Original scientific paper

Received: November, 15.2022.
Revised: November, 30.2022.
Accepted: December, 04.2022.

UDK:
159.937.5.072
159.931.072
10.23947/2334-8496-2022-10-3-37-51

Check for updates

Recognition of Facial Expressions Based on Information From the Areas of Highest Increase in Luminance Contrast

Vitali Babenko[1*], *Daria Alekseeva*[1], *Denis Yavna*[1], *Ekaterina Denisova*[2], *Ekaterina Kovsh*[1], *Pavel Ermakov*[1]

[1]Southern Federal University, Rostov-on-Don, Russian Federation,
e-mail: babenko@sfedu.ru, alexeeva_ds@mail.ru, yavna@fortran.su, katya-kovsh@yandex.ru, paver@sfedu.ru
[2]Don State Technical University, Rostov-on-Don, Russian Federation, e-mail: denisovakeith@gmail.com

Abstract: It is generally accepted that the use of the most informative areas of the input image significantly optimizes visual processing. Several authors agree that, the areas of spatial heterogeneity are the most interesting for the visual system and the degree of difference between those areas and their surroundings determine the saliency. The purpose of our study was to test the hy-pothesis that the most informative are the areas of the image of largest increase in total luminance contrast, and information from these areas is used in the process of categorization facial expressions. Using our own program that was developed to imitate the work of second-order visual mechanisms, we created stimuli from the initial photographic images of faces with 6 basic emotions and a neutral expression. These images consisted only of areas of highest increase in total luminance contrast. Initially, we determined the spatial frequency ranges in which the selected areas contain the most useful information for the recognition of each of the expressions. We then compared the expressions recognition accuracy in images of real faces and those synthe-sized from the areas of highest contrast increase. The obtained results indicate that the recognition of expressions in synthe-sized images is somewhat worse than in real ones (73% versus 83%). At the same time, the partial loss of information that oc-curs due to the replacing real and synthesized images does not disrupt the overall logic of the recognition. Possible ways to make up for the missing information in the synthesized images are suggested.

Keywords: expression recognition, saliency, total luminance contrast, second-order visual filters.

Introduction

It is obvious that different image areas contain different volume of information. Classical experiments of A. Yarbus (Yarbus, 2013) have made it possible to see that the eyes ignore homogeneous areas of the image and, on the contrary, the gaze is directed to the most heterogeneous areas.

Starting from the early levels of visual processing, neurons respond precisely to heterogeneities. So, striate neurons are activated by luminance heterogeneity in their receptive fields (Marat et al., 2013). However, single luminance gradients are only local heterogeneities. When it comes to the perception of scenes or objects, salient regions have significant spatial extent. In this case, the heterogeneity is spatial modulation of luminance gradients (changes in their contrast, orientation, or spatial frequency).

The optimization of the visual perception implies finding and processing the most informative parts of the input image. A number of authors have posited that the areas that differs most from the surroundings are of the greatest interest to the visual system and attract the attention of the observer (Bruce and Tsotsos, 2009; Marat et al., 2013; Perazzi et al., 2012; Xia et al., 2015). Perhaps, mental representations of complex visual stimuli are formed by the information from these areas. The importance of finding the areas of interest determines a large number of studies aimed at finding an algorithm for identifying them and constructing saliency maps. However, a significant part of proposed saliency detection algorithms often is not based on nor considers real brain mechanisms of visual perception (Cheng et al., 2015; Perazzi et al., 2012; Wu, Shi and Lu, 2012).

The human visual system has tools for detecting spatial modulations of luminance gradients in the input image. These are the so-called second-order visual filters (Graham, 2011), which act preattentively. They at a certain spatial interval combine the outputs of striate neurons (first-order filters) with the same

*Corresponding author: babenko@sfedu.ru

© 2022 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

frequency tuning. First-order filters encode information about the carrier (localization, spatial frequency and orientation of luminance gradients). The second-order filters are activated when spatial modulation of the contrast, orientation or spatial frequency of these gradients (envelope) fall within their receptive fields. Moreover, the higher the modulation amplitude, the stronger their reaction. At the same time, it has been shown that different second-order filters respond to different modulations (Yavna, 2012). Since orientation modulations are primarily important for detecting texture boundaries (Solomon and Morgan, 2017), and spatial frequency modulations are important for detecting surface curvatures (Sakai and Finkel, 1995), it is fair to consider the filters selective to contrast modulations to be the first candidate for the role of a segmentation mechanism for real scenes and objects (Açık et al., 2009; Frey, König and Einhäuser, 2007; 't Hart et al., 2013).

The aim of our study was to determine the role of image areas of largest increase in total (non-local) luminance contrast in visual processing using facial expression recognition tasks. The hypothesis was that information from these areas of the image is used in categorization.

We chose faces as visual stimuli due to both their high social significance and multidimensionality, which implies separate processing of variable and invariant facial characteristics. At the same time, face detection and identification is characterized by unique speed (Cauchoix et al., 2014; Willis and Todorov, 2006). The same applies to emotion recognition (Willis and Todorov, 2006; Liu and Ioannides, 2010; Vuilleumier and Pourtois, 2007).

To test our hypothesis, we created gradient operator of total contrast (GOTC), a computer program that simulates the second-order filters and calculates a map of instantaneous values of the non-local contrast modulation function over the entire image (Babenko et al., 2021). These maps make it possible to create stimuli using areas of the raster image with certain modulation values.

To a certain extent, this approach resembles the Bubbles method (Gosselin and Schyns, 2001; Smith et al., 2005). In both approaches the accuracy of expression recognition is studied when fragments of the face image are shown to the subjects. The difference is that in the Bubbles method, the fragments are selected randomly, and in our study, they are selected in accordance with the contrast gain. In addition, the Bubbles technique involves the preliminary learning of the initial set of faces, so observers are working with familiar faces, and this changes the range of effective spatial frequencies (Butler et al., 2010; Lobmaier and Mast, 2007; Smith, Volna and Ewing, 2016). Our approach allows us to use unfamiliar faces, which does not limit the number of stimuli and brings the experimental procedure closer to the real conditions of face perception. In addition, the bubbles technique can not be used to answer the question about the mechanisms for highlighting certain facial features.

Prior to creating stimuli, it was necessary to determine several parameters of the model that simulates how second-order filters work. First of all, we had to choose the spatial frequency ranges in which the contrast modulation should be calculated. Since second-order filters were previously found to form five spatial frequency pathways that are tuned in 1 octave steps (Ellemberg et al., 2006), we decided to follow this scheme.

Secondly, it was necessary to select the parameters of the apertures through which the whole image and its fragments are passed during the formation of facial stimuli. To keep the constant ratio between the carrier and envelope frequencies, the aperture diameter was reduced by a factor of 2 to increase the filtering frequency in cycles per image (CPI) by 1 octave, while the filtering frequency inside the aperture of different diameters remained constant and was equal to 4 cycles per aperture diameter. Such a filtering frequency was due to the data on the optimal ratio of the carrier and envelope frequencies in human perception of contrast modulations (Babenko, Ermakov and Bozhinskaya, 2010; Sun and Schofield, 2011). Similar psychophysical results were also obtained in the analysis of neuronal responses in V2 in primates (Willis and Todorov, 2006). Another aperture parameter is the transfer function. Based on the central subfield profile of the second-order filter the transfer function was set as Gaussian.

