Neuropsychologia 50 (2012) 2830–2838

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

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

The eyes are not the window to basic emotions

Caroline Blais [a,b], Cynthia Roy [a], Daniel Fiset [b], Martin Arguin [a], Frédéric Gosselin [a,*]

[a] Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, QC, Canada H3C 3J7
[b] Département de psychoéducation et de psychologie Université du Québec en Outaouais, C.P. 1250, Succ. Hull, Gatineau, QC, Canada J8X 3X7

ARTICLE INFO

Article history:
Received 5 March 2012
Received in revised form
23 June 2012
Accepted 13 August 2012
Available online 20 August 2012

Keywords:
Facial expression recognition
Facial features
Visual mechanisms
Classification image technique

ABSTRACT

Facial expressions are one of the most important ways to communicate our emotional state. In popular culture and in the scientific literature on face processing, the eye area is often conceived as a very important – if not the most important – cue for the recognition of facial expressions. In support of this, an underutilization of the eye area is often observed in clinical populations with a deficit in the recognition of facial expressions of emotions. Here, we used the Bubbles technique to verify which facial cue is the most important when it comes to discriminating between eight static and dynamic facial expressions (i.e., six basic emotions, pain and a neutral expression). We found that the mouth area is the most important cue for both static and dynamic facial expressions. We conducted an ideal observer analysis on the static expressions and determined that the mouth area is the most informative. However, we found an underutilization of the eye area by human participants in comparison to the ideal observer. We then demonstrated that the mouth area contains the most discriminative motions across expressions. We propose that the greater utilization of the mouth area by the human participants might come from remnants of the strategy the brain has developed with dynamic stimuli, and/or from a strategy whereby the most informative area is prioritized due to the limited capacity of the visuo-cognitive system.

1. Introduction

Interactions with other individuals are an important part of our everyday life, and having the skills to adequately transmit our state of mind and decode that of others is crucial for the success of social communication. The facial expression of emotions is one of the most important ways of communicating those states (Mehrabian, 1968), and the study of the perceptual information available to decode facial expressions, as well as the actual information used by human observers to achieve this task, has preoccupied researchers at least since the publication of Darwin's seminal book *The Expression of Emotions in Man and Animals* (1872). It is now well accepted that the various facial expressions differ from one another in terms of where the information is available across the different facial areas (e.g., Bassili, 1979; Cunningham, Kleiner, Bülthoff, & Wallraven, 2004; Ekman, 1982; Hanawalt, 1944; Nummenmaa, 1964; Nusseck, Cunningham, Wallraven, & Bülthoff, 2008; Plutchik, 1962; Smith, Cottrell, Gosselin, & Schyns, 2005). However, some areas of the face may convey more information than others when it comes to discriminating *all* emotions or a significant subset of them.

In popular culture, the eyes are often portrayed as the most important emotional cue. This is, for instance, reflected in our metaphorical language: "The eyes are the window to the soul", "I could see the fear in his eyes", "His eyes were filled with anger", "There was passion in her eyes", "Love showed in his eyes", "Her eyes welled with tears", and so on. In fact, a considerable part of the literature that emerged from cognitive science research suggests that the eyes are particularly important for face recognition. Research on infant development also suggests that the eye region is "special" from very early on in life. For example, infants show a preference for looking at the eye region in comparison to other facial areas (Hainline, 1978; Haith, Bergman, & Moore, 1977; Maurer, 1985) and prefer to look at faces with direct eye contact (Farroni, Csibra, Simion, & Johnson, 2002). The existence of an innate gaze module, dedicated to the task of detecting the presence of eyes, has even been proposed (Batki, Baron-Cohen, Wheelwright, Connellan, & Ahluwalia, 2000).

Moreover, research on clinical populations has shown that the eye region is processed less efficiently or is processed in an abnormal way in many neurological pathologies leading to social impairments (e.g., Adolphs et al., 2005; Lee, Gosselin, Wynn, & Green, 2011; Spezio et al., 2007a; 2007b). One example of this is the finding that SM, a patient with a bilateral amygdala lesion suffering from a major deficit at categorizing the expression of fear, processes eye information less effectively than do control

* Corresponding author. Tel.: +1 514 343 7550; fax: +1 514 343 5787.
E-mail address: frederic.gosselin@umontreal.ca (F. Gosselin).

0028-3932/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neuropsychologia.2012.08.010

subjects (Adolphs et al., 2005). Furthermore, this patient's performance with the expression of fear returns to normal when she is instructed to look at the eyes. Similarly, schizophrenia patients rely less on the high spatial frequencies in the eye region than control participants when categorizing fear (Lee, Gosselin, Wynn, & Green, 2011). Adults with autism, who show a deficit in the categorization of facial expressions (Humphreys, Minshew, Leonard, & Behrmann, 2007; Harms, Martin, & Wallace, 2010), have also been shown to process the eye region less efficiently than normal participants (Baron-Cohen, Wheelwright, & Jolliffe, 1997; Spezio et al., 2007a; 2007b). Acquired prosopagnosic patients, who have been shown to process the eye region less efficiently than controls (Bukach, Bub, Gauthier, & Tarr, 2006; Bukach, LeGrand, Kaiser, Bub, & Tanaka, 2008; Caldara et al., 2005; Rossion, Kaiser, Bub, & Tanaka, 2009), also suffer from a deficit in discriminating facial expressions of emotions (Humphreys, Avidan, & Behrmann, 2007). Thus, many neuropsychological phenomena associated with a failure in facial expression categorization involve an underutilization of the eyes.

From the observations listed above, it is tempting to conclude that the eyes are more important than any other facial area for the discrimination of facial emotions. However, the studies that have directly addressed the question of which facial features are useful for the discrimination of basic facial expressions of emotions have led to contradictory results. Some have found that the lower part of the face was more important than the upper part of the face (Dunlap, 1927; Ruckmick, 1921) and some have found no greater importance of one part of the face over another (Baron-Cohen et al., 1997; Coleman, 1949; Frois-Wittman, 1930). In a related vein, eye-tracking studies that have examined how ocular fixations are distributed on faces during the recognition of facial expressions have found a roughly equal sampling of the mouth and eye areas (Eisenbarth & Alpers, 2011; Jack et al., 2009). However, eye fixation patterns are partly dissociable from information use (Arizpe, Kravitz, Yovel, & Baker, 2012; Jonides, 1981; Posner, 1980; see however, Rayner, 1998; Deubel & Schneider, 2003; Godjin & Theeuwes, 2003), and the question of interest here is what visual information is actually used to discriminate facial expressions from one another.

