



RESEARCH ARTICLE

REVISED Nonlocal contrast calculated by the second order visual mechanisms and its significance in identifying facial emotions [version 2; peer review: 2 approved]

Vitaly V. Babenko, Denis V. Yavna , Pavel N. Ermakov , Polina V. Anokhina

Department of Psychophysiology and Clinical Psychology, Academy of Psychology and Education Sciences, Southern Federal University, Rostov-on-Don, Russian Federation

V2 First published: 06 Apr 2021, 10:274
<https://doi.org/10.12688/f1000research.28396.1>

Latest published: 29 Aug 2023, 10:274
<https://doi.org/10.12688/f1000research.28396.2>

Abstract

Background: Previously obtained results indicate that faces are /preattentively/ detected in the visual scene *very fast*, and information on facial expression is rapidly extracted at the lower levels of the visual system. At the same time different facial attributes make different contributions in facial expression recognition. However, it is known, among the preattentive mechanisms there are none that would be selective for certain facial features, such as eyes or mouth. The aim of our study was to identify a candidate for the role of such a mechanism. Our assumption was that the most informative areas of the image are those characterized by spatial heterogeneity, particularly with nonlocal contrast changes. These areas may be identified /in the human visual system/ by the second-order visual /mechanisms/ filters selective to contrast modulations of brightness gradients.

Methods: We developed a software program imitating the operation of these /mechanisms/ filters and finding areas of contrast heterogeneity in the image. Using this program, we extracted areas with maximum, minimum and medium contrast modulation amplitudes from the initial face images, then we used these to make three variants of one and the same face. The faces were demonstrated to the observers along with other objects synthesized the same way. The participants had to identify faces and define facial emotional expressions.

Results: It was found that the greater is the contrast modulation amplitude of the areas shaping the face, the more precisely the emotion is identified.

Conclusions: The results suggest that areas with a greater increase in nonlocal contrast are more informative in facial images, and the second-order visual /mechanisms/ filters can claim the role of /filters/ elements that detect areas of interest, attract visual attention and are windows through which subsequent levels of visual processing receive

Open Peer Review

Approval Status

	1	2
version 2 (revision) 29 Aug 2023		 view
version 1 06 Apr 2021	 view	 view

1. **Yuri E. Shelepin**, Russian Academy of Sciences, St. Petersburg, Russian Federation
2. **Tina Tong Liu**, National Institutes of Health, Bethesda, USA

Any reports and responses or comments on the article can be found at the end of the article.

valuable information.

Keywords

face, emotion, saliency, spatial heterogeneity, nonlocal contrast, second-order visual mechanisms



This article is included in the **Software and Hardware Engineering** gateway.

Corresponding author: Denis V. Yavna (yavna@fortran.su)

Author roles: **Babenko VV:** Conceptualization, Funding Acquisition, Investigation, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Yavna DV:** Data Curation, Software, Writing – Original Draft Preparation; **Ermakov PN:** Funding Acquisition, Writing – Review & Editing; **Anokhina PV:** Investigation

Competing interests: No competing interests were disclosed.

Grant information: Supported by Russian Science Foundation (Project No. 20-64-47057).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2023 Babenko VV *et al.* This is an open access article 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.

How to cite this article: Babenko VV, Yavna DV, Ermakov PN and Anokhina PV. **Nonlocal contrast calculated by the second order visual mechanisms and its significance in identifying facial emotions [version 2; peer review: 2 approved]** F1000Research 2023, **10**:274 <https://doi.org/10.12688/f1000research.28396.2>

First published: 06 Apr 2021, **10**:274 <https://doi.org/10.12688/f1000research.28396.1>

REVISED Amendments from Version 1

In Introduction, some statements have been changed and several references have been corrected. Some statements seemed too categorical have been softened. Thus, the term "mechanism" was replaced by "filter"; a more balanced view on the issue of preattentive nature of face recognition was given. A more detailed description of the second-order visual filter model has been given as well.

The section "Methods" has been structured and expanded to make the design of the study more understandable to the reader. The details about the observer's task and the stimuli used was added. In the section "Results", the confidence boundaries have been added to [Figure 2](#). The conclusions have been expanded.

The content of the OSF repository linked to the article has been expanded. All raw results obtained by the authors and the stimuli created by processing photographs from the FERET collection have been provided for free access. The names of the original face images from this collection have been added to [readme.txt](#) file. Also, some additional information about the study design and descriptions of some scripts used have been added to this file.

Any further responses from the reviewers can be found at the end of the article

Introduction

Experiments involving a saccadic task ([Crouzet et al., 2010](#); [Kirchner & Thorpe, 2006](#)), registration of MEG ([J. Liu et al., 2002](#)), ERP ([Cauchoux et al., 2014](#); [Dering et al., 2011](#); [Herrmann et al., 2005](#)) and measurement of intracranial field potentials ([H. Liu et al., 2009](#)) showed that face detection and identification are so fast that we can most probably speak here of a feedforward processing ([Crouzet & Thorpe, 2011](#); [Muukkonen et al., 2020](#); [VanRullen, 2006](#); [Vuilleumier, 2000](#); [Vuilleumier, 2002](#); but see [T. Liu et al., 2022](#)), that is without any involvement of attention ([Entzmann et al., 2021](#); [Kovarski et al., 2017](#); [Reddy et al., 2004](#); [Reddy et al., 2006](#)). However, there is also an opposite point of view ([Pessoa et al., 2002](#); [Schindler & Bublatzky, 2020](#); [Tomasik et al., 2009](#)). This may mean that low level information is used to distinguish a face from the background and to define its characteristics.

Many researchers believe that faces are holistically coded within the low-frequency range, and this description is sufficient not just to detect the face but also to determine its emotional expression ([Calder et al., 2000](#); [Schyns & Oliva, 1999](#); [Tanaka et al., 2012](#); [White, 2000](#)). Meanwhile the classical work by [A.L. Yarbus \(1967\)](#) clearly demonstrated that while viewing a face we fix our eyes at quite definite details. Further eye tracking experiments and experiments with the "bubbles" method showed that not all areas of the face are equally useful for emotion recognition ([Blais et al., 2017](#); [Duncan et al., 2017](#)). Different facial features are significant for the discrimination of different emotions ([Atkinson & Smithson, 2020](#); [Calvo et al., 2014](#); [Eisenbarth & Alpers, 2011](#); [Fiset et al., 2017](#); [Jack et al., 2014](#); [Smith & Merlusca, 2014](#); [Smith & Schyns, 2009](#); [Smith et al., 2005](#); [Wang et al., 2011](#)), these emotions being probably processed at different rates too ([Ruiz-Soler & Beltran, 2012](#)).

The problem is that the lower levels of the human visual system, which are classified as preattentive stages of processing, lack neurons which would be selective to certain facial features. While recent evidence suggests that V1 activity is modulated by the amygdala in the perception of emotional faces ([T. Liu et al., 2022](#)), this feedback is unlikely to be involved in feedforward processing. Nevertheless, there should exist a mechanism permitting the detection of faces automatically and to extract significant information quickly. The aim of this investigation was to identify the possible candidate for the above mechanism.

Realization of the importance of defining those areas of interest in the images that attract visual attention, was the impetus for those research trends aimed at finding the algorithm of formation of saliency maps ([Borji et al., 2013](#); [Judd et al., 2012](#); [Rahman et al., 2014](#)). At the same time, the choice of the attention goal should be based on the principle of information maximization ([Bruce & Tsotsos, 2005](#)).

