

Generating Controlled Image Sets in Cognitive Neuroscience Research

Jean-François Knebel · Ulrike Toepel · Julie Hudry ·
Johannes le Coutre · Micah M. Murray

Accepted: 11 February 2008 / Published online: 13 March 2008
© Springer Science+Business Media, LLC 2008

Abstract The investigation of perceptual and cognitive functions with non-invasive brain imaging methods critically depends on the careful selection of stimuli for use in experiments. For example, it must be verified that any observed effects follow from the parameter of interest (e.g. semantic category) rather than other low-level physical features (e.g. luminance, or spectral properties). Otherwise, interpretation of results is confounded. Often, researchers circumvent this issue by including additional control conditions or tasks, both of which are flawed and also prolong experiments. Here, we present some new approaches for

controlling classes of stimuli intended for use in cognitive neuroscience, however these methods can be readily extrapolated to other applications and stimulus modalities. Our approach is comprised of two levels. The first level aims at equalizing individual stimuli in terms of their mean luminance. Each data point in the stimulus is adjusted to a standardized value based on a standard value across the stimulus battery. The second level analyzes two populations of stimuli along their spectral properties (i.e. spatial frequency) using a dissimilarity metric that equals the root mean square of the distance between two populations of objects as a function of spatial frequency along x- and y-dimensions of the image. Randomized permutations are used to obtain a minimal value between the populations to minimize, in a completely data-driven manner, the spectral differences between image sets. While another paper in this issue applies these methods in the case of acoustic stimuli (Aeschlimann et al., Brain Topogr 2008), we illustrate this approach here in detail for complex visual stimuli.

Keywords Visual perception · Images · Luminance adaptation · Spectral frequency

J.-F. Knebel and U. Toepel contributed equally to this report.

J.-F. Knebel (✉) · U. Toepel · M. M. Murray
The Functional Electrical Neuroimaging Laboratory,
Neuropsychology and Neurorehabilitation Service, Centre
Hospitalier Universitaire Vaudois (CHUV) and Université de
Lausanne (UNIL), 5, Rue Pierre Decker, 1011 Lausanne,
Switzerland
e-mail: jean-francois.knebel@chuv.ch

U. Toepel · M. M. Murray
The Functional Electrical Neuroimaging Laboratory, Radiology
Service, Centre Hospitalier Universitaire Vaudois (CHUV) and
Université de Lausanne (UNIL), Lausanne, Switzerland

J. Hudry · J. le Coutre
Nestlé Research Center, Vers-chez-les-Blanc, Lausanne,
Switzerland

M. M. Murray
EEG Brain Mapping Core, Centre for Biomedical Imaging,
Radiology Service, Centre Hospitalier Universitaire Vaudois
(CHUV) and Université de Lausanne (UNIL), Bugnon 46, 1011
Lausanne, Switzerland
e-mail: Micah.murray@chuv.ch

Introduction

Investigating visual and auditory perception in humans by utilizing non-invasive brain imaging methods such as electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and near-infrared spectroscopy (NIRS) is a rapidly growing field within neuroscience. Many studies have been specifically concerned with categorical processing of object stimuli (e.g. [1–10]). Because real-world objects are often quite difficult and infeasible to present within laboratory

settings (i.e. due to their size, availability, and/or the associated difficulty in controlling their presentation), research on visual and auditory perception most often deals with replicas, including photographs and drawings, of real-world objects. Naturally, these objects and their photographed counterparts do not solely differ in their categorical attributes (e.g. cars vs. animals vs. houses, etc.). Low-level visual features also vary substantially between objects. For example, Delplanque and colleagues [11] recently examined the International Affective Picture System (IAPS; [12]) and found that differences in affective ratings can co-occur with differences in the images' spectral power. In light of these considerations neuroscientific studies of discrimination and categorization certainly need to assure that observed effects are indeed due to the category specificity of an object and not to low-level perceptual features (e.g. photograph angle, luminance, spectral properties). Otherwise, data interpretation will likely be confounded.