Thirdly, the number of apertures at each filtering frequency had to be determined. The entire face image is described by a single aperture with the lowest filtering frequency (in CPI). We decided that since at each next step the filtering frequency should double, the number of selected areas should also double. In this case, the total diameter of apertures remains constant, and the filtering frequency in cycles per image increases by a factor of 2 at each frequency step.

Materials and Methods

Participants

The experiments involved 179 subjects of both sexes in total, Europeans, aged 18 to 30 years. All participants had normal or corrected vision and had no history of neurological or psychiatric disease. The subjects were informed about the upcoming procedure and gave written consent to voluntarily participate in the experiment. The study was approved by the local ethics committee and was performed in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

Equipment

The experimental setup included an x86-64 compatible Ubuntu Linux PC with NVIDIA GeForce GT 730 graphics and Acer VG271U Pbmiipx monitor. Screen resolution was 2560x1440, frame rate was 60 Hz. The monitor was calibrated with a digital luminance meter in grey scale mode. ACM (Adaptive Contrast Management) and HDR (High Dynamic Range) functions have been disabled. The luminance varied from 1 to 225 cd/m2, gamma non-linearity was standard with an exponent of 2.2.

Stimuli

The set of stimulus images of faces with different emotional expression was compiled from open access databases: MMI (Pantic et al., 2005), KDEF (Lundqvist, Flykt and Öhman, 1998), Rafd (Langner et al., 2010) and WSEFEP (Olszanowski et al., 2015). For further processing and preparation of the stimulus material we selected 70 initial full-faced photographs of male and female Caucasian faces with the expression of 6 basic emotions according to P. Ekman (Ekman, 1992) (fear, anger, sadness, disgust, surprise, happiness), and a neutral expression. Each emotion was represented by 10 faces (5 male and 5 female). Different faces were used for different expressions.

First, faces from different databases were equalized in average luminance (50 cd/m^2) and RMS contrast, and size-adjusted to a circle of 880 pixels. Then, each initial image was processed using GOTC that simulates the functioning of the second-order filters set with the same localization and filtering frequency in full range of orientation tunings. The operator is a concentric area with Difference of Gaussians profile. The diameter of the center of this area («window») is equal to the width of the surrounding ring. The filtration frequency in the window was constant and equaled to 4 cycles per window. When the size of the operator was 2 times reduced, the filtering frequency in cycles per image (CPI) doubled. Thus, for an image filtered at a frequency of 4 CPI, the window diameter is equal to the size of the entire image. For an image filtered at a frequency of 8 CPI, the window 2 times decreased and equaled the half the image size, for a filtering frequency of 16 CPI it decreased by 4 times, for 32 CPI - by 8 times and for 64 CPI - by 16 times. The bandwidth of all filters was the same and equaled 1 octave.

The operator window calculates spectral power of the image filtered at a given frequency in CPI. The spectral power of all spatial frequencies perceived by a human was calculated in the surrounding ring and rescaled to average power per 1 octave. The non-local contrast increase in each position was cal-culated as the difference between the total energy in the center of GOTC and on its periphery. The operator scans the entire image and builds a two-dimensional map of the contrast gain.

As a result, 5 saliency maps were generated for each initial image (for 5 filtering frequencies). Then, on each map, the local maxima of the increase in contrast were ranked in descending order of the am-plitude value. Local maxima were selected, starting from the highest, according to the following rule: 1 position was selected at a filtering frequency of 4 CPI, 2 positions were selected at a frequency of 8 CPI, 4 positions were selected at a frequency of 16 CPI, and 8 positions, and on 64 CPI - 16.

After that, we moved on to creating stimuli. First, each initial image was filtered (with a 10th order Butterworth filter) in five one-octave-wide frequency bands with center frequencies of 4, 8, 16, 32, and 64 CPI. Then, a circular aperture with a Gaussian transfer function was placed in the positions previously selected on the saliency maps. An already filtered image of the corresponding spatial frequency was passed through it. The aperture diameter was equal to the diameter of the central region of the gradient operator (at the lowest frequency, the entire image is transmitted; at higher frequencies, progressively smaller fragments of the image are transmitted).

Facial stimuli were created by combining images transmitted through the aperture from different spatial frequency ranges (15 different combinations of frequency ranges were used). As a result, for each initial face image, 15 stimuli were created, consisting of areas of highest increase in non-local contrast. For experiment 1, stimuli were created in a similar way, consisting of areas of the initial image with the smallest increase in contrast.

After performing all calculations, the created stimuli were scaled down to 8.5 ang deg. As a result,

the lowest filtering frequency, equal to 0.5 CPD, approximately corresponded to the frequency tuning of the lowest frequency channel in the human visual system.

Procedure

Prior to the experiment, observers were instructed and looked through the examples of faces expressing basic emotions. In the experiment, stimuli were presented in a random sequence, and their duration was not limited. Viewing distance was 70 cm. The observers were tasked with recognizing facial expressions, choosing 1 of 7 possible responses that characterize emotional expression. The responses were given verbally. The accuracy of recognition for each type of stimuli was calculated as a percentage of correct responses.

Statistical data analysis

ANOVA was used for statistical analysis of the results. Pairwise comparison of the percentages of correct responses by Student's t-test was carried out in the ANOVA procedure as post-hoc tests performed with Holm's correction for multiple comparisons.

Results

Experiment 1. Influence of the magnitude of the increase in the contrast of the regions forming the stimulus on the recognition of facial expression

In one of the previous works, it was shown that the greater the increase in the total contrast in the areas from which the facial stimulus is formed, the more accurately happy (joyful) and neutral faces are distinguished (Babenko et al., 2021). However, since in the present study it was supposed to use a significantly larger number of facial expressions (6 basic emotions according to Ekman and a neutral expression), we considered it necessary to conduct a repeated experiment in which we compared the recognition accuracy of 7 expressions. Now we have limited ourselves to two sets of stimuli created from areas with the largest and the smallest increase in total contrast.

Procedure

Experiment 1 involved 52 observers.

Stimuli were created by combining selected fragments of the initial image in 4 spatial frequency ranges with peak frequencies of 8, 16, 32, and 64 CPI (Fig. 1). Each of the 7 facial expressions was represented by 20 stimuli formed from 5 female and 5 male faces (10 images were created from areas with the lowest non-local contrast modulation, and 10 from areas of highest contrast). A total of 140 stimuli ((10+10)*7) were generated for this experiment.

26 subjects were tasked to categorize facial expression when viewing stimuli created from regions with the lowest non-local contrast gain. The other 26 observers were tasked similarly with stimuli generated from regions with the highest increase in non-local contrast. Each subject was presented with 70 stimuli. One of the possible responses was the "I don't know" answer.

Data analysis was performed using one-way ANOVA (intersubject, repeated measures). The independent variable was the amplitude of contrast modulation of the areas that were used for synthesized stimuli. The dependent variable was the proportion of correct responses in the expression recognition task.

Results

Experiment 1 revealed a statistically significant effect of the contrast of the areas that were used for creating stimuli on the accuracy of expression recognition ($F(1,50) = 699.28$, $p = 0.000$, $\omega 2=0.931$). The performance was significantly higher for stimuli created from areas of the initial image with the highest increase in non-local contrast (max) compared to stimuli created from areas with the lowest increase in contrast (min) (Fig. 2).

