Rigid facial motion influences featural, but not holistic, face processing

Naiqi G. Xiao, Paul C. Quinn, Liezhong Ge, Kang Lee

Article history:
Received 24 July 2011
Received in revised form 18 January 2012
Available online 8 February 2012

Keywords:
Face processing
Featural processing
Holistic processing
Face composite effect

A B S T R A C T

We report three experiments in which we investigated the effect of rigid facial motion on face processing. Specifically, we used the face composite effect to examine whether rigid facial motion influences primarily featural or holistic processing of faces. In Experiments 1–3, participants were first familiarized with dynamic displays in which a target face turned from one side to another; then at test, participants judged whether the top half of a composite face (the top half of the target/foil face aligned or misaligned with the bottom half of a foil face) belonged to the target face. We compared performance in the dynamic condition to various static control conditions in Experiments 1–3, which differed from each other in terms of the display order of the multiple static images or the inter-stimulus interval (ISI) between the images. We found that the size of the face composite effect in the dynamic condition was significantly smaller than that in the static conditions. In other words, the dynamic face display influenced participants to process the target faces in a part-based manner and consequently their recognition of the upper portion of the composite face at test became less interfered with by the aligned lower part of the foil face. The findings from the present experiments provide the strongest evidence to date to suggest that the rigid facial motion mainly influences featural, but not holistic, processing.

1. Introduction

Human face recognition is an extensively studied topic in psychological and neuroscience research (Calder et al., 2011). One shortcoming of the current literature on face perception is that the vast majority of the existing studies have used static facial stimuli. However, in the real world, faces are more likely to be encountered while they are moving. For example, when we meet a friend on the street, what we encounter are continuous facial movements such as nodding, expressions, and so forth. Recently, the effects of facial movements on face processing have started attracting researchers’ attention (Bahrick, Gogate, & Ruiz, 2002; Hill & Johnston, 2001; Lander & Davies, 2007; Otsuka et al., 2009; Pike et al., 1997; Rosenblum et al., 2002; Schiff, Banka, & Galdi, 1986; Simons & Levin, 1998). Evidence suggests that face recognition can be facilitated by providing observers with dynamic facial information. Much of the existing work has been devoted to explain why dynamic information facilitates face recognition (Bulf & Turati, 2010; Christie & Bruce, 1998; Knight & Johnston, 1997; Lander & Bruce, 2000, 2003; Lander, Christie, & Bruce, 1999; O'Toole, Roark, & Abdi, 2002; O'Toole et al., 2011; Roark et al., 2003; Wallis, 2002; Wallis & Bülthoff, 2001).

In contrast to the relatively extensive research on why dynamic face information facilitates face recognition, little research has examined what aspect of face recognition is influenced by dynamic face information. Face recognition is a process by which one matches the information about the identity of an individual face to that stored in memory. Most researchers agree that there are at least three types of information that are crucial for recognizing a face. The first is featural information, which refers to isolated parts of a face such as the eyes, nose, and mouth. The second is configural information, which refers to the spatial relationships among the isolated face features. The third is holistic information, which is the combination of featural and configural information in an unbroken whole, or gestalt (see Lee et al., 2011, for a review).

Although no research has directly investigated the relationship between facial motion and configural, featural, or holistic processing, some studies have provided relevant, albeit indirect, evidence. Researchers have used inverted faces as stimuli to investigate dynamic facial information processing because inversion has been thought to disproportionately disrupt the process of face configural and/or holistic information (Freire, Lee, & Symons, 2000; Maurer, Le Grand, & Mondloch, 2002; Richler, Cheung, & Gauthier, 2011; Yin, 1969). For example, Knappmeyer, Thornton, and Bülthoff (2003) reported that idiosyncratic facial motion influenced
observers’ identification when viewing upright and inverted dynamic faces. Employing an animated average face, Hill and Johnston (2001) tested observers’ abilities to identify individuals from facial motion information alone. The results showed that observers were still able to successfully identify an individual in the motion condition, even when the face was inverted. Although observers were more accurate for upright faces as compared to inverted faces. Lander, Christie, and Bruce (1999, Experiments 1 and 2) tested naming of familiar faces, which were inverted or degraded, and they found that even though the overall discrimination of inverted faces was worse than upright ones, the beneficial effect of the motion was independent of face inversion. Taken together, these findings suggest that the processing of facial dynamic information is not completely disrupted by inversion, which in turn suggests indirectly that facial motion may exert its effect by influencing featural processing.

