The Facespan—the perceptual span for face recognition

Michael Papinutto

Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland

Junpeng Lao

Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland

Meike Ramon

Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland

Roberto Caldara

Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland

Sébastien Miellet

Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland
School of Psychology, University of Wollongong, Wollongong, New South Wales, Australia
Bournemouth University, Faculty of Science & Technology, Cognition and Cognitive Neuroscience Research Centre, Bournemouth, UK

In reading, the perceptual span is a well-established concept that refers to the amount of information that can be read in a single fixation. Surprisingly, despite extensive empirical interest in determining the perceptual strategies deployed to process faces and an ongoing debate regarding the factors or mechanism(s) underlying efficient face processing, the perceptual span for faces—the *Facespan*—remains undetermined. To address this issue, we applied the gaze-contingent *Spotlight* technique implemented in an old-new face recognition paradigm. This procedure allowed us to parametrically vary the amount of facial information available at a fixated location in order to determine the minimal aperture size at which face recognition performance plateaus. As expected, accuracy increased nonlinearly with spotlight size apertures. Analyses of Structural Similarity comparing the available information during spotlight and natural viewing conditions indicate that the Facespan—the minimum spatial extent of preserved facial information leading to comparable performance as in natural viewing—encompasses 7° of visual angle in our viewing conditions (size of the face stimulus: 15.6°; viewing distance: 70 cm), which represents 45% of the face. The present findings provide a benchmark for future investigations that will address if and how the Facespan is modulated by factors such as cultural, developmental, idiosyncratic, or task-related differences.

Introduction

Face processing is a socially and biologically crucial feat achieved with high proficiency by the human visual system. An abundance of empirical studies has addressed various aspects of face processing, including the categorization of gender, age, race, or facial expression, as well as face recognition and identification (e.g., Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Ebner, Riediger, & Lindenberger, 2010; Kelly et

al., 2011; McClure, 2000; Meissner & Brigham, 2001). Isolating the very nature of the information used to achieve such visual categorizations is a major challenge in this field.

Recordings of oculomotor behavior have provided a valuable source of information regarding the relationship between the facial information sampled and observers' behavior. For instance, the robust and systematic triangular pattern of fixations exhibited during face perception (Henderson, Williams, & Falk, 2005; Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012; Van Belle, De Graef, Verfaillie, Rossion, & Lefevre, 2010; Walker-Smith, Gale, & Findlay, 1977; Xivry, Ramon, Lefèvre, & Rossion, 2008; Yarbus, 1967) was long considered invariant and universal. More recent investigations, however, have revealed that individuals' fixation patterns are affected by various factors, such as *task demands* and *experience*.

Preferential sampling of the eye region has been reported during face identity processing, whereas a shift towards lower face regions has been observed when the categorization of facial expression is of interest (e.g., Henderson et al., 2005; Malcolm, Lanyon, Fugard, & Barton, 2008). Such findings indicate a direct relationship between task demands and information diagnosticity. This notion receives further support from a recent study in which fixation patterns could be used to determine the task performed by observers' (Kanan, Bseiso, Ray, Hsiao, & Cottrell, 2015), as well as findings obtained using the response classification technique "Bubbles" (Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002). In short, these studies demonstrate that sampled and used face information is consistently determined by the task performed. More fundamentally, this line of research indicates that, in cognitive terms, visual information should not be described purely in terms of low-level properties, but rather needs to be considered as diagnostic for a given task and observer.

Evidence suggesting experience-dependent modulation of oculomotor behavior stems from both cross-cultural, as well as learning studies. For example, we have previously reported differential oculomotor patterns during face recognition as a function of the culture of the observers: East-Asian observers exhibit a central fixation bias, whereas and Western-Caucasian observers exhibit for an eye-mouth bias (Blais et al., 2008; Kelly et al., 2011), both being consistently displayed despite stimulus inversion (Rodger, Kelly, Blais, & Caldara, 2010). However, in studies using gaze-contingent techniques, we have demonstrated that despite using diverse gaze scanpaths, observers from both cultures rely on the same diagnostic features to perform face recognition (i.e., the eyes and the mouth) with comparable levels of performance (Caldara, Zhou, & Miellet, 2010; Miellet, He, Zhou, Lao, & Caldara,

2012; Miellet, Vizioli, He, Zhou, & Caldara, 2013). Thus, in both cultures, observers preferentially use different face information sampling strategies: Westerners favor a local strategy by sampling foveal information, whereas Easterners exhibit a global strategy to use the same diagnostic information. Finally, observers sample facial information differently from personally familiar and unfamiliar faces, with fixations expressed during processing of familiar individuals being relatively more distributed across all facial features (Van Belle et al., 2010).

While collectively these studies have enhanced our understanding of facial information *sampling*, they do not directly provide information regarding *quantity* of information *use*. For instance, observers may flexibly use information located within either the fovea, or extra-foveally during face identification (Miellet, Caldara, & Schyns, 2011), depending on the landing position of their first fixation. Hence, the following important question remains unanswered: *What is the quantity of information processed at each fixation during face recognition?*

This question finds its analogue in the field of reading, where Woodworth (1938) asked, "*How much can be read in a single fixation?*" (p.721)[1] McConkie and Rayner (1975) addressed this question of the perceptual span in reading using an elegant gaze-contingent moving-window paradigm. The authors reported prolonged reading times when the information outside of the window/fixation location, i.e., parafoveal information, was altered. Moreover, by parametrically varying the size of the gaze-contingent window, they were able to determine the minimal aperture size affording for normal reading performance as observed under unconstrained conditions.[2] Recently, this approach has also been employed in the field of visual scene processing (Nuthmann, 2013).