Figure 1. Examples of stimuli used in experiment 1.

An example of a stimulus created from areas of the initial face image with the highest contrast gain (above). An example of a stimulus created from areas with the lowest gain in non-local contrast (bottom). The regions used to create stimuli were selected in the range of spatial frequencies from 5.6 to 90.2 CPI.

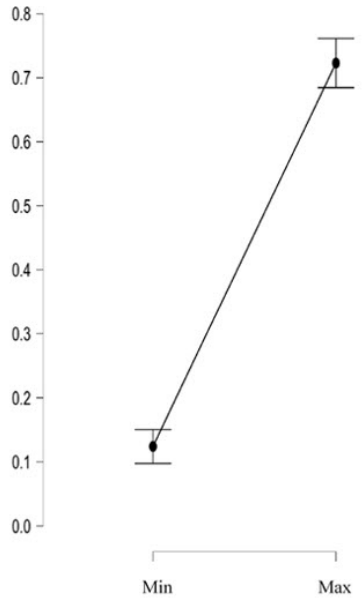

Figure 2. Accuracy of expression recognition depending on the contrast gain in the areas that were used for creating stimuli. "Min" is for stimuli created from areas of the initial image with the lowest non-local contrast increase, "Max" is for stimuli created from areas of highest contrast increase. The y-axis shows the pro-portion of correct responses.

The obtained results indicate that the information contained in the areas of the face image with the highest contrast increase in the range of 4 octaves is useful for recognizing expressions and provides a relatively high accuracy of recognition. In stimuli created from regions with the lowest contrast gain, emotions are correctly determined only at a random decision level.

Experiment 2. Accuracy of expressions recognition in facial stimuli created using the areas of highest increase in contrast with different combinations of spatial-frequency ranges

After it was established that the information contained in the areas of the facial image with the highest contrast gain is useful for expression recognition, it was necessary to understand in which frequency range this information provides the best result for recognizing a particular facial expression.

The majority of researchers agree that the average spatial frequencies are most important for face recognition. However, there is a variety of data on different "effective" ranges: 8-16 CPF (Costen, Parker and Craw, 1996; Gold, Bennett and Sekuler, 1999), 8-13 CPF (Nasanen, 1999), 11-16 CPF (Tanskanen et al., 2005). Collin et al. (2006) extended this range to 25 CPF. At the same time, the role of the general configuration in face recognition was emphasized by many studies (eg, Cheung et al. 2008; Leder and Bruce 2000; Maurer et al., 2002; McKone, 2008). A holistic perception of the face implies its low-frequency description – lower than 8 CPF (Awasthi et al., 2011; Goffaux and Rossion, 2006).

As for facial expression recognition, many authors also prefer configuration information, and hence low spatial frequencies, when solving this problem (e.g., Bombari et al., 2013; Calder et al., 2000; Calvo and Beltrán, 2014; Tanaka et al., 2012; White, 2000). Others, on the contrary, emphasize the role of internal features of the face and, as a result, higher spatial frequencies (Blais et al., 2012; Royer et al., 2018; Smith and Schyns, 2009). The fMRI data also contradicts the notion that low frequency information plays a critical role in the processing of facial expressions (Morawetz et al., 2011). Moreover, C. Deruelle and J. Fagot provide evidence in favor of the priority of high-frequency information in the task of expressions categorization (Deruelle and Fagot, 2005). This contradiction in experimental findings could be caused by the fact that different emotional expressions are encoded by different spatial frequencies (Kumar and Srinivasan, 2011; Pourtois et al., 2005; Stein et al., 2014; Vlamings, Goffaux and Kemner, 2009; Vuilleumier et al., 2003).

Thus, the objective of the second experiment was to determine the frequency ranges for the best recognition accuracy for each of the basic emotions, as well as neutral facial expressions, created from

areas of highest increase in non-local contrast.

Procedure

Experiment 2 involved 78 subjects.

The stimuli were created using the areas of the initial images with the highest increase in the total non-local contrast. Fragments of the face were isolated in five ranges of spatial frequencies with peak frequencies of 4, 8, 16, 32, and 64 CPI. All possible combinations of adjacent frequency ranges were used. A total of 1050 facial stimuli were created (10 initial faces (5 male + 5 female) * 7 facial expressions * 15 combinations of spatial frequencies).

The stimuli were presented in a random sequence. Observers chose one of 7 possible responses after each stimulus was presented.

Results

In experiment 2, we calculated the accuracy of recognition of all basic emotions and neutral facial expressions in stimuli created from areas of highest increase in total nonlocal contrast with different combinations of spatial frequency bands in the stimulus (Table 1).

Table 1

Expression recognition accuracy with different frequency contents of facial stimuli created from areas of highest increase in nonlocal contrast

Stimulus frequency content *	Facial expressions (expression recognition accuracy in percent)							Mean (%)
	fear	anger	sadness	disgust	neutral	surprise	happiness	
4	4,2	6,3	40,4	1,9	33,2	7,1	31,7	17,8
4+8	9,9	35,5	44,9	21,9	44,0	49,9	66,4	38,9
4+8+16	60,5	58,2	51,7	69,7	69,4	69,9	78,1	65,4
4+8+16+32	46,4	55,6	80,3	73,7	84,2	79,7	90,1	72,9
4+8+16+32+64	35,1	50,1	70,0	72,8	85,9	81,8	97,4	70,5
8	16,9	27,8	23,9	17,7	24,1	41,9	27,1	25,6
8+16	41,9	43,7	60,8	69,5	66,9	71,5	72,1	60,9
8+16+32	40,3	47,2	58,6	70,9	81,3	81,7	95,6	67,9
8+16+32+64	61,8	61,0	60,5	76,3	83,1	81,9	95,6	74,3
16	36,8	30,0	36,9	64,0	66,2	60,4	67,1	51,6
16+32	26,8	36,0	48,1	71,4	80,1	75,4	96,5	62,1
16+32+64	37,7	44,9	56,5	74,9	82,6	77,2	93,2	66,7
32	41,9	23,5	31,0	42,6	55,8	56,3	79,7	47,3
32+64	26,7	12,7	35,8	47,6	68,5	47,4	92,2	47,3
64	16,5	9,7	31,4	26,2	50,1	20,8	67,3	31,7

* - here and in the following tables the integration of spatial frequency ranges in the stimuli is shown (the central frequency of the range is in cycles per image)

We began the analysis of the obtained results with an assessment of the accuracy of expression recognition based on a low-frequency holistic description of the face. To do this, we analyzed the percentage of correct responses for those trials when the image of the entire face filtered in the range of 2.8–5.6 CPI (central frequency 4 CPI) was presented as a stimulus. These stimuli were created by filtering the initial images at a specified frequency, through an aperture with a Gaussian transfer function, the diameter of which corresponded to the largest extent of the analyzed image (facial image height). Table 1 shows that in this case the accuracy of expression recognition was 17.8% (the random decision level was 14.2% and the confidence interval ranges was from 10.76% to 27.86% for the 95% significance level). At the same time, our previous findings indicate that if such facial stimuli are presented in a set of other objects created in a similar way, the accuracy of face detection reaches 75%. It suggests that low-frequency information may be sufficient to detect a face, but not enough to differentiate the emotions expressed on it. This confirms the idea that only low-frequency information is not enough for facial expression recognition (e.g., Jennings, Yu and Kingdom, 2017).

