
VISION
RESEARCH
An International Journal for Functional Aspects of Vision
Biochemistry & Cell Biology • Molecular Biology & Genetics
Anatomy, Physiology, Pathology & Pharmacology • Optics, Accommodation & Refractive Error
Circuitry & Pathways • Psychophysics • Perception • Attention & Cognition
Computational Vision • Eye Movements & Visuomotor Control
ISSN 0042-6989 | Volume 51 | Issue 1 | January 2011

Vision Research 54 (2012) 61–67

Contents lists available at SciVerse ScienceDirect

Vision Research

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

# Timing constraints of temporal view association in face recognition

Gabriel Arnold[a], Eric Siéroff[a,b,*]

[a] *Paris Descartes University, Laboratoire de Psychologie et Neuropsychologie Cognitives, France*
[b] *Institut Universitaire de France, France*

ARTICLE INFO

*Article history:*
Received 14 March 2011
Received in revised form 15 November 2011
Available online xxxx

*Keywords:*
Face recognition
Temporal view association
Timing constraints
Motion
Generalization

ABSTRACT

This study tests the hypothesis that the recognition of a face is facilitated when the face has previously been presented in a rapid rather than a slow view sequence. We used a sequential comparison task, in which a first face, rotating back and forth around a left or a right three-quarter view, was followed, after a 1-s delay, by a static view of a second face, with the same or a different viewpoint. We compared rapid (180 ms per view) and slow (720 ms per view) sequences to evaluate the timing constraints of temporal view association, and video and view sequences to evaluate the importance of motion smoothness. Response times were faster for rapid view sequences, showing the importance of perceiving the views in a short temporal window. When the views of a face are perceived in a rapid sequence, attention may be distributed over the entire sequence, leading to a unified representation associating the views. This unified representation facilitates the recognition of the face. Moreover, response times were faster for view sequences than for video sequences, showing no advantage of motion smoothness.

## 1. Introduction

The retinal image of a face changes dramatically with viewpoint, expression, lighting, and other factors. In other words, for each individual face, a large number of dissimilar views are possible. How does the brain recognize faces despite this view dissimilarity? In natural conditions, faces are often perceived in motion, so the different views of each face are often perceived in close temporal sequences. The visual system may benefit from temporal view sequences to associate the views in a unified face representation and to facilitate recognition under different conditions of viewpoint, expression, or lighting (Wallis & Bülthoff, 2001; Wallis et al., 2009). However, the rate of view succession may also play a role in the association of views in a unified representation. The goal of our study was to evaluate the timing constraints of view association from temporal view sequences. We found that face recognition is facilitated when the different views of a face are presented in rapid as opposed to slow view sequences.

### *1.1. Temporal view association and timing constraints*

According to the temporal association hypothesis, the visual system associates different facial views when they are perceived in a temporal sequence (Wallis & Bülthoff, 1999). Indeed, the ability to discriminate between two different faces is impaired when the views of these two faces have previously been presented in the same temporal sequence, because these views may have been erroneously associated in a single face representation (Wallis & Bülthoff, 2001). Other studies have shown that object and face recognition benefits from the perception of views in a temporal sequence. Recognition is facilitated when a face or an object is learned from a temporally coherent sequence of views rather than one single view (Pike et al., 1997; Pilz, Thornton, & Bülthoff, 2006; Thornton & Kourtzi, 2002) or multiple but temporally independent views (Balas & Sinha, 2008; Kourtzi & Shiffrar, 1999; Liu, 2007; Pilz, Thornton, & Bülthoff, 2006). In monkeys, the view selectivity of inferotemporal neurons can be modified by the temporal correlation of views (Miyashita, 1988; Miyashita, Date, & Okuno, 1993). Individual inferotemporal neurons have been found to show selective response to specific colored fractal images. However, after learning sequences of random colored fractal images, these same neurons showed some additional response to the temporal-neighboring images of the sequence. This temporal-neighboring effect might be explained by the spatial proximity of the neurons coding for specific images perceived in a temporal sequence. A strong possibility is that invariant object representations are learned by a neural network, which uses spatiotemporal correlations to map different views of objects onto topographic representations (Michler, Eckhorn, & Wachtler, 2009).

The association of views from temporal view sequences may have a time limit. One way to investigate the timing constraints of temporal view association is to vary the speed of view sequences, that is, the duration of each view in the sequence.

---

[*] Corresponding author. Address: Paris Descartes University, Centre Henri Piéron, Equipe Psychologie et Neuropsychologie du Vieillissement, 71 Avenue Edouard Vaillant, 92100 Boulogne-Billancourt, France.