In respect of the human visual system, one can only speak of the preattentive processes actualized within low-level vision and able of "bottom-up" attention control. It is clear that attention is attracted to what changes in time (on- and off-reactions) and in space (changes in luminance). For the saliency problem, the latter is the most important. Indeed, there are specialized cells for finding brightness gradients in the visual system, and these are striate neurons ([Hubel & Wiesel, 1962](#)). However, these can only find local heterogeneities. To find areas of interest, there should exist mechanisms beyond local operations. Yet we first have to answer the question about the characteristics of these nonlocal areas of interest. In recent years, there appeared a viewpoint stating that the image areas whose information content differs from the surroundings are of the greatest interest for the visual system ([Baldi & Itti, 2010](#); [Hou et al., 2013](#); [Itti & Baldi, 2009](#)). This refers to difference in low-level feature distribution in the field of view ([Itti et al., 1998](#)), while salience in this case is determined by the degree of total difference of features within the analyzed area from features in the surrounding area ([Bruce & Tsotsos, 2009](#); [Gao & Vasconcelos, 2007](#); [Perazzi et al., 2012](#)).

Important is that the human visual system can find space heterogeneities of brightness gradients (see review [Graham, 2011](#)). This operation is implemented by so-called second-order filters. They are localized mainly in the ventral extrastriate regions ([Larsson et al., 2006](#)) and unite the outputs of simple striate neurons according to a certain rule. Two successive stages of linear filtering are separated by a rectifying non-linearity (for a more detailed introduction to the model of the second-order filters, see [Kingdom et al., 2003](#)). In this case, the description of the carrier by the first-order filters is transformed by the second-order filters into a description of the envelope. The receptive fields of the second-order filters are organized in such a way that these elements do not respond to homogeneous textures, but are activated when the texture has modulations of contrast, orientation, or spatial frequency of brightness gradients.

So far these processes have been predominantly studied and considered as an instrument of segmentation of textures (e.g. [Graham & Sutter, 2000](#); [Graham & Wolfson, 2004](#); [Kingdom *et al.*, 2003](#); [Schofield & Yates, 2005](#)). Here we raise the question whether the second-order visual filters can be of use in segmenting natural images and finding in them those saliency areas that are used for categorization. Our expectation was to obtain the answer through the task of detecting faces in a series of successively presented objects and determining their emotional expression.

It was shown earlier that the second-order filters are specific to the modulated visual feature, i.e. whether it is contrast, orientation or spatial frequency of brightness gradients ([Babenko & Ermakov, 2015](#); [Kingdom *et al.*, 2003](#)). Then it was revealed that modulations of contrast take priority in competition for attention ([Babenko *et al.*, 2020](#)). All this enabled us to work out a hypothesis stating that areas of maximum modulation of nonlocal contrast contain information helpful in identifying emotional facial expressions. To test this hypothesis, we developed a software program (gradient operator of nonlocal contrast) imitating operation of the second-order visual filters and calculating the space distribution of contrast modulation amplitude in the input image.

Methods

Participants. A total of 38 students between the ages of 19 and 21 took part in this investigation. All the participants had normal or corrected to normal vision and reported no history of neurological or psychiatric disorders. All the research participants were informed about the purpose and 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 conducted 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 design of the experiment, the methodological approach, the conditions of confidentiality and use of the consent of participants were performed according to the Code of Ethics of Southern Federal University (SFU; Rostov-on-Don, Russia) and approved ethically by the Academic Council of the Academy of Psychology and Pedagogy of SFU, on 25 March, 2020.

Stimuli. Initial digitized photos of faces and objects brought to a single size (8 ang.deg.), medium brightness (35 cd/m²) and RMS contrast (0.45), were processed by the nonlocal contrast gradient operator. A total energy of the image filtered at a frequency of 4 cycles per a diameter of this central area with a 1 octave bandwidth, was calculated in the center of the operator's concentric area. In the peripheral part of the operator (a ring whose width equaled the central area diameter), the spectral power of the entire range of spatial frequencies perceived by a person was calculated, per 1 octave on average.

The contrast modulation amplitude amounted to the difference of values of the power spectrum obtained in the operator's

central and periphery areas. Operators of various diameters were used, and for each operator we defined those areas where the total contrast was maximum different from the surroundings, i.e. had the highest modulation amplitude.

The algorithm of stimuli formation is shown in [Figure 1](#). An initial image example can be seen on the left. Then there are spatial frequencies in cycles per image (cpi) for which space distribution of the total nonlocal contrast was defined. On the right, one may see 3D maps of space distribution of contrast modulation amplitude when using operators of various diameters. The next column demonstrates the same maps in a 2D format. Red dots on them show local maximum apexes. While processing the image with the gradient operator of the largest size with its central area diameter making one half of the image size, we selected 2 maximums, after which, in the course of operator diameter two-fold reduction, selected were 4, 8 and 16 maximums correspondingly. A round aperture with a Gaussian transfer function transmitting four image cycles (hereinafter this aperture is referred to as a "window") was placed within positions found this way. Areas of maximum contrast modulation amplitude were combined in a new image (the right column). The total diameter of the areas found at different spatial frequencies equaled the diameter of a conventional circle with the initial image fit to it. Stimuli were the faces synthesized from the areas extracted at one spatial frequency (examples can be seen in the right column of [Figure 1](#)), as well as those resulting from the combination of these images within one aggregate image (i.e. containing all the previously used spatial frequency ranges).

To create stimuli, we used 56 initial images of faces and 235 initial images of natural objects. Two sets of object images and one set of face stimuli, 120 each, were formed. Objects and faces repeated in different sets contained different spatial frequencies. Photos of faces in frontal view (actually the angle is slightly different) were taken from [FERET Database](#) collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office ([Phillips *et al.*, 1998](#); [Phillips *et al.*, 2000](#)). This database was created with the consent of participants and contains photographs of men and women of different races with different emotional facial expressions. We used part of the images from the database provided to us in full accordance with Color FERET Database Release Agreement. In fact, we used the "bubbles" method ([Gosselin & Schyns, 2001](#)), yet unlike the traditional approach with the aperture located at random, the aperture of our research was placed in definite, previously pre-estimated positions which corresponded to the areas with a definite modulation value of the total nonlocal contrast.

Then the same way we formed stimuli consisting of areas with the minimum contrast modulation amplitude, as well as images consisting of areas with a modulation having the medium amplitude between the closest minimums and maximums.

Study Design. We employed a one-way design for independent samples having a three-level factor "Amplitude of

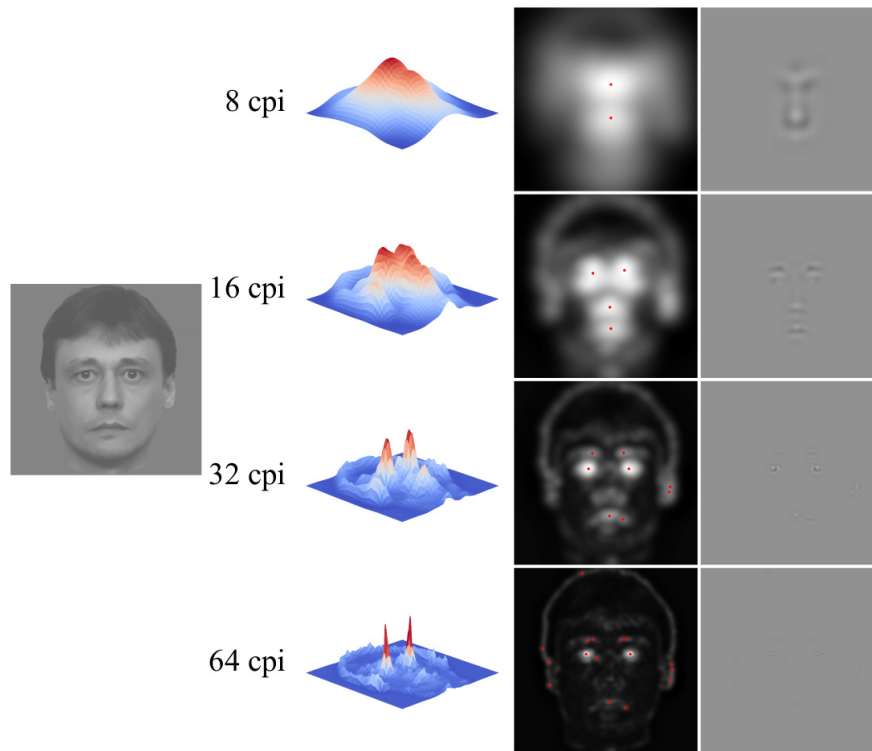


Figure 1. Algorithm of facial stimuli formation.