A different but related question was addressed by He et al. (2015) and Näsänen and Ojanpää (2004) who investigated the number of faces that can be recognized in a single fixation during a search task. Their research question differs from ours in that He et al. (2015) and Näsänen and Ojanpää (2004) used the number of faces as metric of the visual span for face recognition, whereas we considered the information within a single face during recognition for the Facespan.

In the present study we sought to determine the currently unknown perceptual span for faces—the *Facespan*. Following the aforementioned reasoning, we aimed to specify the spatial extent across which facial information is accrued within a single fixation during face recognition. In other words we asked, *what is the minimum quantity (in terms of spatial extent) of information needed at a fixation to achieve normal face recognition performance?* To this end we employed a gaze-contingent moving-window technique, adapted

for face stimuli, presented in the context of an old-new face recognition task. The facial information available to observers was restricted by using the *Spotlight* technique (for details, see e.g., Caldara, Zhou, & Miellet, 2010; Miellet, Vizioli, He, Zhou, & Caldara, 2013) with Gaussian apertures dynamically centered on their fixations, the size of which varied parametrically. Our goal was to ascertain the spotlight size that would allow subjects to reach a performance similar to that observed during natural face viewing. We reasoned that benchmarking the Facespan in this manner would provide the basis for investigating how information intake is impacted by factors such as those related to the stimulus, observer, or task-dependency.

Method

Participants

Two hundred and twenty young adults (173 females, $M = 21.47$, $SD = 2.50$) from the University of Fribourg, Switzerland, participated in this study. All participants had normal or corrected vision, were Western-Caucasian, and received course credit for participation. All participants gave informed consent; the protocol was approved by the ethical committees of the Department of Psychology of Fribourg University, Switzerland.

Materials

Stimuli were obtained from the Karolinska Directed Emotional Faces (KDEF; Lundqvist, Flykt, & Ohman, 1998) and the Asian Face Image Database (AFID; Bang, Kim, & Choi, 2001) and consisted of 84 Eastern-Asians and 84 Western-Caucasians identities containing equal numbers of males and females. The images were 688 pixels in size vertically and 702 pixels horizontally, subtending 15.3° and 15.6° of visual angle, respectively. Face stimuli were aligned with respect to the eye and mouth positions, normalized for luminance, and no distinctive external features or facial hair. Images were viewed at a distance of 70 cm, which is representative of a natural distance during human interactions (Hall, 1966), and were presented on a 1920 × 1080 pixels gray background displayed on a View-Pixx/3D LCD monitor (120 Hz refresh rate).

Apparatus

Eye movements were recorded at a sampling rate of 1000 Hz with the SR Research Desktop-Mount Eye-Link 2K eye tracker (with a chin and forehead rest),

which has an average gaze position error of about 0.5° and a spatial resolution of 0.01°. The eye-tracker had a linear output over the range of the monitor used. Although viewing was binocular, only the dominant eye was tracked. The experiment was implemented in Matlab (R2009b, The MathWorks, Natick, MA), using the Psychophysics toolbox (PTB-3) (Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) and EyeLink Toolbox extensions (Cornelissen, Peters, & Palmer, 2002; Kleiner et al., 2007). Calibrations of eye fixations were conducted at the beginning of the experiment using a nine-point fixation procedure as implemented in the EyeLink API (see EyeLink Manual) and using Matlab software. Afterwards, calibrations were validated with the EyeLink software and repeated when necessary until reaching an optimal calibration criterion. At the beginning of each trial, participants were instructed to fixate a cross at the center of the screen to perform a drift correction. If this exceeded 1°, a new calibration was performed to ensure optimal recording quality.
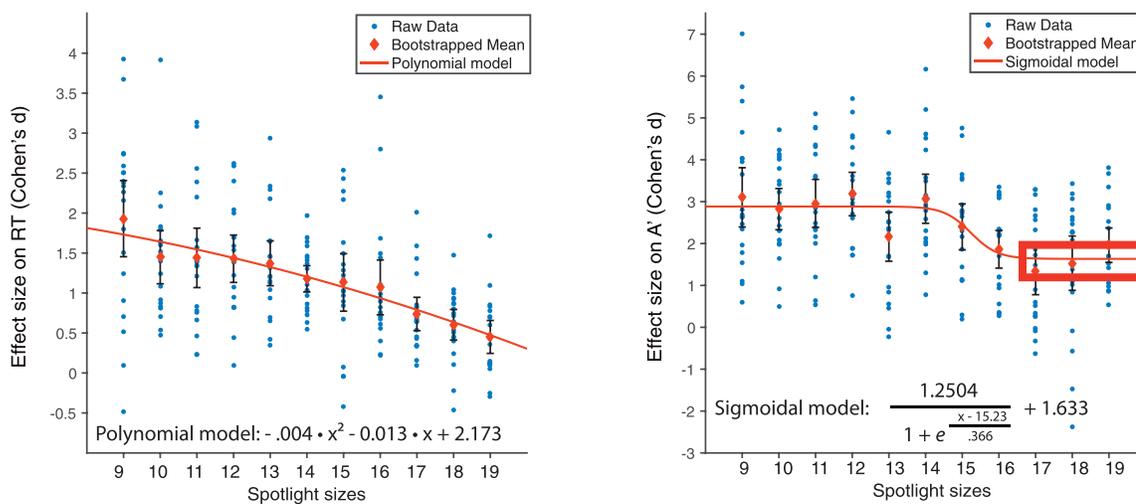
Spotlight