*E-mail address:* eric.sieroff@parisdescartes.fr (E. Siéroff).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.12.001

A benefit of rapid sequences has been shown: different parts of a face are better integrated when they are presented in rapid sequences than in slow ones (Anaki, Boyd, & Moscovitch, 2007; see also Singer & Sheinberg, 2006). Three horizontal segmented parts of full-face views were presented for 17 ms in a sequence, with short (200 ms), long (700 ms), or null intervals after each part. Inversion and misalignment effects, reflecting facial part integration, were found only at short and null intervals, that is, when the different parts were presented in a short temporal window. According to Anaki, Boyd, and Moscovitch (2007), the different parts presented in a sequence are temporarily stored and combined in a short-term visual buffer. This buffer is limited in time and thus requires a rapid succession of views. Another related explanation of the benefit of rapid sequences focuses on the attentional operations involved in rapid and slow sequences. Several studies using rapid serial visual presentations have shown that the processing of two targets in a sequence is improved when the targets are separated by more than 400–500 ms (for a review, see Shapiro, Arnell, & Raymond, 1997). In slow view sequences, because viewers may have time to focus their attention on each view, the representations they form relate to the individual views rather than to their association. Conversely, in rapid view sequences, the presence of the views in a short temporal window may allow the attention to be distributed over the entire sequence of views, facilitating the association between views. Thus, rapid view sequences may facilitate a unified representation of the face and its maintenance in a short-term visual buffer.

Rapid sequences have also been found to be superior when full faces are processed from different viewpoints. Busey and Zaki (2004) found that faces are better recognized from novel symmetrical viewpoints after rapid (180 ms for each view) than slow (720 ms for each view) sequences rotating back and forth in depth around a learning viewpoint. In this long-term memory task, faces were first learned in sequences of five different views rotating around one viewpoint showing the left side of the face (−70° or −35°), and then recognized from the same viewpoint, the full-face view (0°), or the symmetrical viewpoints showing the right side of the face (+70° or +35°). The advantage of rapid sequences was found with novel symmetrical viewpoints for both −70° and −35° sequences, and also with the novel +35° viewpoint for −70° sequences. However, there was also an advantage of slow sequences with the novel −70° viewpoint for −35° sequences. According to the authors, motion was smoother in rapid than in slow sequences, facilitating the recovery of 3-D information (i.e., depth information) via structure-from-motion processing (see below). They suggest that such facilitation may occur only with symmetrical viewpoints. Full-face views may provide little 3-D information, and learned views may have been recognized with a simple view-based matching process.

### 1.2. Role of motion in the processing of temporal view sequences

The superiority of rapid view sequences may result from an advantage of motion in face recognition (Busey & Zaki, 2004). Indeed, the presentation of an ordered view sequence of a face may give the impression of motion, and motion is more apparent, or smoother, in rapid than in slow sequences. According to the representation enhancement hypothesis, facial motion – specifically, rigid motion – contributes to recognition by facilitating the perception of the 3-D structure of the face (O'Toole, Roark, & Abdi, 2002) via structure-from-motion processing (Ullman, 1979; Wallach & O'Connell, 1953). The 3-D structure of a face or an object may be an important cue to recognition because it is an invariant feature (Biederman, 1987; Marr & Nishihara, 1978). Structure-from-motion processing refers to the computation of local depth information by comparing the relative distances between features in the different 2-D projections of a face or an object in motion. This integration of form and motion information involves a neural network that includes several regions of the ventral and dorsal streams (Farivar, Blanke, & Chaudhuri, 2009; Orban et al., 1999; Sarkheil et al., 2008).

Note, however, that most of the studies cited above used not real motion but only apparent motion in view sequences. For example, in an attempt to evaluate the role of motion in dynamic face or object recognition, ordered view sequences have been compared to random view sequences in which the temporal view proximity is preserved but the perception of apparent motion is disrupted. However, this method has led to conflicting results. Some studies found an advantage of ordered over random sequences (Vuong & Tarr, 2004; Wallis & Bülthoff, 2001), some did not find any difference (Liu, 2007), and others even found an advantage of random over ordered sequences (Harman & Humphrey, 1999). So it remains unclear whether the association of face views on the basis of temporal proximity depends on the smoothness of the sequence.

### 1.3. Presentation of the study

The aim of the present study was to evaluate the timing constraints of temporal view association and the role of view association vs. motion smoothness in the rapid view sequence superiority (described by Busey and Zaki (2004)), by comparing view and video sequences with different speeds. Like Anaki, Boyd, and Moscovitch (2007), we used a sequential comparison (short-term memory) task, because this kind of task produces few errors and allows the recording of response latencies. A learning face sequence, rotating back and forth around a three-quarter view, was followed by a static test view of the same face or a different one. For the view sequences, the presentation times for each view were the same as in the study by Busey and Zaki: 180 ms for each view in the rapid sequences and 720 ms in the slow sequences. For the video sequences, the rotation speed was the same as the apparent speed of the view sequences.

The superiority of rapid view sequences may be explained by the presence of the views in a short temporal window because the views follow one another more quickly in rapid than in slow sequences. If this explanation is valid, we should see an overall advantage for rapid over slow view sequences. The rapid sequence superiority may also be explained by apparent motion, because apparent motion is smoother in rapid than in slow view sequences; motion is also smoother in video than in view sequences. If this explanation applies, we should find a general advantage for video over view sequences, which could be stronger for slow sequences because, according to Busey and Zaki (2004), slow view sequences do not produce an apparent motion.

In the present study, we used only sequences rotating around the 45° view (which is frequently used in the literature) unlike Busey and Zaki (2004), who used 35° and 70° views. We were using an additional smoothness factor (view or video sequences) and wanted to limit the duration of the experiment and the number of repetitions of each face. We used sequences of faces rotating around the left (−45°) and right (+45°) three-quarter views, again unlike Busey and Zaki, who used only left three-quarter views.

