Possible significance of spatial heterogeneities of local visual features for face perception [version 1; peer review: 1 approved, 1 approved with reservations]

Vitaly V. Babenko, Daria S. Alekseeva, Denis V. Yavna
Academy of Psychology and Pedagogics, Southern Federal University, Rostov-on-Don, 344038, Russian Federation

Abstract
Second-order visual filters are the mechanisms which preattentively combine the rectified outputs of first-order filters (the linear striate neurons). This allows them to select the image areas which are characterized by spatial heterogeneity of the local visual features. The aim of our research is to determine whether information from these areas may be sufficient to detect unfamiliar faces and to distinguish their gender. In our experiments we used digital photos of real living things or artificial objects and faces. All these images were adjusted to an average luminance, contrast and size (7 angle degree) and were processed to extract the areas which differ the most in contrast, orientation, and spatial frequency in each of the six spatial frequencies (0.5, 1, 2, 4, 8, and 16 cpd). The other image parts were adjusted to the background. The obtained pictures were presented in a random sequence. The observer had to say what he/she saw after each presentation. When a face was presented the observer’s answer could be assigned to one of the categories: ‘it is not clear’, ‘head’, ‘human face’, ‘male / female’. We found that the information contained in the image areas with a spatial heterogeneity of the local features is sufficient not only for detecting a face, but also for distinguishing its gender. The best results were obtained at a carrier frequency of 2 cpd. The results were a little bit worse at 0.5 and 1 cpd. However, the information extracted from the high-frequency half of the spectrum was significantly less useful. The obtained results allow us to suggest that the information transmitted by the second-order visual filters may be used for pattern recognition.

Keywords
visual filters, spatial heterogeneity, pattern recognition
Introduction
The issue of visual image formation has a long history. Until recently, there hasn’t been proposed a theory that could explain everything on that matter. In visual neuroscience 3 points of view on image formation have prevailed. According to the first one, an image is a holistic description in which the most typical object of its class is taken as a standard. This is called ‘a template theory’. The second point of view also came from a holistic image description, but takes an average description of an object of its class as standard. This is called ‘a prototype theory’. The third theory assumes that every image can be described using the summation of its features. This is called ‘a feature theory’. Until now it’s remained unknown how could the invariance of holistic descriptions be provided and what could be presented as separation features.

No matter what point of view is closer to the truth, it is now stated for sure that the initial visual processing is a parallel local description of an input, which results in breaking a scene into a quantity of fragments, which are known as primitives. These are the gradients of luminance of various localization, orientation and spatial frequency. The operation begins in retina and ends in the visual cortex.

But this is only the start of visual processing. Image formation inevitably includes grouping of primitives, attributed to a single object. At first, the theory of the integration of features, according to which the mechanism of bounding is selective attention, was popular. Lately though, the number of tasks had been described to solve which spatial grouping implements preattentively. It is, for example, perceiving of second order movement and texture separation. These operations can be done by the so-called ‘second-order mechanisms’ that preattentively (following a certain rule) bind outputs of first-order mechanisms (the linear striate neurons). The following studies proved an existence of such mechanisms and determined its properties. Considering the second order movement to be a laboratory phenomenon, the ability to quickly divide textures is very important in everyday life.

Is the role of the second order mechanisms limited by the task of texture separation? Considering that these mechanisms can distinguish special modulations of local features, the attempts were made to establish the considerable role of this information in perceiving complex scenes and objects. Analysis of the natural pictures showed that notably the first and the second order features spatially overlap. As a result the second order features were determined to delete the ambiguities from interpretation of the change of luminance (the first order features).

Meanwhile, it’s quite rational to assume that these modulations could contain important information concerning an object’s forms and their details. Considering this, the goal of our study – to determine whether information concerning spacial heterogeneity – could be useful in identifying all the images and faces among the ‘not faces’ in particular.

Initiating the task ahead, we cannot ignore the fact that the early visual processing is operated by the system of parallel paths that are set to different spacial frequencies. It is known that, when it comes to tasks of identifying faces, these frequencies are not the same. There’s also the probability that the results of the processing by particular spacial channel are united in certain combinations.

Preparing the test images, we followed the assumptions about the organization of the second order mechanisms, which are displayed in the ‘filter-rectify-filter’ model. According to it, the outputs from adjacent linear filters (the first-order filters) with the same frequency and orientation set are united by the certain algorithm in the second-order filters. In other words, in case of the second-order mechanisms, those filters that differ only in localization in field of view unite. Such filters with different resolutions pass those regions of image that differ in heterogeneity to contrast, orientation and spacial frequency. We followed the assumption that these regions, due to their heterogeneity, contain the important information and could be viewed as the ‘regions of interest’.