modulation” (min, med, max). The percentage of correct identification of facial expressions was the dependent variable. The sample size was determined based on Anova’s power = 0.8 and expected Cohen’s $f > 0.5$ effect size. The minimum expected effect size was determined based on the results of the preview of the prepared images performed by the researchers themselves.

The first group of observers (13 people) saw faces composed of areas of minima, objects of the first set composed of areas of maxima, and objects of the second set composed of intermediate areas. The second group of observers (12 people) saw faces composed of maxima, objects of the first set composed of intermediate regions, and objects of the second set composed of regions of minima. The third group (13 people) were presented with faces composed of intermediate regions, objects of the first group composed of minima regions, and objects of the second set composed of maxima regions. Faces and objects were shown mixed, the order of presentation was random.

Procedure. The observers were demonstrated synthesized images of Caucasian and Asian faces (male and female) with neutral and happy facial expressions. These randomly alternated with synthesized images of objects of different categories, the probability of faces within the chains of consequent stimuli making 33%. The observers’ task was to categorize any presented image as accurately as possible. The observer had

to inform about the appearance of a face and possibly define its expression (the answer “I don’t know” was allowed). Exposure time was not limited. The percentage of correct recognitions of facial expressions for the images formed from the areas of different contrast modulation amplitudes, was calculated.

In order to anonymize the identity of the observers, all names were encrypted by md5 algorithm, and initial raw data files were saved on the local disk storage with limited access.

Results

First, we compared task solution effectiveness where the face images had been formed from maximum nonlocal contrast areas belonging to the narrow spatial frequency range. It is worth reminding that the lesser the diameter of the areas, the higher the spatial frequency (cpi) contained in them and the greater the general number of the areas found. Where synthesized face images contained space frequencies of just one range of 1 octave, the general result of facial expression recognition was low (Figure 2). The performance was higher for the stimuli formed from the areas with the maximum increase in contrast having the central spatial frequency of 16 cpi. Somewhat lower were the values of 32 cpi frequency, and much lower these were for the lowest and the highest frequency ranges. The obtained distribution generally agrees with the data suggesting that the medium spatial frequency range expressed in cycles per face is more important in face recognition (Boutet *et al.*, 2003;

Collin *et al.*, 2006; Näsänen, 1999; Parker & Costen, 1999; Tanskanen *et al.*, 2005; see also review Ruiz-Soler & Beltran, 2006).

However, our main purpose was to test the hypothesis stating that the most informative image areas are those with the greatest increase in nonlocal contrast using the example of faces of different emotional expressions.

To answer this question, we compared the effectiveness of task solution for the faces formed from the areas of different contrast modulation amplitudes: maximum, minimum and medium (Figure 3). The stimuli were combined from the areas found in all the applied spatial frequency ranges.

It was found that in the task of identifying the facial emotional expression the result approximately improves from 5% to 61% with the increase in the modulation amplitude of the total contrast in those fragments from which the stimulus is formed (see Figure 4).

Using ANOVA (JASP software, RRID:SCR_015823) has revealed the statistical significance of the dependence obtained (see Table 1). The Levene’s test calculation showed a need to use homogeneity corrections.

The obtained effect is very high (Cohen’s $f = 1.3$). Post Hoc analysis with the application of Tukey’s test with Bonferroni and Holm’s corrections (see Table 2) also showed that accuracy with

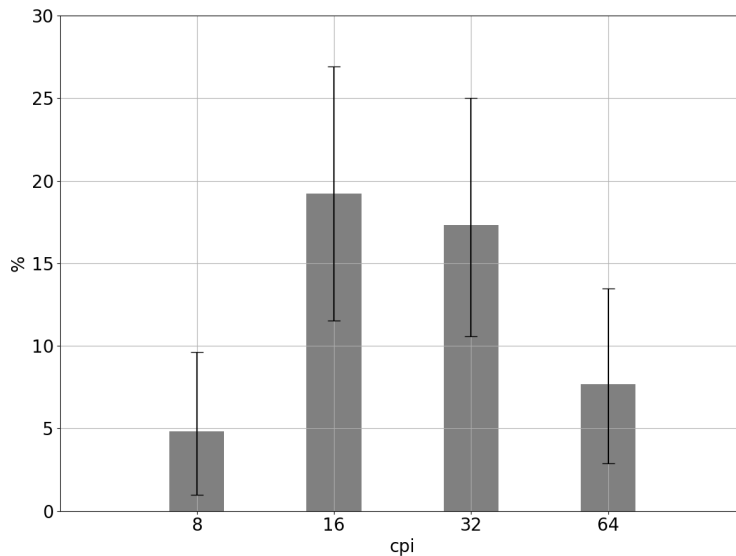


Figure 2. Comparison of the accuracy of distinguishing emotional expressions of faces collected from areas with maximum nonlocal contrast, containing different spatial frequencies. Axis X shows the central spatial frequency of the areas from which the face stimulus was synthesized. Vertical lines represent the 95% binomial confidence intervals.

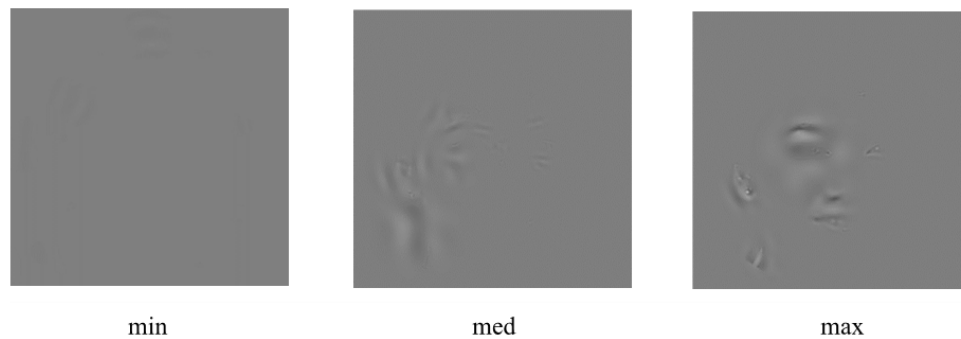


Figure 3. Examples of the face images formed from the areas of minimum (min), medium (med) and maximum (max) nonlocal contrast modulation amplitude.