Finally, in order to evaluate view generalization, the static test face was presented from the same viewpoint as the learning face (the central three-quarter view) or from a novel viewpoint (the full-face view or the symmetrical three-quarter view). The elaboration of a unified representation may facilitate face recognition under different viewpoint conditions, that is, for novel viewpoints as well as learned ones. If the presentation of the views in a short temporal window facilitates the elaboration of a unified representation, rapid sequences should be superior for all the different test viewpoints. In the same way, if the perception of apparent motion facilitates the elaboration of an abstract representation based on the invariant features of the face (3-D structure), we should observe an advantage of video sequences over view sequences for all the different viewpoints.

## 2. Method

### 2.1. Participants

Ninety-six psychology students at Paris Descartes University (85 females and 11 males, mean age = 21.2, range = 18–40) participated in the experiment. All participants were naive to the experiment and unfamiliar with the presented faces. They reported normal or corrected-to-normal vision, and were all right-handed according to the Edinburgh Inventory (Oldfield, 1971). All participants provided informed consent and received course credits for participating in the experiment.

### 2.2. Design

The participants completed 144 trials, which included 6 repetitions of each of the 24 conditions resulting from the combination of the speed of the learning sequence (rapid, slow), smoothness of the learning sequence (view sequences, video sequences), Test viewpoint (same, full-face, symmetrical three-quarter), and Face identity (same, different) factors. The side of the learning face (left, right) was a between-subjects factor (to limit the duration of the experiment and the number of face repetitions).

### 2.3. Stimuli

Twenty-six actors (14 males and 12 females) were filmed in the same room and under the same lighting conditions. Each actor sat on a rotating chair, placed in front of a plain black wall. Two spotlights placed on the left and the right side of the actor and oriented toward a white polystyrene plate, which reflected the light toward the actor, lighted them indirectly. Each actor was filmed at shoulder height and wore a black t-shirt and a black scarf to hide the shoulders and hair. The chair was manually rotated, with a constant speed, from the left (−90°) to the right profile (+90°), and from the right to the left profile, to obtain a semicircular rotation in depth of the head. The rotation of the chair started a few degrees before and ended a few degrees after the 90° angle, to avoid motion acceleration and deceleration during the recording.

Full-color view and video sequences were built with Adobe After Effects CS4. Each sequence of the learning face showed a head rotating back and forth around the left (−45°) or the right (+45°) three-quarter view, ±15° away from this view, resulting in a 30°-amplitude rotation. Each slow sequence began with the 45° view, then included a first rotation toward the 30° or 60° view, followed by a second rotation backward, and ended with the 52.5° or 37.5° view, respectively. Each rapid sequence consisted of four times as many rotations as the slow sequence. The total presentation time was the same (5760 ms) for rapid and slow sequences. The view sequences were created with five views extracted from the semicircular sequence, with the following angles: ±60°, ±52.5°, ±45°, ±37.5°, and ±30° (see Fig. 1). The presentation time of each view was 180 ms in rapid sequences and 720 ms in slow sequences. For video sequences, the original rotation speed was slightly modified in order to have the same total presentation time as in the view sequences. Motion was completely linear before and after changing direction and the rotation speed was about 40° per second in rapid and 10° per second in slow sequences. The functional frame rate was 25 frames/s. Thus, each frame was presented for 40 ms and there were 144 frames in each (rapid or slow) video sequence. The angular difference between frames was about 1.6° in rapid video sequences and about 0.4° in slow video sequences. Video sequences were phenomenologically smoother than view sequences, even for rapid sequences. Three different viewpoints, left three-quarter (−45°), right three-quarter (+45°), and full-face (0°), were used as test views.

The faces were about 7.6 cm high and 7.3 cm wide and appeared on a black background in a 12.9 cm high × 16.2 cm wide frame (480 × 600 pixel resolution). Two horizontal black bands, 2.7 cm high, were placed at the top and bottom of the image to block out the shoulders and the top of the head.

### 2.4. Procedure

The experiment was run with SuperLab 4 on an iMac G5 with a 24″ color monitor with a 1920 × 1200 pixel resolution.

Participants sat at a distance of 57 cm from the screen. The sequence for each trial consisted of a learning face followed by a test face (see Fig. 2). Each trial began with a blank screen for 1000 ms. Then a rotation sequence (learning face) showing a head rotating back and forth around the left or the right three-quarter view was presented for 5760 ms. There were four types of sequences combining the speed (rapid, slow) and smoothness (video, view) factors with two rotation directions (a first rotation toward the 30° or the 60° view). After a blank screen for 1000 ms, the test face was presented, with three different viewpoints (same three-quarter, full-face, or symmetrical three-quarter), until the participant responded. The test face had either the same identity as the learning face or a different one. After the response, there was an inter-trial interval of 1000 ms before the beginning of the next trial. Faces were presented at the center of the screen and subtended about 7.6° vertically and 7.3° horizontally in a 12.9° × 16.2° black frame.

Participants were asked to indicate, as accurately and as rapidly as possible, whether the test face was the same person as the learning face, by pressing two different keys on the keyboard with the index and middle fingers of their right hand. Half of the participants responded *same* with the index and *different* with the middle finger, and the other half did the opposite. Response times (RTs) and errors were recorded.