Brain imaging methods, in particular those with a high temporal resolution like EEG and MEG, are prone to data misinterpretation caused by low-level visual attributes. For example, the temporal dynamics of emotional influences on face processing can vary, in part, due to the level of control of low-level visual attributes (see [13] for discussion of some impacts of ERP data analysis methods). While categorization effects within ~ 80 ms after stimulus onset have been reported [14], similar effects are only observed at ~ 130 ms when stimuli are highly controlled for in their physical features [15]. Moreover, some studies have challenged the specificity of a face-selective EEG component, the N170, by relating it to variance in the pixel arrangements between images conveying facial stimuli as opposed to other objects ([16, 17], though see [18]). Although both these and other studies have reported faces versus objects ERP differences in the P100 component (i.e. substantially earlier than the N170), such effects could be driven by inter-category luminance differences and/or spectral variation (e.g. [19–21] for effects of luminance and spatial frequency on ERPs).

In order to circumvent problems of low-level physical differences between stimulus conditions, researchers might choose to introduce additional control conditions or tasks in their experimental designs, both of which are imperfect and also prolong experiments (e.g. [3, 9, 22, 23] for some examples). In this technical report, we present some new approaches for controlling classes of visual stimuli intended for use in cognitive neuroscience. These methods can also be readily extrapolated to auditory stimuli and other applications (see [24]).

Our approach is comprised of two levels. A first level serves to equate individual stimuli in terms of their mean intensity (i.e. the luminosity of images), while also

addressing potential confounds of image saturation. Many neuroscientific studies on visual processing indeed equalize image luminosity, yet details on the utilized procedures are seldom given. Thus, one objective of the present technical report was to provide newcomers with the necessary tools and mathematical formulae. The second level of our approach aims at controlling two sets of stimuli with respect to their spectral properties (i.e. spatial power spectra; see also [11]), without requiring a priori filtering of the stimuli. The two levels should be applied consecutively since the luminosity treatment can have some influences on the spatial frequency properties of an image. As will be made clear below, because the luminance treatment adjusts all images to one standard value, this value will indeed stay stable even after the definition of image subsets obtained from the second level of our approach. We present analyses based on a photographic image database we developed to investigate the time course of food categorization.

Materials And Methods

The images for the analyses were obtained in the following way. Top-view photographs of food items were taken in front of equal backgrounds from identical angles. The photographs were subdivided into high- versus low-fat food classes by means of official nutritional databases. All images were sized to 300×300 pixels. Yet, non-squared images can be utilized as well when equal in the number of pixels between images.

Intensity Treatment

The intensity treatment serves to control adaptively the luminosity of images which can be mathematically defined as luminance or grayscale value, respectively. In the case of sounds, volume would be the analogue measure. In a first step, all images are defined in color space in terms of the YUV color model (e.g. [25]), which is that used for the PAL British standard format. In this model, the Y component represents the luminance of an image. The U and V components, on the other hand, represent the chrominance of an image.

The luminance (=grayscale value) of a pixel is influenced by the color levels red (R), green (G) and blue (B) in varying proportions, such that $Y = 0.299 \times R + 0.587 \times G + 0.114 \times B$. In this way, the Y component represents a particular image in its original dimensions as a matrix of its luminance values. In order to compare the luminance properties between images, we calculated one numerical value per image. The value \bar{Y} expresses the weighted

a

$$j \in [0;255]$$

$$0=\text{black}, 255=\text{white}$$

$$\bar{Y} = \sum_{j=0}^{255} \left(\frac{\overbrace{h_j}^{\text{normalized histogram}}}{\sum_{i=0}^{255} h_i} \cdot j \right)$$

b

$$N = \bar{Y}_{\text{image}} - \text{STDVAL}$$

$$IM_{\text{new}} = IM_{\text{old}} - N$$

Cycle until N=0

$$\bar{Y}_{\text{new}} = \sum_{j=0}^{255} \left(\frac{h_{j,\text{new}}}{\sum_{i=0}^{255} h_{i,\text{new}}} \cdot j \right)$$

Fig. 1 (a) This equation shows how one numerical value per image is obtained that represents its luminance. b. Illustration of the iterative cycle applied to adapt the luminance across images (See text for details about the equations)

luminance of an image (see Fig. 1a). In order to obtain the value \bar{Y} , we applied an equation that serves to calculate the mathematical expectation of a normalized histogram [26].