We used 11 spotlight aperture sizes from 9° to 19° of visual angle by step of 1° (see Figure 1A). These apertures were centered dynamically on the observers' fixations. The target face was presented at fixation location and was progressively masked by an average face template with retinal eccentricity. More specifically, at the center of the spotlight, the alpha channel had a value of zero, corresponding to complete transparency of the average template and full access to the target face. The alpha channel value increased with distance from the fixation location according to a Gaussian function. For each spotlight, the standard deviation of the corresponding Gaussian mask was 10.73% of the spotlight size. Values below 0.004 were set to 0, corresponding to complete opacity of the average template and no information from the target face. We used progressive masking to avoid both extrafoveally attracting participants' attention and abruptly dividing facial features by a hard aperture border. However, this progressive masking also involves challenges in terms of quantifying the information preserved within the Gaussian aperture. In the section "Data-driven reconstruction of the Facespan," we discuss how we overcame these challenges by introducing a measurement of the Facespan that goes beyond the mere diameter of the gaze-contingent window. This was done by taking into account acuity drop-off, fixation locations, and similarity between the target face and the average template. The information out with the spotlight was an average face composed from all stimuli from databases. This average face did not provide any useful information for recognition, but
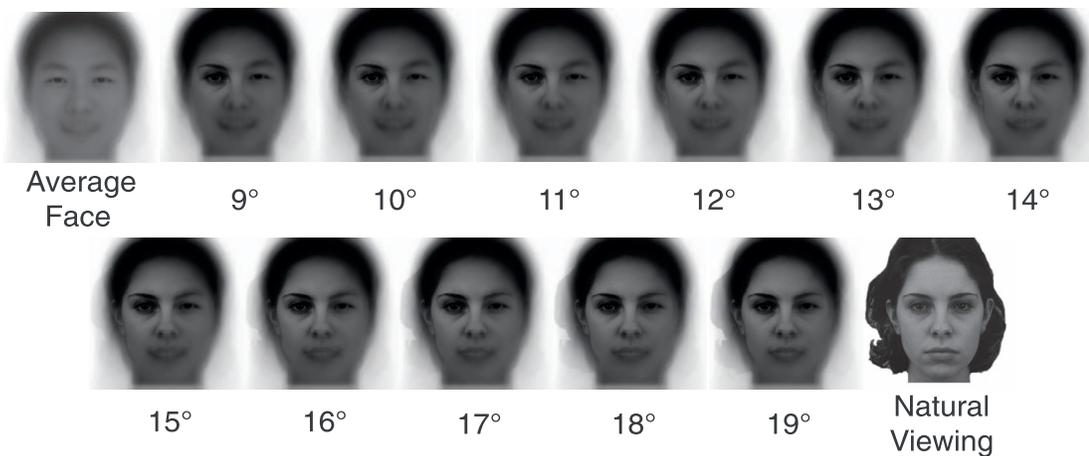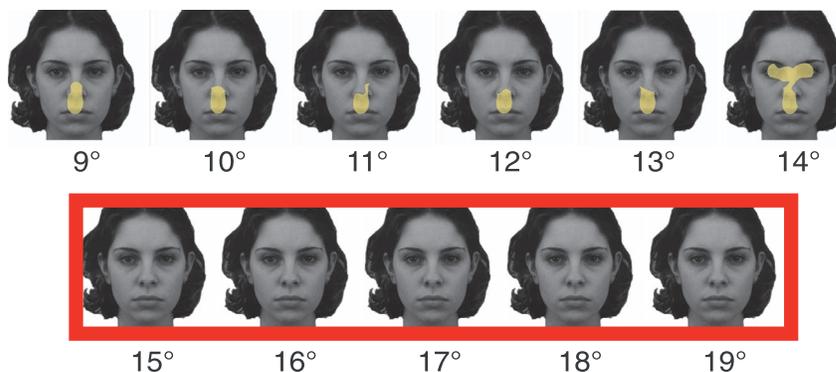
Figure 1. (A) The average face used as mask; an example stimulus with different spotlight sizes centered on the left eye and under natural viewing. (B) Impact of spotlights on RT (left) and A′ (right). The performance plateau in accuracy is highlighted by the red rectangle. (C) Effect of spotlights on fixation pattern. Highlighted in yellow are the areas fixated longer for the spotlight compared to natural viewing condition. The rectangle highlights the absence of a significant effect of spotlight (compared to natural viewing) on fixation patterns. Face images were retrieved from the Karolinska Directed Emotional Faces database (KDEF, http://www.emotionlab. se/resources/kdef).

allowed observers to program natural saccades. In terms of gaze-contingent display updating, it took 1 ms to receive a sample from the eye-tracker, less than 1 ms to calculate texture including the background and the Gaussian mask, and 2–57 ms to refresh the screen. Thus, the display was updated on average every 9 ms (between 3 and 58 ms), which eliminated any subjective impression of flickering.

Procedure

Each observer performed both natural viewing and spotlight conditions. For the spotlight condition, each observer was randomly assigned one of the eleven spotlight sizes. Participants started with the calibration procedure described in the apparatus section. They were then presented with a training session to familiarize them with the gaze-contingent moving-window display and informed about the experimental procedure. This consisted of two blocks (natural viewing, and spotlight condition), each involving presentation of a series of faces to be learned and subsequently recognized. Each of the blocks was divided into two subblocks (one per ethnicity—Western-Caucasian and East-Asian), resulting in a total of four subblocks. In each subblock participants had to learn 14 face identities (seven males, seven females), which randomly displayed neutral, happy, or disgusted facial expressions. After a 30 s break, they were presented with a series of 28 faces (14 faces from the learning phase and 14 new faces; 14 males and 14 females) and instructed to indicate as quickly and as accurately as possible whether each face had already been presented in the learning phase or not by pressing keys on the keyboard with the index of their left and right hand. Reaction times and accuracy were collected and analyzed for the purpose of this experiment; stimulus presentation duration was terminated by subjects' responses.