However, inversion at best only provides an indirect test of whether motion leads to either featural, configural, or holistic face processing because face processing is done typically in the canonical upright orientation. There is also recent evidence to suggest that inversion may not disrupt face configural/holistic face processing disproportionately more than featural information (Yovel & Kanwisher, 2004). A more direct test of holistic/configural information processing (or more precisely holistic processing) is the composite effect (McKone, 2008; Mondloch & Maurer, 2008; Mondloch et al., 2007; Young, Hellawell, & Hay, 1987). A composite face is composed of two face parts (upper and lower), which belong to different individuals. With a composite face, the two face parts can be vertically aligned to form a face or misaligned so that the two face parts are perceived as separate. When perceivers are asked to identify one of the face parts, the composite effect is observed: identification performance is worse in the aligned condition as compared to the misaligned condition. The basis for this effect is that holistic processing interferes with the identification of the upper face part from the lower face part in the aligned condition when the upper and lower parts of the face can be fused together to form a face gestalt; such interference is lessened when the two parts are misaligned because misalignment prevents the upper and lower parts to be integrated into a whole face. Thus, it is widely accepted that the observation of such a face composite effect indicates the presence of holistic processing, whereas the lack of it suggests featural processing.

In the present study, we specifically aimed to examine whether dynamic faces influence either holistic face processing or featural processing primarily, with the use of the composite face paradigm. It should be noted that faces can move in at least two ways. One is called rigid motion whereby the face moves, but remains unchanged in shape (e.g., raising eyebrows or smiling); another is called elastic motion whereby the face transforms in shape (e.g., smiling, chewing). Most of the facial movements in the natural world involve both types of motion. However, it has been argued that each type of facial motion may serve a different function and influence face recognition via a different mechanism (Bruce, 1994; Lander & Bruce, 2003). We limited the facial motion in the present investigation to the rigid format so as to eliminate the possible confound engendered by the facial motion types.

In the current study, we first familiarized participants with either a rigidly moving face or multiple static images of an individual’s face. Then, in the recognition phase, we showed a static composite face whose upper and bottom parts came from different people. The two parts were either aligned or misaligned with each other. The task was to decide whether the upper part of the face belonged to the familiarized face. A face composite effect would result if participants had more difficulty recognizing the top face part in the aligned condition than the misaligned one because in the aligned condition the top and bottom parts of the face were seen fused together, resulting in difficulty in decomposing the face into top and bottom parts. We reasoned that if facial dynamic information would influence featural face processing more than holistic processing, we should observe a decreased face composite effect (the featural influence hypothesis). This is because the rigid facial motion may influence observers to learn the upper and lower parts of the target face in a part-based manner, and as a result their recognition of the upper face part of the target face at test would become less influenced by the aligned lower part of the foil face. Alternatively, if rigid motion would influence holistic face processing more than featural processing, an increased face composite effect should result. This is because the rigid facial motion may influence observers to process the upper and lower parts of a target as a gestalt and consequently their recognition of the upper face part will become more influenced by the aligned lower face part (the holistic influence hypothesis). It should be noted, however, that rigid dynamic facial motion might influence either holistic or featural processing, but such influence might not necessarily lead to better face recognition overall.

2. Experiment 1

2.1. Method

2.1.1. Participants

Thirty-two Chinese undergraduates (22 females) participated in the experiment. All participants were of the same race as the face stimuli and had normal or corrected-to-normal vision. None of participants were familiar with any of the individuals depicted in the face stimuli. Participants in this and the following experiments gave informed consent prior to their participation.

2.1.2. Materials

The individuals who were filmed to form the stimuli are referred to as “the models” and the individuals who were tested in the experiment are referred to as the “participants”. There were eighteen models (ten males and eight females); all the models were Chinese, aged 19–24.

To create the stimuli to be used in the present study, each model was asked to sit on a 360°-revolving chair. A digital camera was placed 3 m from the chair, and the height of the camera was adjusted to keep the model’s face in the center of the image. Two light sources were placed at each side of the camera to avoid any shadow on the model’s face. For each model, nine pictures were taken. One of them was the front view (90°), which was used as the testing stimulus. The other eight pictures depicted different viewpoints of the face, from 0° to 180° at the interval of 26°. These pictures were used to create a familiarization stimulus for this model’s face. During the photo taking, models were asked to maintain a neutral expression and to keep their head still. The original color pictures were converted into gray-scale format with the image size of 640 × 480 pixels.

2.1.2.1. Familiarization stimulus. The familiarization stimulus was comprised of eight non-front view pictures of the same person to create an animation of a head rotating back and forth. We used one profile view picture only once and the other seven of the eight non-front view pictures twice, which gave rise overall to 15 pictures in each familiarization stimulus. The picture sequence would be 1–2–3–4–5–6–7–8–7–6–5–4–3–2–1, in which 1 and 8 mean two opposite facing profile-view face images (0° or 180°), and the others mean face images from different angles between 0° and 180°. The duration of each picture was 80 ms, with no interval between consecutive pictures, which gave the overall presenting time of 15 pictures × 80 ms = 1200 ms. Two types of
familiarization stimuli were created in the present experiment: dynamic and multi-static stimuli for each model’s face. As shown in Fig. 1, the dynamic stimulus consisted of the 15 face pictures of the same face shown in the natural sequence (from 0 to 180 degrees and then 180° to 0°) such that it appeared the person’s face rotated from the left to the right and then back. Accordingly, in the dynamic stimulus, the first and last pictures always showed the face in a profile view. In contrast, the multi-static stimulus was made up of the same pictures as those used in the multi-dynamic stimuli except the presentation order of the pictures was randomized. This control stimulus was created to remove the rigid dynamic facial motion from the dynamic stimulus while keeping the static information identical between the two stimulus types. Nevertheless, we showed the first and last pictures of the multi-static stimulus in the same profile view as the dynamic stimulus. Thus, both types of the stimuli began and ended in the same way to ensure that any difference in recognition performance would be attributable to differences in the nature of dynamic motion, not those in the initial and last views of the stimuli. It should also be noted that the frontal view of the face was never shown in the familiarization stimulus.

