the integration process is a whole face and not isolated facial features.

The finding that holistic face perception could arise even when face parts are presented sequentially supposedly challenge theories that emphasize the *gestalt* nature of face recognition processes (e.g., Farah, 2004; Tanaka & Farah, 2003). These accounts claim that face recognition does not involve part decomposition, and that both perception and representation of a face is unparsed (Farah, Tanaka, & Drain, 1995). However, gestalt theories could be reconciled with the present findings if they postulate that a unified, template-like, representation is formed at more advanced visual processing stages after integration has occurred. It is only at these later stages that the differences between holistic aspects of face processing and analytic aspects of object perception begin to emerge.

Iconic memory, consisting of visible and informational persistence, is the possible candidate for the locus of this visual buffer. Previous behavioural studies have shown that the temporal boundaries of informational persistence are about 300 ms (Coltheart, 1980; Loftus, Duncan, & Gehrig, 1992). In the current study, integration was observed at 400 ms if no interfering stimulus intervened. These values also are within the range reported in functional neuroimaging that show sustained neural activity in occipito-temporal regions, which are claimed to reflect the persistence of a trace of previously presented stimuli (Mukamel, Harel, Hendler, & Malach, 2004). This finding is further supported by results showing continued firing of neurons in the macaque superior temporal sulcus, which persist for about 100 ms after stimulus disappearance (Keysers, Xiao, Földiák, & Perrett, 2005). Despite the slight variation between studies, variations whose significance has yet to be determined, their common feature points to a mechanism in lateral occipital cortex or superior temporal sulcus which is capable of maintaining visual information for several hundred milliseconds.

Interestingly, behavioural studies of matrices pattern integration have shown that informational persistence does not support integration and it is found only in visible persistence (e.g., Di Lollo, 1980). One explanation for this divergence is the differences in paradigms. In the present study, the face parts were not spatially overlapping, whereas in studies that did not find any integration, the stimuli overlapped. It is therefore possible that the second matrix replaced the first one and prevented integration, just as the interfering stimulus in Experiment 2 did.

Another discrepancy between face and object (or pattern) integration lies in VSTM whose temporal boundaries are several seconds long (Phillips, 1974). Recent studies have reported temporal integration of patterns in

VSTM (Brockmole et al., 2002; Hollingworth et al., 2005), whereas we failed to find integration at intervals longer than 400 ms. This incongruity may stem from the processes occurring in VSTM. Evidence suggests that the integration found in VSTM depends on the formation of an abstract representation, consisting of the relative, rather than absolute, location of the stimulus and its components. This is also consistent with the prefrontal cortex involvement in working memory tasks (e.g., Curtis & D'Esposito, 2003). In our task, the absence of integration in longer intervals implies that integration depends on the strict alignment of low-level features.

The present findings are compatible with recent studies that address the issue of the temporal aspects of face processing and arrive at the same conclusions that configural face perception does not rely on simultaneous presentation of the face. Singer and Sheinberg (2006) have found that face parts, namely inner and outer halves (Exp. 1) or top and bottom parts (Exp. 2) separated by 160 ms and 80 ms intervals, respectively, yield the face composite effect. This effect, where two complementary parts of different faces interfere with the identification of either part, has been considered as strong evidence for the holistic nature of face processing. In a similar vein, Vinette et al., (2004) used the *Bubbles* procedure developed by Gosselin and Schyns (2001), and found that different facial parts are used effectively as time unfolds; between 47 and 94 ms the left eye is informative, although later both eyes are used effectively (see also McCabe, Blais, & Gosselin, 2005 for a theoretical formulation of these ideas). Additional studies examining the role of different spatial frequency bands to face recognition have shown that low-frequency bands are used earlier in the course of face perception ($<$ 100 ms after stimulus onset), and at later stages high spatial frequencies are active as well (Parker & Costen, 1999). These findings, supporting a "coarse-to-fine analysis" hypothesis in face perception, originally suggested by Sergent (1986), also accentuate the fact that face perception is a process in which different types of information dynamically interact over time to form a unified facial representation (see Goffaux & Rossion, 2006, for review and current findings, and Moscovitch, Scullion, & Christie, 1976, for temporal effects in laterality of face perception).

The need for gradual extraction of sensory information from a facial stimulus, exhibited by these studies, necessitates both a mechanism that temporarily maintains the visual input and a process that integrates it over time, concepts that have been the focus of the present study. The introduction of these concepts may also clarify other aspects of face perception as they suggest that disorders of face perception, such as prosopagnosia, could result also from rapid decay of information in the temporary buffer, or deficits in the implementation of integration processes. These issues are currently under investigation.

