



# 2022 NEUROTECHNOLOGY AND SENSORY COMMUNICATION

Russian Academy of Sciences  
I.P. Pavlov Institute of Physiology  
«Neuroiconica Assistive» Company Limited

# NEUROTECHNOLOGY AND SENSORY COMMUNICATION

St.–Petersburg, 2022

## **Reviewers:**

**Mikhail Kupriyanov**, D. Sc., Professor, Head of Scientific and Educational Areas, Head of the Department of Computer Science and Engineering, St.Petersburg Electrotechnical University "LETI".

**Sergei Stafeev**, D Sc, Professor, Professor at Faculty of Physics and Engineering, Chief Expert at ITMO.KIDS Technopark, Head of Office of Educational Psychology, Professor of Institute for cognitive studies St. Petersburg State University

## **Authors:**

Alekseenko S., Antipova A., Babenko V., Basova O., Belokopytov A., Bobkov D., Boiko A., Bondarko V., Chikhman V., Danilova M., Fadeev D., Glasman K., Grinenko E., Egorova V., Ermakov P., Gerasimov A., Gorbenkova O., Gorelik S., Gracheva M., Iourov I., Krasnoschekova E., Kharazova A. Khaskelberg E., Khlynov R., Kolotii A., Korotaev V., Kotelin V., Kupriyanov I., Lebedev V., Lekhnitskaya P. Leshchinskiy V. Malashin R., Meleshenko E., Maximov P., Merkulyeva N., Mikhalkin A., Mikheev S., Morozov S., Murav'eva S., Naumova D., Nazaralieva E., Neroeva N., Nikolaev D., Rodionova I., Ryzhova V., Schemeleva O., Shebut D., Shelepin E., Shelepin Yu., Shevtsov M., Sizova O., Skuratova K., Solnushkin S., Timofeev V., Tsapenko I., Tsvetkov E., Vasiliev P., Vershinina E., Veshchitskii A., Vorsanova S., Volkov V., Yavna D., Yurov Yu., Zabrodskaya Yu., Zhuravleva A., Zueva M.

**NEUROTECHNOLOGY AND SENSORY COMMUNICATION, St.- Petersburg, 2022,**  
VVM Publishing llc., 2022. 318 p.

**ISBN: 978-5-9651-1483-2**

**Editor: Yuri Shelepin**

UDC (Universal Decimal Classification) 001 Science and knowledge in general, 004 Computer science and technology, 159.91 Psychophysiology, 612.8 Neurophysiology and Sensory perception.

The title of the book reflects traditional integration between sensory physiology and communication technology. Neurotechnology as a part of information technology was based on sensory physiology, decision making and planning the movements for goal achievement. Neurotechnology is a part of the creation of the artificial intelligence systems, the convergence of man and autonomous intelligent artificial devices with targeted activities. This book reflects a wide range of investigations from biological and artificial neural networks to rehabilitation neurotechnologies for neurological patients, with sensory and cognitive dysfunctions. The most part of this book has been discussed by the authors at the IEEE International Conference «Video and Audio Signal Processing in the Context of Neurotechnologies» November 7–11, 2022 at I.P. Pavlov's Institute of physiology RAS.

The edition of this book got the financial support from «Neuroiconica Assistive» Company Limited. St. Petersburg, Russia,

Synthesis of the image on the cover: "Touching Invisible– The Enigma of Sensory Communication", was carried out by Evgeny Shelepin using Midjourney 5 AI.