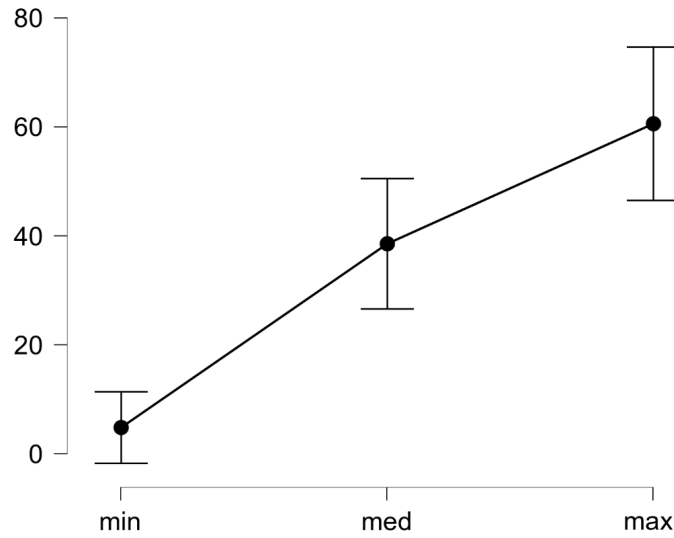


Figure 4. Dependence between the accuracy of recognizing the face emotional expression and the nonlocal contrast modulation amplitude in the areas which have made the stimulus. The abscissa shows the modulation amplitude (see the text explanations).

Table 1. Comparison of the accuracy in recognizing face emotional expressions (in the correct answers percentage) for the images with different nonlocal contrast modulation amplitude using ANOVA.

Homogeneity Correction	Cases	Sum of Squares	df	Mean Square	F	p	$\eta^2 p$
None	amplitude	20497.269	2.000	10248.635	30.332	< .001	0.634
	Residuals	11825.921	35.000	337.883			
Brown-Forsythe	amplitude	20497.269	2.000	10248.635	30.274	< .001	0.634
	Residuals	11825.921	27.843	424.740			
Welch	amplitude	20497.269	2.000	10248.635	37.026	< .001	0.634
	Residuals	11825.921	20.665	572.265			

Note. Type III Sum of Squares

Table 2. Post Hoc comparison of the accuracy of recognizing facial expressions for the images with different contrast modulation amplitudes.

		Mean Difference	SE	t	p_{tukey}	p_{bonf}	p_{holm}
max	med	22.035	7.359	2.995	0.014	0.015	0.005
max	min	55.769	7.210	7.735	< .001	< .001	< .001
med	min	33.734	7.359	4.584	< .001	< .001	< .001

Note. P-value adjusted for comparing a family of 3

which the observers recognize emotions in the faces formed from the areas of different contrast modulation amplitudes, significantly grows with the amplitude increase.

Thus the obtained results have verified our hypothesis stating that the face image areas of the greatest increase of total

nonlocal contrast contain information which can be used by the visual system in recognizing emotional expressions.

Discussion

We used the task of recognizing face emotional expressions in order to demonstrate that the areas of the greatest nonlocal

contrast modulation amplitude might possibly be the most informative ones, hence they may be used in categorizing face expressions. Meanwhile the same areas may be revealed by the second-order visual filters.

It should be noted that in recent years there have been publications of a number of model studies where the assessment of the image area aggregate energy is making the basis of the algorithm of segmenting the scenes and selecting objects from the background (Cheng *et al.*, 2011; Fang *et al.*, 2012; Perazzi *et al.*, 2012). These calculation operations demonstrate really good effectiveness, yet they generally have little in common with the true-life filters in the human visual system.

In our study we too proceeded from the assumption that space heterogeneities of the image energy might contain helpful information. Yet the most important item of our work is that we propose a definite physiological process able of detecting these areas of interest in the image. The developed gradient operator calculating the nonlocal contrast modulation amplitude imitates the functioning of the second-order visual filters with different spatial-frequency tunings. Moreover, we tried to maximally approximate these operators' parameters to the well-known characteristics of the second-order filters. Thus, for example the spatial frequency (in cycles per "window") passed from the extracted areas is constant for the "windows" of all the used sizes. This emphasizes the presence of a fixed ratio of the frequency tunings of the first- and second-order filters (Dakin & Mareschal, 2000; Sutter *et al.*, 1995) and thus ensures the invariance of the description when changing the scale. We have also used a "window" resizing step which provides a change step in the spatial frequency passed by the "windows", this step equaling 1 octave, which roughly corresponds to the step in the change of the spatial-frequency tuning of pathways in the human visual system (Wilson & Gelb, 1984). The bandpass of the second-order filters also corresponds to the given bandwidth of our operator and is equal to 1 octave (Landy & Henry, 2007). We have used the Gaussian envelope in passing the extracted image area, thus imitating the spatial characteristics of the filters at the human visual system input. We have defined that a "window" transmits namely four cycles of the input image. This value is also based on the previously obtained results (Babenko *et al.*, 2010).

At the same time there were parameters whose optimality remains doubtful to us. So, for example, the number of identified areas grows exponentially in cases where the operator's size reduces, this chain starting from two "windows". We have proceeded from the requirement that the total diameter of the identified areas should be equal to the diameter of the whole image. In this case the spatial frequency of the synthesized face may be easily calculated in cycles per image. However, in reality there might be some other number of areas identified at each frequency that is optimal. No doubt, increase in their number will lead to an improved result. Neither did we introduce eccentricity correction since we assumed that in natural conditions saliency maps may also be formed by the human visual system with the use of eye movements. However, the data

concerning the time of facial expression perception might indicate that one fixation is sufficient for this (L. Liu & Ioannides, 2010; Pourtois *et al.*, 2010; see also reviews George, 2013; Vuilleumier & Pourtois, 2007), although another opinion also exists (Duncan *et al.*, 2019; Eimer & Holmes, 2007; Eimer *et al.*, 2003; Erthal *et al.*, 2005; Okon-Singer *et al.*, 2007; Pessoa *et al.*, 2002).

Nevertheless, it is impossible to take into account every parameter of the processes providing search for areas of interest in the image and can hardly put into question the conclusion that the information content of the facial image reflecting its emotional expression increases with the growth of the nonlocal contrast amplitude of areas which form this image.

It is also noteworthy that the areas of a maximum nonlocal contrast amplitude can generally be found specifically around the eyes and the mouth (see Figure 1 and Figure 3), i.e. those parts of the face that are considered to be most informative in conveying emotional signals (Bombari *et al.*, 2013; Eisenbarth & Alpers, 2011; Yu *et al.*, 2018).

However, another question arises. Nonlocal luminance contrast has been effective in the task of discriminating facial expressions, but will it be a salience feature in another tasks, such as gender or race determination, for example? And what about non-facial image recognition? The answer to this question should be given by future experiments. However, since we consider the process of finding areas with the largest increase in nonlocal contrast as preattentive, its result should not depend on the visual task. Preattentive processing only offers a set of image areas, information from which can be used by higher processing levels with the help of attention. The strongest saliencies automatically attract 2–3 initial fixations. These few hundred milliseconds are enough to recognize an emotional facial expression (Du & Martinez, 2013). Subsequent top-down attention allows for more information.

Conclusions

The obtained experimental results have supported the hypothesis stating that the image areas of the greatest increase in the nonlocal contrast contain information that contributes to the identification of emotional facial expressions. The second-order visual filters are able to find such information.

We also suppose that the second-order visual filters that highlight the image areas with the highest modulation amplitude of nonlocal contrast are able to attract visual spatial attention; these filters are the windows through which subsequent processing levels receive significant information.

Data availability

Underlying data

Open Science Framework: Nonlocal contrast calculated by the second order visual mechanisms and its significance in identifying facial emotions, <https://doi.org/10.17605/OSF.IO/KGRWA> (Yavna, 2021).