Our primary aim here is to discover which facial information is the most important when it comes to discriminating a significant subset of facial emotions. The data that will be analyzed in this paper is part of a larger project in which we examined many dimensions of the visual information extraction strategies employed for the discrimination of static and dynamic facial expression of emotions using the Bubbles technique (Gosselin & Schyns, 2001). Here, we will focus on the spatial and the temporal dimensions of the data. Our secondary aim is to examine if the use of information varies between static and dynamic stimuli. We will also present a novel analysis of the data from Smith et al. (2005), who have used static Bubbles to verify which facial information is the most important to discriminate *each* expression of basic emotion from one another.

2. Methods

2.1. Participants

Forty-one Caucasian participants (14 males; 24.2 years old on average) with normal or corrected-to-normal visual acuity took part in the experiment with static stimuli, and 59 different participants (30 males; 23.9 years old on average) took part in the experiment with dynamic stimuli. All procedures were carried out with the ethics approval of the Université de Montréal.

2.2. Materials and stimuli

Stimuli were displayed on a calibrated high-resolution CRT monitor with a refresh rate of 60 Hz. The experimental program was written in Matlab, using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The stimulus width subtended 5.72 degrees of visual angle.

The stimuli were created using a validated database composed of highly recognizable pictures and videos of 10 Caucasian actors' faces spontaneously expressing one of eight states (i.e., anger, disgust, fear, happy, pain, sad, surprise, and a neutral expression; Roy et al., 2007). The database as well as normative data are available at the following address: http://mapageweb.umontreal.ca/gosselif/STOIC.rar. The videos consisted of 15 frames, displayed at a rate of 30 Hz, starting with a neutral expression that naturally deployed into one of the eight expressions mentioned above and ending with the apex of the expression. They lasted a total duration of 500 ms. The static stimuli consisted of the apex of the facial expressions mentioned above, displayed for a duration of 500 ms. All the stimuli were gray-scaled and their luminance was normalized. To minimize the head movements that occurred while the actors made the facial expressions, the stimuli were also spatially aligned frame by frame to ensure that the eyes and nose were located at about the same spatial coordinates across frames and stimuli.

To reveal the visual information useful for the discrimination of facial expressions, we used the Bubbles technique. The Bubbles technique consists of randomly sampling the visual information contained in a stimulus, such that, on each trial, a different subset of this information is rendered available to the participant. The performance of the participant with these subsets of information indicates which parts of the stimulus are most useful in performing the task. Here, we sampled the static expressions on the space (i.e., x, y coordinates of the face) and on the spatial frequency dimensions; and we sampled the dynamic expressions on the space, spatial frequency and time dimensions.

For each trial, the creation of a bubblized stimulus went as follows: first, the image of a facial expression was decomposed into five spatial frequency bands (128–64, 64–32, 32–16, 16–8, 8–4 cycles/image, or 86–43, 43–21.5, 21.5–10.8, 10.8–5.4, 5.4–2.7 cycles/face; the remaining low frequency bandwidth served as a constant background; see Fig. 1a, top row) using the Laplacian pyramid (Burt & Adelson, 1983). With dynamic stimuli, the spatial frequency decomposition was performed on each frame of the videos (see Fig. 1b, top row, for an example with the third spatial frequency band). Then, independently for each spatial frequency band, the bubbles' locations (i.e., a bubble is a Gaussian aperture through which the information is visible) were randomly selected (see Fig. 1a and b, middle row). On the space dimension, the size of the bubbles (FWHM: 14.1, 28.3, 56.5, 113.0, and 226.1 pixels) was adjusted as a function of the frequency band so that each bubble revealed 1.5 cycles of spatial information. Because the size of the bubbles increased as the spatial scale became coarser, the number of bubbles differed across scales to keep the size of the sampled area constant across frequency bands. The size of the bubbles also varied as a function of spatial frequency on the time dimension (i.e., with dynamic bubbles), such that their duration increased as the spatial frequency band increased (FWHM: 14.1, 28.3, 56.5, 113.0, and 226.1 pixels on the space dimension and 7.3, 6.1, 5.1, 4.2 and 3.5 frames on the time dimension). This was done to take into account the faster processing of lower spatial frequencies (Hughes, Fendrich, & Reuter-Lorenz, 1990; Parker, Lishman, & Hughues, 1992). A pointwise multiplication was then performed between the bubbles' masks and the filtered images (see Fig. 1a and b, bottom row). Finally, the information revealed by the bubbles was fused across the five frequency bands to produce an experimental stimulus (Fig. 1a, bottom row, rightward picture).

2.3. Procedure

Each participant completed 4000 trials divided into experimental sessions comprising 160 trials each. On each trial, the sequence of events went as follows: a fixation point was first displayed in the center of the screen for 200 ms and was immediately replaced by the stimulus (i.e., a bubblized image or video of a facial expression). The stimulus was displayed for 500 ms and was then replaced by a homogenous grey screen that remained visible until the participant responded. The participant was instructed to press on the keyboard key that corresponded to the facial expression he had perceived. Responses were not restricted by time pressure. No accuracy feedback was provided. The accuracy was maintained at 56% correct on average (i.e., halfway between chance and perfect performance) across all expressions by adjusting the total number of bubbles on the stimulus on a trial-by-trial basis using QUEST (Watson & Pelli, 1983). We used a constant number of bubbles across expressions because we did not want this parameter to become a cue for the recognition. A threshold of 56% correct (midway between chance and perfect performance) was chosen to make sure that the performance would reach neither ceiling for facial expressions that are easier to recognize, such as happy, nor floor for facial expressions that are more difficult to recognize, such as fear.