Methods
Apparatus
The stimuli were displayed on a 17” LG Flatron 775FT monitor hosted by a PC (amd64-compatible) with an NVIDIA GeForce 7300 SE graphical subsystem running Debian GNU/Linux 7.2 (wheezy). The screen resolution was 1152 × 864 pixels with a refresh rate of 75 Hz. The monitor luminance was calibrated by a digital photometer (manufactured by ‘TKA’, St. Peterburg, Russia) using 256 gray levels.

Stimuli
The digital photographs of real objects and faces were used as initial images. All images were previously adjusted in size (7 angle deg.). The average luminance of the stimulus equaled the luminance of the background and was 19 kd/m². Initial images were processed in such a way that the areas which were different from the surroundings in contrast, orientation and spatial frequency in 6 frequency ranges corresponding to the frequency tuning of human visual pathways were extracted. The object size was such that its maximum length along any axis corresponded to 0.5 period of the SOF (the window diameter) which was tuned to the lowest carrier frequency (3 cpi).

The sequence of the computations for the preparation of the test images reproduced the operation sequence in the basic model ‘filter-rectify-filter’:

1. The initial image linear filtration (by FOFs).
   The FOF core is a two-dimensional Gabor function. FOF bandpass is 2 octaves. 6 peak spatial frequency with an increment of 1 octave (from 4 to 128 cpi) and 6 orientations with an increment of 30 deg. (from 0 to 150 deg.).

2. Rectification.
   The rectification was realized by square-rooting of the sum of squares of the FOFs’ outputs forming the quadrature pair.

3. The linear filtering of the 36 obtained images (6 spatial frequencies × 6 orientations) by the SOFs.
The SOF' core is a two-dimensional Gabor function with 1 period which is 8 times longer than 1 period of the combined FOFs. The orientation tunings of the FOF and SOF were the same.

4. Orientation integration.

6 values corresponding to 6 FOF orientations were obtained for each pixel. Then the maximum of these values was attached to each pixel.

5. Finding the local peaks at the SOFs' outputs.

The local maximums were found in each of the six matrices of the SOFs' outputs (6 spatial frequencies of the carrier).

6. Windows allocation.

Each maximum in each ‘frequency slice’ became the window center through which the information from the FOFs was allowed to pass. The window’s diameter was 0.5 of the period of the SOFs forming this frequency slice.

7. Filling of the windows.

Each window was filled with the image, obtained by FOFs at corresponding frequency. The pixels’ luminance was decreased by Gaussian from the window center to the periphery. The image was filled with the background outside the window. In the case of overlapping of the windows the pixels got the major luminance value.

Results

The information was allowed to pass only through one ‘window’, centered relative to the face (Figure 1A,C) when the initial (real) images were processed by the SOFs tuned to the lowest frequency of the carrier (0.5 cpd) (we denote these filters as F1). The result of the processing is shown in Figure 1B,D. Looking at the presented images the observers determined the gender in 87.9% and gave a more general response “a face” only in 11.4%.

If the initial images were processed by the SOFs tuned to a higher frequency of the carrier (1 cpd) (F2) the information from only a part of a face could be transmitted through one window (Figure 2A,C).

![Figure 1. The face processing at the carrier frequency 0.5 cpd. A, C – the initial images. The circles are the windows through which the filtered image is allowed to pass. B, D – the test (processed) images. There are only 'the regions of interest' at the frequency of filtering.](image1)

![Figure 2. The face processing at the carrier frequency 1 cpd.](image2)

Procedure

The subjects were seated at the distance of 1.15 m from the monitor that was randomly showing the previously made stimulus. Looking at the queue image, the observer needed to tell what he saw. The time of showing wasn’t limited. The images based on photo of man’s and women’s face (unfamiliar) were shown in the queue of the ‘not-faces’ images. The subject’s responses to the above-mentioned images could’ve been categorized in one of three existing categories (‘head’, ‘human’s face’, ‘man or women’), or said that it hadn’t been noted at all in case of a wrong or missing answer. Questions that could lead the observer to the right answer were not asked.

