Facial expression processing in neuronal nets: architecture and large-scale neural networks reconstruction of the human brain under conditions of indeterminacy

O.V. Zhukova, *Vision Physiology Department, I.P. Pavlov Institute of Physiology, Russian Academy of Sciences, St-Petersburg, 199034, Russia,* , K. Yu. Malakhova, *Vision Physiology Department, I.P. Pavlov Institute of Physiology, Russian Academy of Sciences, St-Petersburg, 199034, Russia*

***Abstract*—The research problem consisted of using the following methods of neuroiconics: digital image processing, functional magnetic-resonance tomography (fMRI) of brain activity, psychophysical and neurotechnology methods to determine under conditions of indeterminacy the minimum information needed to recognize faces. Indeterminacy was exemplified at the semantic level by a pattern of weak facial manifestations of emotions that bear a strong emotional burden and at the physical level by noise of various intensities imposed upon the image. Difference-recognition thresholds were established as a result of analyzing this research, and it was shown that there is correlation between the number of altered pixels in the test images and the correct answers of the observers. The number of altered pixels in the test images of faces being compared is proportional to the correct answers when a decision is being made in the psychophysical studies and to the number of altered voxels in the images being compared in maps of the brain responses from the data of the fMRT studies. in a neurotechnological series of studies it was shown that there are fundamental limitations inherent to natural and artificial neural networks. The probability of recognizing a smile is thus fairly high under ordinary conditions, but it decreases under indeterminacy conditions (threshold and noisy images) both in humans and in artificial neural networks.**

***Index Terms*—fMRI, recognition, facial expression, conditions of indeterminacy, minimum information, pixels, voxels, biological and artificial neural networks**

I. Introduction

There is special interest in pattern-recognition tasks in recognizing a person's facial expression. Under threshold conditions, the interpretation becomes ambiguous when the facial expression changes, a classical example of which is the indeterminacy of La Gioconda's smile. Leonardo da Vinci brought the variation of the curve of her lips, imitating a smile, to the threshold value. With threshold values of a signal (a smile) under the influence of external noise of the visual system, her facial expression produces indeterminacy in the observer (is she smiling or isn't she?). Thus, the research problem consisted of using the following methods of neuroiconics: digital image processing, functional magnetic-resonance tomography (fMRT) of brain activity, psychophysical and neurotechnology methods to determine under conditions of indeterminacy the minimum information needed to recognize faces with different intensities of the facial expression.

This work was supported by the Program of Fundamental Scientific Research of State Academies for 2013–2020 (GP -14, section 63).

O.V. Zhukova is with Vision Physiology Department, I.P. Pavlov Institute of Physiology, Russian Academy of Sciences, St-Petersburg, 199034, Russia (e-mail: volgazhukova@gmail.com).

K. Yu. Malakhova is with Vision Physiology Department, I.P. Pavlov Institute of Physiology, Russian Academy of Sciences, St-Petersburg, 199034, Russia (e-mail: katerina.malahova@gmail.com).

II. Methods

A. Methods and instruments of a Neurophysiological Series of Studies.

Twenty nine volunteers took part in the psychophysical research. Fourteen persons with visual acuity of at least unity, right handed, with no neurological pathologies, took part in the fMRT. All the volunteers agreed to participate in the experiment.

361 images of faces of virtual people were used as stimuli. Two forms of emotions were studied, gladness and sadness, and two forms of rotation, to the right and the left of the same face. The emotion and rotation-morphing scales consisted of ten gradations. The other intensities of the attribute were the same as on the intensity scale for the attribute of emotion. A set of stimuli of a low degree of the emotional intensity of gladness and sadness was thus obtained, since most of the gradations are displaced into the region of minimal values. Three ways were used to present an image with additive noise (covered 30%, 50%, 70% of the image). For the fMRT research series, only four intensity gradations of rotation and emotion were used, and noise was superimposed on the stimulus that covered 50% of the image of the faces. In terms of this study, all image manipulations are carried out on the basis of the FACEGEN software (Singular Inversions, Canada).