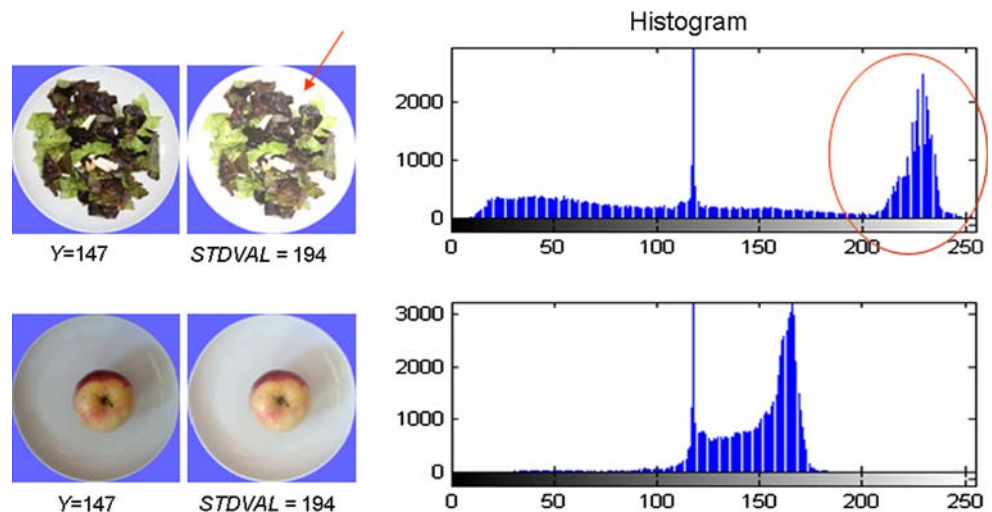
In this equation, the factor j conveys the specific grayscale value which can range from 0 (=black) to 255 (=white). The factor h_j describes how often this particular grayscale value j is present in the luminance matrix Y of one image. The value h_j is normalized by dividing it by the absolute sum of pixels h_i per image. In order to account for the arbitrary amount of pixels along the black-white dimension (0–255) the result of the normalization is multiplied with the factor j . Intrinsically, this weighting by multiplication with the factor j assumes that the luminance

increases between black and white in a linear way. The value \bar{Y} is calculated as the sum of all weighted values per image, resulting in one representative numerical value \bar{Y} per image.

Figure 1b illustrates the procedure for a set of images. First, the \bar{Y} value for each image is computed. In succession, one standard value has to be defined based on the obtained \bar{Y} values across a set of images. In general, the particular standard value used can be defined by the experimenter (for example, it may be the mean or median \bar{Y} value across all images). Second, this standard value (STDVAL) is subtracted from the \bar{Y} value of each original image (\bar{Y}_{image}) resulting in a negative or positive value N . If the value N for an image is positive, the particular image conveys a higher luminance (i.e. is brighter) than the standard value. By contrast, obtaining a negative value N would indicate that an image is lower in luminance than the defined standard. In succession, a new image (IM_{new}) is created by subtracting the N value from the original image (IM_{old}). Consecutively, the mathematical expectation of a normalized histogram (see Fig. 1) is recalculated to obtain the new \bar{Y} value of each image. This new \bar{Y} value is again submitted to the computation of the N value. The cycle is repeated until the value of N equaled 0 for all images that were submitted to the algorithm.