Twenty-four different faces (12 females, 12 males) were used in the experimental session. For each participant, each face was presented 6 times (corresponding to the three test viewpoints and the two face identities), and appeared with only one of the four types of learning sequences (rapid view, slow view, rapid video, or slow video). A Latin square design over all participants was used to present all the different faces with the four types of learning sequence.

The experimental session was divided into two blocks of 72 trials separated by a short break of 1 min. Before the experiment, participants performed a practice session of 12 trials with two faces that were different from the faces used in the experimental session. The total duration of the experiment was 30 min.

## 3. Results

### 3.1. Response times

We calculated the mean RT and standard deviation of each condition for each participant. RTs corresponding to errors and exceeding the mean by more than 2 standard deviations were excluded from the RT analysis. As same and different trials may involve different processes, our experimental predictions relate only to same trials. An ANOVA was calculated with the side of the learning face (left, right) as between-subjects factor, and speed

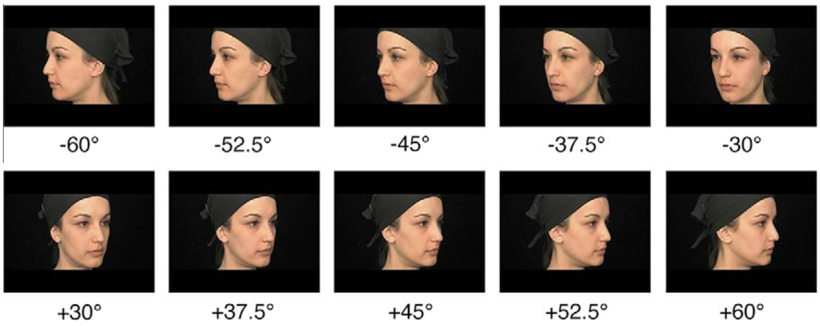

**Fig. 1.** Examples of the five views appearing in a view sequence. *Top row*: views centered on the left three-quarter view (−45°). *Bottom row*: views centered on the right three-quarter view (+45°).

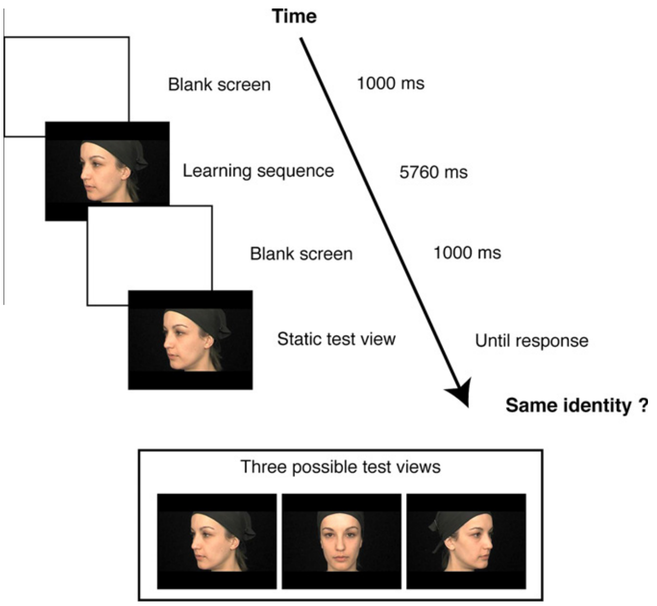

**Fig. 2.** Description of a trial. The learning face was either a rapid or a slow sequence and either a video or a view sequence. After a 1-s delay, it was followed by a static test view with the same identity as the learning face or a different one and with the same viewpoint or a different one.

(rapid, slow), smoothness (video, view), and Test viewpoint (same, full-face, symmetrical) as within-subject factors. An alpha level of .05 was used. See Table 1 for the complete (same and different) RT data.

There was a significant effect of speed [$F(1,94) = 4.18$; $p < .05$; $\eta^2 = .04$]. As expected, RTs were faster in response to rapid (747 ms, SD = 180 ms) than to slow sequences (765 ms, SD = 200 ms). There was also a significant effect of smoothness [$F(1,94) = 6.37$; $p < .05$; $\eta^2 = .06$], showing faster RTs for view (748 ms, SD = 191 ms) than for video sequences (764 ms, SD = 190 ms). In addition, there was a significant interaction between speed and smoothness [$F(1,94) = 6.16$; $p < .05$; $\eta^2 = .06$]: RTs were faster for rapid (731 ms, SD = 173 ms) than for slow view sequences (764 ms, SD = 205 ms) [$F(1,94) = 9.29$; $p < .01$; $\eta^2 = .09$], but no difference was found between rapid (763 ms, SD = 187 ms) and slow (765 ms, SD = 196 ms) video sequences (see Fig. 3).