3. Results

3.1. Classification plane and classification volume

A mean of 144.3 bubbles (SD: 119.6) and of 241.3 bubbles (SD: 253.6) were necessary to maintain the average performance at

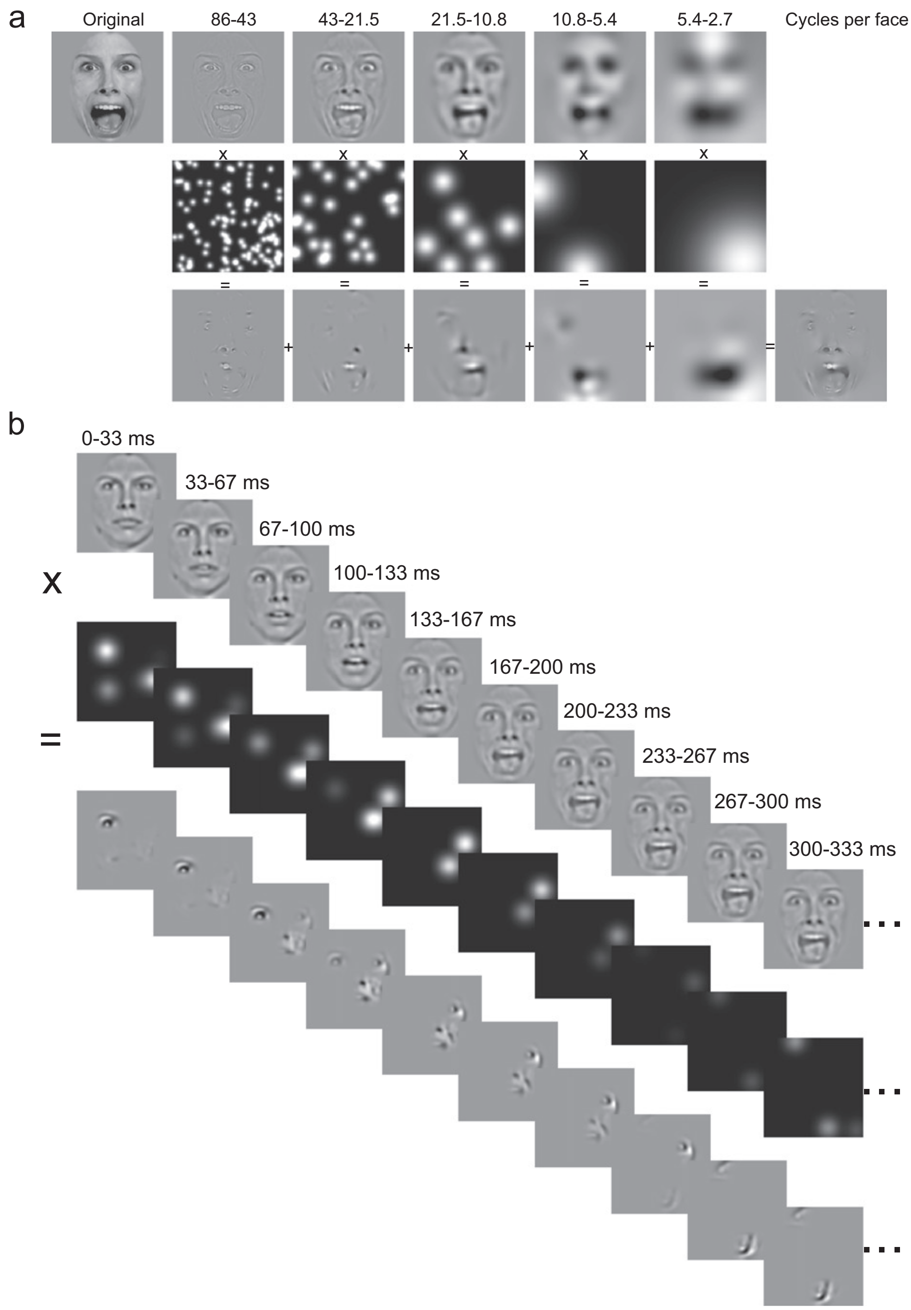

Fig. 1. Illustration of the creation of a bubblized stimulus. (a) Procedure with static stimuli. A facial expression image was decomposed into five spatial frequency bands (top row). Bubbles were then placed at random locations separately for each spatial frequency band (middle row). The information revealed by the bubbles (bottom row) was then fused across the five frequency bands to produce an experimental stimulus. (b) With dynamic stimuli, the procedure to create a bubblized stimulus was very similar to the procedure for static stimuli, except that the decomposition into five spatial frequency bands was performed on each frame. Here, we illustrate the procedure for the third frequency band and the first 10 frames of the video. The same procedure was repeated on each frequency band. Each frame of the video was decomposed into five frequency bands (top row represents the third frequency band). Bubbles were then placed randomly at different locations and frames, separately for each spatial frequency band (middle row). The bottom row illustrates the information revealed by the bubbles on each frame.

56% for the static and the dynamic version of the experiment, respectively.

The visual information useful in categorizing facial expressions of emotions was determined using an analysis procedure that amounts to a multiple linear regression on the bubbles masks (explanatory variables) and on the participant's response accuracy (predictor variable). In other words, for each participant, each facial expression, and each spatial frequency band, a weighted sum of all the bubbles centers was calculated, using the accuracies transformed into z-score values as weights. This resulted in 3D volumes (or 4D volumes in the dynamic version of the experiment) of regression coefficients that will be referred to as classification volumes. These classification volumes were then summed across participants, leading to one classification volume per expression and per spatial frequency. These classification volumes were transformed into z-scores. We used the voxels (i.e., volume elements) outside the face area to calculate the mean and standard deviation of the distribution of the null hypothesis. Since we were particularly interested in verifying the location of the most useful visual information across all facial expressions in a categorization task, we then summed the classification volumes across the eight expressions tested and across the five spatial frequency bands, and we normalized the resulting classification volumes by dividing them by the square root of 40 (i.e., 8 expressions $\times$ 5 spatial frequency bands). We then smoothed the 2D classification volume (or classification planes) using a Gaussian window with a FWHM of 28.3 pixels (equivalent to the spatial extent of the bubbles that revealed the information at second finest scale) and the 3D classification volume using a Gaussian window with a FWHM of 28.3 pixels on the spatial dimension and of 6.1 frames on the temporal dimension (equivalent to the temporal extent of the bubbles that revealed the information at second finest scale); and we transformed one last time the pixels (or voxels) of the classification plane (or volume) into z-scores, using the pixels (or voxels) outside the face area to calculate the mean and standard deviation of the distribution of the null hypothesis. In order to determine if the facial information significantly correlated with accuracy, we applied the *Pixel* test ($p < 0.05$, $Z_{crit} = 3.76$ and 4.31 for the classification plane and classification volume, respectively) to the classification images and movies. The statistical threshold provided by this test corrects for multiple comparisons while taking the spatial correlation inherent to structured images into account (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005). The classification plane and classification volume are displayed in Fig. 2a and b, respectively. The non-significant pixels are depicted in gray, and the different colors indicate the z-score values of the significant pixels (and voxels).

A quick visual inspection of the classification plane and volume reveals that the eye and mouth regions are the most important facial areas. To further characterize the information available in the classification plane and classification volume in terms of facial features, we conducted a region-of interest (i.e., ROI) analysis on six facial areas (i.e., the eyes, the eyebrows, the frown lines, the nose, the nasolabial folds, and the mouth; see inset in Fig. 3). In the first part of this analysis, we were interested in verifying the relative importance of the different face areas and in comparing the result obtained for dynamic and static stimuli. We therefore collapsed the classification volume on the temporal dimension.