Figure 2 illustrates the outcome of the luminance adaptation for two images from our database. Both images are adapted to a standard value of 194. The outcome for the two images also illustrates a limitation to the algorithm. When the histogram of an image, i.e. the distribution of pixels, is rather widely dispersed across the range of values 0–255 with a concentration of pixels close to the 255 value (i.e. white), the rendering of this image to the standard value can result in a further pixel concentration towards the 255 value. Thus, a saturation effect for the image can occur, which leads to a bright appearance of the image. The histogram of the pixel values within the example image in

Fig. 2 The left panel displays sample images before and after intensity treatment. The right panel displays the histograms of these images (the x-axis represents the intensity value of a given pixel and the y-axis the number of pixels) and how the images can become faded out when distributions are wide while also containing a concentration near extreme values of Y (See text for details on how this issue can be addressed)



the upper panel of Fig. 2 exhibits a widely distributed pixel range, but a high concentration of pixels between the values 220 and 250. Luminance adaptation for the images according to the previously calculated mean standard value substantially fades out the image. The problem can be overcome by lowering the standard value for all images to be rendered or by using the median. Therefore, using the mean \bar{Y} value as a standard value might not always be satisfactory. Alternatively, images with high pixel concentrations towards the 255 value could be excluded. Note that the latter option would forcibly reduce the number of images available for the experiment.

Spectral Distance Optimization

The spectral analysis ensures equal arrangement of spectral properties within the visual objects themselves. In contrast to intensity properties, the spectral properties of an object cannot be readily altered without potentially impacting the recognizability, as this would change the overall arrangement of pixels within an image and the resultant appearance. The same constraint also applies to the spectral phase within an image as it plays a predominant role in image appearance [27].

In some studies the spatial features are, however, specifically modified to examine the functional role of high versus low spatial frequencies, e.g. for the processing of the emotional valence of facial expressions [15, 28]. Here, we chose to closely match (sub)groups of images (i.e. images from experimental condition A with the images from condition B) in terms of their spatial frequencies to gain maximal physical approximation or minimal dissimilarity, respectively. Such notwithstanding, approaches that filter the spectral properties of images [29, 30] can gain even better or perfect approximations of similarity between image groups. However, these approaches alter the overall appearance of images by eliminating certain frequency bands, which our approach seeks to avoid. Thus, while the images here have not been filtered before applying the spectral distance optimization to ensure the quasi-natural appearance of objects in an image, the experimenter interested in the functional role of selective spatial frequencies could precede the analysis with the application of a filter.

A mathematically simple way to achieve maximal “aliqueness” for subgroups of images is the Dissimilarity equation, which we have modified from the one often used in the analysis of EEG and MEG datasets to identify whether the topographies of responses differ (c.f. [13, 31, 32]). Dissimilarity as such bears no physical significance; rather, it is a singular measure of the difference along a given dimension (e.g. topography of an ERP or spatial frequency in the case of images) without any quantification

of the variance in this difference. For this reason, Dissimilarity is interpretable when a distribution of values is generated based on permutations of a dataset (here, images whereas topographic maps in the case of EEG/ERP).

Inputs for the Dissimilarity equation are the mean spectra of two subgroups of images (i.e. two experimental conditions) selected from among a larger population of images. The equation compares the mean values between conditions using the root mean square of the difference between the spectra obtained for each subgroup. The values u and v represent the mean spatial power spectra for two groups of images at a given location within the image space. This calculation yields a matrix that is then summed for all points (or coordinates) of the matrix. Finally, the result is normalized by dividing it by the factor, k , which represents the size of the image.

The subgroups yielding the lowest Dissimilarity value from a range based on all other possible subgroups of images can then be identified. The range of Dissimilarity values is based on numerous trial iterations. That is, different mean spectra are obtained by iteratively (re)selecting new subgroups of images from each original group and calculating their dissimilarity values. The subgroups yielding the lowest Dissimilarity value is selected to form the materials for the experiment, as these constitute the sets of images that are the spectrally most similar. This procedure is schematized in Fig. 3.