2.1.2.2. Testing stimulus. The testing composite face stimuli were created from the frontal faces of the individuals whose faces from other angles between 0° and 180° made up the familiarization stimuli. The front view faces were first split into top and bottom halves, with the cut occurring a little below the eyes. To create a target composite face, the upper face half of one model was combined with the bottom half of a different face. For each target composite face, there were two versions in terms of the alignment of the upper and lower face halves. As shown in Fig. 2, for aligned composite faces, the bottom half face was adjusted to fuse well with the top half. To attain the best fusion at the nose and cheek, a few adjustments were made (e.g., changing size, contrast, and brightness). The misaligned composite face was created by moving the bottom half to the right by approximately half a face width. We also created two types of foil composite faces, aligned and misaligned ones, respectively. They were made in the exact same way as the target composite face except that the inner features (i.e., eyes, eyebrows, and top half of the nose) of the top halves of the face had never been seen before by the participants. For each pair of foil and target faces, the external face features (e.g., hair style and face contour) were kept the same to ensure that participants’ face recognition was based on the more individuating aspects of the faces such as their eyes, nose, mouth, and the configurations among them.

2.1.3. Procedure

Participants were tested individually, seated approximately 70 cm from a computer screen with the visual angle of 27° in a quiet and fully lit room. The experimental procedure was operated by E-Prime 2.0 software on a desktop computer. The screen resolution was 1024 × 768 pixels.

Each trial began with a fixation (+) displayed at the center of the screen (500 ms). After the offset of the fixation, participants saw a dynamic stimulus (1200 ms) in the dynamic condition, and they were instructed to remember the face seen. The static condition was the same as the dynamic one except that the multi-static stimulus was used. These two conditions were run as two blocks separately. The order of the two blocks was counterbalanced between participants.

To avoid the faces presented in the first block from influencing the identification of the faces used in the second block, we used different faces in these two blocks: the 18 models’ faces were divided equally into two stimulus sets. For each participant, one of the two sets was shown in the dynamic format, while the other was shown in the multi-static format. The allocation of faces into the two sets was counterbalanced across participants.

Once the familiarization stimulus was terminated, a visual mask was presented for 500 ms to eliminate the effect of afterimage. A static aligned or misaligned test composite face was displayed following the visual mask, the alignment of which was selected randomly. Participants decided whether or not the upper half of the composite face belonged to the person seen in the familiarization stimulus by key pressing ("1" as "the same", "2" as "different"), and their response initiated the next trial. Because most of the existing studies examining the face composite effect only...
varied the top half of the face while keeping the lower half constant (e.g., McKone, 2008; Mondloch & Maurer, 2008; Mondloch et al., 2007), to ensure comparability between the present and previous studies, we also limited the changes to the top half of the face. There were overall 576 trials, equally divided across 2 (dynamic or multi-static) × 2 (aligned or misaligned) × 4 conditions. In each condition combination, half of the trials presented the target composite faces (“yes” trials), and the other half presented the foil composite faces (“no” trials).

A set of eight practice trials was administered prior to the commencement of the experiment to familiarize participants with the experimental procedure and stimulus formats. The faces used in the practice trials were never used in the experiment, and data from the practice trials were not analyzed.

2.2. Results

Table 1 shows the means and standard deviations of accuracy, $A'$, and $B_0$ for each condition. The $A'$ and $B_0$ were computed using the formulae provided by Donaldson (1992). Both accuracy and $A'$ provided similar results. However, we focused on the $A'$ data not only to avoid redundancy but also because the accuracy measure may contain response biases, whereas $A'$ does not (Donaldson, 1992). Indeed, as shown in Table 1, similar to the findings of previous studies (e.g., Gauthier, Klaiman, & Schultz, 2009; Le Grand et al., 2004), participants appeared to be more biased in their response to the aligned trials than to the misaligned trials. Because the aims of the present experiment were to test whether rigid facial motion influences featural or holistic face processing in face recognition, we must use a recognition performance measure with the influence of response biases already controlled for. For this reason, in this experiment and in both subsequent experiments, we focused on the $A'$ measure for data analyses.