This project contains the following underlying data:

- emotions.csv contains main data,
- emotions.jasp contains main statistics,
- raw_results folder contains raw anonymized response logs,
- calc_result2.py and give_me_res.sh from the raw_results folder are scripts for processing response logs (*.json) and creating faces.csv file,
- faces.jasp contains statistics of emotion recognition at all frequencies,
- stimuli.tar.bz contains all the stimuli used in the study

- readme.txt contains some additional information and comments

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

Software availability

Source code available from: https://github.com/dvyavna/2ord_contrast

Archived source code as at time of publication: <https://doi.org/10.17605/OSF.IO/5YZGW> (Yavna, 2021).

License: MIT

References

- Atkinson AP, Smithson HE: **The Impact on Emotion Classification Performance and Gaze Behavior of Foveal versus Extrafoveal Processing of Facial Features.** *J Exp Psychol Hum Percept Perform.* 2020; **46**(3): 292–312. [PubMed Abstract](#) | [Publisher Full Text](#)
- Babenko VV, Ermakov PN: **Specificity of Brain Reactions to Second-Order Visual Stimuli.** *Vis Neurosci.* 2015; **32**: E011. [PubMed Abstract](#) | [Publisher Full Text](#)
- Babenko VV, Ermakov PN, Bozhinskaya MA: **Relationship between the Spatial-Frequency Tunings of the First- and the Second-Order Visual Filters.** *Psikhol Zh.* 2010; **31**(2): 48–57.
- Babenko VV, Yavna DV, Rodionov EG: **Contributions of Different Spatial Modulations of Brightness Gradients to the Control of Visual Attention.** *Neurosci Behav Physiol.* 2020; **50**(8): 1035–42. [Publisher Full Text](#)
- Baldi P, Itti L: **Of Bits and Wows: A Bayesian Theory of Surprise with Applications to Attention.** *Neural Netw.* 2010; **23**(5): 649–66. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Blais C, Fiset D, Roy C, et al.: **Eye Fixation Patterns for Categorizing Static and Dynamic Facial Expressions.** *Emotion.* 2017; **17**(7): 1107–19. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bombardi D, Schmid PC, Mast MS, et al.: **Emotion Recognition: The Role of Featural and Configural Face Information.** *Q J Exp Psychol (Hove).* 2013; **66**(12): 2426–42. [PubMed Abstract](#) | [Publisher Full Text](#)
- Borji A, Sihite DN, Itti L: **Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study.** *IEEE Trans Image Process.* 2013; **22**(1): 55–69. [PubMed Abstract](#) | [Publisher Full Text](#)
- Boutet I, Collin C, Faubert J: **Configural Face Encoding and Spatial Frequency Information.** *Percept Psychophys.* 2003; **65**(7): 1078–93. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bruce NDB, Tsotsos JK: **Saliency Based on Information Maximization.** In: *NIPS.* 2005. [Reference Source](#)
- Bruce NDB, Tsotsos JK: **Saliency, Attention, and Visual Search: An Information Theoretic Approach.** *J Vis.* 2009; **9**(3): 5.1–24. [PubMed Abstract](#) | [Publisher Full Text](#)
- Calder AJ, Rowland D, Young AW, et al.: **Caricaturing Facial Expressions.** *Cognition.* 2000; **76**(2): 105–46. [PubMed Abstract](#) | [Publisher Full Text](#)
- Calvo MG, Fernández-Martín A, Nummenmaa L: **Facial Expression Recognition in Peripheral versus Central Vision: Role of the Eyes and the Mouth.** *Psychol Res.* 2014; **78**(2): 180–95. [PubMed Abstract](#) | [Publisher Full Text](#)
- Cauchoix M, Barragan-Jason G, Serre T, et al.: **The Neural Dynamics of Face Detection in the Wild Revealed by MVPA.** *J Neurosci.* 2014; **34**(3): 846–54. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng MM, Zhang GX, Mitra NJ, et al.: **Global Contrast Based Salient Region Detection.** *2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011.* 2011; 409–16. [Publisher Full Text](#)
- Collin CA, Therrien M, Martin C, et al.: **Spatial Frequency Thresholds for Face Recognition When Comparison Faces Are Filtered and Unfiltered.** *Percept Psychophys.* 2006; **68**(6): 879–89. [PubMed Abstract](#) | [Publisher Full Text](#)
- Crouzet SM, Kirchner H, Thorpe SJ: **Fast Saccades toward Faces: Face Detection in Just 100 Ms.** *J Vis.* 2010; **10**(4): 16.1–17. [PubMed Abstract](#) | [Publisher Full Text](#)
- Crouzet SM, Thorpe SJ: **Low-Level Cues and Ultra-Fast Face Detection.** *Front Psychol.* 2011; **2**: 342. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dakin SC, Mareschal I: **Sensitivity to Contrast Modulation Depends on Carrier Spatial Frequency and Orientation.** *Vision Res.* 2000; **40**(3): 311–29. [PubMed Abstract](#) | [Publisher Full Text](#)
- Dering B, Martin CD, Moro S, et al.: **Face-Sensitive Processes One Hundred Milliseconds after Picture Onset.** *Front Hum Neurosci.* 2011; **5**: 93. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Du S, Martinez AM: **Wait, are you sad or angry? Large exposure time differences required for the categorization of facial expressions of emotion.** *J Vis.* 2013; **13**(4): 13. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Duncan J, Dugas G, Brisson B, et al.: **Dual-Task Interference on Left Eye Utilization during Facial Emotion Perception.** *J Exp Psychol Hum Percept Perform.* 2019; **45**(10): 1319–1330. [PubMed Abstract](#) | [Publisher Full Text](#)
- Duncan J, Gosselin F, Cobarro C, et al.: **Orientations for the Successful Categorization of Facial Expressions and Their Link with Facial Features.** *J Vis.* 2017; **17**(14): 7. [PubMed Abstract](#) | [Publisher Full Text](#)
- Eimer M, Holmes A: **Event-Related Brain Potential Correlates of Emotional Face Processing.** *Neuropsychologia.* 2007; **45**(1): 15–31. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Eimer M, Holmes A, McGlone FP: **The Role of Spatial Attention in the Processing of Facial Expression: An ERP Study of Rapid Brain Responses to Six Basic Emotions.** *Cogn Affect Behav Neurosci.* 2003; **3**(2): 97–110. [PubMed Abstract](#) | [Publisher Full Text](#)
- Eisenbarth H, Alpers GW: **Happy Mouth and Sad Eyes: Scanning Emotional Facial Expressions.** *Emotion.* 2011; **11**(4): 860–65. [PubMed Abstract](#) | [Publisher Full Text](#)
- Entzmann L, Guyader N, Kauffmann L, et al.: **The Role of Emotional Content and Perceptual Saliency During the Programming of Saccades Toward Faces.** *Cogn Sci.* 2021; **45**(10): e13042. [PubMed Abstract](#) | [Publisher Full Text](#)
- Erthal FS, de Oliveira L, Mocaiber I, et al.: **Load-Dependent Modulation of Affective Picture Processing.** *Cogn Affect Behav Neurosci.* 2005; **5**(4): 388–95. [PubMed Abstract](#) | [Publisher Full Text](#)