Facial expressions differed between learning and recognition phases in order to ensure identity learning, rather than trivial image matching. Each trial started with the presentation of a central fixation cross allowing the experimenter to check that the calibration was still accurate (toleration of an error of a maximum of 0.5° of visual angle). If the calibration was not sufficiently accurate, a new one was performed. Therefore, calibration was validated prior to each trial. This calibration checking was followed by a final central fixation cross used as the drift correction. Afterwards, a face was presented at a random location on the computer screen in order to avoid anticipatory oculomotor strategies.

Data analyses

Behavioral performance measures included reaction times (RT) and A-Prime (A′; Stanislaw & Todorov, 1999). Participants whose face recognition performance did not exceed chance level were discarded (14% of the sample tested; significance threshold determined at 61% using a bootstrapping approach). Moreover, trials for which participants exhibited RT larger than 2.5 standard deviations of their average RT were considered outliers and discarded from analysis. Individuals Cohen's d effect sizes (Cohen, 1988) were then calculated to assess the impact of each spotlight size on performance.

Saccades and fixations were determined based on angular velocity using a custom algorithm. Individual saccade velocity thresholds were set by expert experimenters and ranged between 30°/s and 80°/s ($M = 40.16$, $SD = 9.20$). When the velocity threshold exceeded 100°/s, participants were discarded (cf., Holmqvist, 2011) as having noisy eye-movement data (15% of the sample tested). Fixations that were spatially ($< 0.3°$) and temporally (< 20 ms) too close were merged. Previous studies have not revealed any effect of the experimental phase (learning vs. recognition), correctness of the answer (correct vs. incorrect trials), or race of face stimuli (Western-Caucasians vs. Eastern-Asians; Blais et al., 2008; Caldara et al., 2010; Miellet et al., 2012). In light of these findings and to ensure compatibility with the common procedure in the field, only the eye-movements of trials with correct responses were analyzed and data from Eastern-Asian and Western-Caucasian face stimuli were collapsed (see Figure A in the Supplemental data for demonstration of the lack of effect of stimulus race). Variables describing the general oculomotor behavior (number of fixations and fixation durations) were computed. Statistical fixation maps were computed with the *i*Map toolbox (version 4.01, Caldara & Miellet, 2011; Lao, Miellet, Pernet, Sokhn, & Caldara, 2016; Miellet, Lao, & Caldara, 2014). The *i*Map4 performed pixel-wise Linear Mixed Models (LMM) on the smoothed fixation maps with subject as a random effect, and condition (spotlight vs. natural viewing) as a fixed effect. A spatial cluster test based on bootstrapping was then used to assess statistical significance of the linear contrasts between the natural viewing and spotlight conditions. The trimmed mean of beta values, i.e., the model coefficients of the fixed effect (in this case they are the conditional mean difference of fixation duration in seconds) within the face region in the fixation maps, was then extracted for each spotlight size in order to assess the impact of size on the distribution of fixations.

Finally, stimuli from the candidate spotlight size (smallest spotlight size leading to performance similar to natural viewing) were compared to natural viewing
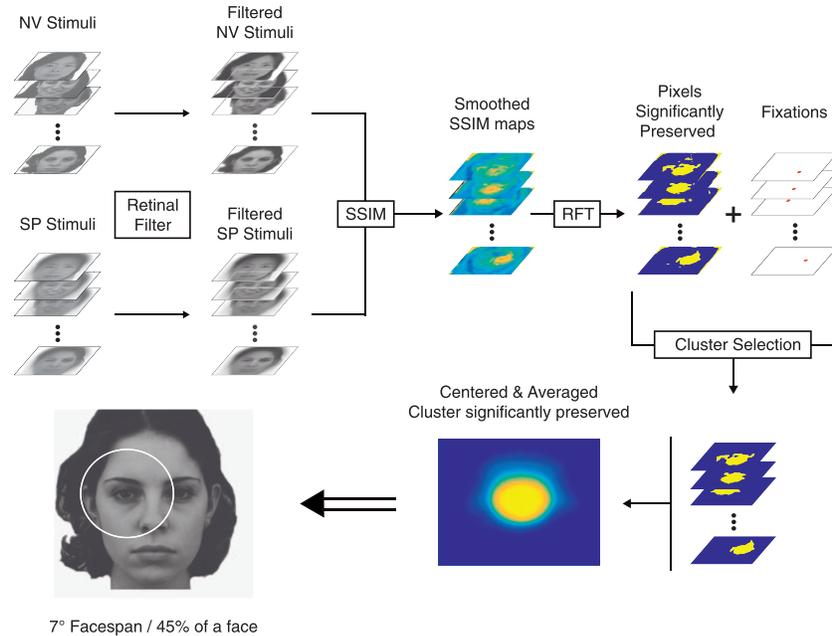
Figure 2. Pipeline of the Facespan reconstruction. On the top left, stimuli from recognition phases for both spotlight and natural viewing were passed through a retinal filter. To assess their similarity, SSIM was used and resulted in SSIM maps. Afterwards, the pixel-test (RFT) was used to assess the area significantly preserved from natural viewing in the 17° spotlight. Clusters of interest were then selected, centered and averaged. The averaged cluster significantly preserved was estimated to contain 7° of information, representing the Facespan. Face images were retrieved from the Karolinska Directed Emotional Faces database (KDEF, http://www.emotionlab.se/resources/kdef).