Taking into account the data confirming the global precedence effect (Goffaux et al., 2011; Peyrin et

al., 2010), we studied how the accuracy of recognition changes with a gradual expansion of the bandwidth, starting from the lowest frequency range (2.8-5.6 CPI), by adding more and more high-frequency ranges (function 1 in Figure 3). As expected, expanding the range of spatial frequencies improves the results. The most noticeable performance increase was observed when expanding the range from 1 to 3 octaves. The addition of the 5th octave no longer affected the accuracy for this task.

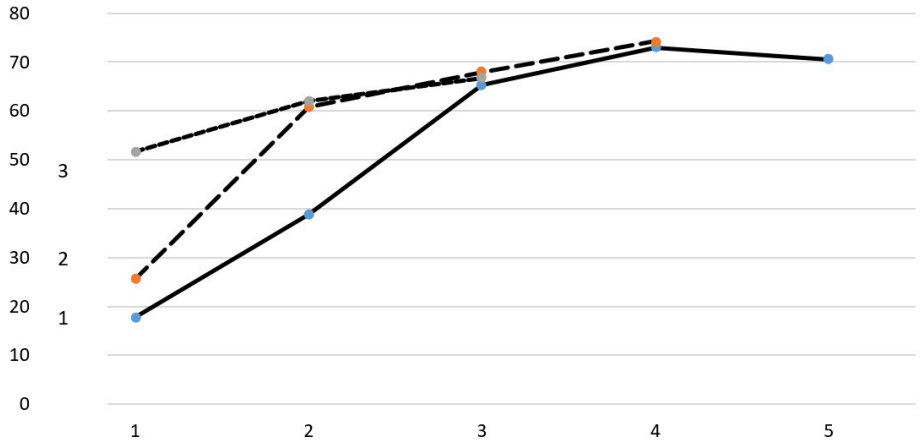

Figure 3. Accuracy of expression recognition with expanding the range of spatial frequencies that used for the facial stimuli. For function 1, the expansion of the frequency range starts from a frequency of 4 CPI, for function 2 - from 8 CPI, for function 3 - from 16 CPI. On the x-axis is the width of the frequency band of the stimulus in octaves. The y-axis shows the percent of correct responses.

Functions 1 and 2 in Figure 3 overlap when the bandwidth becomes equal to 3 octaves. However, the initial increase for function 2 was more significant. The difference is especially noticeable at a bandwidth of 2 octaves. If the spatial frequency increment starts from a higher frequency range (11.3-22.6 CPI, the center frequency is 16 CPI), a significant difference between this curve and the previous ones arises already for a frequency band of 1 octave (function 3 in Figure 3).

It has been shown that any range of spatial frequencies three octaves wide is sufficient for relatively efficient (about 70% correct responses) differentiation of expressions in facial stimuli created from areas of highest contrast gain. The comparison of the obtained functions was performed using two-way Repeated Measures ANOVA with Greenhouse-Geisser correction (main effects: Band Width (1, 2 and 3 octaves) and Start Frequency (4, 8 and 16 CPI), as well as their interaction). It revealed that a significant increase in the performance with the expansion of the frequency band of the stimuli towards higher spatial frequencies ($F(1.699, 130.852) = 1804.298$, $p<0.0000$, $\omega2=0.824$) depends on the frequency from which the band expansion begins ($F(1.661, 127.934) = 519.873$, $p<0.0000$, $\omega2=0.584$). Significantly more information about facial expression is contained precisely in the range with a central frequency of 16 CPI and 1 octave width, in comparison with other frequency ranges of the same width (Table 2). And the increase in performance occurs faster when expanding the range, starting from this frequency ($F(3.479, 130.852) = 246.979$, $p<0.0000$, $\omega2=0.472$).

Table 2

Comparison of expression recognition accuracy for stimuli with a bandwidth of 1 octave

frequency content of compared stimuli (central frequency in CPI)	expression recognition accuracy (%)	Student's t-test (t)	significance level adjusted for multiple comparisons (P_{Holm})
4 / 8	17.8 / 25.6	7.453	0.000
4 / 16	17.8 / 51.6	32.316	0.000
8 / 16	25.6 / 51.6	24.863	0.000

However, if we track how the accuracy of expression recognition changes with the expansion of the frequency range not only towards an increase, but also towards a decrease in the spatial frequency, then we will get a somewhat unexpected result. For different emotions, the optimal direction of the frequency range expansion is evidently different (Table 3).

Table 3

Comparison of recognition accuracy of different expressions for stimuli with a bandwidth of 2 octaves

Facial expressions	8+16 / 16+32 %	Student's t-test (t)	significance level adjusted for multiple comparisons (PHolm)
fear	**41.9** / 26.8	6.615	0.000
anger	**43.7** / 36.0	3.364	0.068
sadness	**60.8** / 48.1	5.550	0.000
disgust	69.5 / **71.4**	0.841	1.000
surprise	71.5 / **75.4**	1.628	1.000
neutral	66.9 / **80.1**	5.774	0.000
happiness	72.1 / **96.5**	10.708	0.000

Higher accuracy values in comparison pairs are shown in bold.

The table shows that for the happiness and a neutral facial expression, it is really more optimal to add a higher spatial frequency to the range of 11.3 - 22.6 CPI the information. For the recognition of emotions of negative valence (fear, anger, sadness), it turned out to be more effective to expand the frequency range towards lower spatial frequencies. Moreover, this is less typical for anger than for other negative emotions. At the same time, for disgust and surprise, the expansion in both directions turned out to be almost equivalent.

Considering that the range with the central frequency of 16 CPI turned out to be the most informative (see Figure 3), we can assume that information from this range is processed first. This information may be sufficient to hypothesize a probable facial expression, and the results of this preliminary analysis determine the direction of further expansion of the frequency range.

This assumption does not contradict the thesis about the sequential processing of spatial frequencies from lower to higher ones, but at the same time, it is consistent with the data on the possibility of flexible use of early perceptual representation by top-down control. This allows the visual system to selectively use different spatial frequencies depending on how useful they are for solving a particular problem (Flevaris and Robertson, 2016; Oliva and Schyns, 1997).

We then moved on to the main question in experiment 2: what combination of frequency ranges is most effective for recognizing each of the expressions? The result of this analysis is shown in Table 4.

Table 4

Combinations of spatial-frequency ranges in facial stimuli formed from areas of highest contrast gain, providing the best result of expression recognition

frequency content of stimuli	Facial expressions (expression recognition accuracy in percent)						
	fear	anger	sadness	disgust	neutral	surprise	happiness
4+8+16+32	46.4	55.6	**80.3**	73.7	84.2	79.7	90.1
4+8+16+32+64	35.1	50.1	70.0	72.8	**85.9**	81.8	**97.4**
8+16+32+64	**61.8**	**61.0**	60.5	**76.3**	83.1	**81.9**	95.6

Higher accuracy values in comparison pairs are shown in bold.

It is shown that for different facial expressions, the optimal combinations of spatial frequencies in the stimulus differs. So for better recognition of a neutral facial expression and happiness, the full frequency range, that is, all 5 octaves, is more preferable. To recognize other emotions, a band of 4 octaves is enough. However, for stimuli expressing sadness, the effective range is shifted to a lower spatial frequency, while for other emotions it is shifted to a higher frequency region. It should also be noted that for the negative emotions (fear, anger, sadness) the optimum is quite clear (significant differences were obtained according to Student's test), and for other expressions it is not so obvious.