# CONTENTS

Chapter 1. How is the 3-dimensional space represented in the primary visual cortex in strabismus? <i>Alekseenko S. V.</i> .....	5
Chapter 2. Comparative analysis of neural network and human attention patterns in image classification. <i>Antipova A.</i> .....	12
Chapter 3. Stimulus display with extra wide color gamut for color discrimination thresholds assessment by strict substitution method. <i>Belokopytov A.</i> .....	24
Chapter 4. Perception of Classical Architecture: Psychophysical Researches and Aesthetics. <i>Bondarko V., Danilova M., Solnushkin S., Chikhman V., Shelepin Yu.</i> ...	36
Chapter 5. Cross-modal interaction in the integral assessment of image and sound quality. <i>Glasman K., Grinenko E., Shelepin Yu.</i> .....	43
Chapter 6. Genomics of cognitive functions: from molecular instability to mental disability. <i>Iourov I. Y.</i> .....	127
Chapter 7. Monitoring of psychophysical reactions of a person in the course of professional activity. <i>Gorelik S., Khaskelberg E., Leshchinskiy V.</i> .....	137
Chapter 8. Visual perception of microscopic objects. <i>Gerasimov A., Shevtsov M., Bobkov D., Zabrodsкая Yu., Nazaralieva E., Kolotii A., Yurov Yu., Vorsanova S., Iourov I.</i> .....	152
Chapter 9. Assessment of color discrimination thresholds by the strict substitution method. <i>Gracheva M., Belokopytov A., Timofeev V., Basova O., Maximov P., Nikolaev D.</i> .....	162
Chapter 10. Non-visual eye movements during memorization, long-term and short-term memory retrieval on the blank display. <i>Lekhnitskaya P.</i> .....	176
Chapter 11. Least action classifiers. <i>Malashin R., Boiko A.</i> .....	187
Chapter 12. The effect of transcranial Direct Current Stimulation on the electrical activity of the humans' cerebral cortex during facial expression emotion recognition. <i>Meleshenko E. A., Egorova V. A., Ermakov P. N., Gorbenkova O. A.</i> .....	203
Chapter 13. Features of Visual Information Processing in Patients with Schizophrenia. <i>Murav'eva S. V., Schemeleva O. V., Lebedev V. S., Vershinina E. A.</i> ...	217
Chapter 14. Functional near-infrared brain spectroscopy for ergonomics and clinical practice. <i>Rodionova I., Korotaev V., Shebut D., Ryzhova V., Khlynov R., Mikheev S., Shelepin Yu., Morozov S., Vasiliev P.</i> .....	224
Chapter 15. Features of the perception of emotionally colored images by adolescents and young people depending on the presence of social scenes on them. <i>Skuratova K., Shelepin E., Naumova D.</i> .....	235
Chapter 16. Cortical mechanisms of speech comprehension: from index sign to symbolic sign. <i>Sizova O. B.</i> .....	242
Chapter 17. Neural mechanisms of emotional memory. <i>Tsvetkov E., Krasnoschekova E., Kharazova A.</i> .....	253

Chapter 18. Analysis of the spinal neuronal populations in condition of high background neuropil staining using the “Cell Annotation Software.” <i>Veshchitskii A., Mikhalkin A., Merkulyeva N.</i> .....	271
Chapter 19. Hybrid algorithm for optical and SAR image registration based on U-Net neural network and keypoint-based refinement. <i>Volkov V.</i> .....	280
Chapter 20. The role of spatial modulations of local visual features in a saliency map formation in humans. <i>Yavna D., Babenko V., Kupriyanov I.</i> .....	292
Chapter 21. Visual Rehabilitation Technologies Based on Neuroplasticity. <i>Zueva M. V., Neroeva N. V., Kotelin V. I., Zhuravleva A. N., Tsapenko I. V., Fadeev D. V.</i>	306

# Chapter 1. How is the 3-dimensional space represented in the primary visual cortex in strabismus?

*Alekseenko S. V.*

*Pavlov Physiology Institute Russian academy of sciences, St.Petersburg, Russia*

## Abstract

Deviation of the eye from the normal position (strabismus) causes a loss of stereoscopic vision, which based on visual cortical neurons that are sensitive to the disparity of images on two retinas. Such neurons encountered in the primary visual cortex, where they tuned to the position of objects relative to the bifixation point. The aim of this paper is to create maps of projections of dot-objects of 3D space in this cortical area patients with strabismus caused by optical and anatomical disorders.