stimuli. The goal here was to assess how much information was preserved in the spotlight condition compared to natural viewing when considering observers' fixations. Indeed, it is not straightforward to infer the perceptual span for face recognition based on our gaze-contingent manipulation. We needed to quantify how much information (in terms of spatial extent) in the critical spotlight was identical to the natural viewing condition. In order to address this question, we needed to consider four challenges. First, the spotlight is a Gaussian aperture, so the target and average faces information blend progressively. Second, visual acuity drops off with retinal eccentricity. Thus, small extrafoveal variability in high spatial frequencies between the spotlight and natural viewing stimuli might not be captured by the visual system. Third, the similarity between spotlight and natural viewing stimuli might depend on how dissimilar a given target face (presented centrally) is from the average face (displayed extrafoveally). Finally, the similarity between spotlight and natural viewing stimuli might depend on the actual fixation location. To address these challenges, we used a procedure that aimed at mimicking early constraints of the visual system (e.g., acuity drop-off with eccentricity), while also considering targets' typicality (how similar to the average face a specific target face is), as well as fixation locations. First, for each participant and stimulus, we convolved a retinal filter

(Targino Da Costa & Do, 2014) on the spotlight of interest and natural viewing stimuli according to fixation locations in the critical spotlight condition (see Figure 2 for the Facespan reconstruction pipeline). The retinal filter parameters were distance to the screen (700 mm in our viewing conditions), stimulus size in pixels, and the lossy parameter ($\Delta = 25$, chosen to include visual perceptual lost according to human sight). We then used the Structural SIMilarity index (SSIM; Wang, Bovik, Sheikh, & Simoncelli, 2004) to quantify, independently for each target face, the similarity between spotlight and natural viewing stimuli after retinal filtering at each of the fixation locations in the critical spotlight condition. The SSIM uses luminance, contrast, and structure of two images in order to assess similarity of both images pixel by pixel. In the next step, we used the pixel-test (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005) to assess significance in the SSIM maps corresponding to each target face and fixation location. The pixel-test (Chauvin et al., 2005), which is based on the Random Field Theory (RFT), allowed highlighting the information that was significantly preserved between spotlight and natural viewing conditions (pixels that were significantly most similar). We used the following parameters for the pixel-test as recommended by Chauvin et al. (2005): sigma = 20, cluster test threshold = 2.7, $p = 0.05$. The Random Field Theory provides the probability of observing a cluster

of pixels exceeding a threshold in a smooth Gaussian random field while taking into account the spatial correlation inherent to the data set. More details about SSIM and the pixel-test can be found in Wang et al. (2004) and Chauvin et al. (2005), respectively. We finally centered and averaged, across stimuli, observers and fixations, the areas of information preserved by the critical spotlight manipulation (significant SSIM according to pixel-test after retinal filtering on spotlight and natural viewing stimuli). The rationale of this approach is that the amount of preserved information, corresponding to the minimal spotlight from which additional information does not improve performance, is the Facespan.

Results

Behavior

First, we investigated the individual effect sizes (Cohen's *d*) of spotlight compared to natural viewing on performance. Large values indicate that, for a given participant, performance advantage, in terms of A-prime (A′), for natural viewing as compared to the spotlight condition is larger than the variability between trials irrespective of condition. In contrast, small effect sizes indicate that the difference between viewing conditions for a given participant is not much larger (or smaller for effect sizes < 1) than the variability between trials across conditions. An effect size which does not significantly differ from 0 indicates that performance does not differ across viewing conditions.

We observed that effect sizes decreased parametrically with increasing spotlight size. However, spotlights always impacted on performance regardless of their size. A-posteriori analyses revealed that log-average luminance was higher for stimuli in the natural viewing, compared to spotlight conditions (Natural viewing = 0.27; Spotlight = 0.39). In contrast, spotlight conditions did not differ from each other in terms of luminance ($M = 0.39$, $SD = 0.002$). The difference in luminance between natural viewing and spotlight conditions might be due to our use of the alpha-blending function from the Psychotoolbox (Screen: "BlendFunction"). Thus, the fact that the effect of spotlight on performance was always significant might arise from differences in luminance between spotlight and natural viewing conditions. Importantly, the critical spotlight that we wanted to identify does not necessarily correspond to an identical performance between natural viewing and spotlight condition. It rather corresponds to the spotlight size from which additional information available in larger spotlights does not further improve

performance. Performance data was fitted with various models from the simplest linear regression to sigmoidal or polynomial models. A two degree polynomial model offered the best fit, $R^2 = 0.94$, $F(8, 11) = 60$, $p < 0.001$, with the effect size of RT depending on spotlight size (Figure 1B). The effect sizes of A′ were best described by a sigmoidal model, $R^2 = 0.79$, $F(7, 11) = 126$, $p < 0.001$.

The first plateau of the sigmoidal model indicates that the effect sizes for A′ remained constant for spotlight sized 9° to 14°, indicating that until 14° additional information does not noticeably increase performance. The second plateau of the sigmoidal model shows smaller effect sizes for performance (A′). Crucially, this second plateau also indicates that, from a 17° spotlight size onward, additional information available in larger spotlights did not further improve performance.

Eye-movements

Contrasts between the spotlight and the natural viewing conditions performed individually for each spotlight size on the number of fixations and fixation durations (paired *t* tests, Bonferroni corrected) did not reveal any consistent differences. In contrast, differential fixation maps generated by *i*Map4 (Lao et al., 2016) revealed differences in the spatial distribution of fixation durations between the spotlight and natural viewing conditions for spotlight sizes ranging from 9° to 14°. This range of spotlights exhibits areas with significantly longer fixation durations in the spotlight than in the natural viewing condition. Crucially *i*Map4 does not show any significant differences across the whole stimulus space between spotlight and natural viewing conditions for spotlights of, or larger than, 15° (Figure 1C).