Finding the optimal combination of spatial-frequency ranges for each facial expression allowed us to move on to experiment 3.

Experiment 3. Testing the possibility of effective expressions recognition in facial stimuli created with the areas of highest contrast gain.

The results obtained indicate that the information from the areas of the face with the highest contrast gain is indeed useful for expression recognition. However, the question remains how much the solution to this problem depends on whether the subject uses all the information about the face, or only information from areas of highest increase in non-local contrast. To do this, under the same experimental conditions, it was necessary to compare the accuracy of expression recognition in photographic images of real faces (unfiltered) and in faces formed from fragments selected in the optimal spatial frequency ranges for each emotion.

Procedure

Experiment 3 involved 49 subjects.

Synthesized facial stimuli expressing fear, anger, disgust, and surprise included frequencies of 8, 16, 32, and 64 CPI. Stimuli expressing sadness were created from the ranges with central frequencies of 4, 8, 16, and 32 CPI. Stimuli with a neutral expression and happiness were created from fragments identified in the range of five octaves: 4, 8, 16, 32 and 64 CPI. The set of real face images used as stimuli did not overlap with the set of initial images used to create the synthesized stimuli. A total of 70 synthesized and unfiltered facial images were used (10 faces x 7 expressions).

The stimuli were presented in a random sequence. The exposure time was not limited. After training, the subjects were asked to make a decision on each presented stimulus as quickly as possible and press the key. Pressing the key removed the image. That way it allowed us to measure the decision time. Then the subjects gave a verbal response and it was recorded by the experimenter. As before, the range of possible responses was limited to 7 expressions.

Results

The results obtained in experiment 3 are shown in Figure 4. In general, the average accuracy of expression recognition was expectedly somewhat higher when perceiving natural facial images (83% correct responses) compared to synthesized stimuli (73%). For real images, the decision time was also shorter (by 290 ms on average).

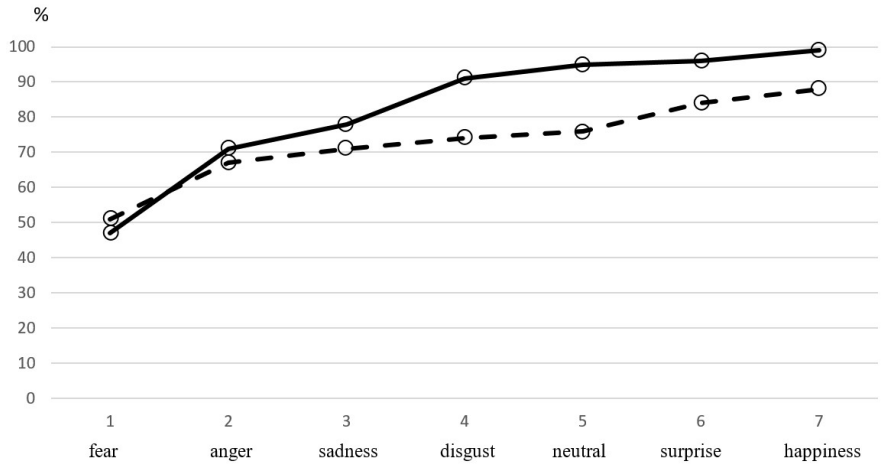

Figure 4. Accuracy of expression recognition in real (continuous line) and synthesized (dotted line) faces.

For statistical analysis of the obtained data we used a two-way Repeated Measures ANOVA (main effects: Expression (7 expressions) and Stimulus Type (real and synthesized), as well as their interaction). It was confirmed that the recognition accuracy of different expressions is different for both real and synthesized facial stimuli ($F(3.284, 157.609)=68.276$, $p<0.0000$, $\omega2=0.530$, Greenhouse-Geisser corrected). The accuracy of expression recognition for different types of stimulus differs significantly ($F(1, 48)=110.154$, $p<0.0000$, $\omega2=0.351$). The curves from Figs. 4 are also different ($F(4.755, 228.233)=8.911$, $p<0.0000$, $\omega2=0.101$, Greenhouse-Geisser corrected). The last of these differences is determent by the fact that for disgust, surprise, happiness and neutral expression the recognition accuracy is higher for real face images, while fear, anger and sadness are actually recognized with the same accuracy as in

synthesized images (Table 5).

Table 5

Comparison of recognition accuracy of different expressions for real and synthesized facial stimuli

Facial expressions	Stimuli real / synthesized %	t	PHolm
fear	47 / 51	-1.424	1.000
anger	71 / 67	1.582	1.000
sadness	78 / 71	2.610	0.246
disgust	91 / 74	6.487	0.000
surprise	95 / 76	7.357	0.000
neutral	96 / 84	4.430	0.000
happiness	99 / 88	4.746	0.000

Higher accuracy values in comparison pairs are shown in bold.

The accuracy of expression recognition in real and synthesized facial stimuli somewhat differs. At the same time, real and synthesized faces formed the same sequence of gradual increase in recognition accuracy in a series of expressions (see Fig. 4). Statistical analysis using rank correlation coefficient showed that these are similar functions (Kendall's τb (47) = 1, p = 0.000). This may indicate that the natural course of the information processing is not disturbed when a real face is replaced with a synthesized image created from fragments with the highest contrast gain. However, there is enough information to recognize emotions of negative valence in synthesized stimuli, but not enough for recognition of other expressions. That suggests that in the synthesized facial stimuli some important information is missing.

Discussions

The ability of the human visual system to process huge amount of information in a very short time is determined by the ability to find "useful" areas in the input image. This step can be based on the search for spatial heterogeneities in the image using the second-order visual mechanisms. To simulate the operation of these mechanisms and to test the usefulness of the information extracted by them in the expression recognition, we created the gradient operator of total non-local contrast (GOTC). Two variables determine the overall contrast: the contrast of the single luminance gradients and the number of gradients in a given area of the image. Moreover, the second variable make a greater contribution to the total signal energy. Therefore, regions of interest first of all are the areas with the largest accumulation of luminance gradients.

The design of the created operator reflects the main properties of second-order visual filters: the multichannel nature of the second-order mechanism (a set of operators of different sizes); bandpass filtering of carrier and the certain relationship between the carrier and envelope frequencies (the operator size has inverse relation with filtering frequency in CPI); opponent organization of the filter, which makes it possible to encode the amplitude of the contrast modulation (concentric organization of the GOTC); weighting function of the filter receptive fields (Gaussian transfer function aperture). The stimuli we used were created using this gradient operator.

In experiment 1, we showed that the recognition of 7 basic emotions in facial expressions has relatively high level of accuracy when it is based on the information of different spatial frequencies from areas of highest increase in non-local contrast (about 75% of correct responses). At the same time, facial stimuli created from areas with the lowest contrast gain turned out to be absolutely ineffective in terms of solving this problem (recognition accuracy was at a random decision level). Together with the previously published results (Babenko et al., 2021), this indicated that the informativeness of the image area is determined by the degree of its difference in the total contrast from the surroundings.