Subjects

A total number of 70 students (9 men and 61 women) aged between 17 and 21 took part in this experiment. All the participants had normal or corrected to normal vision and no history of neurological or psychiatric disorders had been reported. The participants did not take any medicines just before or during the study tests. All the participants of the research were informed about the purpose and the procedures of the experiment; they all signed a consent form that outlined the risks and benefits of participating in the study and indicated that they believed in the safety of the investigation. The study was realized in accordance with the ethical standards consistent with The Code of Ethics of the World Medical Association (Declaration of Helsinki) and approved by the local ethics committee.
The information was transmitted through 2 windows because there were 2 local maximums at the F2 outputs. The result of the processing may be seen in Figure 2B,D. We should mention that the observers' results were a little worse than the previous ones. Now the gender was defined in 75.7% and the response ‘a face’ was given in 20%.

The SOFs spatially integrating the higher frequency signals (2 cpd) (F3) passed information through the windows which size was about 0.25 of the face (Figure 3A,C). As a result, the test images were formed, shown in Figure 3B,D. In this case the performance was again improved. The observers determined the gender in 94.3%.

A further reducing of the SOFs’ size while increasing the carrier frequency led only to deterioration of the performance (Figure 4, Figure 5, Figure 6).

All obtained results are summarized in the Table 1.

<table>
<thead>
<tr>
<th>Carrier frequency (cpd)</th>
<th>Category (percentage of answers)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>head</td>
</tr>
<tr>
<td>0.5 (F1)</td>
<td>0.71</td>
</tr>
<tr>
<td>1.0 (F2)</td>
<td>2.14</td>
</tr>
<tr>
<td>2.0 (F3)</td>
<td>0.00</td>
</tr>
<tr>
<td>4.0 (F4)</td>
<td>0.71</td>
</tr>
<tr>
<td>8.0 (F5)</td>
<td>0.00</td>
</tr>
<tr>
<td>16.0 (F6)</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table 1. Face categorization using the image areas with a spatial heterogeneity of the local visual features.
Integration of the information which was extracted by three SOFs from the low-frequency half of the spectrum (F1+F2+F3) (Figure 7A) did not improve the performance comparing with using only F3 (92.1% versus 94.3% respectively).

But this operation (F1+F2+F3) makes it to identify the person if the initial image is a familiar face\textsuperscript{25} (Figure 7B).

<table>
<thead>
<tr>
<th>Dataset 1. Frequencies of different types of the responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://dx.doi.org/10.5256/f1000research.5975.d41499">http://dx.doi.org/10.5256/f1000research.5975.d41499</a></td>
</tr>
<tr>
<td>Each cell contains the frequency of a certain type of response from the participants to the type of stimuli as shown.</td>
</tr>
</tbody>
</table>

Discussion
The issue of ‘face perceiving’ can be divided conditionally into two parts: the allocation of the useful information (the reduction of redundancy) and the building a sizer of the selected information (the recognition). Our research concerns first part of this issue.

Among all the known algorithms of finding the ‘regions of interest’ only the small part could be viewed as neural\textsuperscript{26–32}. These algorithms can be divided into the modular and the net. In case of the first ones the weight is given, in case of the second ones it is formed during the training of the net. The approach used by us in our work is based on the modular architecture of the earlier levels of processing, finishing with the automatic allocation of the useful information from an incoming image.

In our research the first order filters formed six copies of an incoming image with different definition, and the second order filters were used as windows, which were at the maximum level of difference from surroundings by contrast, orientation or spatial-frequency.

The received results show that the identification of a face is more effective on the carrier frequency 2 cpd. This conforms with the other authors’ data that showed that the identification of faces is faster and more precise if the frequencies of the middle range are used\textsuperscript{16,17,33–37}. So what is new in our information?

We’ve shown that not the whole face is informative, but only its regions with spatial heterogeneity. Meaning, in task of detection of a human face and the definition of the gender information of a whole face is significantly redundant. It does not contradict with the data that processing of a human face is holistic\textsuperscript{38,39}. It’s just that the integrated information concerning its most informative areas could be enough for the holistic description of it.

If 100% would be the sum of all second-order filters activated by our images we can presume that at frequency of 2pcd the volume of selected information would be 1%. Reminder, this amount of information is enough to determine gender confidently.

To identify a familiar face, determination of regions of interest on one of the frequencies is not enough. The summation of the low-frequency half of spectrum is necessary at least (Figure 8). High-frequency information is not crucial for the gender determination and identification, but useful in perceiving details and delicate differentiation.