Data-driven reconstruction of the Facespan

As noted above, eye movements exhibited during natural viewing and spotlight conditions did not differ for spotlights equal to or larger than 15°. Moreover, A′ indicated that facial information contained by spotlights larger than a 17° Gaussian aperture did not lead improved performance. In short, from eye movement patterns and performance, it appears that, on average, face information outside of a 17° Gaussian aperture around the fixation location is not used for face recognition as tested in the present experiment. As described in the data analyses section, we quantified how much information (in terms of spatial extent) in the spotlight was identical to the natural viewing condition, for a 17° spotlight. The average area of

preserved information was contained in a circle of 7° of visual angle.[3]

In summary, performance and fixation patterns revealed that face information outside a 17° Gaussian aperture is not used for face recognition. SSIM after retinal filtering and RFT indicate that face information projecting in the central 7° of the 17° Gaussian aperture was preserved compared to natural viewing. Thus, we conclude that the average perceptual span for face recognition (as tested in the current experiment), the Facespan, is 7° of visual angle given the viewing conditions used in our paradigm.

Discussion

In this study we aimed to determine the perceptual span for faces—the *Facespan*. To this end we employed a gaze-contingent paradigm in which the quantity of available information on the target face—the spotlight or aperture size—varied parametrically. Importantly, we reconstructed the available information inside these apertures in a data-driven fashion. This allowed us to quantify the extent of information preserved in the spotlight compared to natural viewing conditions, while taking into account the spotlight size, acuity drop-off, fixation locations, and dissimilarity between the target faces and the average face template.

With regards to observers' performance and oculomotor behavior during the old/new recognition task, our findings suggest that the *Facespan*—the information extent required at each fixation for normal face recognition—is 7° of visual angle given our experimental condition. For a fixation located on the nose, this perceptual span covers almost the entire area of internal features; for a fixation positioned on a different feature, it encompasses that feature along with its surrounding information (for visual representation, see Figure 2). Another way of expressing the Facespan is in terms of percentage of the face size, which in this study represents 45% of the face. Interestingly this value is very similar to that reported by Kwon, Liu, and Chien (2016). In a study manipulating spatial-frequency content and gaze-contingent window size, these authors found that at least 50% of facial information was required for observers to exhibit efficient face recognition performance. The similarity between the percentage of necessary information reported by Kwon et al. (2016) and the Facespan we measured here is remarkable, especially considering the numerous differences between both studies: face sizes (4° and 2° vs. 15.5° respectively), tasks (identification vs. recognition), familiarity (famous vs. unfamiliar faces), and presence of extrafoveal information (homogeneous mask vs. average face template mask). The most critical differ-

ence between both studies is that Kwon et al. (2016) measured the revealed information across several fixations (cf., Miellet et al., 2013), whereas the Facespan represents the information necessary at a single fixation.

Concerning observers' oculomotor behavior, fixation patterns varied as a function of aperture size for spotlights of 9°–14° as demonstrated in Figure 1C; beyond this aperture size, no further changes were observed. This is consistent with the idea that access to face information from a wider visually central region progressively allows for a more typical oculomotor behavior (see also Miellet et al., 2012). Interestingly, subjects' performance increased once fixation patterns were stable, supporting the idea that typical eye-movements are crucial to achieve optimal face recognition (Henderson et al., 2005). Finally, performance plateaued once at least an individual feature was visible at each fixation.

With a face displayed at a distance used during natural human interaction (about 80 cm), a Facespan of 7° allows the covering of the entire face within three fixations that would be directed towards these facial features. The perceptual span for face recognition is in line with the perceptual span for reading (McConkie & Rayner, 1975) and for visual scene search (Nuthmann, 2013). Indeed, the Facespan falls between these two perceptual spans (the perceptual span for reading: ~4.5°; the perceptual span for face recognition: ~7°; and the perceptual span for scene search: ~8°). This observation could be explained in terms of the density of to-be-processed information under each of those tasks, which is highest for text and lowest for scenes. This idea, that the perceptual span depends on information density, is supported by findings in reading. For instance, in Chinese where each character carries more linguistic information than in English, the perceptual span is much smaller (one character to the left of the fixated character and three characters to its right, according to Inhoff & Liu, 1997). Using the gaze-contingent parafoveal magnification technique, Miellet, O'Donnell, and Sereno (2009) demonstrated that the perceptual span is more accurately defined in terms of number of characters (quantity of linguistic information), rather than visual angle. This hypothesis is also consistent with evidence showing that the perceptual span for reading is modulated by text difficulty (Henderson, Ferreira, Rayner, & Pollatsek, 1990), as well as reader characteristics, such as expertise (Rayner, 1986; Rayner, Slattery, & Belanger, 2010), age (Rayner, 1986), and disorder (Rayner, 1983). From this perspective, information is not considered in absolute terms but rather follows the concept of diagnostic information and thus varies depending on viewing conditions, strategies, and task at hand (see, for instance, Oliva & Schyns, 1997).