It should be noted that in addition to response accuracy (Mondloch & Maurer, 2008), correct response latency (McKone, 2008; Young, Hellawell, & Hay, 1987) has been used to test the composite effect in the literature. In those studies using response latency, participants’ response accuracy was kept at the ceiling level through extensive pretest training (e.g., McKone, 2008) or using famous faces (e.g., Young, Hellawell, & Hay, 1987). In contrast, when faces were unfamiliar or not previously trained, response accuracy was used (e.g., Anaki, Boyd, & Moscovitch, 2007; Mondloch & Maurer, 2008). Because our participants were unfamiliar with the target faces, we chose response accuracy as measured by $A'$ as the main index of the face composite effect.

A 2 (condition: dynamic vs. multi-static) × 2 (alignment: aligned vs. misaligned) repeated measures ANOVA revealed a significant main effect for alignment, $F(1,31) = 24.91$, $MSE = 0.02$, partial $\eta^2 = .45$, $p < .001$: participants recognized the upper face parts in the misaligned composite faces ($M = .87$, $SD = .10$) better than those in the aligned ones ($M = .85$, $SD = .11$). This outcome suggests the presence of a face composite effect (Young, Hellawell, & Hay, 1987), indicating that the irrelevant, bottom face half interfered with the recognition of the top half when the two halves formed a face gestalt.

As shown in Fig. 3, even though there was no significant difference between dynamic and static conditions ($p = .435$), the crucial interaction between condition and alignment was significant, $F(1,31) = 7.10$, $MSE = 0.01$, partial $\eta^2 = .19$, $p = .012$. This finding suggests that the magnitude of the composite effect was modulated by the familiarization stimulus format. With the multi-static familiarization format, participants recognized the upper face parts significantly better in the misaligned condition ($A' = .87$) than in the aligned condition ($A' = .83$, $p < .001$). In contrast, with the dynamic format, participants recognized equally well in the aligned and misaligned conditions ($A' = .86$ for the aligned condition, .87 for the misaligned condition, $p = .121$).

2.3. Discussion

Here we used the face composite effect to test the effect of rigid facial motion on face processing. We compared participants’ recognition performance of the upper part of a frontal view face after they were familiarized with the face in either a dynamic or static display. In the multi-static condition, we found that although the participants never saw the frontal view of the familiarization face,

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aligned</td>
<td>Misaligned</td>
</tr>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit</td>
<td>.66 (.21)</td>
<td>.78 (.15)</td>
</tr>
<tr>
<td>False alarm</td>
<td>.15 (.17)</td>
<td>.18 (.16)</td>
</tr>
<tr>
<td>ACC</td>
<td>.75 (.13)</td>
<td>.80 (.11)</td>
</tr>
<tr>
<td>$A'$</td>
<td>.83 (.12)</td>
<td>.87 (.09)</td>
</tr>
<tr>
<td>$B_0$</td>
<td>.44 (.53)</td>
<td>.12 (.59)</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit</td>
<td>.65 (.24)</td>
<td>.83 (.16)</td>
</tr>
<tr>
<td>False alarm</td>
<td>.11 (.09)</td>
<td>.15 (.13)</td>
</tr>
<tr>
<td>ACC</td>
<td>.77 (.12)</td>
<td>.84 (.10)</td>
</tr>
<tr>
<td>$A'$</td>
<td>.85 (.10)</td>
<td>.90 (.07)</td>
</tr>
<tr>
<td>$B_0$</td>
<td>.46 (.51)</td>
<td>.13 (.71)</td>
</tr>
<tr>
<td><strong>Experiment 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit</td>
<td>.55 (.22)</td>
<td>.78 (.19)</td>
</tr>
<tr>
<td>False alarm</td>
<td>.06 (.08)</td>
<td>.11 (.09)</td>
</tr>
<tr>
<td>ACC</td>
<td>.74 (.11)</td>
<td>.83 (.10)</td>
</tr>
<tr>
<td>$A'$</td>
<td>.85 (.07)</td>
<td>.90 (.06)</td>
</tr>
<tr>
<td>$B_0$</td>
<td>.80 (.31)</td>
<td>.28 (.55)</td>
</tr>
</tbody>
</table>

Fig. 3. Mean $A'$ of each condition in Experiments 1–3. Error bars represent 1 SE.
they readily recognized its upper part in the frontal view. This result suggests that providing participants static images of a face from multiple angles was sufficient for them to form a representation of the face in the frontal view. Furthermore, we found that after being familiarized with static images of a face from multiple angles, participants recognized the upper face parts better when they were misaligned, than when they were aligned, with the lower distracting face parts. This finding replicated the robust face composite effect that has been typically observed with the familiarization of only one single frontal face.

In contrast, a more crucial finding was that participants identified equally well in the aligned and misaligned conditions when familiarized with a dynamic stimulus that showed the same face in rigid motion. The disappearance of the face composite effect in the dynamic condition indicated that the judgment of the upper face part was not affected by the lower distracting part. This finding suggests that by familiarizing a face when it moves rigidly, the top part of the face became easier to be decomposed from the face gestalt consisting of the top part of the familiarized face and a lower part of a different face.