We then analyzed the possibility of using a low-frequency representation of the entire face in expression differentiation tasks. In previous studies, we have shown that stimuli generated by the operator with a central area that matched the full size image were recognized as faces in a series of other stimuli with high accuracy (about 75%). When in experiment 2 the task was transformed and it was required not only to detect a face, but to differentiate the emotions in facial expressions, the result decreased significantly - to about 18% of correct responses (when a random decision level was 14%). This result is

consistent with widely accepted assumption that face processing should be considered as consecutive steps of face detection and individualization (Comfort and Zana, 2015). However, at the second stage of the processing the low-frequency description is no longer enough. Higher spatial frequencies provide additional information about the internal features of the face, which are very important for its configurative description (Goffaux, 2009; Piepers and Robbins, 2012).

In experiment 2, we studied the accuracy of expression recognition in facial stimuli with different combinations of fragments isolated in different ranges of spatial frequencies. We found that the most effective frequency range is the 11.3-22.6 CPI band with a center frequency of 16 CPI. And while this result is not consistent with the idea of the low spatial frequencies importance in the perception of faces, it is consistent with the data indicating that the frequencies of the middle range are most important in identifying faces. It is noteworthy that in this frequency range (11.3–22.6 CPI), the GOTC more often singled out the eyes and mouth as areas of interest in the initial images, which are known to be very important for conveying emotionally significant information.

However, unlike the experiments with Bubbles technique, we did not aim to determine the independent contribution of each frequency range to expression recognition, since the perception of a face is not a simple sum of its components (Jack et. al., 2012, but see Gold, Mundy and Tjan, 2012). It was important for us to determine the range of spatial frequencies for each expression that provides the best accuracy of recognition.

Our results certainly do not provide an unambiguous answer to the question of how information from different spatial frequency pathways is combined. Previously published results in this area have also been somewhat controversial. There is data on that the visual system processes spatial frequencies in a certain sequence, from low to high (Gao and Bentin, 2011). At the same time, flexible top-down selection of spatial frequency channels can significantly optimize the visual processing (Flevaris and Robertson, 2016). It is also impossible to exclude the possibility of simultaneous processing of all frequencies. Considering the above, our results clearly indicate the frequency range that contains the most useful information about facial expressions and which would be the most reasonable to start processing with (11.3-22.6 CPI). The conclusion that this information can determine the strategy for further integration of spatial frequencies is also supported by the fact that for emotions of negative valence it is more optimal to add information from a lower frequency ranges, and for other facial expressions from a higher frequency ranges.

Different frequency ranges turned out to be effective for different expressions. For the best recognition of neutral and joyful facial expressions, all 5 octaves were required. This result is consistent with the data on neutral facial expression containing a complete set of basic expressions (Lee and Kim, 2008), and that the expression of happiness is encoded by both low and high spatial frequencies (Becker et al., 2012). Our data showed that in sadness recognition, 4 octaves were enough (without the highest frequency range). To recognize fear, anger, disgust and surprise, 4 octaves were also enough, but without the lowest-frequency range.

So, as a result of the experiment 2, we have determined in what ranges of spatial frequencies the areas of the greatest contrast gain should be extracted in order to provide the best recognition accuracy of a particular expression. Now it was necessary to make sure that this is exactly the information that is used by the visual system when recognizing the expression of real faces. To do this, in experiment 3 we compared the accuracy of recognition of each expression in the perception of the images of real faces and stimuli formed from the optimal combination of selected fragments. Indeed, synthesized images were recognized somewhat worse than real ones (73% versus 83%).

It is interesting to note that the decrease in the recognition accuracy for the synthesized stimuli was not found for the expressions of negative valence. In these cases, these fragmentary images of faces were perceived with approximately the same accuracy as real ones. Such peculiarity of the recognition of negative expressions is consistent with the data on that the perception of such emotions is associated with the activation of special mechanisms (Shaw et al., 2011; Stein et al., 2014; Vuilleumier et al., 2003). However, this does not dismiss the question of the insufficiency of the information contained in the selected areas for the recognition of other emotions. It became obvious that some of the useful information in the synthesized stimuli is missing. Probably the same is evidenced by the increase in reaction time. In fact, this was expected.

Even though choosing the operator parameters we tried to rely on literature data, we had to make the choice arbitrarily in a number of cases. This concerns the number of areas that stand out in each of the frequency ranges, for example. An increase in their number, especially at high spatial frequencies, will be expected to improve the recognition rate. Another aspect that can affect accuracy of expression recognition is the filtering frequency in cycles per aperture. Previous research suggests that the optimal carrier-envelope ratio in second-order filters is 1 to 8 (Babenko, Ermakov and Bozhinskaya, 2010; Peng

and Schofield, 2011). However, this result was obtained in the tasks with modulated textures and not faces. Obviously, even a slight increase in the filtering frequency (for example, from 4 to 4.5 cycles per aperture) can improve the accuracy of expression recognition.

The most interesting finding that we would like to emphasize is that numerous studies have shown that people recognize different expressions with different efficiency, and the recognition accuracy for different expressions form a certain sequence. Fear is recognized with the worst accuracy, and happiness with the best. In experiment 3, as in previous studies, we found a certain sequence of the increase in accuracy of expression recognition for images of real faces. And it was repeated with synthesized images created from the areas of the greatest contrast gain. This may be evidence that the replacement of a real image by a fragmented one, although accompanied by some general decrease in recognition accuracy, does not violate the general logic of the processing.

Conclusions

The obtained results indicate that the informative content of image areas can be determined by the difference between these areas and their surroundings in terms of such a physical parameter as the total non-local contrast. Moreover, the greater this difference, the higher the informational significance of these fragments. This seemingly unexpected result can be explained by the fact that the greatest contribution to the value of the total contrast is made not so much by the contrast of each single luminance gradients, but by the total number of gradients in the analyzed image area. And since each gradient is a kind of visual information unit, the more gradients it contains, the more informative this area would be.

We established that information from the areas of highest increase in contrast is necessary for facial expression recognition. Moreover, this information is sufficient for recognition of basic expressions with a very high accuracy.

These areas are characterized by spatial modulation of luminance gradients and they can be extracted from the input image by second-order visual filters. Thus, these filters are good candidates to be viewed as mechanism of selecting the areas of interest.

Since the signal at the filter output is proportional to the amplitude of the modulation, those that are more activated than their neighbors gain an advantage, due to the lateral interaction between the filters. The locations of these filters form a saliency map, in which priorities for selective attention are distributed in accordance with the amplitude of the modulation.

At the same time, the filters themselves, drawing attention to certain areas of the image, can actually play the role of windows through which information from these areas of the visual field is transmitted to post-attentive levels of processing.

Thus, the results obtained allow us to draw the following conclusions:

- Information from image areas of highest increase in luminance contrast is necessary and sufficient for recognition of basic facial expressions.
- The second-order visual filters extract the salient regions of the image, and a signal value at the filter output determines its priority for attention.
- The receptive fields of the second-order filters act as windows for the attention to extract information, which is then transferred to post-attentive levels of processing.

Acknowledgements

The study was carried out with the financial support of the Russian Science Foundation (project 20-64-47057).

Conflict of interests

The authors declare no conflict of interest.