First, n images from the total population available for either condition are randomly selected. The mean spatial power spectrum for each of these subgroups of images is then calculated separately. Next, the Dissimilarity is calculated. The input (=image name) and output (=Dissimilarity) value of the equation are stored. Consecutively, the procedure undergoes a random number of

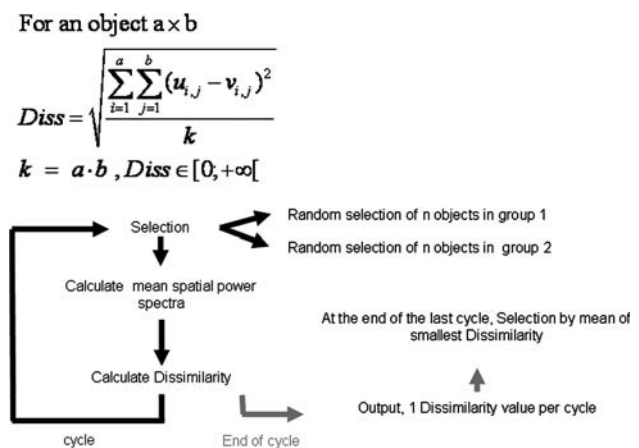
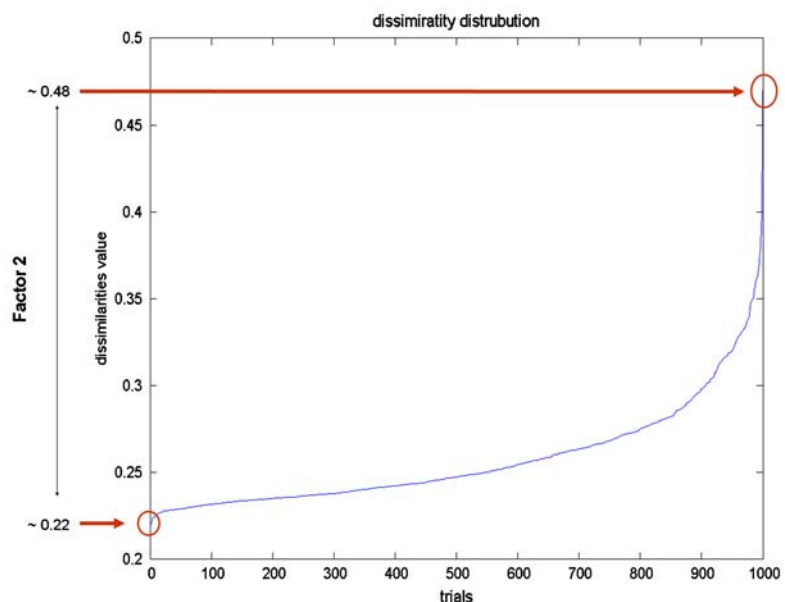
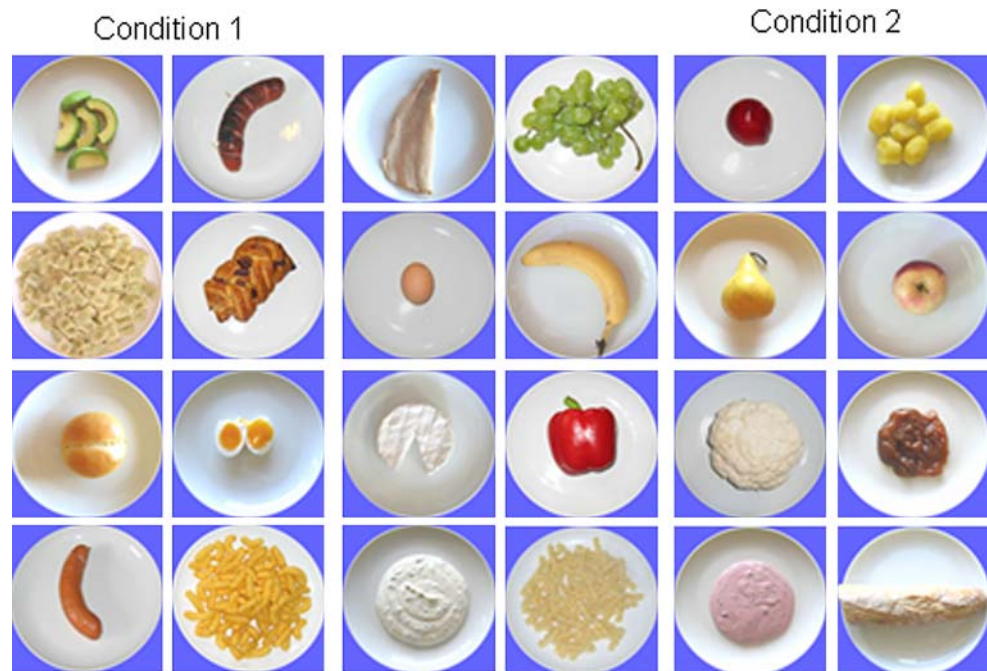