It should be noted, however, that in the current experiment, we used sequentially vs. randomly-ordered multiple-image displays to test the effect of rigid facial dynamic information on face recognition. This comparison is vulnerable to the critique that the dynamic stimuli differed from the multi-static stimuli in both the dynamic information and image display order. It is therefore possible that the results could be explained by the difference in display order rather than by the existence of rigid facial dynamic information. We tested this alternative account in Experiment 2.

3. Experiment 2

The aim of Experiment 2 was to test whether the disappearance of the face composite effect in the dynamic condition of Experiment 1 was due to the dynamic information in the rigid facial motion stimulus or simply the display order of the static images. To address this issue, we used a new type of multi-static stimulus. Unlike those used in the multi-static condition of Experiment 1, we used multi-static stimuli that had the same images, presentation duration of each image, and display order as those used in the dynamic stimuli. The only difference was the interval between the face images for each familiarization stimulus. We varied the interval randomly from 100 ms to 700 ms (with an average of 400 ms).

By doing so, we completely removed any possible cues for perceiving apparent motion between the face images for each familiarization face stimulus. However, the sequence of each face image in the multi-static stimulus and the presentation time of each face image were identical to those of the dynamic stimulus. Through this manipulation, we could determine if either display order or dynamic motion information was responsible for the disappearance of the face composite effect seen in Experiment 1. If the display order was the reason for the disappearance, participants should perform equally well in the aligned and misaligned conditions when the test composite face was preceded by the new multi-static familiarization stimulus. Alternatively, if the dynamic facial motion was crucial for the disappearance of the face composite effect, the new multi-static condition should continue to produce a robust composite effect, whereas we should continue to fail to observe this effect in the dynamic condition.

3.1. Method

3.1.1. Participants

Eighteen undergraduates (4 males) participated in the experiment. All participants were of the same race as the face stimuli and had normal or corrected-to-normal vision. None of the participants took part in the prior experiment or knew the individuals depicted in the stimuli.

3.1.2. Materials and procedure

The materials used in Experiment 2 were the same as in Experiment 1, except for the duration of the interval between the face images in the multi-static stimuli. We used random intervals ranging from 100 ms to 700 ms (mean = 400 ms). The overall duration of the multi-static stimuli ranged from 2600 ms to 11,000 ms, with the mean of 6800 ms and the standard deviation of 46 ms. Theoretically, 96% of the trials’ duration should fall between 6708 ms (Mean – 2 × SD) and 6892 ms (Mean + 2 × SD), and the duration in less than six trials (4%) would fall outside this range.

The procedure was the same as that used in Experiment 1: it was comprised of two blocks, one for the dynamic condition, and the other for the random interval multi-static condition. The order of the two blocks was counterbalanced across participants. The aligned and misaligned trials were presented randomly.

3.2. Results

The means and standard deviations of response accuracy, A’, and B0 are listed in Table 1. Like the analyses in Experiment 1, we focused on the A’ data. A 2 (condition) × 2 (alignment) repeated measures ANOVA was performed on A’.

It revealed a significant main effect for alignment, \( F(1,17) = 13.87, \text{MSE} = 0.02, \eta^2 = .25, p = .003 \), suggesting that the face composite effect was different in the dynamic and static conditions. Post hoc comparisons (Bonferroni corrected, alpha = .0125) showed a robust face composite effect in the random interval multi-static condition. However, no face composite effect in the dynamic condition. Also, the recognition was not significantly different between the dynamic and multi-static conditions in the aligned trials or in the misaligned trials.

3.3. Discussion

In the current experiment, we created a new set of multi-static stimuli. We inserted random intervals between the face images used in the dynamic stimuli to remove the motion information while keeping the images and their display order identical to those used in the dynamic stimuli. We replicated the findings of Experiment 1: in the new random interval multi-static condition, we still observed a robust face composite effect as we did in Experiment 1; in contrast, in the dynamic condition identical to that in Experiment 1, we failed to observe a significant face composite effect: participants showed equivalently strong recognition of the upper parts of the familiarized faces from the aligned and misaligned composite faces. Taken together, these findings suggest that the rigid facial motion, not simply the image display order itself, led to the disappearance of the face composite effect.

Although the random interval multi-static condition matched the dynamic condition in terms of the presentation order of the face images, the very fact that the interval between the face images varied was a concern. Because of the random presentation, the face images appeared irregularly and unevenly in the multi-static condition, which differed from the dynamic stimuli that presented the face images regularly and smoothly. It might be the regularity of the face images from one angle to another, not the motion information itself, which led to the disappearance of the face composite effect in the dynamic condition. To address this issue, we conducted Experiment 3.
4. Experiment 3

Experiment 3 used exactly the same materials and procedures as Experiment 2. The only difference was that the interval between each two images in the multi-static stimuli was fixed at 400 ms, which was the mean interval of the random interval multi-static stimuli used in Experiment 2. The 400 ms intervals also effectively removed the apparent motion, while keeping the display order and the static information identical to those used in the dynamic condition. Additionally, the face images now appeared regularly, unlike those in the random interval multi-static condition of Experiment 2. Thus, if dynamic motion was indeed responsible for the disappearance of the face composite effect found in Experiments 1 and 2, we should continue to observe the participants performing equally well in the aligned and misaligned conditions in the dynamic condition. Moreover, we should continue to obtain a robust face composite effect in the fixed interval multi-static condition.