3.2. *ROI analysis without the time dimension*

We kept only the portion of the static and dynamic "classification planes" that corresponded to the highest 5% regression coefficients. This ensured that the same number of pixels was considered for the ROI analysis on the data from the static and the dynamic version of the experiment. We then calculated, separately for the static and the dynamic classification planes, the proportion of the total number of these pixels that fall on each facial feature, and divided this proportion by the total number of pixels in that feature, thus normalizing the proportion for feature size (e.g., Gibson, Lazareva, Gosselin, Schyns, & Wasserman, 2007). The results of this analysis are summarized in Fig. 3. For both static and dynamic facial expressions of emotions, the mouth area is more important than the eye area (i.e., 8.56 and 15.51 times more important on average for the static and the dynamic stimuli, respectively). To make sure that the average classification planes reflected the strategy of most participants rather than only a few participants, we created 1000 classification planes using random subsamples of 20 participants (results are robust to changes in the size of this subsample of participants). We calculated the average of the z-score values for the mouth area and for the eye area, and we calculated the ratio of these two values (mouth/eyes). The ratios were higher than one (i.e., mouth > eyes) on 99.6% of the

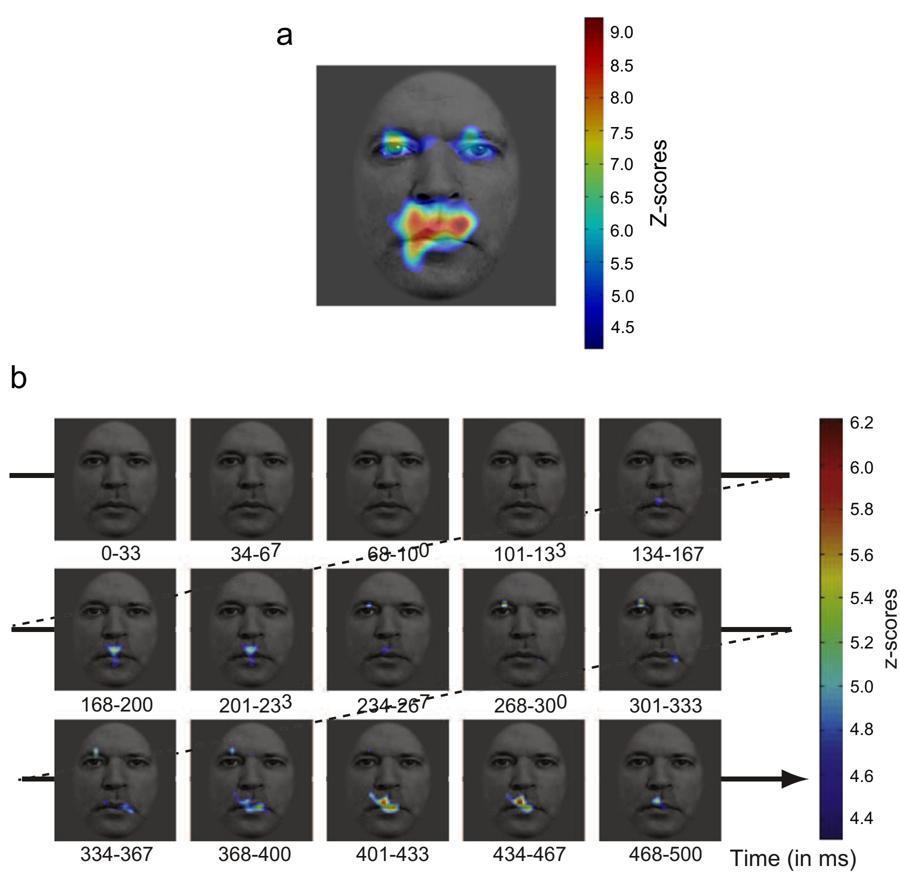

Fig. 2. Classification plane obtained with the static stimuli (a) and classification volume obtained with the dynamic stimuli (b). The areas depicted in color were significantly correlated with accuracy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

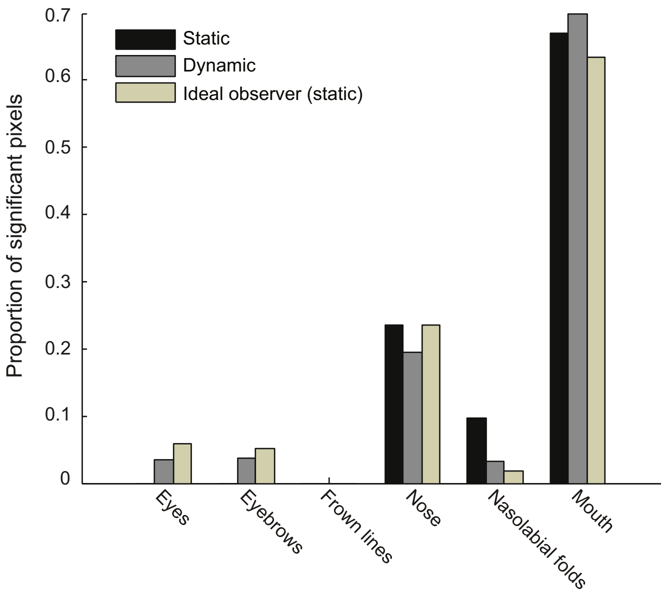

Fig. 3. ROI analysis calculated on the classification plane (i.e., static stimuli) and on the classification volume (i.e., dynamic stimuli). The time frames of the classification volume were collapsed in order to compare the relative importance of each area for the static and dynamic stimuli. The figure indicates the probability that a significant pixel falls on each ROI. The delimitation of each ROI is shown in the inset of Fig. 4.