## Introduction

Changes in eyes alignment (strabismus) at an early age affect the visual representation of the position of objects in the near surrounding 3-dimensional space. This is due to the fact that stereoscopic vision is lost or decreased, the squinting eye eventually becomes amblyopic, the parameters of eye movements and crowding zones change, an asymmetry of optokinetic nystagmus appears, etc. (Daw, 2006). The stereoscopic vision based on visual cortical neurons interconnections that are sensitive to the disparity of images on two retinas, which occurs due to the distance between the eyes (Barlow et al., 1967; Nikara et al., 1968; Poggio, Fischer, 1977; Freeman & Ohzawa, 1990; Prince et al., 2002). Such binocular neurons estimate an image mismatch of 10–20 arc seconds (Parker et al., 2016). Disparate neurons first encountered in the primary visual cortex, where they are tuned to the position of objects relative to the gaze bifixation point (Cumming, Parker, 1999; Roe et al., 2007; Gardner et al., 2008; Alekseenko, 2015). The signals from these neurons used the oculomotor system to change the position of the fixation plane and to fuse the monocular images perceived separately by the left and right eyes into a single visual percept (binocular fusion).

Ophthalmologists distinguish many forms of strabismus: convergent, divergent, vertical, concomitant and incomitant (paralytic), torsional etc. This is understandable, because the position of the eyes in the orbits, the direction of gaze depends on the state of many anatomical structures (the lens, the cornea, six eye muscles, four oculomotor nuclei) and in each of them, as well as in the structures of their management, disturbances leading to strabismus are possible.

Despite the variety of strabismus forms, clinical methods of correcting the position of the eye are provided by different procedures of glasses or contact lenses supplemented, eye muscle surgery, refractive surgery, and by functional training (Сергиевский, 1951; Рожкова, Матвеев, 2007; Жанро, Дюкре, 2022).

These methods of treatment directly indicate that there are fundamentally different optical, anatomical and functional defects in concomitant strabismus. The manifestation of these defects, leading to mismatch of information coming from the two retinas, is manifested at different child ages. If congenital concomitant strabismus is caused by disorders leading to anatomical defects such as strength of the external eye muscles or different refraction of eyes, then strabismus will be visible during the first 3.5–6 months, i.e., during beginning formation of stereoscopic vision (Fox et al., 1980; Petrig et al., 1981, Atkinson, 1984; Braddick et al., 1980). If the child has accommodative disorders leading to different refraction of the eyes due to differences in the curvature of the lenses, corneas, then strabismus will develop at the

age of 3.5–6 months. In this postnatal period, monocular structures that provide visual acuity of the eyes, cortical structures of perception of information from different eyes (ocular dominance columns (Hubel, Wiesel, 2006)), as well as mechanisms for controlling vergent and versional eye movements develop. By the age of stereo vision formation, their development reaches a level sufficient to guide the optical axes of the eyes to a common point—bifixation, which is necessary for projecting images of objects on the corresponding points of the retinas (Daw, 2006).

The occurrence of strabismus, as well as its treatment, primarily leads to a change in the structure of projections of objects in the primary visual cortex. Maps of cortical representation of objects in the cases of strabismus with anatomical or optical defects have not yet been demonstrated.

**The purpose** of this work is to create and compare maps of projections of objects to the primary visual cortex in individuals with various forms of strabismus and to assess the possibilities for the formation of binocular neurons in this cortical area.