The subjects were instructed to determine the facial expression or rotation of the head of the virtual person. The choice was recorded by the subject pressing the left or right mouse button

In the psychophysical series, the stimulus images were presented on a notebook screen (Sony Vaio VPC-F13S1R/B, clock frequency 2667 MHz). In the fMRT series, the visual stimuli were presented by means of a BENQ PB 8250 XGA multimedia projector located outside the chamber (frame-scan rate 60 Hz) on a semitransparent screen. A Siemens-Symphony high-field magnetic-resonance tomograph (1.5 T) with 40-mT/m gradients was used. Photographs with a size of 2.5° appeared at the center of the screen. The technique used to obtain images of the activation patterns was echo planar imaging (EPI), which made it possible to obtain an image of the entire brain (21 cuts) with a 64 × 64-pixel array in 3 s. The subjects held the computer mouse in their right hand. The presentation time of the stimuli was 1000 ms.

B. Methods and instruments of a Neurotechnological Series of Studies.

In the neurotechnological series of the study the VGG Face deep neural network was used as the basic model, trained to the task of recognizing well-known people on a data network composed of 2.6 million images collected from the Internet. This network was chosen because of the assumed relevance of the attributes chosen during the identification of faces to the task of recognizing facial expressions. The neural network was trained on a specialized data set of smaller size that includes images of virtual and real faces with different degrees of facial expression and different rotations of the head. The training data set consisted of two classes of 500 stimuli each. The first group included images of faces with a glad expression, and the second group included images of faces with a sad expression. In the process of training the network, the data were broken up into two groups: a training set, used for retraining the network, and a test set, needed for monitoring the training quality. Images of faces presented in portrait paintings were chosen as a control sample by expert evaluation; an example is the image of La Gioconda by Leonardo da Vinci. The problem of this series of experiments was to learn howan artificial neural network trained to recognize a smile and sadness behaves under threshold (indeterminacy) conditions.

III. Results

1) It is shown that there is a complex system that jointly operates the zones of the brain, each of which makes its own specific contribution to the accomplishment of mental processes. We demonstrated the activation of the classical cortical "face" areas (fusiform gyrus, BA37) and in the frontal (BA10, BA9, BA8, BA32), parietal (BA40, BA7), and premotor (BA6) cortices of the human brain. Each of these zones of the brain can be involved in the implementation of various functions, depending on the instruction and the experimental conditions. So, the results actually showed that the dorsal visual pathways are activated in the task of recognizing the rotation of the head, and the "emotional regions" are activated in the task of recognizing a facial expression. It should be noted brain activation patterns in our experiments contain 'mirror neurons'. They were activated not only in the process of observation and imitation, but in the process of evaluating the actions of another person and determination of emotions.

2) Difference-recognition thresholds were established as a result of analyzing this research, and it was shown that there is correlation between the number of altered pixels in the test images and the correct answers of the observers. Thus, the number of altered pixels in the test images of faces being compared is proportional to the correct answers when a decision is being made in the psychophysical studies and to the number of altered voxels in the images being compared in maps of the brain responses from the data of the fMRT studies. Important to emphasize, that the subjects had difficulty recognizing facial expressions and rotations under indeterminacy conditions and this caused a slower reaction rate. The probability of recognizing the facial expression decreased to the level of random guesswork—50%.

3) Testing the recognition of the facial expression by means of the artificial neural network showed a result of categorizing the images of faces with a pronounced smile equal to 97%. Under conditions of indeterminacy or a scarcely noticeable smile, the recognition probability decreased to the level of random guesswork—50% (is it a smile or not?). For instance, when the indeterminate facial expression of La Gioconda was discriminated, it was shown that her face expressed a smile with probability 0.69.

IV. Conclusion

The methods of visualizing and analyzing the activation maps of large-scale neural networks thus made it possible to study the processes within biological neural networks and to compare them with neurotechnological data. Because artificial neural networks are based on certain features borrowed from living visual systems, similar results were obtained in certain tasks with respect to accurate recognition of test images. The probability of recognizing a smile is thus fairly high under ordinary conditions, but it decreases under indeterminacy conditions (threshold and noisy images) both in humans and in artificial neural networks. For instance, the recognition of a smile in La Gioconda's facial expression by a person and by an artificial neural network occurs with probability 0.69. We assume that the most important operating principle in both networks is a matched-filtering mechanism as a measure of how well the presented image corresponds to a pattern learned by the neural network—in particular, a smile.

Acknowledgment

The study was carried out with the support and active participation of the lab head and staff of the laboratory of physiology of vision of the I. P. Pavlov Institute of Physiology, Russian Academy of Sciences, St. Petersburg, Russia: Yu.E. Shelepin, P. P. Vasilyev S.V. Pronin, as well as employees of the National Medical Research Center V.A. Almazov, St. Petersburg, Russia: V. A. Fokin, and A. V. Sokolov.