4.1. Method

4.1.1. Participants

Nineteen undergraduates (7 males) participated in the experiment. All participants shared the same race with the face stimuli and had normal or corrected-to-normal vision. None of the participants took part in the prior experiments or knew the individuals depicted in the stimuli.

4.1.2. Materials and procedure

The dynamic stimuli were the same as those used in Experiment 1. A new type of multi-static stimulus was created by inserting 400 ms intervals between face images from 0° to 180° angles, while keeping the images and their display order identical to the dynamic stimuli. Thus, the duration of the multi-dynamic stimulus was 15 pictures × 80 ms = 1200 ms, whereas the duration for the new multi-static stimulus was 15 pictures × 80 ms + 14 intervals × 400 ms = 6800 ms. The composite faces were identical to those used in Experiment 1. The procedure was also the same as that used in Experiment 1. The whole experiment was divided into two blocks, one for the dynamic condition, and the other for the 400 interval multi-static condition. The order of these two blocks was counterbalanced across participants. The aligned and misaligned trials were presented randomly.

4.2. Results

Table 1 shows the means and standard deviations of accuracy, A', and B0 for each condition. As was the case in Experiments 1 and 2, we focused on the A' data. A 2 (condition) × 2 (alignment) repeated measures ANOVA on A' revealed a main effect of alignment: F(1,18) = 4.65, MSE = 0.01, partial η² = .21, p = .045. Participants were more accurate in the misaligned trials (M = .88, SD = .07) than the aligned ones (M = .85, SD = .06).

More importantly, as we found in Experiment 1, the crucial interaction between condition and alignment was significant, F(1,18) = 13.40, MSE = 0.01, partial η² = .43, p = .002. This result indicated that the amount of interference from the lower face part was different for the dynamic and static conditions. Multiple comparisons (Bonferroni correction, alpha = .0125) between the aligned and misaligned trials in the 400 ms interval multi-static condition revealed a significant composite effect (p < .001) whereby the participants recognized the upper face parts better when they were misaligned with the lower parts of the foil faces. In contrast, in the dynamic condition, no significant difference was found between the aligned and misaligned composite faces. Further, the recognition performance for the dynamic and static conditions in the aligned trials and that for the two conditions in the misaligned trials were not significantly different.

These results replicated those observed in Experiment 1, suggesting that dynamic information rather than the display order is the cause of the disappearance of the face composite effect in the dynamic condition.

To examine the consistency of findings in the three experiments, we combined participants’ A’ results from the three experiments with experiment as the between subject factor. Not surprisingly, a 2 (condition) × 2 (alignment) × 3 (experiment) mixed ANOVA showed a significant main effect of test face alignment [F(1,66) = 32.96, partial η² = .33, p < .001] and interaction between condition × alignment [F(1,66) = 26.54, partial η² = .29, p < .001]. We did not find any significant main effect of experiment or interactions that included experiment (p > .185), indicating the main findings to be consistent across the three experimental manipulations.

4.3. Discussion

The current results successfully replicated the findings from Experiment 2: the dynamic stimuli failed to produce a face composite effect. In contrast, in the fixed interval multi-static condition, we continued to observe the robust face composite effect. There was no difference in results between the multi-static conditions of Experiments 2 and 3. Thus, the interval duration randomization did not exert any impact on the size of the face composite effect. This finding suggests that both the fixed interval at 400 ms and randomized intervals were equally effective in removing dynamic facial motion information. Further, because the results in the static conditions of Experiments 1–3 were not significantly different from each other, the presentation order of the face images in the multi-static stimuli also appeared not to affect the size of the face composite effect in the static conditions. Since the dynamic condition was identical across the three experiments, the consistent finding suggests that the effect of facial motion on the face composite effect is highly robust. These results taken together suggest that dynamic facial information may more likely lead to featural, rather than holistic, processing whereby it allows participants to process faces in a feature-based manner, which would reduce the holistic interference from the irrelevant face part.

5. General discussion

We conducted three experiments to examine directly the effect of rigid dynamic facial information on face recognition. More specifically, we tested two opposing hypotheses, the holistic influence hypothesis and the featural influence hypothesis. The former suggests that dynamic facial information leads to the processing of the face more holistically, whereas the latter suggests that dynamic facial information leads to face processing in a featural or part-based manner. We capitalized on the classic face composite effect that when seeing a composite face made up of the upper part of one person and the lower part of another, the upper part is better recognized in the misaligned condition than in the aligned condition. We used this effect to test these hypotheses because the effect has been widely accepted as reflecting holistic processing and the aligned condition fuses the upper and lower parts into a face gestalt, which makes it difficult to recognize the upper face part independently of the lower face part.