The side of the learning face was not significant but there was a significant Side × Speed × Smoothness interaction [$F(1,94) = 4.91$; $p < .05$; $\eta^2 = .05$] (see Fig. 4). There was a significant Speed × Smoothness interaction when the learning face was rotating around the left three-quarter view [$F(1,94) = 11.03$; $p < .01$; $\eta^2 = .11$]. RTs were significantly faster for rapid (725 ms, SD = 152 ms) than slow (763 ms, SD = 187 ms) view sequences [$F(1,94) = 6.33$; $p < .05$; $\eta^2 = .06$], although there was no significant speed effect for video sequences [$F(1,94) = 2.01$; $p$ = ns; $\eta^2 = .02$]. For learning faces rotating around the right three-quarter view, there was no Speed × Smoothness interaction [$F(1,94) < 1$; $p$ = ns; $\eta^2 = .00$]. The speed effect was not significant for video sequences [$F(1,94) = 2.73$; $p$ = ns; $\eta^2 = .03$], and there was only a tendency toward faster RTs in rapid (738 ms, SD = 191 ms) than in slow view sequences (765 ms, SD = 223 ms) [$F(1,94) = 3.22$; $p < .08$; $\eta^2 = .03$].

There was a significant Test viewpoint effect [$F(2,188) = 104.65$; $p < .001$; $\eta^2 = .53$]. RTs were faster with the same (689 ms, SD = 148 ms) than with the full-face viewpoint (758 ms, SD = 186 ms) [$F(1,94) = 62.31$; $p < .001$; $\eta^2 = .40$], and faster with the full-face than with the symmetrical viewpoint (821 ms, SD = 210 ms) [$F(1,94) = 52.38$; $p < .001$; $\eta^2 = .36$]. No other effect or interaction was significant.

### 3.2. Errors

The global error percentage was only 3% (with 3.4% omissions, and 2.5% false alarms). We conducted an ANOVA on the global error percentage for all data (including same and different responses) with the same factors as in the RT analysis. Again, an alpha level of .05 was used. See Table 2 for results.

Only Test viewpoint had a significant effect [$F(2,188) = 21.81$; $p < .001$; $\eta^2 = .19$], with the lowest error percentage when the test viewpoint was the same three-quarter view (1.4%, SD = 3.6%), and no difference between full-face (3.8%, SD = 5.7%) and symmetrical three-quarter (3.7%, SD = 5.9%) viewpoints. There was no other effect or interaction.

## 4. Discussion

### 4.1. The superiority of rapid view sequences and the timing constraints on view association

In the present study, we investigated the timing constraints on the processing of temporal view sequences of faces. Like Busey and Zaki (2004), but using a different method, we found a superiority of rapid over slow view sequences. They used a long-term recognition task, whereas we used a short-term memory task. Thus, the recognition of faces learned from view sequences may have the same timing constraints for short and long retention intervals. However, we found that rapid view sequences were superior independently of the test viewpoint, whereas Busey and Zaki found this superiority only with symmetrical viewpoints. Our task induced very few errors, allowing the recording of response latencies. One possibility is that the recording of response latencies may be more sensitive to the evaluation of perceptual processes than the memory task and accuracy analysis used by Busey and Zaki.

**Table 1**
Mean response times in ms and standard deviations (SD) for same and different trials as a function of side of the learning face (left, right), speed (rapid, slow), smoothness (video, view), and test viewpoint (same, full-face, symmetrical).

| | Video sequence | | | | View sequence | | | |
|---|---|---|---|---|---|---|---|---|
| | Rapid | | Slow | | Rapid | | Slow | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| *Same trials* | | | | | | | | |
| Left | | | | | | | | |
| Same | 704 | 146 | 671 | 116 | 651 | 106 | 693 | 121 |
| Full-face | 744 | 147 | 754 | 167 | 727 | 149 | 768 | 185 |
| Symmetric | 840 | 210 | 801 | 190 | 796 | 161 | 829 | 218 |
| Right | | | | | | | | |
| Same | 692 | 163 | 713 | 144 | 688 | 169 | 700 | 197 |
| Full-face | 767 | 193 | 792 | 234 | 740 | 197 | 769 | 207 |
| Symmetric | 832 | 205 | 858 | 241 | 785 | 198 | 826 | 248 |
| *Different trials* | | | | | | | | |
| Left | | | | | | | | |
| Same | 698 | 128 | 719 | 138 | 717 | 122 | 745 | 156 |
| Full-face | 806 | 237 | 794 | 201 | 788 | 231 | 758 | 161 |
| Symmetric | 689 | 124 | 714 | 137 | 740 | 158 | 739 | 138 |
| Right | | | | | | | | |
| Same | 692 | 130 | 713 | 136 | 700 | 142 | 707 | 121 |
| Full-face | 815 | 269 | 790 | 179 | 740 | 166 | 767 | 203 |
| Symmetric | 709 | 156 | 714 | 137 | 707 | 131 | 720 | 147 |

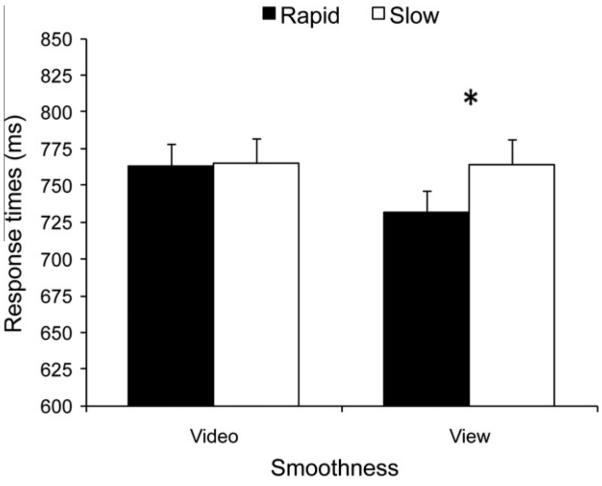

**Fig. 3.** RTs in ms as a function of speed and smoothness. Error bars correspond to standard errors of the mean. $^*p < .05$.