- Fang Y, Lin W, Lee BS, et al.: **Bottom-Up Saliency Detection Model Based on Human Visual Sensitivity and Amplitude Spectrum.** *IEEE Trans Multimedia.* 2012; **14**(1): 187–98.
[Publisher Full Text](#)
- Fiset D, Blais C, Royer J, et al.: **Mapping the Impairment in Decoding Static Facial Expressions of Emotion in Prosopagnosia.** *Soc Cogn Affect Neurosci.* 2017; **12**(8): 1334–41.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gao D, Vasconcelos N: **Bottom-up Saliency Is a Discriminant Process.** In: *2007 IEEE 11th International Conference on Computer Vision.* 2007; 1–6.
[Publisher Full Text](#)
- George N: **The Facial Expression of Emotions.** In *The Cambridge Handbook of Human Affective Neuroscience.* edited by J. Armony and P. Vuilleumier. Cambridge: Cambridge University Press. 2013; 171–97.
[Reference Source](#)
- Gosselin F, Schyns PG: **Bubbles: a technique to reveal the use of information in recognition tasks.** *Vision Res.* 2001; **41**(17): 2261–71.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Graham NV: **Beyond Multiple Pattern Analyzers Modeled as Linear Filters (as Classical V1 Simple Cells): Useful Additions of the Last 25 Years.** *Vision Res.* 2011; **51**(13): 1397–1430.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Graham N, Sutter A: **Normalization: Contrast-Gain Control in Simple (Fourier) and Complex (Non-Fourier) Pathways of Pattern Vision.** *Vision Res.* 2000; **40**(20): 2737–61.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Graham N, Wolfson SS: **Is There Opponent-Orientation Coding in the Second-Order Channels of Pattern Vision?** *Vision Res.* 2004; **44**(27): 3145–75.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Herrmann MJ, Ehlis AC, Ellgring H, et al.: **Early Stages (P100) of Face Perception in Humans as Measured with Event-Related Potentials (ERPs).** *J Neural Transm (Vienna).* 2005; **112**(8): 1073–81.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hou W, Gao X, Tao D, et al.: **Visual Saliency Detection Using Information Divergence.** *Pattern Recognit.* 2013; **46**(10): 2658–69.
[Publisher Full Text](#)
- Hubel DH, Wiesel TN: **Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex.** *J Physiol.* 1962; **160**(1): 106–54.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Itti L, Baldi P: **Bayesian Surprise Attracts Human Attention.** *Vision Res.* 2009; **49**(10): 1295–1306.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Itti L, Koch C, Niebur E: **A Model of Saliency-Based Visual Attention for Rapid Scene Analysis.** *IEEE Trans Pattern Anal Mach Intell.* 1998; **20**(11): 1254–59.
[Publisher Full Text](#)
- Jack RE, Garrod OGB, Schyns PG: **Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time.** *Curr Biol.* 2014; **24**(2): 187–92.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Judd T, Durand F, Torralba A: **A Benchmark of Computational Models of Saliency to Predict Human Fixations.** *Undefined.* 2012; Retrieved December 10, 2020 (/paper/A-Benchmark-of-Computational-Models-of-Saliency-to-Judd-Durand/daef3f3dc4190927c063ae94c12437cf82a6d1c20).
[Reference Source](#)
- Kingdom FAA, Prins N, Hayes A: **Mechanism Independence for Texture-Modulation Detection Is Consistent with a Filter-Rectify-Filter Mechanism.** *Vis Neurosci.* 2003; **20**(1): 65–76.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kirchner H, Thorpe SJ: **Ultra-Rapid Object Detection with Saccadic Eye Movements: Visual Processing Speed Revisited.** *Vision Res.* 2006; **46**(11): 1762–76.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kovarski K, Latinus M, Charpentier J, et al.: **Facial Expression Related vMMN: Disentangling Emotional from Neutral Change Detection.** *Front Hum Neurosci.* 2017; **11**: 18.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Landy MS, Henry CA: **Critical-Band Masking Estimation of 2nd-Order Filter Properties.** *Perception.* 2007; **36**(Suppl.): 61.
[Reference Source](#)
- Larsson J, Landy MS, Heeger DJ: **Orientation-selective adaptation to first- and second-order patterns in human visual cortex.** *J Neurophysiol.* 2006; **95**(2): 862–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liu H, Agam Y, Madsen JR, et al.: **Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex.** *Neuron.* 2009; **62**(2): 281–90.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liu TT, Fu JZ, Chai Y, et al.: **Layer-specific, retinotopically-diffuse modulation in human visual cortex in response to viewing emotionally expressive faces.** *Nat Commun.* 2022; **13**(1): 6302.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liu J, Harris A, Kanwisher N: **Stages of Processing in Face Perception: An MEG Study.** *Nat Neurosci.* 2002; **5**(9): 910–16.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Liu L, Ioannides AA: **Emotion Separation Is Completed Early and It Depends on Visual Field Presentation.** *PLoS One.* 2010; **5**(3): e9790.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Muukkonen I, Ölander K, Numminen J, et al.: **Spatio-temporal dynamics of face perception.** *NeuroImage.* 2020; **209**: 116531.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Näsänen R: **Spatial Frequency Bandwidth Used in the Recognition of Facial Images.** *Vision Res.* 1999; **39**(23): 3824–33.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Okon-Singer H, Tzelgov J, Henik A: **Distinguishing between Automaticity and Attention in the Processing of Emotionally Significant Stimuli.** *Emotion.* 2007; **7**(1): 147–57.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Parker DM, Costen NP: **One Extreme or the Other or Perhaps the Golden Mean? Issues of Spatial Resolution in Face Processing.** *Current Psychology.* 1999; **18**(1): 118–27.
[Publisher Full Text](#)
- Perazzi F, Krähenbühl P, Pritch Y, et al.: **Saliency Filters: Contrast Based Filtering for Salient Region Detection.** In: *2012 IEEE Conference on Computer Vision and Pattern Recognition.* 2012; 733–40.
[Publisher Full Text](#)
- Pessoa L, McKenna M, Gutierrez E, et al.: **Neural Processing of Emotional Faces Requires Attention.** *Proc Natl Acad Sci U S A.* 2002; **99**(17): 11458–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Phillips PJ, Moon H, Rizvi SA, et al.: **The FERET Evaluation Methodology for Face-Recognition Algorithms.** *IEEE Trans Pattern Anal Mach Intell.* 2000; **22**(10): 1090–1104.
[Publisher Full Text](#)
- Phillips PJ, Wechsler H, Huang J, et al.: **The FERET Database and Evaluation Procedure for Face-Recognition Algorithms.** *Image Vis Comput.* 1998; **16**(5): 295–306.
[Publisher Full Text](#)
- Pourtois G, Spinelli L, Seeck M, et al.: **Temporal Precedence of Emotion over Attention Modulations in the Lateral Amygdala: Intracranial ERP Evidence from a Patient with Temporal Lobe Epilepsy.** *Cogn Affect Behav Neurosci.* 2010; **10**(1): 83–93.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rahman S, Rochan M, Wang Y, et al.: **Examining Visual Saliency Prediction in Naturalistic Scenes.** In: *2014 IEEE International Conference on Image Processing (ICIP).* 2014; 4082–86.
[Publisher Full Text](#)
- Reddy L, Reddy L, Koch C: **Face Identification in the Near-Absence of Focal Attention.** *Vision Res.* 2006; **46**(15): 2336–43.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Reddy L, Wilken P, Koch C: **Face-Gender Discrimination Is Possible in the near-Absence of Attention.** *J Vis.* 2004; **4**(2): 106–17.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ruiz-Soler M, Beltran FS: **Face Perception: An Integrative Review of the Role of Spatial Frequencies.** *Psychol Res.* 2006; **70**(4): 273–92.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ruiz-Soler M, Beltran FS: **The Relative Saliency of Facial Features When Differentiating Faces Based on an Interference Paradigm.** *J Nonverbal Behav.* 2012; **36**(3): 191–203.
[Publisher Full Text](#)
- Schindler S, Bublatzky F: **Attention and emotion: An integrative review of emotional face processing as a function of attention.** *Cortex.* 2020; **130**: 362–386.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schofield AJ, Yates TA: **Interactions between Orientation and Contrast Modulations Suggest Limited Cross-Cue Linkage.** *Perception.* 2005; **34**(7): 769–92.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schyns PG, Oliva A: **Dr. Angry and Mr. Smile: When Categorization Flexibly Modifies the Perception of Faces in Rapid Visual Presentations.** *Cognition.* 1999; **69**(3): 243–65.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smith ML, Cottrell GW, Gosselin F, et al.: **Transmitting and Decoding Facial Expressions.** *Psychol Sci.* 2005; **16**(3): 184–89.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smith ML, Merlusa C: **How Task Shapes the Use of Information during Facial Expression Categorizations.** *Emotion.* 2014; **14**(3): 478–87.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smith FW, Schyns PG: **Smile through Your Fear and Sadness: Transmitting and Identifying Facial Expression Signals over a Range of Viewing Distances.** *Psychol Sci.* 2009; **20**(10): 1202–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sutter A, Sperling G, Chubb C: **Measuring the Spatial Frequency Selectivity of Second-Order Texture Mechanisms.** *Vision Res.* 1995; **35**(7): 915–24.
[PubMed Abstract](#) | [Publisher Full Text](#)