References

Açık, A., Onat, S., Schumann, F., Einhäuser, W., & König, P. (2009). Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. *Vision research, 49*(12), 1541-1553. https://doi.org/10.1016/j.visres.2009.03.011

Awasthi, B., Friedman, J., & Williams, M. A. (2011). Faster, stronger, lateralized: Low spatial frequency information supports face processing. *Neuropsychologia, 49*(13), 3583-3590. https://doi.org/10.1016/j.neuropsychologia.2011.08.027

Babenko, V. V., Ermakov, P. N., & Bozhinskaya, M. A. (2010). Relationship between the Spatial-Frequency Tunings of the First- and the Second-Order Visual Filters. *Psikhologicheskii Zhurnal, 31*(2), 48-57. Retrieved from https://www.elibrary.ru/item.asp?id=14280688 (in Russ.)

Babenko, V. V., Yavna, D. V., Ermakov, P. N., & Anokhina, P. V. (2021). Nonlocal contrast calculated by the second order visual mechanisms and its significance in identifying facial emotions. F1000 *Research, 10*, 274. https://doi.org/10.12688/f1000research.28396.1

Babenko, V., Yavna, D., Vorobeva, E., Denisova, E., Ermakov, P., & Kovsh, E. (2021). Relationship Between Facial Areas With the Greatest Increase in Non-local Contrast and Gaze Fixations in Recognizing Emotional Expressions. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE), 9*(3), 359–368. https://doi.org/10.23947/2334-8496-2021-9-3-359-368

Barabanshchikov, V. A. (2012). *Ekspressii lits i ikh vospriyatiye* [Facial expressions and their perception]. Moscow: Izdvo «IPRAN» [IPRAS Publishing House]. (in Russ.)

Barabanshchikov, V. A., Hoze E.G. (2013) Vospriyatiye ekspressiy spokoynogo litsa [Perception of expressions of a neutral face]. *Mir psikhologii* [World of Psychology], 1:203-223 Retrieved from https://www.elibrary.ru/item.asp?id=18907610 (in Russ.)

Becker, D. V., Neel, R., Srinivasan, N., Neufeld, S., Kumar, D., & Fouse, S. (2012). The vividness of happiness in dynamic facial displays of emotion. *PLoS One, 7*(1), e26551. https://doi.org/10.1371/annotation/f0519e8c-f347-4950-b7e8-3e9cbc3ec2a9

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia, 50*(12), 2830-2838. https://doi.org/10.1016/j.neuropsychologia.2012.08.010

Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *Quarterly Journal of Experimental Psychology, 66*(12), 2426-2442. https://doi.org/10.1080/17470218.2013.789065

Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of vision, 9*(3), 5-5. https://doi.org/10.1167/9.3.5

Butler, S., Blais, C., Gosselin, F., Bub, D., & Fiset, D. (2010). Recognizing famous people. *Attention, Perception, & Psychophysics, 72*(6), 1444-1449. https://doi.org/10.3758/APP.72.6.1444

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human perception and performance, 26*(2), 527. https://doi.org/10.1037/0096-1523.26.2.527

Calvo, M. G., & Beltrán, D. (2014). Brain lateralization of holistic versus analytic processing of emotional facial expressions. *Neuroimage, 92*, 237-247. https://doi.org/10.1016/j.neuroimage.2014.01.048

Cauchoix, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The neural dynamics of face detection in the wild revealed by MVPA. *Journal of Neuroscience, 34*(3), 846-854. https://doi.org/10.1523/JNEUROSCI.3030-13.2014

Cheng, M. M., Mitra, N. J., Huang, X., Torr, P. H., & Hu, S. M. (2014). Global contrast based salient region detection. *IEEE transactions on pattern analysis and machine intelligence, 37*(3), 569-582. https://doi.org/10.1109/TPAMI.2014.2345401

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception and Performance, 34*(6), 1327-1336. https://doi.org/10.1037/a0011752

Collin, C. A., Therrien, M., Martin, C., & Rainville, S. (2006). Spatial frequency thresholds for face recognition when comparison faces are filtered and unfiltered. *Perception & psychophysics, 68*(6), 879-889. https://doi.org/10.3758/BF03193351

Comfort, W. E., & Zana, Y. (2015). Face detection and individuation: Interactive and complementary stages of face processing. *Psychology & Neuroscience, 8*(4), 442. https://doi.org/10.1037/h0101278

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & psychophysics, 58*(4), 602-612. https://doi.org/10.3758/BF03213093

Deruelle, C., & Fagot, J. (2005). Categorizing facial identities, emotions, and genders: Attention to high-and low-spatial frequencies by children and adults. *Journal of experimental child psychology, 90*(2), 172-184. https://doi.org/10.1016/j.jecp.2004.09.001

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion, 6*(3-4), 169-200. https://doi.org/10.1080/02699939208411068

Ellemberg, D., Allen, H. A., & Hess, R. F. (2006). Second-order spatial frequency and orientation channels in human vision. *Vision Research, 46*(17), 2798-2803. https://doi.org/10.1016/j.visres.2006.01.028

Flevaris, A. V., & Robertson, L. C. (2016). Spatial frequency selection and integration of global and local information in visual processing: A selective review and tribute to Shlomo Bentin. *Neuropsychologia, 83*, 192-200. https://doi.org/10.1016/j.neuropsychologia.2015.10.024

Frey, H. P., König, P., & Einhäuser, W. (2007). The role of first-and second-order stimulus features for human overt attention. *Perception & Psychophysics, 69*(2), 153-161. https://doi.org/10.3758/BF03193738

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological bulletin, 134*(5), 662-676. https://doi.org/10.1037/0033-2909.134.5.662

Gao, Z., & Bentin, S. (2011). Coarse-to-fine encoding of spatial frequency information into visual short-term memory for faces but impartial decay. *Journal of Experimental Psychology: Human Perception and Performance, 37*(4), 1051-1064. https://doi.org/10.1037/a0023091

Goffaux, V. (2009). Spatial interactions in upright and inverted faces: Re-exploration of spatial scale influence. *Vision research, 49*(7), 774-781. https://doi.org/10.1016/j.visres.2009.02.009

Goffaux, V., & Rossion, B. (2006). Faces are" spatial"--holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human perception and performance, 32*(4), 1023-1039. https://doi.org/10.1037/0096-1523.32.4.1023

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex, 21*(2), 467-476. https://doi.org/10.1093/cercor/bhq112

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychological science, 23*(4), 427-434. https://doi.org/10.1177/0956797611427407