REFERENCES

Anaki, D., Boyd, J., & Moscovitch M. (2007). Temporal integration in face perception: Evidence of configural processing of temporally separated faces. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1–19.

Barton, J. J. S., Deepak, S., & Malik, N. (2003). Attending to faces: Change detection, familiarization, and inversion effects. *Perception*, *32*, 15–28.

Brockmole, J. R., Wang, R. F., & Irwin, D. E. (2002). Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 315–334.

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. *Perception and Psychophysics*, *65*, 1296–1306.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics*, *27*, 183–228.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*, 415–423.

Davies, T. N., & Hoffman, D. D. (2002). Attention to faces: A change-blindness study. *Perception*, *31*, 1123–1146.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, *109*, 75–97.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107–117.

Farah, M. J. (2004). *Visual agnosia*. Cambridge, MA: MIT Press.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 628–634.

Fraser, I. H., Craig, G. L., & Parker, D. M. (1990). Reaction time measures of feature saliency in schematic faces. *Perception*, *19*, 661–673.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial": Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1023–1039.

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

Haig, N. (1986). Exploring recognition with interchanged facial features. *Perception*, *15*, 235–247.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements, visual memory, and scene representation. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 356–383). Oxford, UK: Oxford University Press.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 113–136.

Hollingworth, A. (2006). Visual memories for natural scenes: Evidence from change detection and visual search. *Visual Cognition*, *14*, 781–807.

Hollingworth, A., Hyun, J. S., & Zhang, W. (2005). The role of visual short-term memory in empty cell localization. *Perception and Psychophysics*, *67*, 1332–1343.

Humphreys, G. W., Hodsoll, J., & Campbell, C. (2005). Attending but not seeing: The "other race" effect in face and person perception studied through change blindness. *Visual Cognition*, *12*, 249–262.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*, 420–456.

Irwin, D. E. (1992). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146–165). New York: Springer-Verlag.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science*, *5*, 94–100.

Irwin, D. E., & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.

Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 343–360.

Keysers, C., Xiao, D. K., Földiák, P., & Perrett, D. (2005). Out of sight but not out of mind: The neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology*, *22*, 316–332.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology*, *53*, 513–536.

Leder, H., Candrian, G., Hubber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception*, *30*, 63–73.

Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., & Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Research*, *45*, 2287–2297.

Loftus, J. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 530–549.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science*, *6*, 255–260.

McCabe, E., Blais, C., & Gosselin, F. (2005). Effective categorization of objects, scenes, and faces through time. In C. Lefebvre & H. Cohen (Eds.), *Handbook of categorization in cognitive science* (pp. 767–791). Amsterdam: Elsevier.

Moscovitch, M., Scullion, D., & Christie, D. (1976). Early versus late stages of processing and their relation to functional hemispheric asymmetries in face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 401–416.

Moscovitch, M., Winocour, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, *9*, 555–604.

Mukamel, R., Harel, M., Hendler, T., & Malach, R. (2004). Enhanced temporal non-linearities in human object-related occipito-temporal cortex. *Cerebral Cortex*, *14*, 575–585.

Murray, J. E. (2004). The ups and downs of face perception: Evidence for holistic encoding of upright and inverted faces. *Perception*, *33*, 387–398.

Palermo, R., & Rhodes, G. (2003). Change detection in the flicker paradigm: Do faces have an advantage? *Visual Cognition*, *10*, 683–713.

Parker, D. M., & Costen, N. P. (1999). One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. *Current Psychology*, *18*, 118–127.

Peterson, M. A., & Rhodes, G. (2003). Introduction: Analytic and holistic processing—The view through different lenses. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 3–19). Oxford, UK: Oxford University Press.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, *16*, 283–290.

Psychological Software Tools, Inc. (2000). E-prime (version 1.0b) [Computer software]. Pittsburgh, PA: Author.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition*, *47*, 25–57.

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science*, *12*, 94–99.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *257*, 9–15.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews*, *1*, 63–75.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 904–915.

Sergent, J. (1986). Microgenesis of face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. M. Young (Eds.), *Aspects of face processing* (pp. 17–33). Dordrecht, The Netherlands: Martinus Nijhoff.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*, 16–20.

Singer, J. M., & Shinberg, D. L. (2006). Holistic processing unites face parts across time. *Vision Research*, *46*, 1838–1847.

Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 53–74). Oxford, UK: Oxford University Press.

Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes! *Cognitive Science*, *28*, 289–301.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*, *98*, 4800–4804.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.

Manuscript received November 2005
Manuscript accepted November 2006
First publised online April 2007