Tanaka JW, Kaiser MD, Butler S, *et al.*: **Mixed Emotions: Holistic and Analytic Perception of Facial Expressions.** *Cogn Emot.* 2012; **26**(6): 961–77.
[PubMed Abstract](#) | [Publisher Full Text](#)

Tanskanen T, Näsänen R, Montez T, *et al.*: **Face Recognition and Cortical Responses Show Similar Sensitivity to Noise Spatial Frequency.** *Cereb Cortex.* 2005; **15**(5): 526–34.
[PubMed Abstract](#) | [Publisher Full Text](#)

Tomasik D, Ruthruff E, Allen PA, *et al.*: **Nonautomatic emotion perception in a dual-task situation.** *Psychon Bull Rev.* 2009; **16**(2): 282–8.
[PubMed Abstract](#) | [Publisher Full Text](#)

VanRullen R: **On Second Glance: Still No High-Level Pop-out Effect for Faces.** *Vision Res.* 2006; **46**(18): 3017–27.
[PubMed Abstract](#) | [Publisher Full Text](#)

Vuilleumier P: **Faces Call for Attention: Evidence from Patients with Visual Extinction.** *Neuropsychologia.* 2000; **38**(5): 693–700.
[PubMed Abstract](#) | [Publisher Full Text](#)

Vuilleumier P: **Facial Expression and Selective Attention.** *Curr Opin Psychiatry.* 2002; **15**(3): 291–300.
[Reference Source](#)

Vuilleumier P, Pourtois G: **Distributed and Interactive Brain Mechanisms**

during Emotion Face Perception: Evidence from Functional Neuroimaging. *Neuropsychologia.* 2007; **45**(1): 174–94.
[PubMed Abstract](#) | [Publisher Full Text](#)

Wang HF, Friel N, Gosselin F, *et al.*: **Efficient Bubbles for Visual Categorization Tasks.** *Vision Res.* 2011; **51**(12): 1318–23.
[PubMed Abstract](#) | [Publisher Full Text](#)

White M: **Parts and Wholes in Expression Recognition.** *Cogn Emot.* 2000; **14**(1): 39–60.
[Publisher Full Text](#)

Wilson HR, Gelb DJ: **Modified Line-Element Theory for Spatial-Frequency and Width Discrimination.** *J Opt Soc Am A.* 1984; **1**(1): 124–31.
[PubMed Abstract](#) | [Publisher Full Text](#)

Yarbus AL: **Eye Movements and Vision.** New York: Plenum Press, 1967.
[Publisher Full Text](#)

Yavna D: **Nonlocal contrast calculated by the second order visual mechanisms and its significance in identifying facial emotions.** 2021.
<http://www.doi.org/10.17605/OSF.IO/5YZGW>

Yu D, Chai A, Chung STL: **Orientation Information in Encoding Facial Expressions.** *Vision Res.* 2018; **150**: 29–37.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 08 September 2023

<https://doi.org/10.5256/f1000research.153731.r201230>

© 2023 Liu 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. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.



Tina Tong Liu

National Institutes of Health, Bethesda, USA

The authors made an effort to revise the manuscript to enhance readability and to promote openness and transparency.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Vision, face recognition, emotional processing, primary visual cortex, neuroplasticity

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.

Version 1

Reviewer Report 20 June 2023

<https://doi.org/10.5256/f1000research.31420.r179353>

© 2023 Liu 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. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.



Tina Tong Liu

National Institutes of Health, Bethesda, USA

In this manuscript, the authors aim to identify a mechanism in human visual perception that would be selective for facial features/attributes, which are known to contribute to facial expression recognition. The authors hypothesized that the most informative areas of an image are those characterized by spatial heterogeneity, particularly with nonlocal contrast changes. To test this hypothesis, the authors first developed a software program to create three variants of a single face, each with maximum, minimum, and medium contrast modulation amplitudes. Then, the authors applied the "bubbles" method and ask the participant to judge the emotional expressions of faces. The study found that the greater the contrast modulation amplitude of the areas shaping the face, the more accurately participants were able to identify the emotion, revealing that nonlocal contrast can be a diagnostic feature for facial emotion recognition.

While this manuscript addresses a very interesting question, it entails a number of noticeable limitations regarding the 6 review criteria:

1. Is the work clearly and accurately presented and does it cite the current literature?

1.1 I would like to suggest that the authors pay closer attention to citation accuracy.

For instance, the sentence below could be revised to reflect the appropriate techniques used in the cited studies:

"Experiments involving a saccadic task (Crouzet *et al.*, 2010; Kirchner & Thorpe, 2006) and registration of event-related potentials (Cauchoux *et al.*, 2014; Dering *et al.*, 2011; Herrmann *et al.*, 2005; Liu *et al.*, 2002; Liu *et al.*, 2009; Pitcher *et al.*, 2007)..."

Instead of measuring ERP, Liu *et al.* (2002) used MEG, Pitcher *et al.* (2007) used TMS, and Liu *et al.* (2009) specifically measured intracranial field potentials, not event-related potentials.

Another example:

"Indeed, there are specialized mechanisms for finding brightness gradients in the visual system, and these are striate neurons (Hubel & Wiesel, 1962; Hubel & Wiesel, 1968)."

These two seminal papers focused on the preferences of V1 neurons for orientation, direction of movement, and spatial frequency, not brightness gradients. It is important to note that V1 neurons do not "find" sinusoidal gratings in the visual scene. The central concept revolves around receptive fields—specific regions in visual space that have an impact on the activity of V1 neurons.

1.2 The authors may need to exercise caution when making negative claims that can be readily disproven.

For example, "The problem is that the lower levels of the human visual system lack neurons which would be selective to certain facial features" may not hold true based on recent evidence. Recent evidence show that human primary visual cortex is sensitive to emotional expressions of faces (Bo *et al.*, 2021¹, Liu *et al.*, 2022²).

Also, "It is clear that the evenly lit areas are of no interest to the visual system." The authors should be aware of the existence of multiple ON and OFF regions within receptive fields, before making this claim.

1.3 I would encourage the authors to provide a more balanced view of the literature.

The claim that face processing is pre-attentive is not universally accepted. It is crucial for the authors to acknowledge the contradictory nature of prior studies regarding the topic of fast face detection and pre-attentive processing. Please include citation of Pessoa *et al.* (2002)³ for a balanced view of the literature.

1.4 The authors should aim to provide a clearer description of the second-order visual processes as proposed by Graham (2011), rather than simply referring to it as a "mechanism." "Mechanism" is a very strong word, meaning the underlying processes, principles, or components that are responsible for this phenomenon. Instead of using the term "second-order visual mechanism," it would be more appropriate to adhere to Graham's (2011) terminology and refer to it as "second-order processes" or "higher-order processes."