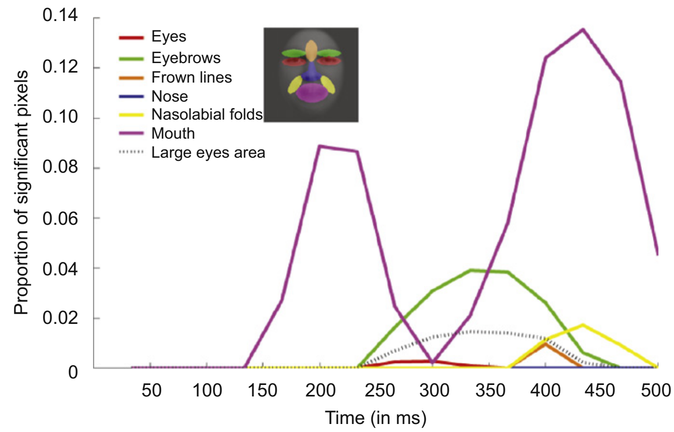

Fig. 4. ROI analysis across time on the classification volume. The figure indicates, for each time frame, the proportion of significant pixels on each ROI.

dynamic classification images and on 100% of the static classification images. We also verified that this result was not an artifact of the smoothing applied on our classification planes by performing the same analysis on unsmoothed classification planes. The mouth remained more important than the eye area on 96.7% of the dynamic classification planes and on 100% of the static classification planes.

3.3. *ROI analysis with the time dimension*

We then examined in what order the different facial features became useful for the categorization of dynamic facial expressions. We calculated for each frame of the 3D classification volume, the proportion of the total number of significant pixels (i.e., as determined with the *Pixel* test) that fell on each facial feature across the frames and divided this proportion by the total number of pixels in that feature. The relative importance of each facial feature across time is displayed in Fig. 4. The mouth is the first area to become useful for the discrimination of facial expressions, around 100 ms after stimulus onset. The left eye and the eyebrows are the second areas to become useful (around 234 ms after stimulus onset), followed by the frown lines and the nasolabial folds (around 368 ms after stimulus onset). Except for a short period of time between 268 and 333 ms, the mouth remains the most useful area throughout the stimulus presentation. We also looked at the relative importance of the mouth and the "metropolitan" eye area by grouping the eyes, the frown lines, and the eyebrows (see the dotted black curve in Fig. 4). Again, except for a short period of time around 300 ms after stimulus onset, the mouth remains the most useful area throughout stimulus duration.

Thus, in both ROI analyses, the mouth is the most useful area when it comes to discriminating facial expressions of emotions. It is conceivable that the mouth was particularly informative in the set of stimuli we used, or that the parameters used in our experiments affected the participants' strategy, and that our results overestimate the importance of the mouth area. To test this, we reanalyzed the data of Smith et al. (2005) with the same procedure as described above. These authors also applied the Bubbles method (i.e., the same version we used in our study with the static stimuli) in a discrimination task of facial expressions of emotions. However, they used a completely different set of stimuli (the California Facial Expressions – i.e., CAFE – database; Dailey, Cottrell, & Reilly, 2001) with a slightly different subset of facial expressions (i.e., they did not use the pain expression). Moreover, they adjusted the performance differently from what we did. They manipulated the number of bubbles such that accuracy was approximately equal (i.e., accuracy threshold of 75%) across the seven facial expressions. Here, we decided to allow accuracy to vary across the expressions to prevent the number of bubbles from becoming a cue for discriminating between facial expressions of emotions (i.e., the average accuracy was not controlled separately for each expression, but was instead controlled across expressions; the accuracy threshold was of 56%). Another difference between the experiment of Smith et al. (2005) and ours was the stimulus duration. They displayed the stimuli until the participant's response, whereas we displayed the stimuli for 500 ms. Because the aim of Smith et al. (2005) was to verify which information was useful in categorizing each facial expression, they analyzed each expression and spatial scale independently. Here, we re-analyzed their data to reveal the information that was useful across all expressions irrespective of spatial scale.

3.4. *Reanalysis of Smith et al. (2005)*

Similarly to our study, their experiment provided one classification plane per expression and per frequency band. We therefore added the five frequency bands and the seven expressions of their classification planes, we smoothed the resulting classification plane using a Gaussian window with a FWHM of 28.3 pixels, and we z-scored it using the pixels outside the face area to calculate the mean and standard deviation of the distribution of the null hypothesis. The resulting classification plane is presented in Fig. 5. The pixels with the highest regression coefficients (i.e., we kept the same number of pixels as displayed in Fig. 2a) are displayed in colors. These results, which are quite consistent with our own, clearly show that the mouth area is favored over the eye area during the categorization of facial expressions.

4. Discussion

We examined what visual information is most useful in discriminating the basic facial emotions and pain from one

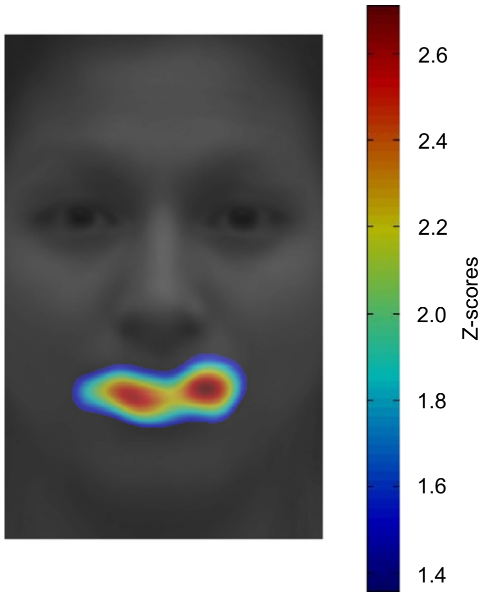

Fig. 5. Classification plane obtained with the data of Smith et al. (2005). The pixels depicted in color were those with the highest z-score values (i.e., we display the same number of pixels that used for the ROI analysis on the static version of our study). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

another using the *Bubbles* technique. We observed a clear preference for the mouth for both static and dynamic stimuli. The results obtained for the discrimination of dynamic facial expressions also show that the mouth remains the most useful area throughout the stimulus duration. We also replicated our finding in a re-analysis of the data from Smith et al. (2005) despite very different experimental parameters: the stimulus set used, the variance of performance across expressions (i.e., constant in the experiment from Smith and colleagues, varying in our experiments), and the stimulus duration.

Our findings may appear inconsistent with the eye movement literature, which shows a roughly equal sampling of the eye and mouth areas (e.g., Eisenbarth & Alpers, 2011; Jack et al., 2009), but they need not be. There is a partial dissociation between where the eye fixations land and what visual information is actually processed (Arizpe et al., 2012; Jonides, 1981; Posner, 1980; see however, Rayner, 1998; Deubel & Schneider, 2003; Godjin & Theeuwes, 2003). The eyes are smaller than the mouth, and are therefore represented by higher spatial frequencies. They may thus need to be processed within the fovea, since the high density of cones found in this area of the retina makes it more suitable for processing high spatial frequencies. The mouth, which is represented by lower spatial frequencies, may be adequately processed in parafoveal regions and may thus also be processed while the eye fixations land close to the eye area. In other words, when the eye fixations fall on the eye area, both the eye and the mouth area may be processed, whereas when the eye fixations fall on the mouth area, only the mouth area may be processed. This could explain why the proportion of fixations falling on the eye and on the mouth area is similar even if the mouth area is more useful than the eye area for the recognition of facial expressions.