Fig. 3 Dissimilarity equation (upper panel) and illustration of the computation cycle that enables the comparison of Dissimilarity values between the different choices of images for each stimulus condition (lower panel)

Fig. 4 Output of the computation cycle depicted in Fig. 3, which shows the distribution of Dissimilarity values across trial cycles, and the corresponding image set that yielded the lowest Dissimilarity value



cycles each of which is based on new selections of n images per group, saving input and output properties as above. It is important to note that this procedure cannot be run on *all* of the original images, because the obtained Dissimilarity values would always be unchanged and no meaningful assertions could be made. After a sufficient number of iterations (e.g. 1,000) the experimenter can identify the groups of images for use in the experiment whose comparison yielded the lowest dissimilarity value.

Figure 4 illustrates an example of the output of the algorithm. In this example, each of the original 2 groups of images comprised 50 different photographs. For each cycle where Dissimilarity is calculated a subgroup of 12 of these

50 images is selected. The upper graph displays the Dissimilarity values on the y-axis, and the number of computed cycles on the x-axis. It becomes evident that one particular choice of photographs per group yields the lowest dissimilarity (~ 0.22), whereas another choice results in a Dissimilarity value twice as high (~ 0.48). The lower panel of Fig. 4 shows the groups of images from each condition that yielded the lowest Dissimilarity value and therefore match most closely in terms of spatial frequencies.

Importantly, in contrast to the intensity treatment, the spectral distance optimization does not alter the physical properties of the images. Rather, it identifies the subgroup

of images that are the closest spectrally from among the possible options (the bounds of which are defined by the available stimulus set). Consequently, the outcome of the spectral distance optimization is only true in a specific space (i.e. the specific content of groups of objects). Consequently, this method cannot reveal that the objects between groups are physically identical, but rather only that these groups of images are the least different.

Conclusion

Low-level differences in visual features often constitute a major caveat in the interpretation of neuroscientific studies of object processing. Here we present some intuitive and mathematically straightforward methods for controlling luminance and spectral properties within and between stimulus conditions. These methods are not solely applicable to visual feature control but can also be extrapolated to acoustic properties.

Acknowledgments Financial support has been provided by the Leenaards Foundation (2005 Prize for the Promotion of Scientific Research to MMM). Researchers interested in obtaining the MatLab code for the procedures should either contact Jean-François Knebel or Micah Murray.