Unlike the previous studies that tested the face composite effect with the use of only a single static frontal face image after familiarizing participants with the same image, we used multiple static
face images of the same person from multiple angles. We used them to create a familiarization face stimulus that was either dy-
namic (appearing to move side-to-side) or static (appearing one
image at a time without motion). Also, we never used the frontal
image in the familiarization stimuli but tested participants on
the frontal view of the upper part of the face. Regardless of these
changes, in three experiments, we found a robust face composite
effect in the multi-static condition: Participants reliably recognized
the upper face part better in the misaligned trials than in the
aligned trials. This effect is so robust that it can be observed even
when the exact image of the frontal view face has never been seen
before and must be inferred from static images of the same face
from other angles. In contrast, we found when comparing the per-
formance between the aligned and misaligned trials in the dy-
namic condition of the three experiments, unlike in the multi-
static condition, no significant difference was found between the
two types of trials. In other words, in the dynamic condition, there
was no face composite effect. Further, this robust static-dynamic
difference was found regardless of how we presented the static
face images temporally in the multi-static condition.

It should also be noted that in the present study, all familiariza-
tion and test stimuli used to assess the face composite effect were
identical in the multi-static and dynamic conditions. Hence, the
static-dynamic difference was rooted in the nature of the familiar-
ization stimuli. Because the face images used to create the dynamic
and multi-static familiarization stimuli were identical, the robust
difference between the dynamic and multi-static conditions could
only be attributed to the naturally flowing motion information
present in the dynamic face stimuli but not in the multi-static
stimuli. These findings taken together suggest that after familiar-
ization with the multi-static face stimuli, the upper and lower
parts of the composite face in the aligned trials are more difficult
to decompose than those in the misaligned trials, because the upper
and lower face parts have formed a gestalt when aligned.

In contrast, during familiarization, the rigid motion appears to
influence participants to process the target face in a part-based
manner. In other words, they learn the upper and lower parts of
the target face separately. Consequently, during the test, the recog-
nition of the upper part of the test face becomes less interfered
with by the lower part of the test face. Thus, the upper part of
the test face is recognized equally well regardless of whether it is
aligned or misaligned with the interfering lower face part.

This part-based processing suggestion is consistent with the
findings of several prior studies. For example, both Hill and
Johnston (2001) and Lander, Christie, and Bruce (1999, Experi-
ments 1 and 2) found that face inversion was less pronounced
when faces are shown dynamically than when they are static dis-
plays. As inversion has been thought to disrupt disproportionally
face holistic processing (McKone, 2004), the reduced inversion ef-
fect suggests that facial motion may exert its influence on face rec-
ognition by shifting it to part-based processing. However, as
pointed out earlier, because there is much controversy regarding
whether inversion indeed disrupts holistic more than featural pro-
cessing (Yovel & Kanwisher, 2004), the suggestion that facial mo-
tion affects specifically featural processing has remained
unconfirmed (Hill & Johnston, 2001; Lander, Christie, & Bruce,
1999). The present findings thus provide the first direct evidence
to confirm this suggestion and thus support the featural influence
hypothesis.

Our findings are relevant to the representation enhancement
hypothesis proposed by O’Toole, Roark, and Abdi (2002) and Roark
et al. (2003). This hypothesis posits that rigid motion helps to build
an enhanced representation, which in turn may sometimes lead to
more accurate recognition of faces learned dynamically than faces
learned in static. In addition to its potential benefits in enhancing
recognition accuracy, the face representation formed in motion is
thought to also be more viewpoint flexible, because it incorporates
three-dimensional face structural information. Our data imply that
one of the outcomes of such viewpoint flexibility is that rigid mo-
tion makes participants more inclined to learn about a dynamic
face in a part-based manner, with the consequence that the recog-
nition of the upper part of the target face is less affected by the
interfering lower part of the foil face.

Why does rigid facial movement influence observers to process
faces in a part-based manner? An attention cuing explanation may
provide a reasonable account. This account places emphasis on two
important characteristics of rigid facial motion: sequentially dis-
played face images and the temporal continuation of these images.
Our findings from Experiment 1 can be explained by viewing strat-
yegy differences between the dynamic and multi-static conditions.
As acknowledged by a number of researchers, visual spatial atten-
tion can be guided by both the exogenous and endogenous cues
(Correa, Lupiáñez, & Tudela, 2006; Correa et al., 2006; Folk,
Remington, & Johnston, 1992; Theeuwes, 1991). In the dynamic
condition, due to the face view changing sequentially, observers
could predict the content of the upcoming image (e.g., viewpoint
and spatial position) by the appearance of the image previous to
this image (Friedman, Vuong, & Spetch, 2009; Stone, 1998, 1999;
Vuong & Tarr, 2004). That is, each image could act as a cue for
the following image. By this cuing effect, participants could appro-
priately and smoothly allocate attention to specific face parts
on the images. Because the present task was to recognize only the
upper part of face, participants could allocate their attention re-
sources only to the upper part of the moving familiarization face
while ignoring the lower part. Also, because the target to be recog-
nized was a frontally facing upper part of a face, observers could
allocate more attention to the upper parts of the moving familiar-
ization face in viewpoints that were closer to the frontal view.
This accounting is consistent with the idea that not all of the views
in the multiple-dynamic faces were equally attended, with some
images receiving more attention, while others received less atten-
tion (Friedman, Vuong, & Spetch, 2009; Harman & Humphrey,