Why is the mouth area more important than any other facial area for the accurate categorization of facial expressions of basic emotions? One conceivable explanation for this finding is that the mouth is the most informative area of the face (i.e., it contains more signal). To verify this possibility, we submitted an *ideal observer* – a model observer that uses all the available information optimally – to the same static facial discrimination task as our human observers (e.g., see Smith et al., 2005). The classification plane of the model observer reveals which areas of the face are informative in discriminating the expressions from one another.

4.1. Ideal observer analysis on static facial expressions

On each trial, the model observer was presented with a stimulus of the experiment that we conducted with the human observers. The same mask of bubbles was applied to the face and Gaussian white noise was added to the stimulus in order to keep the average accuracy at the same level as the one used with the humans, i.e., 56%. The amount of noise was adjusted on a trial-by-trial basis using QUEST (Watson & Pelli, 1983). The same mask of bubbles was also applied to all the other faces of the stimulus set, and the ideal observer calculated the correlation between the target stimulus presented and every other face. The facial expression of the face that had the highest correlation with the target stimulus was the model's response. We then computed the classification plane of the ideal observer using the same procedure as explained in Section 3 (see Fig. 6).

The model observer mostly used the mouth and eye areas, confirming that these face areas are the most informative when it comes to discriminating the expressions included in our study from one another. Most importantly, the ideal observer shows that the mouth area contains more information than the eye area (see also Fig. 3). This may explain at least in part why this area was used most by human participants. However, a more rigorous analysis of the similarities and differences in the visual extraction strategies of human observers and of the model observer reveals that the relative utilization of the mouth and the eyes is different for the ideal observer and for the human participants. Indeed, the ratio of the proportion of diagnostic pixels (i.e., top 5% pixels) that fell on the mouth vs. on the eye area was much greater for human participants (i.e., 8.56 and 15.51 on average for the static and dynamic stimuli, respectively) than for the ideal observer (i.e., 3.08). To test if the ratio of the mouth vs. eyes utilization was statistically significantly higher for the humans than for the ideal observer, we created 1000 classification planes using random subsamples of 20 participants (results are robust to changes in the size of this subsample of participants), and calculated the ratios of the average of the z-scores in the mouth area and in the eye area. We compared the ratio found in each of these classification planes to the one found for the ideal observer, and performed

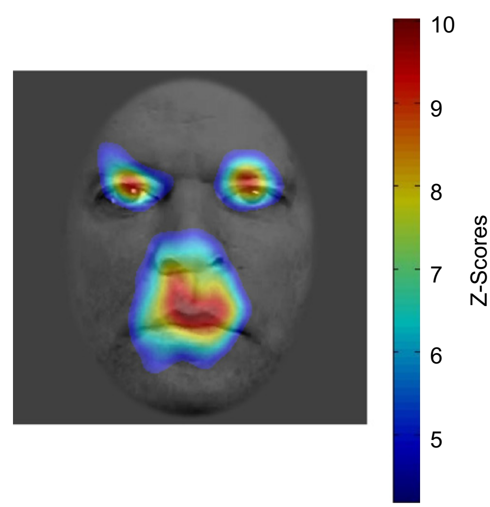

Fig. 6. Classification plane obtained by the ideal observer with the static stimuli. The areas depicted in color were significantly correlated with accuracy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a sign test on the result of this comparison. This analysis indicated a significant difference between the ratio of the mouth and eyes utilization for the humans and for the ideal observer ($p < 0.001$ for both static and dynamic stimuli). Again, we made sure that this result was not an artifact of the smoothing of our classification planes by repeating the analysis on unsmoothed classification planes: A significant difference was again found between the ratio of the mouth and eyes utilization for the humans and for the ideal observer ($p < 0.001$). Thus, pixel-wise, informativity does not account entirely for the human preference for the mouth. What else then could explain this preference?

One possibility comes from the inherently dynamic nature of facial expressions. It may be the case that the movements of the mouth contain much more information for the discrimination of natural facial expressions than the movements of any other facial area. Moreover, the human brain may have learned to use these motion cues to discriminate facial expressions and remnants of this strategy may influence how humans recognize static facial expressions. To test the first part of this hypothesis, we measured how the amplitude of the movements in different areas of the expressive face varies across our dynamic stimuli.

4.2. Motion analysis on the dynamic facial expressions

We first calculated, using a three-step search method (Koga, Linuma, Hirano, Iijima, & Ishiguro, 1981) with a spatial granularity of 10×10 pixels, the surface-based motion occurring between each step of two frames in our dynamic faces (the results of this analysis are robust to parameter changes). This resulted in one motion vector for each 10×10 area of our stimuli. We then calculated the amplitude of those vectors and, finally, we calculated the variance of these amplitudes for each area across all the stimuli and averaged these values across all frames (see Fig. 7). The variance values were transformed into z-scores using the average and the standard deviations of the variances across all the facial areas. The more variance there is in an area across all stimuli, the more this area gives information about the expression portrayed. It is clear, from this analysis, that the mouth area is by far the most informative for the categorization of all the expressions.

Therefore, the greater utilization of the mouth area in comparison to the eye area by the human observers could be explained in part by the mouth area conveying most of the movement information in dynamic stimuli. As proposed above, it is possible that the brain has elaborated a strategy over the course of its development that is suitable for dynamic expressions, since they appear in an ecological environment. Whenever the brain is exposed to a facial expression categorization task, a similar strategy is used no matter the state – static or dynamic – of the expressions. We are not suggesting that facial expression recognition relies only on motion cues, but that the importance of movement for the memory representations of facial expressions of emotions should not be underestimated. If this hypothesis is true, the facial expressions would benefit from being represented by neurons that integrate both the movement and the shape of an object. Interestingly, the superior temporal sulcus (STS), a region of the cortex that has been proposed as a site of integration for these two kinds of information (Vaina, Solomon, Chowdhury, Sinha, & Beliveau, 2001), is involved in the processing of both static (Furl, van Rijsbergen, Treves, Friston, & Dolan, 2007; Haxby, Hoffman, & Gobbini, 2000; Narumoto, Okada, Sadato, Fukui, & Yonekura, 2001; Pessoa & Padmala, 2007; Tsuchiya, Kawasaki, Oya, Howard, & Adolphs, 2008) and dynamic facial expressions (Ishai, 2008; Said, Moore, Engell, Todorov, & Haxby, 2010).