Thus been said, we chose areas of image that differ most from the surroundings in contrast, orientation and spacial frequency to be the most informative. The second order filters that we used form maps of convexity for every carrier frequency. As a result we have the “embedded maps”. At the lowest of the used frequencies one of the filters selects face as a whole. The following maps select areas of a face that are smaller and smaller. If the object approaches or retreats, so that its size changes in a certain range, the embedded maps stay the same. The difference would be only if the object approaches the same regions of interest would be allocate by second-order filters, which are tuned at lower carrier frequency, and if it retreats – at the higher one.

The smaller the window, which transmits the information, the higher its definition is. As a result, the same portions of information are transmitted through every window. If size of an object or it’s turns is changing, general capacity and nature of information transmitted by second-order filters stays the same.

Conclusions
Note that the information allocated with the given algorithm is useful for perceiving faces, the following hypothetical model of second-order filters image forming can be proposed. The face describing is simultaneous in a number of definition levels. At the relatively low level a face is described as a whole. With the higher

![Figure 7. The results of windows' combination at the frequencies 0.5, 1, and 2 cpd. A – the unfamiliar face from our experiment, B – the familiar face.](image-url)
The result of the familiar faces’ processing at the carrier frequency 2 cpd. At the left – Einstein, at the right – princess Diana.

definition transmits information concerning large objects of a face. Every higher level describes even smaller details. Wherein the given information allocates and transmits with parallel frequency channels. As a result, a hierarchical description of a face formed parallel, according to automatic algorithm. Wherein, the system of decision making can not use all available information. Elaborations will be made until specific visual task will be resolved.

Figure 8. The result of the familiar faces’ processing at the carrier frequency 2 cpd. At the left – Einstein, at the right – princess Diana.

The obtained results allow to assume that second-order filters are suitable candidates to the role of mechanism of convexity map formatting, and information they allocate can be used to form a face image.

Data availability
F1000Research: Dataset 1. Frequencies of different types of the responses, 10.5256/f1000research.5975.d41499.

Author contributions
Babenko V.V. is the author of idea and method. He analyzed the obtained results and wrote the manuscript. Alekseeva D.S. prepared the initial images and conducted the study. Yavna D.V. created the computer model of the second order visual filters end the experimental software, formed the test stimuli, and designed the article.

Competing interests
No competing interests were disclosed.

Grant information
This work was financially supported by the Ministry of education and science of Russia (Agreement 1741).

References
Reference Source

PubMed Abstract | Publisher Full Text

Publisher Full Text

Reference Source

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text | Free Full Text

PubMed Abstract | Publisher Full Text

Publisher Full Text

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text

Reference Source

PubMed Abstract | Publisher Full Text

Data Source
Open Peer Review

Current Peer Review Status: ✔️❓

Version 1

Reviewer Report 09 April 2015

https://doi.org/10.5256/f1000research.6392.r7885

© 2015 Bachmann T. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Talis Bachmann
Laboratory of Cognitive Neuroscience, Institute of Public Law, University of Tartu, Tartu, Estonia

This paper includes several interesting ideas (e.g., how different response categories by subjects allow to know what is salient in the test image, using spatial heterogeneity areas as the basic image processing strategy). However, in its present form it could not be accepted by a standard peer review based scientific journal. Among several problems the following stand out.

1. The language is poor; the manuscript should be thoroughly edited by a native speaker.

2. Authors should carefully and convincingly persuade a reader why their study is valid for real face perception. As it stands now, they artificially create an image of some object (e.g., face) with spurious new visual characteristics, then they use these images for a perception task by human subjects and then they generalize for normal visual perception. It seems they do not study real perception of untransformed objects belonging to certain category, but study perception of specific artificially created stimuli.

3. In the Introduction (e.g., first two paragraphs) needed specific references are absent. There are many minor points to be corrected (e.g., kd pro cd, Procedure too vaguely described, etc.)

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 06 March 2015

https://doi.org/10.5256/f1000research.6392.r7766
Alexander Latanov
Faculty of Biology, Moscow State University, Moscow, Russian Federation

The manuscript represents very interesting study on second-order visual filters that determine perception of face feature. These filters are associated with second-order neuronal populations that combine the outputs of striate neurons coding the primary visual features. The authors assume that second-order filters represent the mechanism of salience map formation. This map constitutes the basis for the recognition of human faces.
I recommend this article for indexation.

However the manuscript needs to be slightly corrected before indexation.
- What is the difference between abbreviation ‘cpi’ and ‘cpd’?
- Sometimes author use the term ‘spacial frequency’ and sometimes the term ‘spatial frequency’.
- It seems that the next two sentences represent the result of splitting of whole sentence: ‘Wherein, the system of decision making can not use all available information. Elaborations will be made until specific visual task will be resolved’.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.