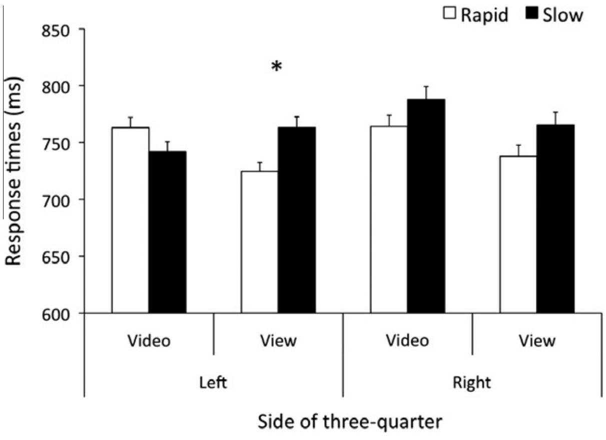

**Fig. 4.** RTs in ms as a function of side, speed and smoothness. Error bars correspond to standard errors of the means. $^*p < .05$.

**Table 2**
Mean global error percentages and standard deviations (SD) as a function of side (left, right), speed (rapid, slow), smoothness (video, view), and test viewpoint (same, full-face, symmetrical).

| | Video sequence | | | | View sequence | | | |
|---|---|---|---|---|---|---|---|---|
| | Rapid | | Slow | | Rapid | | Slow | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| *Left* | | | | | | | | |
| Same | 1.0 | 2.8 | 1.4 | 4.0 | 1.9 | 4.6 | 1.4 | 3.6 |
| Full-face | 5.4 | 6.3 | 3.1 | 4.4 | 3.6 | 4.8 | 2.8 | 5.3 |
| Symmetric | 4.5 | 5.9 | 2.8 | 4.7 | 3.1 | 6.3 | 4.5 | 7.3 |
| *Right* | | | | | | | | |
| Same | 1.7 | 5.4 | 1.7 | 4.2 | 1.2 | 3.0 | 0.9 | 2.6 |
| Full-face | 4.2 | 6.0 | 4.2 | 6.4 | 3.6 | 7.1 | 3.6 | 4.8 |
| Symmetric | 3.1 | 5.6 | 5.2 | 7.0 | 2.4 | 4.2 | 3.8 | 5.4 |

The superior processing of rapid over slow sequences may be explained by the temporal association hypothesis, which posits that the visual system associates several facial views when they are perceived in a close temporal sequence (Wallis & Bülthoff, 2001). Face recognition was facilitated when the views of a face were perceived in a short temporal window (advantage of rapid over slow sequences), revealing the existence of timing constraints on temporal view association. The superiority of rapid view sequences can be explained by the involvement of different attentional processes in rapid and slow view sequences. In rapid sequences, the presence of the views in a short temporal window may allow attention to be distributed over the entire sequence, facilitating the association of views in a unified representation. In these conditions, the sequence can be encoded as a whole. Conversely, in slow view sequences, attention has time to focus on each individual view, disrupting the encoding of the sequence as a whole. This view is similar to the view of Anaki, Boyd, and Moscovitch (2007), who hypothesize that the rapid succession of views may allow them to be associated in a short-term visual buffer, facilitating the development of a unified representation of the face. Rapid view sequence superiority (in our results) and rapid part sequence superiority (Anaki, Boyd, & Moscovitch, 2007) show the importance of temporal proximity, with its timing constraints, for the integration of facial information in a unified representation.

In one case, facial information concerns views, and in the other case it concerns parts, but the processes may be similar.

Another possible explanation of our results is that the comparison process in short-term memory may have been facilitated in rapid sequences because the delay between the views presented in the learning sequence and the test view was, on average, shorter in rapid than in slow sequences. According to the delay hypothesis, we should also have found a superiority of rapid over slow video sequences, because the delays were the same in view and video sequences. However, rapid sequences were superior only for view sequences. It is possible that the presentation time of each view in video sequences was too short, preventing the generation of a stable short-term memory trace of the views. Still, in our experiment, the last view of each sequence was the 52.5° or 37.5° view, both of which are very similar to the 45° view. If the delay hypothesis is correct, we should have found a superiority of rapid view sequences only for learned 45° views and not for novel full-face or symmetrical views. We obtained a rapid sequence superiority effect independently of the test viewpoint. This result was expected because all viewpoints should benefit from the formation of a unified representation, based on the temporal association of views. So, even though we cannot totally exclude the possibility that the delay between the learning and the test views plays a role (and it would be interesting to systematically vary this delay in future studies), we believe that our results are best explained by the timing constraints on the view association.