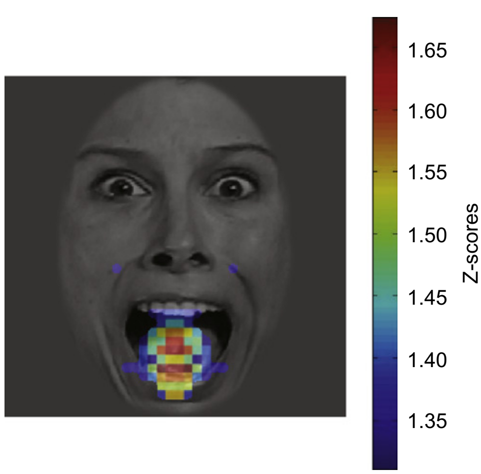

Fig. 7. Relative variance of the amplitude of motion in different areas of a face across all the dynamic stimuli tested (i.e., 80 stimuli: 8 expressions × 10 identities). The pixels depicted in color were the 5% pixels with the highest z-score values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.3. A strategy for a system with a limited capacity

Another potential explanation for the greater utilization of the mouth area by the human observers in comparison to the ideal observer may be the limited capacity of the human visual system (Levin & Simons, 1997; Simons & Rensink, 2005). Indeed, with limited capacity, a strategy whereby the most informative area is favored at the expense of other areas may be selected. The more resources are available, the more the other areas – for instance the second most informative area, the eyes – receive processing. This could also explain, at least in part, why an underutilization of the eye area has often been reported in clinical populations that show a deficit in facial expression recognition. Patients suffering from a brain lesion or from brain dysfunction related to facial expression recognition most likely have less visual resources to devote to facial expression discrimination than does the healthy population. They could have just enough resources to process the mouth but not the mouth and the eyes. Since the mouth is very informative, the patients are capable of performing the task. However, since the eyes also convey crucial information, they are impaired compared to healthy individuals. This last proposition is congruent with the relative difficulty of revealing a deficit in basic emotion recognition in the autistic population (Adolphs, Sears, & Piven, 2001; Baron-Cohen et al., 1997; Grossman, Klin, Carter, & Volkmar, 2000; Ogai et al., 2003; Ozonoff, Pennington, & Rogers, 1990; Prior, Dahlstrom, & Squires, 1990; Spezio et al., 2007a,b; Teunisse & de Gelder, 1994; Volkmar, Sparrow, Rende, & Cohen, 1989) and, therefore, the need to use very sensitive tasks—for instance the facial expression megamix (Humphreys, Minshew, Leonard, & Behrmann, 2007). Interestingly, this population has been shown to underutilize the eye area compared to healthy individuals and to rely more on the mouth area during the processing of the facial expression of basic emotions (Spezio et al., 2007a; 2007b).

Of course, the two explanations proposed above, in Sections 4.2 and 4.3, are speculative and more research will be needed to understand why the mouth area is so important for the recognition of the facial expression of basic emotions. It will also be important to use techniques other than Bubbles because each technique employed to probe the use of visual information may interact with this use of information.

4.4. Conclusion

Even if our results show that humans use the mouth area more than the eye area to discriminate the basic facial expressions from one another, the importance of the eye area should not be underestimated. Our participants effectively used the eye area, though less so than the mouth area. In fact, the eye area is the most important visual cue for the recognition of fear (Adolph et al., 2005; Smith et al., 2005; Gosselin, Spezio, Tranel, & Adolphs, 2011). Tasks using composite facial expressions (e.g., smiling mouth with angry eyes) show that the top and the bottom parts of facial expressions interact to create the final percept of the facial emotion. This is consistent with our finding that both the eye and the mouth areas are useful for the recognition of facial expressions. Moreover, research suggests that the eye area becomes more important when recognizing complex mental states (Baron-Cohen et al., 1997). Thus, our aim here is not to negate the importance of the eye area in the field of Social Neuroscience but, rather, to rehabilitate the importance of the mouth area for the recognition of facial expressions.

Aknowledgements

We would like to thank Marie Smith for kindly giving us access to her data. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (CRSNG) to Frédéric Gosselin and by a scholarship from the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT) to Caroline Blais.

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P. G., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*, 68–72.

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience*, *13*, 232–240.

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: difficulties with eye movements as a measure of information use. *PLoS One*, *7*, e31106, http://dx.doi.org/10.1371/journal.pone.0031106.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults and adults with autism or Asperger syndrome. *Visual Cognition*, *4*, 311–331.

Bassili, J. N. (1979). Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*, 2049–2058.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, *23*, 223–229.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Bukach, C. M., Bub, D. N., Gauthier, I., & Tarr, M. J. (2006). Perceptual expertise effects are not all or none: spatially limited perceptual expertise for faces in a case of prosopagnosia. *Journal of Cognitive Neuroscience*, *18*, 48–63.

Bukach, C. M., LeGrand, R., Kaiser, M., Bub, D. N., & Tanaka, J. W. (2008). Preservation of mouth region processing in two cases of prosopagnosia. *Journal of Neuropsychology*, *2*, 227–244.

Burt, P., & Adelson, E. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, *31*, 532–540.

Caldara, R., Rossion, B., Mayer, E., Smith, M., Gosselin, F., & Schyns, P. G. (2005). Does prosopagnosia take the eyes out from faces? Evidence for a defect in the use of diagnostic facial information in a single-case. *Journal of Cognitive Neuroscience*, *17*, 1–15.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*, 659–667.

Coleman, J. (1949). Facial expressions of emotion. *Psychological Monographs*, *63*, 1–296.

Cunningham, D.W., Kleiner, M., Bülthoff, H.H., & Wallraven, C. (2004). The components of conversational facial expressions. *Proceedings of the first symposium on applied perception in graphics and visualization*, http://dx.doi.org/10.1145/1012551.1012578.

Dailey M., Cottrell G.W., Reilly J. (2001). California facial expressions, CAFE. *Unpublished digital images*, University of California, San Diego, Computer Science and Engineering Department.

Darwin, C. (1965). *The expression of the emotions in man and animals*. London: John Murray.

Deubel, H., & Schneider, W. X. (2003). Delayed saccades, but not delayed manual aiming movements, require visual attention shifts. *Annals of the New York Academy of Sciences*, *1004*, 289–296.

Dunlap, K. (1927). The role of the eye muscles and mouth muscles in the expression of emotions. *Genetic Psychology Monographs*, *2*, 199–233.