Following this idea, we emphasize that the Facespan should not be considered as an absolute quantity. Inasmuch as the perceptual span for reading is not absolute, but instead flexible, the Facespan reported here should be considered as an average benchmark obtained under the aforementioned specific viewing conditions and task. Mirroring previous observations for the perceptual span in reading, we suggest that the Facespan may vary as a function of factors including— but not limited to—e.g., stimulus or observer characteristics, and task demands. Indeed, previous research in face processing has demonstrated cultural (Blais et al., 2008; Caldara et al., 2010; Miellet et al., 2012, Miellet et al., 2013), idiosyncratic (Kanan et al., 2015; Mehoudar, Arizpe, Baker, & Yovel, 2014; Peterson & Eckstein, 2013; Ramon et al., 2016) and familiarity (Van Belle et al., 2010) biases in information sampling. Moreover, initial fixation location on the face also impacts on sampling strategy (Miellet et al., 2011). These findings suggest that the Facespan might also be flexibly modulated by such factors. However, further studies are necessary to understand precisely when and how the Facespan is modulated.

We suggest that characterizing the Facespan at the individual level, as well as across cultures may broaden our understanding of these well-established, systematic differences in perceptual biases and processing. Specifically, we propose a direct relationship between individual preferences in information sampling (global vs. local) and an individual's *Facespan*, which in turn determines the facial representation generated based on the visual input. In other words, we propose that observers exhibiting global sampling strategies (i.e., favoring fixations on the center of the face) will exhibit a broader Facespan, which allows sampling diagnostic feature (eyes and mouth) from central fixation. In contrast, observers exhibiting local sampling strategies (i.e., favoring fixations on eyes and mouth) will show a spatially more constrained Facespan. As demonstrated in Miellet et al. (2013), the sampling strategies determine the information available to the brain.

Conclusions

In the present study, we determined the perceptual span for face recognition—the *Facespan*. Our data show that given our viewing conditions, observers require information encompassing 7° of visual angle to exhibit face recognition performance comparable to that observed during natural viewing. Importantly, similarly to the perceptual span in reading, the Facespan should not be considered as fixed, but rather flexible. Our findings provide a benchmark for information sampling during face processing, which will

leverage further investigations into intra- and interindividual differences and how the perceptual span for face processing is modulated by constraints such as culture, development, or task.

Keywords: facespan, face perception, eye movements, gaze-contingent, spotlight, perceptual span

Footnotes

[1] See also McConkie and Rayner (1975), who improved upon the operationalization of this question by asking "*How far into the periphery are specific aspects of the visual stimulus typically acquired and used during fixations in reading?*"(p. 578).

[2] For English the perceptual span extends from three characters to the left of fixation (approximately the beginning of the fixated word) to 14 characters to the right of fixation (McConkie & Rayner, 1975). Importantly in active reading, involving eye-movements, the span is mainly constrained by higher level processing limitations rather than low-level visual constraints (Miellet, O'Donnell, & Sereno, 2009). Note that a number of studies (Kwon & Legge, 2012; Legge, Mansfield, & Chung, 2001; Pelli et al., 2007; Yu, Cheung, Legge, & Chung, 2007) support the notion of a visual span largely limited by low-level processing. However, crucially, these studies used rapid presentations and identification of meaningless trigrams in absence of eye-movements (Rapid Serial Visual Presentation). The processes involved might have been different from those present in active reading of meaningful sentences or texts, for instance visual remapping and contextual influence. Thus, these

studies are likely to underestimate the higher level influences in play in natural reading while overestimating the low-level influences. The concept of visual span presented in those studies might differ in nature from the perceptual span measured in active reading of sentences (see Legge, 2007).

[3] Note that the Facespan estimate is robust to systematic manipulations of the pixel-test parameters, spanning the values found in the literature. The parameters manipulation had a minimal influence on the estimated Facespan whose values ranged from 6.48° to 7.29° of visual angle (42% to 47% of the face).

References

Bang, S., Kim, D., & Choi, S. (2001). *Asian face image database*. Korea: Postech.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE*, *3*(8), doi.org/10.1371/journal.pone.0003022.

Caldara, R., & Miellet, S. (2011). iMap: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, *43*(3), 864–878, doi.org/10.3758/s13428-011-0092-x.

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "Spotlight" reveals universal information use for face recognition. *PLoS ONE*, *5*(3), e9708, doi.org/10.1371/journal.pone.0009708.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9):1, 659–667, doi:10.1167/5.9.1. [PubMed] [Article]

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New Jersey: Erlbaum.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617, doi.org/10.3758/BF03195489.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362, doi.org/10.3758/BRM.42.1.351.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271, doi.org/10.1016/S0042-6989(01)00097-9.

Hall, E. T. (1966). *The hidden dimension*. New York: Doubleday, doi.org/10.1017/CBO9781107415324.004.

He, Y., Scholz, J. M., Gage, R., Kallie, C. S., Liu, T., Legge, G. E., . . . J.-Y., Z. (2015). Comparing the visual spans for faces and letters. *Journal of Vision*, *15*(8):7, 1–13, doi:10.1167/15.8.7. [PubMed] [Article]

Henderson, J. M., Ferreira, F., Rayner, K., & Pollatsek, A. (1990). Effects of foveal processing Difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 41–429, doi.org/10.1037/0278-7393.16.3.417.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106, doi.org/10.3758/BF03195300.

Holmqvist, K. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

Hsiao, J., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, *19*(10), 998–1006, doi.org/10.1111/j.1467-9280.2008.02191.x.

Inhoff, A. W., & Liu, W. (1997). The perceptual span during the reading of Chinese text. In H. C. Chen (Ed.), *Cognitive processing of Chinese and related Asian languages* (pp. 243–266). Hong Kong, China: The Chinese University of Hong Kong Press.

Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, *108*, 67–76, doi.org/10.1016/j.visres.2015.01.013.

Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Developmental Science*, *14*(5), 1176–1184, doi.org/10.1111/j.1467-7687.2011.01067.x.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception 36 ECVP Abstract Supplement*, *36* (ECVP Abstract Supplement), doi.org/10.1068/v070821.