### 4.2. Timing constraints on view association or motion smoothness?

Another potential explanation of the superiority of rapid sequences is the representation enhancement hypothesis, which posits that rigid motion facilitates the perception of the face's 3-D structure, which is useful for recognition (O'Toole, Roark, & Abdi, 2002). According to the representation enhancement hypothesis, smooth motion should improve face recognition and we should have obtained faster RTs in video than in view sequences. The video sequences are phenomenologically smoother than the view sequences, even when they are rapid. However, not only did we find that the use of video sequences did not facilitate face recognition, but the RTs were faster in view than in video sequences, specifically for rapid sequences. Thus, our results do not support the representation enhancement hypothesis, at least in the conditions we used, but are compatible with the temporal association hypothesis. Note that this does not exclude a role for motion in face processing, and video sequences have been found to induce superior processing in other types of perceptual processes such as expression recognition (Lander, Chuang, & Wickham, 2006) or action anticipation (Thornton & Hayes, 2004).

We found a superiority of view over video sequences in rapid sequences. This superiority may be explained by a perceptual constraint on the visual system. The number of presented views is larger (25 frames/s) and the presentation time of each view is shorter (40 ms) in a video sequence than in a view sequence (180 ms). It is possible that the presentation time of each view was too short in video sequences, disrupting the processing of each view. A single view of a face can be processed with a very brief presentation time (Rolls & Tovee, 1994), but the processing of a stream of views may require a longer presentation time for each one. As Keysers and Perrett (2002) indicated, the quality of object representation and the ability to recognize an object are reduced when images of objects are presented at a rate faster than one image every 200 ms. Thus, the disadvantage of rapid video compared to rapid view sequences may be explained by the cost of processing many views with very brief processing time for each one in video sequences compared to the processing of a limited number of views with sufficient processing time in view sequences. Temporal view association may require a rapid succession of views, but the processing of each view needs sufficient time. One question is which view duration is best for the temporal integration of views as well as the processing of each individual view. We chose a 180-ms view duration, similarly to Busey and Zaki (2004) and close to the 200 ms used by Anaki, Boyd, and Moscovitch (2007) and Keysers and Perrett (2002). A cell recording study in monkeys has also shown that a sequence of short video clips of body actions was processed as a unified body action when each clip was no longer than 200 ms (Singer & Sheinberg, 2010). However, shorter view durations have been used for efficient temporal view association (80 ms for Balas & Sinha, 2008; 100 ms for Wallis & Bülthoff, 2001). Thus, for temporal view association, the optimal duration of each view may be between 100 ms and 200 ms, but further studies would be helpful.

Note that, in order to have the same total presentation time for each type of sequence, we presented the rapid sequences with four rotations and the slow sequences with only one rotation. The association of the different views in a unified representation may have been facilitated by the repetition of the rotation. However, if repetition played a role, we should also have found a superiority of rapid over slow video sequences, but we did not find any differences between these two conditions. Thus, the repetition of the rotation cannot totally explain our results.

### 4.3. Viewpoint effects

In our study, globally faster RTs were obtained with the same test viewpoint as the learning face. As well, RTs were globally faster for the full-face test view than for the symmetrical three-quarter test view. So, when faces are presented from different viewpoints, recognition performance may decrease as the angular difference between the learning view and the test view increases. However, some researchers have found that the test view that is symmetrical to the learning view has an advantage over other novel test views (Busey & Zaki, 2004; Troje & Bülthoff, 1996, 1998). We did not observe such an advantage. Possibly, the use of natural, asymmetrical faces may explain the absence of this advantage for the symmetrical presentation (Busey & Zaki, 2004).

Another possibility is that the use of rotation-in-depth sequences in a sequential comparison task biased the processing of the test view toward a mental rotation process (Shepard & Metzler, 1971), advantaging the full-face view because the angular difference between the presented view and the full-face view is smaller than that between the presented view and the symmetrical view. In our study, participants knew that the test face always showed the same viewpoint as the learning face (three-quarter view) or it extrapolated viewpoints in the same direction, that is, the full-face or symmetrical three-quarter views. Thus, despite the fact that there was no biased direction in the sequences we used (back-and-forth rotation), participants may have "continued" the rotation in the direction of the extrapolated views. Our angle effect may be comparable to the advantage of novel object views that prolong rather than reverse the direction of a rotation sequence (Friedman, Vuong, & Spetch, 2009, 2010; Vuong & Tarr, 2004). These direction effects may be explained by the encoding of the spatiotemporal sequence, that is, a sequence with a particular direction, as a dynamic characteristic of faces and objects. There is a growing body of evidence suggesting that representations of faces and objects are dynamic and that spatiotemporal signatures may be encoded and used to aid recognition (Freyd, 1987; O'Toole, Roark, & Abdi, 2002; Stone, 1999). The neural networks engaged in object and face recognition may integrate form and dynamic information via interactions between regions of the ventral and dorsal streams (Farivar, Blanke, & Chaudhuri, 2009; O'Toole, Roark, & Abdi, 2002; Sarkheil et al., 2008).

Finally, we found differences between the two sides of the learning face. The effect of speed was numerically similar for view sequences rotated around the left (38 ms) and the right (27 ms) three-quarter views, but was significant only for left view sequences. The interaction between speed and smoothness was also significant only for the left sequences. This result might be explained by facial asymmetry or by hemispheric differences (the main facial feature are located in the left side of the space in the left three-quarter view). However, the side × speed interaction was not significant, and it is difficult to conclude that there was any specific effect. Yet, because other studies have also found an advantage of left over right three-quarter views (Siéroff, 2001; Yamamoto et al., 2005), more research on this question seems necessary.