However, in the multi-static condition of Experiment 1, the im-
age display order was randomized, which made it impossible to
predict the face viewpoint and position of certain face part posi-
tions in the next image. This in turn means that it would be diffi-
cult to shift attention from the face part on the present image to
the same part on the next image. Because of the brief exposure
time and the cost of attention shifting, focusing on individual face
parts would be a less optimal processing strategy relative to a
holistic viewing strategy, resulting in the dynamic-static difference
in the face composite effect.

This explanation suggests that image display order is a very
important factor in rigid facial movement; however, it is not suffi-
cient to explain the effect of rigid motion per se. For the multi-static
stimuli used in Experiments 2 and 3, the face composite effect
was still robustly observed, even though face images were dis-
played in the same sequential order as those in the dynamic con-
dition. For these multi-static images, just like the dynamic images,
each sequentially displayed image should provide a cue to the
position of specific face parts in the next image. We suspect that
the reason why the same cue did not lead to the same effect
was because it was inhibited during the interval between each im-
age. As other researchers have suggested, visual attention cues can
produce both facilitation and inhibition effects (Lupiáñez et al.,
2001; Posner & Cohen, 1984). Which of these two effects is ob-
served is in part determined by the temporal separation between
the cue and the target. Generally, facilitation effects dominate
when the SOA (stimulus onset asynchrony) is less than 300 ms,
and are replaced by inhibition effects when the SOA is longer than
300 ms (Klein, 2000; McAuliffe & Pratt, 2005). For the multi-static
stimuli used in the present experiments, the interval between images was 400 ms, which is likely to produce inhibition that dis-tracts attention from the cued position. This distraction prevented attention from focusing on the task relevant face part, which made it difficult for participants to use a stable featural processing strat-egy and in turn might have led them to resort to a holistic process-ing strategy that is known to be a more dominant mode of process-ing among adult face processing experts.

In summary, the present study provided the first direct evi-dence to suggest that rigid dynamic facial information influences featural or part-based face processing by making it easier to decompose the face parts from a face gestalt. To achieve this effect, one must observe a face from multiple angles presented in a natu-ral temporal sequence such that the face appears to move coher-ently. One of the direct consequences of this viewing is that the faces are processed more in a part-based manner.

However, it should be noted that the face composite effect is only one of many phenomena one can capitalize on to test the role of rigid facial motion on face featural or holistic processing (e.g., the Tanaka–Farah effect that face parts are easier to recognize in a face gestalt than in isolation if the parts are learned in a face ge-stalt). Additionally, the present study only focused on the contrast between face featural and holistic processing. It is unclear as to whether rigid dynamic facial information also leads to featural pro cessing during face recognition more than the processing of the spacing information among major internal facial parts such as eyes, nose, and mouth (i.e., configural information). Further work is thus needed to ascertain whether the observed effect of dynamic facial information on face featural processing here can be generalized be-yond the current face composite effect. Future studies should also use paradigms other than the face composite effect to examine the influence of rigid motion on face processing. Because the task de-mand of the present study called for attention to facial parts, par-ticipants might capitalize on the information provided by the moving face to process face parts. However, it is unclear whether the influence of rigid facial motion will change when the task de-mand is to process the face as a whole. Moreover, although the present findings suggest strongly that rigid facial motion influ ences featural processing, this does not necessarily lead to an improvement in overall face recognition performance as evidenced by the results of Experiments 1 through 3: although the dynamic condition allowed participants to decompose the faces into parts better, it did not have an overall performance advantage over the static condition in recognizing the upper portion of the face. Last but not least, as mentioned earlier, facial dynamics not only in-volves rigid motion but also elastic motion. It is still unclear as to whether elastic facial motion facilitates or hinders the processing of facial featural and holistic information during face recognition and what specific roles the two types of dynamic information play in face recognition specifically and face processing more generally. These questions await investigation in the near future.

Acknowledgments

This research is supported by grants from the Natural Science and Engineering Research Council of Canada, NIH (R01 HD046526), the Natural Science Foundation of China (60910006, 31028010, 31070908), the Zhejiang Provincial Natural Science Foundation of China (Y2100997), and the Zhejiang Qianjiang Talent Foundation (QJC1002010 101371-M). We thank two anonymous reviewers for their comments.

References