Ekman, P. (1982). *Emotion in the human face* ((2nd ed.)). Cambridge, UK: Cambridge University Press.

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion*, *11*, 860–865.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*, 9602–9605.

Frois-Wittman, J. (1930). The judgement of facial expression. *Journal of Experimental Psychology*, *13*, 113–151.

Furl, N., van Rijsbergen, N. J., Treves, A., Friston, K. L., & Dolan, R. J. (2007). Experience-dependant coding of facial expression in superior temporal sulcus. *Proceedings of the National Academy of Science*, *104*, 13485–13489.

Gibson, B., Lazareva, O. F., Gosselin, F., Schyns, P. G., & Wasserman, E. A. (2007). Non-accidental properties underlie shape recognition in mammalian and non-mammalian vision. *Current Biology*, *17*, 336–340.

Godjin, R., & Theeuwes, J. (2003). Parallel allocation of attention prior to the execution of saccade sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 882–896.

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

Gosselin, F., Spezio, M. L., Tranel, D., & Adolphs, R. (2011). Asymmetrical use of eye information from faces following unilateral amygdala damage. *Social Cognitive and Affective Neuroscience*, *6*, 330–337.

Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. R. (2000). Verbal bias in recognition of facial emotions in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *41*, 213–225.

Hainline, L. (1978). Developmental changes in visual scanning of face and non-face patterns by infants. *Journal of Experimental Child Psychology*, *25*, 90–115.

Haith, M., Bergman, T., & Moore, M. (1977). Eye contact and face scanning in early infancy. *Science*, *198*, 853–855.

Hanawalt, N. (1944). The role of the upper and lower parts of the face as the basis for judging facial expressions: II. In posed expressions and "candid camera" pictures. *Journal of General Psychology*, *31*, 23–36.

Harms, M., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychological Review*, *20*, 290–322.

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed neural human system for face perception. *Trends in Cognitive Science*, *4*, 223–233.

Hughes, H. C., Fendrich, R., & Reuter-Lorenz, P. A. (1990). Global versus local processing in the absence of low spatial frequencies. *Journal of Cognitive Neuroscience*, *2*, 272–282.

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research*, *176*(2), 356–373.

Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007). A fine-grained analysis of facial expression processing in high-functioning adults with autism. *Neuropsychologia*, *45*, 685–695.

Ishai, A. (2008). Let's face it: it's a cortical network. *NeuroImage*, *40*, 415–419.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, *19*, 1543–1548.

Jonides, J. (1981). Voluntary versus automatic control over the minds eyes movement. In J. L. A. Baddeley (Ed.), *Attention and Performance IX*. Hillsdale, NJ: Erlbaum.

Koga, T., Linuma, K., Hirano, A., Iijima, Y., & Ishiguro, T. (1981). Motion compensated interframe coding for video conferencing. *Proceedings of national telecommunication conference*., New Orleans, 5.3.1–5.3.5.

Lee, J., Gosselin, F., Wynn, J. K., & Green, M. F. (2011). How do schizophrenia patients use visual information for decoding facial emotion? *Schizophrenia Bulletin*, *37*, 1001–1008.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, *4*, 501–506.

Maurer, D. (1985). Infants' perception of facedness. In: T. Field, & M. Fox (Eds.), *Social perception in infants*. Norwood, NJ: Ablex.

Mehrabian, A. (1968). Communication without words. *Psychology Today*, *2*, 53–56.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Cognitive Brain Research*, *12*, 225–231.

Nummenmaa, T. (1964). *The language of the face (Jyvaskyla studies in education, psychology, and social research)*. Finland: Jyvaskyla.

Nusseck, M., Cunningham, D. W., Wallraven, C., & Bülthoff, H. H. (2008). The contribution of different facial regions to the recognition of conversational expressions. *Journal of Vision*, *8*, 1–23.

Ogai, M., Matsumoto, H., Suzuki, K., Ozawa, F., Fukuda, R., Uchiyama, I., Suckling, J., Isoda, H., Mori, N., & Takei, N. (2003). fMRI study of recognition of facial expressions in high-functioning autistic patients. *Neuroreport*, *14*, 559–563.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1990). Are there emotion perception deficits in young autistic children? *Journal of Child Psychology and Psychiatry*, *31*, 343–361.

Pessoa, L., & Padmala, S. (2007). Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cerebral Cortex*, *17*, 691–701.
Parker, D. M., Lishman, J. R., & Hughues, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–160.
Pelli, D. G. (1997). The video toolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
Plutchik, R. (1962). *The emotions: facts, theories, and a new model*. New York: Random House.
Posner, M. I. (1980). Orienting of attention. *Quaterly Journal of Experimental Psychology*, *32*, 3–25.
Prior, M., Dahlstrom, B., & Squires, T. (1990). Autistic children's knowledge of thinking and feeling states in other people. *Journal of Child Psychology and Psychiatry*, *31*, 587–601.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
Rossion, B., Kaiser, M. D., Bub, D., & Tanaka, J. W. (2009). Is the loss of diagnosticity of the eye region of the face a common aspect of acquired prosopagnosia? *Journal of Neuropsychology*, *3*, 69–78.
Roy, S., Roy, C., Fortin, Ethier-Majcher, Belin, & Gosselin (2007). A dynamic facial expression database. *Journal of Vision*, *7*, 944.
Ruckmick, C. A. (1921). A preliminary study of the emotions. *Psychological Monographs*, *30*, 30–35.
Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, *10*, 1–12.
Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present and future. *Trends in Cognitive Sciences*, *9*, 16–20.
Smith, M., Cottrell, G., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions of emotions. *Psychological Science*, *16*, 184–189.
Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007a). Abnormal use of facial information in high-functioning autism. *Journal of Autism and Developmental Disorders*, *37*, 929–939.
Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007b). Analysis of face gaze in autism using Bubbles. *Neuropsychologia*, *45*, 144–151.
Teunisse, J.-P., & de Gelder, B. (1994). Do autistic children have a generalized face processing deficit? *International Journal of Neuroscience*, *77*, 1–10.
Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M. A., & Adolphs, R. (2008). Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. *PLoS One*, *3*(12), e3892, http://dx.doi.org/10.1371/journal.pone.0003892.
Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Beliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Science*, *98*, 11656–11661.
Volkmar, F. R., Sparrow, S. S., Rende, R. D., & Cohen, D. J. (1989). Facial perception in autism. *Journal of Child Psychology and Psychiatry*, *30*, 591–598.
Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113–120.