In conclusion, our results showed that the rapid succession of the views, rather than the smoothness of the motion, facilitates face recognition, and this applies to novel as well as learned views. With rapid view sequences, attention may be distributed over the entire sequence, leading to a unified representation associating the different facial views. The advantage of rapid view sequences over rapid video sequences also shows that temporal view association may require sufficient time for each facial view to be processed. In real life, head motions are often faster than the rapid video sequences we used. However, there are also frequent pauses, and these pauses may play an important role in face processing. The superiority of rapid view sequences may also reveal some constraints on the neural mechanisms underlying face processing.

## Acknowledgments

We thank Stéphane Hain and Anaïs Andreassian for their help creating the video databank. We also thank Zofia Laubitz for correcting our English.

This work was supported by an Agence Nationale de la Recherche Grant, ANR-07-BLAN-0051, to E.S.

## References

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

Balas, B., & Sinha, P. (2008). Observing object motion induces increased generalization and sensitivity. *Perception, 37*, 1160–1174.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Busey, T. A., & Zaki, S. R. (2004). The contribution of symmetry and motion to the recognition of faces at novel orientations. *Memory & Cognition, 32*, 916–931.

Farivar, R., Blanke, O., & Chaudhuri, A. (2009). Dorsal-ventral integration in the recognition of motion-defined unfamiliar faces. *Journal of Neuroscience, 29*, 5336–5342.

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review, 94*, 427–438.

Friedman, A., Vuong, Q. C., & Spetch, M. L. (2009). View combination in moving objects: The role of motion in discriminating between novel views of similar and distinctive objects by humans and pigeons. *Vision Research, 49*, 594–607.

Friedman, A., Vuong, Q. C., & Spetch, M. L. (2010). Facilitation by view combination and coherent motion in dynamic object recognition. *Vision Research, 50*, 202–210.

Harman, K., & Humphrey, G. K. (1999). Encoding "regular" and "random" sequences of views of novel three-dimensional objects. *Perception, 28*, 601–615.

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences, 6*, 120–125.

Kourtzi, Z., & Shiffrar, M. (1999). The visual representation of three-dimensional, rotating objects. *Acta Psychologica, 102*, 265–292.

Lander, K., Chuang, L., & Wickham, L. (2006). Recognizing face identity from natural and morphed smiles. *The Quarterly Journal of Experimental Psychology, 59*, 801–808.

Liu, T. (2007). Learning sequence of views of three-dimensional objects: The effect of temporal coherence on object memory. *Perception, 36*, 1320–1333.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 200*, 269–294.

Michler, F., Eckhorn, R., & Wachtler, T. (2009). Using spatiotemporal correlation to learn topographic maps for invariant object recognition. *Journal of Neurophysiology, 102*, 953–964.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature, 335*, 817–820.

Miyashita, Y., Date, A., & Okuno, H. (1993). Configural encoding of complex visual forms by single neurons of monkey temporal cortex. *Neuropsychologia, 31*, 1119–1131.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Orban, G. A., Sunaert, S., Todd, J. T., Van Hecke, P., & Marchal, G. (1999). Human cortical regions involved in extracting depth from motion. *Neuron, 24*, 929–940.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences, 6*, 261–266.

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition, 4*, 409–437.

Pilz, K. S., Thornton, I. M., & Bülthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research, 171*, 436–447.

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.

Sarkheil, P., Vuong, Q. C., Bülthoff, H. H., & Noppeney, U. (2008). The integration of higher order form and motion by the human brain. *Neuroimage, 42*, 1529–1536.

Shapiro, K. L., Arnell, K. M., & Raymond, J. E. (1997). The attentional blink. *Trends in Cognitive Sciences, 1*, 291–296.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Siéroff, E. (2001). Feature processing and superiority of three-quarter views in face recognition. *Brain and Cognition, 46*, 272–276.

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

Singer, J. M., & Sheinberg, D. L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *Journal of Neuroscience, 30*, 3133–3145.

Stone, J. V. (1999). Object recognition: View-specificity and motion-specificity. *Vision Research, 39*, 4032–4044.

Thornton, I. M., & Hayes, A. E. (2004). Anticipating action in complex scenes. *Visual Cognition, 11*, 341–370.

Thornton, I. M., & Kourtzi, Z. (2002). A matching advantage for dynamic human faces. *Perception, 31*, 113–132.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*, 1761–1771.

Troje, N. F., & Bülthoff, H. H. (1998). How is bilateral symmetry of human faces used for recognition of novel views? *Vision Research, 38*, 79–89.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Vuong, Q. C., & Tarr, M. J. (2004). Rotation direction affects object recognition. *Vision Research, 44*, 1717–1730.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*, 205–217.

Wallis, G., Backus, B. T., Langer, M., Huebner, G., & Bülthoff, H. H. (2009). Learning illumination- and orientation-invariant representations of objects through temporal association. *Journal of Vision, 9*, 1–8.

Wallis, G., & Bülthoff, H. H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences, 3*, 22–31.

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

Yamamoto, M., Kowatari, Y., Ueno, S., Yamane, S., & Kitazawa, S. (2005). Accelerated recognition of left oblique views of faces. *Experimental Brain Research, 161*, 27–33.