Kwon, M., & Legge, G. E. (2012). Spatial-frequency requirements for reading revisited. *Vision Research*, *62*, 139–147, doi.org/10.1016/j.visres.2012.03.025.

Kwon, M., Liu, R., & Chien, L. (2016). Compensation for blur requires increase in field of view and viewing time. *PLOS ONE*, *11*(9), 1–20, doi.org/10.1371/journal.pone.0162711.

Lao, J., Miellet, S., Pernet, C., Sokhn, N., & Caldara, R. (2016). iMap4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling. *Behavior Research Methods*, 1–17, doi.org/10.3758/s13428-016-0737-x.

Legge, G. E. (2007). *Psychophysics of reading in normal and low vision*. Portions of this research (MNREAD acuity charts) were presented at the OSA Noninvasive Assessment of the Visual System, 1993, Monterey, CA. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. Retrieved from http://psycnet.apa.org/psycinfo/2006-20917-000

Legge, G. E., Mansfield, J. S., & Chung, S. T. (2001). Psychophysics of reading. XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, *41*(6), 725–743. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11248262

Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska directed emotional faces (KDEF). [CD ROM]. Solna, Sweden: Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, doi.org/10.1017/S0048577299971664.

Malcolm, G. L., Lanyon, L. J., Fugard, A. J. B., & Barton, J. J. S. (2008). Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects. *Journal of Vision*, *8*(8):2, 1–9, doi:10.1167/8.8.2. [PubMed] [Article]

McClure, E. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424–453, doi.org/10.1037/0033-2909.126.3.424.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586, doi.org/10.3758/BF03203972.

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of Vision*, *14*(7):6, 1–11, doi:10.1167/14.7.6. [PubMed] [Article]

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35, doi.org/10.1037/1076-8971.7.1.3.

Miellet, S., Caldara, R., & Schyns, P. G. (2011). Local Jekyll and global Hyde: The dual identity of face identification. *Psychological Science*, *22*(12), 1518–1526, doi.org/10.1177/0956797611424290.

Miellet, S., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When East meets West: Gaze-contingent Blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research*, *5*(2), 1–12, doi.org/10.16910/JEMR.5.2.5.

Miellet, S., Lao, J., & Caldara, R. (2014). An appropriate use of iMap produces correct statistical results: A reply to McManus (2013) "iMAP and iMAP2 produce erroneous statistical maps of eye-movement differences." *Perception*, *43*(5), 451–457, doi.org/10.1068/p7682.

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science*, *20*(6), 721–728, doi.org/10.1111/j.1467-9280.2009.02364.x.

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in Psychology*, *4*(February), 34, doi.org/10.3389/fpsyg.2013.00034.

Näsänen, R., & Ojanpää, H. (2004). How many faces can be processed during a single eye fixation? *Perception*, *33*(1), 67–77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15035329

Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition*, *21*(7), 803–837, doi.org/10.1080/13506285.2013.832449.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*(1), 72–107, doi.org/10.1006/cogp.1997.0667.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442, doi.org/10.1163/156856897X00366.

Pelli, D., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*(2):20, 1–36, doi:10.1167/7.2.20. [PubMed] [Article]

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences, USA*, *109*(48), E3314–E3323, doi.org/10.1073/pnas.1214269109.

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science*, *24*(7), 1216–1225, doi.org/10.1177/0956797612471684.

Ramon, M., Miellet, S., Dzieciol, A. M., Konrad, B. N., Dresler, M., & Caldara, R. (2016). Super-memorizers are not super-recognizers. *PloS One*,

11(3), e0150972, doi.org/10.1371/journal.pone.0150972.

Rayner, K. (1983). Eye movements, perceptual span, and reading disability. *Annals of Dyslexia*, *33*(1), 163–173, doi.org/10.1007/BF02648003.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, *41*(2), 211–236, doi.org/10.1016/0022-0965(86)90037-8.

Rayner, K., Slattery, T. J., & Belanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin and Review*, *17*(6), 834–839, doi.org/10.3758/PBR.17.6.834.

Rodger, H., Kelly, D. J., Blais, C., & Caldara, R. (2010). Inverting faces does not abolish cultural diversity in eye movements. *Perception*, *39*, 1491–1503, doi.org/10.1068/p6750.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409, doi.org/10.1111/1467-9280.00472.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149, doi.org/10.3758/BF03207704.

Targino Da Costa, A. L. N., & Do, M. N. (2014). A retina-based perceptually lossless limit and a gaussian foveation scheme with loss control. *IEEE Journal on Selected Topics in Signal Processing*,

8(3), 438–453, doi.org/10.1109/JSTSP.2014.2315716.

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefevre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision*, *10*(5):10, 1–13, doi:10.1167/10.5.10. [PubMed] [Article]

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception*, *6*(3), 313–326, doi.org/10.1068/p060313.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612, doi.org/10.1109/TIP.2003.819861.

Woodworth, R. S. (1938). *Experimental psychology*. New York, NY: Holt.

Xivry, J.-J. O., Ramon, M., Lefèvre, P., & Rossion, B. (2008). Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. *Journal of Neuropsychology*, *2*(1), 245–268, doi.org/10.1348/174866407X260199.

Yarbus, A. L. (1967). *Eye movements and vision*. Boston, MA: Springer, doi.org/10.1007/978-1-4899-5379-7.

Yu, D., Cheung, S.-H., Legge, G. E., & Chung, S. T. L. (2007). Effect of letter spacing on visual span and reading speed. *Journal of Vision*, *7*(2):2, 1–10, doi:10.1167/7.2.2. [PubMed] [Article]