References

1. Thorpe SJ, Fize D, Marlot C. Speed of processing in the human visual system. *Nature* 1996;381:520–2.
2. Mouchetant-Rostaing Y, Giard MH, Bentin S, Aguera PE, Pernier J. Neurophysiological correlates of face gender processing in humans. *Eur J Neurosci* 2000;12:303–10.
3. VanRullen R, Thorpe SJ. The time course of visual processing: from early perception to decision-making. *J Cogn Neurosci* 2001;13:454–61.
4. Carmel D, Bentin S. Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition* 2002;83:1–29.
5. Goffaux V, Gauthier I, Rossion B. Spatial scale contribution to early visual differences between face and object processing. *Brain Res Cogn Brain Res* 2003;16:416–24.
6. Delorme A, Rousselet GA, Mace MJ, Fabre-Thorpe M. Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Res Cogn Brain Res* 2004;19:103–13.
7. Michel CM, Seeck M, Murray MM. The speed of visual cognition. *Suppl Clin Neurophysiol* 2004;57:617–27.
8. Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* 2005;25:5148–58.
9. Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S. Rapid brain discrimination of sounds of objects. *J Neurosci* 2006;26:1293–302.
10. Harel A, Ullman S, Epshtein B, Bentin S. Mutual information of image fragments predicts categorization in humans: electrophysiological and behavioral evidence. *Vision Res* 2007;47:2010–20.
11. Delplanque S, N'diaye K, Scherer K, Grandjean D. Spatial frequencies or emotional effects? A systematic measure of spatial frequencies for IAPS pictures by a discrete wavelet analysis. *J Neurosci Methods* 2007;165:144–50.
12. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical Report A-6. Gainesville, FL: University of Florida; 2005.
13. Murray MM, Brunet D, Michel CM. Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr* 2008. doi:10.1007/s10548-008-0054-5.
14. Pizzagalli D, Regard M, Lehmann D. Rapid emotional face processing in the human right and left brain hemispheres: an ERP study. *Neuroreport* 1999;10:2691–88.
15. Pourtois G, Dan ES, Grandjean D, Sander D, Vuilleumier P. Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: time course and topographic evoked-potentials mapping. *Hum Brain Mapp* 2005;26:65–79.
16. Itier RJ, Taylor MJ. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cereb Cortex* 2004;14:132–42.
17. Thierry G, Martin CD, Downing P, Pegna AJ. Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat Neurosci* 2007;10:505–11.
18. Bentin S, Taylor MJ, Rousselet GA, Itier RJ, Caldara R, Schyns PG, Jacques C, Rossion B. Controlling interstimulus perceptual variance does not abolish N170 face sensitivity. *Nat Neurosci* 2007;10:801–2.
19. Johannes S, Münte TF, Heinze HJ, Mangun GR. Luminance and spatial attention effects on early visual processing. *Brain Res Cogn Brain Res* 1995;2:189–205.
20. McCourt ME, Foxe JJ. Brightening prospects for early cortical coding of perceived luminance: a high-density electrical mapping study. *Neuroreport* 2004;15:49–56.
21. Butler PD, Martinez A, Foxe JJ, Kim D, Zemon V, Silipo G, Mahoney J, Shpaner M, Jalbrzikowski M, Javitt DC. Subcortical visual dysfunction in schizophrenia drives secondary cortical impairments. *Brain* 2007;130:417–30.
22. Doniger GM, Foxe JJ, Murray MM, Higgins BA, Snodgrass JG, Schroeder CE, Javitt DC. Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *J Cogn Neurosci* 2000;12:615–21.
23. Rossion B, Jacques C. Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage* 2007. doi:10.1016/j.neuroimage.10.011.
24. Aeschlimann M, Knebel J-F, Murray MM, Clarke S. Emotional pre-eminence of human vocalizations. *Brain Topogr* 2008. doi:10.1007/s10548-008-0051-8
25. Wyszecki G, Stiles WS. Color science: concepts and methods, quantitative data and formulae. New York: John Wiley & Sons; 1982.
26. Pascal B, Huygens C. *Treaty of mathematics* “De rationibus in ludo aleae”. 1657.
27. Oppenheim AV, Lim JS. The importance of phase in signals. *Proc IEEE* 1981;69:529–41.
28. Vuilleumier P, Armony JL, Driver J, Dolan RJ. Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat Neurosci* 2003;6:624–31.
29. Nasanen R. Spatial frequency bandwidth used in the recognition of facial images. *Vis Res* 1999;39:3824–33.
30. Dakin SC, Hess RF, Ledgeway T, Achtman RL. What causes non-monotonic tuning of fMRI response to noisy images? *Curr Biol* 2002;12:R476–7.
31. Lehmann D, Skrandies W. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol* 1980;48:609–21.
32. Skrandies W. EEG/EP: new techniques. *Brain Topogr* 1993;5:347–50.