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision research, 39*(21), 3537-3560. https://doi.org/10.1016/S0042-6989(99)00080-2
Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision research, 41*(17), 2261-2271. https://doi.org/10.1016/S0042-6989(01)00097-9
Graham, N. V. (2011). Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years. *Vision research, 51*(13), 1397-1430. https://doi.org/10.1016/j.visres.2011.02.007
Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology, 148*(3), 574-591. https://doi.org/10.1113/jphysiol.1959.sp006308
Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences, 109*(19), 7241-7244. https://doi.org/10.1073/pnas.1200155109
Jennings, B. J., & Yu, Y. (2017). The role of spatial frequency in emotional face classification. *Attention, Perception, & Psychophysics, 79*(6), 1573-1577. https://doi.org/10.3758/s13414-017-1377-7
Kumar, D., & Srinivasan, N. (2011). Emotion perception is mediated by spatial frequency content. *Emotion, 11*(5), 1144-1151. https://doi.org/10.1037/a0025453
Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and emotion, 24*(8), 1377-1388. https://doi.org/10.1080/02699930903485076
Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The quarterly journal of experimental psychology Section A, 53*(2), 513-536. https://doi.org/10.1080/713755889
Lee, H. S., & Kim, D. (2008). Expression-invariant face recognition by facial expression transformations. *Pattern recognition letters, 29*(13), 1797-1805. https://doi.org/10.1016/j.patrec.2008.05.012
Li, G., Yao, Z., Wang, Z., Yuan, N., Talebi, V., Tan, J., ... & Baker, C. L. (2014). Form-cue invariant second-order neuronal responses to contrast modulation in primate area V2. *Journal of Neuroscience, 34*(36), 12081-12092. https://doi.org/10.1523/JNEUROSCI.0211-14.2014
Liu, L., & Ioannides, A. A. (2010). Emotion separation is completed early and it depends on visual field presentation. *PloS one, 5*(3), e9790. https://doi.org/10.1371/journal.pone.0009790
Lobmaier, J. S., & Mast, F. W. (2007). Perception of novel faces: The parts have it!. *Perception, 36*(11), 1660-1673. https://doi.org/10.1068/p5642
Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 91*(630), 2-2. https://doi.org/10.1037/t27732-000
Marat, S., Rahman, A., Pellerin, D., Guyader, N., & Houzet, D. (2013). Improving visual saliency by adding 'face feature map'and 'center bias'. *Cognitive Computation, 5*(1), 63-75. https://doi.org/10.1007/s12559-012-9146-3
Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences, 6*(6), 255-260. https://doi.org/10.1016/S1364-6613(02)01903-4
McKone, E. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 310-327. https://doi.org/10.1037/0096-1523.34.2.310
Morawetz, C., Baudewig, J., Treue, S., & Dechent, P. (2011). Effects of spatial frequency and location of fearful faces on human amygdala activity. *Brain research, 1371*, 87-99. https://doi.org/10.1016/j.brainres.2010.10.110
Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision research, 39*(23), 3824-3833. https://doi.org/10.1016/S0042-6989(99)00096-6
Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive psychology, 34*(1), 72-107. https://doi.org/10.1006/cogp.1997.0667
Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Frontiers in psychology, 5*, 1516. https://doi.org/10.3389/fpsyg.2014.01516
Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005, July). Web-based database for facial expression analysis. In 2005 *IEEE international conference on multimedia and Expo* (pp. 5-pp). IEEE. https://doi.org/10.1109/ICME.2005.1521424
Perazzi, F., Krähenbühl, P., Pritch, Y., & Hornung, A. (2012, June). Saliency filters: Contrast based filtering for salient region detection. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 733-740). IEEE. https://doi.org/10.1109/CVPR.2012.6247743
Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., ... & Vuilleumier, P. (2010). The neural substrates and timing of top–down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Journal of cognitive neuroscience, 22*(12), 2768-2780. https://doi.org/10.1162/jocn.2010.21424
Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms "holistic,""configural," and "relational" in the face perception literature. *Frontiers in psychology, 3*, 559. https://doi.org/10.3389/fpsyg.2012.00559
Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., & Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: Time course and topographic evoked-potentials mapping. *Human brain mapping, 26*(1), 65-79. https://doi.org/10.1002/hbm.20130
Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., ... & Fiset, D. (2018). Greater reliance on the eye region predicts better face recognition ability. *Cognition, 181*, 12-20. https://doi.org/10.1016/j.cognition.2018.08.004
Sakai, K., & Finkel, L. H. (1995). Characterization of the spatial-frequency spectrum in the perception of shape from texture. *JOSA A, 12*(6), 1208-1224. https://doi.org/10.1364/JOSAA.12.001208
Shaw, K., Lien, M. C., Ruthruff, E., & Allen, P. A. (2011). Electrophysiological evidence of emotion perception without central attention. *Journal of Cognitive Psychology, 23*(6), 695-708. https://doi.org/10.1080/20445911.2011.586624
Smith, F. W., & Schyns, P. G. (2009). Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distances. *Psychological Science, 20*(10), 1202-1208. https://doi.org/10.1111/j.1467-9280.2009.02427.x
Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological science, 16*(3), 184-189. https://doi.org/10.1111/j.0956-7976.2005.00801.x
Smith, M. L., Volna, B., & Ewing, L. (2016). Distinct information critically distinguishes judgments of face familiarity and identity. *Journal of Experimental Psychology: Human Perception and Performance, 42*(11), 1770-1779. https://doi.org/10.1037/

xhp0000243
Solomon, J.A., & Morgan, M.J. (2017). Orientation-defined boundaries are detected with low efficiency. *Vision Research, 138*, 66-70. https://doi.org/10.1016/j.visres.2017.06.009
Stein, T., Seymour, K., Hebart, M. N., & Sterzer, P. (2014). Rapid fear detection relies on high spatial frequencies. *Psychological science, 25*(2), 566-574. https://doi.org/10.1177/0956797613512509
Sun, P., & Schofield, A. J. (2011). The efficacy of local luminance amplitude in disambiguating the origin of luminance signals depends on carrier frequency: Further evidence for the active role of second-order vision in layer decomposition. *Vision research, 51*(5), 496-507. https://doi.org/10.1016/j.visres.2011.01.008
't Hart, B.M., Schmidt, H.C.E.F., Roth, C., & Einhäuser, W. (2013). Fixations on objects in natural scenes: dissociating importance from saliency. *Frontiers in Psychology, 4*.- Article 455.- 9p. https://doi.org/10.3389/fpsyg.2013.00455
Tanaka, J. W., Kaiser, M. D., Butler, S., & Le Grand, R. (2012). Mixed emotions: Holistic and analytic perception of facial expressions. *Cognition & Emotion, 26*(6), 961-977. https://doi.org/10.1080/02699931.2011.630933
Tanskanen, T., Näsänen, R., Montez, T., Päällysaho, J., & Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cerebral Cortex, 15*(5), 526-534. https://doi.org/10.1093/cercor/bhh152
Vlamings, P. H., Goffaux, V., & Kemner, C. (2009). Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information?. *Journal of vision, 9*(5), 1-13. https://doi.org/10.1167/9.5.12
Vuilleumier, P., & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia, 45*(1), 174-194. https://doi.org/10.1016/j.neuropsychologia.2006.06.003
Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature neuroscience, 6*(6), 624-631. https://doi.org/10.1038/nn1057
White, M. (2000). Parts and wholes in expression recognition. *Cognition & Emotion, 14*(1), 39-60. https://doi.org/10.1080/026999300378987
Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science, 17*(7), 592-598. https://doi.org/10.1111/j.1467-9280.2006.01750.x
Wu, J., Qi, F., Shi, G., & Lu, Y. (2012). Non-local spatial redundancy reduction for bottom-up saliency estimation. *Journal of Visual Communication and Image Representation, 23*(7), 1158-1166. https://doi.org/10.1016/j.jvcir.2012.07.010
Xia, C., Qi, F., Shi, G., & Wang, P. (2015). Nonlocal center–surround reconstruction-based bottom-up saliency estimation. *Pattern Recognition, 48*(4), 1337-1348. https://doi.org/10.1016/j.patcog.2014.10.007
Yarbus, A. L. (2013). *Eye movements and vision*. Springer. https://doi.org/10.1007/978-1-4899-5379-7
Yavna, D. V. (2012). Psikhofiziologicheskiye osobennosti zritel'nogo vospriyatiya prostranstvenno modulirovannykh priznako [Psychophysiological features of visual perception of spatially modulated features]. *PhD Thesis.* Rostov-on-Don (in Russ.)