2. Is the study design appropriate and does the work have academic merit?

Based on the information provided, it seems that the stimuli used in the study consisted of 40 initial images of faces and 240 initial images of natural objects. However, it is unclear whether the same task of judging facial expressions was applied to both types of stimuli (faces and natural objects).

The authors may need to address the potential limitations regarding the limited number of emotional expressions used in the study. With only two expressions (neutral and happy) presented throughout the experiment, there is a possibility that participants may have identified the diagnostic features associated with these expressions.

The authors described all the faces have a frontal view, but the sample face using the bubbles procedure in Figure 3 does not have a frontal view.

3. Are sufficient details of methods and analysis provided to allow replication by others?

The authors may wish to consider breaking down the Method section into subsections to enhance the readability and organization of the manuscript, such as Participants, Study Design, Stimuli, Procedure, etc.

4. If applicable, is the statistical analysis and its interpretation appropriate?

Without detailed information about the specific experimental design, it is challenging to evaluate the appropriateness of using a one-way ANOVA versus a repeated-measures ANOVA.

There is no standard deviation or error in Figure 2. Simply presenting the mean is not acceptable.

5. Are all the source data underlying the results available to ensure full reproducibility?

I appreciate the authors' efforts to promote openness and transparency, However, to ensure a comprehensive replication of the study, it is important for the authors to provide access to all the processed stimuli, rather than just a limited number of sample images on <https://osf.io/5ygzgw/files/osfstorage>. The emotions.csv file provided in the OSF repository is not sufficient or helpful for replication purposes either. It is recommended that the authors provide PII-stripped (Personally Identifiable Information), anonymized raw data in the OSF registry.

6. Are the conclusions drawn adequately supported by the results?

It is misleading to say "confirmed the hypothesis" or "proved the statistical significance". In the context of null hypothesis testing framework, the stats only allow the authors to reject the null hypothesis.

I am not sure if I understand the scope of the conclusions. Is the importance of nonlocal contrast is limited only to facial emotion recognition or if it extends to other aspects such as general face recognition or object recognition in a broader sense?

References

1. Bo K, Yin S, Liu Y, Hu Z, et al.: Decoding Neural Representations of Affective Scenes in Retinotopic Visual Cortex. *Cereb Cortex*. 2021; **31** (6): 3047-3063 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Liu TT, Fu JZ, Chai Y, Japee S, et al.: Layer-specific, retinotopically-diffuse modulation in human visual cortex in response to viewing emotionally expressive faces. *Nat Commun*. 2022; **13** (1): 6302 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Pessoa L, McKenna M, Gutierrez E, Ungerleider LG: Neural processing of emotional faces requires attention. *Proc Natl Acad Sci U S A*. 2002; **99** (17): 11458-63 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Vision, face recognition, emotional processing, primary visual cortex, neuroplasticity

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 28 Jul 2023

Denis Yavna

Dear Reviewer

We sincerely thank you for your attention to our manuscript and your valuable comments. We tried to take into account all of them and made changes to the text. Let me also make some comments.

1.1 I would like to suggest that the authors pay closer attention to citation accuracy.

Thank you, your comments have been taken into account. The only thing we have left unchanged is the reference (Hubel & Wiesel, 1962).

You are correct that "These two seminal papers focused on the preferences of V1 neurons for orientation, direction of movement, and spatial frequency..." However, it is in the article by Hubel & Wiesel (1962) that the authors first state: "The most effective stimulus configurations, dictated by the spatial arrangements of excitatory and inhibitory regions, were long narrow rectangles of light (slits), straight-line borders between areas of different brightness (edges), and dark rectangular bars against a light background". This is a direct indication that striate neurons respond most strongly to brightness gradients (bands or edges).

1.2 The authors may need to exercise caution when making negative claims that can be readily disproven.

We have included citations of relevant papers in the text. However, we are not sure that we have the right to insert citations of papers published after our manuscript was submitted. We also clarified what we mean by "lower levels of the human visual system" in the text. As for the On- and Off-reactions of the visual neurons, of course they take place. However, in the context of the problem of saliency, these reactions are not considered. At the same time, turning on or off the light stimulus can also attract the observer's attention. Therefore, we have made an appropriate addition to the text.

1.3 I would encourage the authors to provide a more balanced view of the literature.

Indeed, "The claim that face processing is pre-attentive is not universally accepted." We have added links.

1.4 The authors should aim to provide a clearer description of the second-order visual processes

We have included in the text a more detailed description of the second-order filter model. As for the term "mechanism", you correct: "Mechanism" is a very strong word." Where possible, we have replaced the word "mechanism".

2. Is the study design appropriate and does the work have academic merit?

A) it is unclear whether the same task of judging facial expressions was applied to both types of stimuli (faces and natural objects).

We have included in the text an explanation of how the subjects' task was formulated.

B) there is a possibility that participants may have identified the diagnostic features associated with these expressions.

We agree with you that such a possibility existed. Therefore, we tried to minimize it. To do this, the subjects had to categorize all presented stimuli (not just faces). Faces appeared much less frequently than non-faces and differed, in addition to expression, in sex, race, and frequency content. The subjects were not warned in advance which facial expressions would be used. They themselves had to determine the facial expression if they detected a face in a series of other stimuli.

C) The authors described all the faces have a frontal view, but the sample face using the bubbles procedure in Figure 3 does not have a frontal view.

We used face images from a database that labeled them as frontal. However, there were some variations of the angle. We considered this to be a positive factor, since it introduced additional variability into the incentives.

3. Are sufficient details of methods and analysis provided to allow replication by others?

Based on your recommendation, we have structured the "Method" section into subsections.

4. If applicable, is the statistical analysis and its interpretation appropriate?

We have made appropriate corrections to the text.

5. Are all the source data underlying the results available to ensure full reproducibility?

Based on your recommendation, we have provided access to all source data.

6. Are the conclusions drawn adequately supported by the results?

We have made corrections in accordance with your comments.

Once again, we sincerely thank you for your attention to our manuscript and for your valuable comments.

Competing Interests: No competing interests were disclosed.

Reviewer Report 06 May 2021

<https://doi.org/10.5256/f1000research.31420.r82785>

© 2021 Shelepin Y. 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.



Yuri E. Shelepin

Pavlov Institute of Physiology, Russian Academy of Sciences, St. Petersburg, Russian Federation

This article solves the problem of the most informative areas of the face images. Authors argue that these areas are in the greatest increase of non-local contrast. A model of the second-order visual mechanism is used, with the help of which the image areas with the highest, lowest and intermediate amplitude of the total contrast modulation are extracted. The answers of the observers in the task of distinguishing the emotional expressions in face images created from areas with different modulation amplitudes are analyzed.

Evidence is given that the second-order visual filters can play the role of preattentive operators, highlighting the most informative image areas. Overall, the finding of the current study is important and interesting, and analysis is reliable. However, I would like to highlight two problems:

1. The second-order visual mechanisms should be described in the Introduction in more detail, since the journal has a broader audience.
2. The authors used the task of distinguishing facial expressions and obtained a result confirming the informational significance of the image areas with the highest total brightness contrast. However, different information is useful for different visual tasks. At the same time, are the preattentive mechanisms specific to the visual task? Is the proposed algorithm for finding the interest areas effective in the task of discriminating emotions and will it be equally effective in other visual tasks? It would be useful to address this issue in the Discussion.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Vision perception and pattern recognition

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.

Author Response 12 May 2021

Denis Yavna

Thank you very much for your attention to our study and its positive assessment. We will add necessary information to the Introduction and Discussion.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research