## Methods

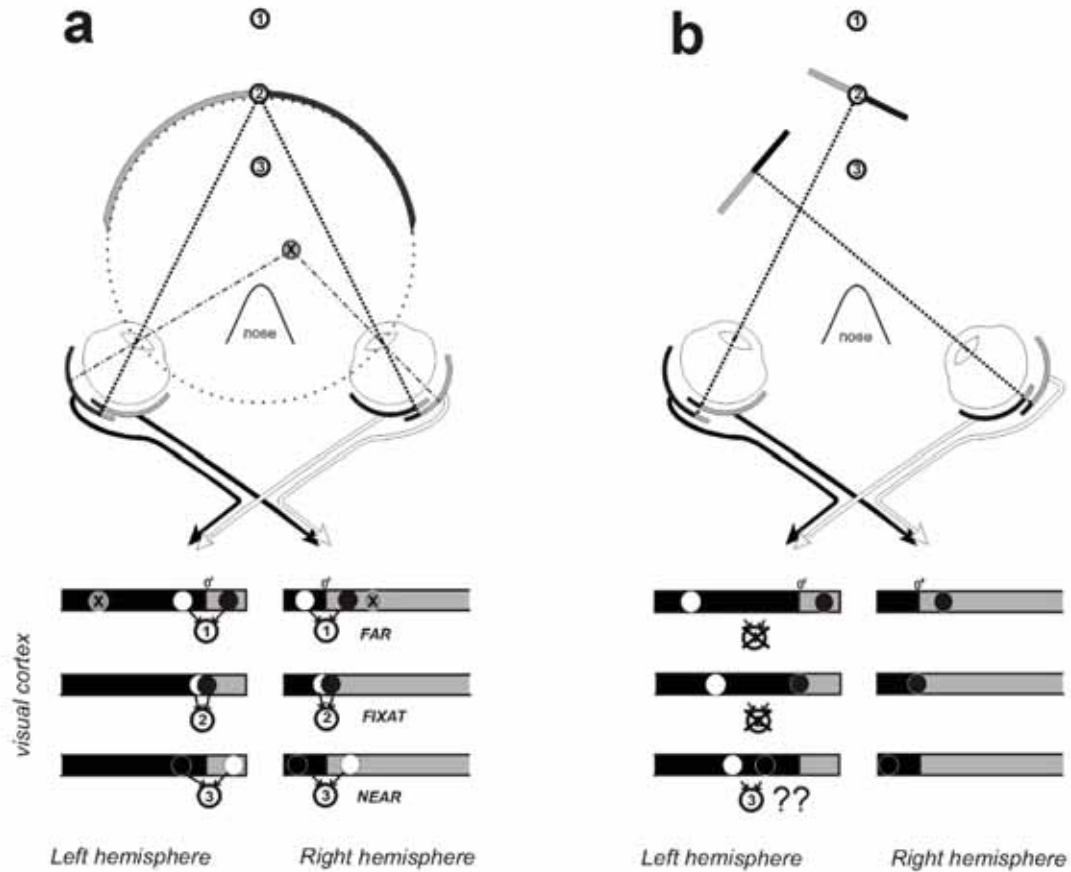
Based on the known data on the course of visual pathways from the retina, the position of dot-objects projections via both eyes to the primary cortical area was determined. This mapping method was described earlier (Alekseenko, 2015). For a brief illustration of the method, the projections of the dot-object X are shown in fig. 1a. The optical axes of the two eyes intersect at the point of gaze bifixation through which the fixation plane passes. The visual pathways from the retina are divided into two streams directed to different hemispheres. Due to the separation of the retinal paths, one visual hemifield of both eyes and a small part of the opposite hemifield, which is the projection from the retinal zone of naso-temporal overlap, are retinotopically represented in each hemisphere.

As can be seen in fig. 1a, information about object X projected onto the temporal halves of both retinas is transmitted along the ipsilateral fibres of these pathways. The lower part of the fig. 1a shows the localization of the projections of the object X on rectangular diagrams of parts of cortical area in both hemispheres in which the horizontal meridian of the visual field is represented. The projection of object X on the left retina is located farther from the optical axis of this eye than on the right retina. Accordingly, monocular projection of object X in the left hemisphere is farther from the projection of the boundary between visual hemifields than in the right hemisphere. However, dot-objects projected onto the retinal zone of the naso-temporal overlap presented in both hemispheres.

## Results

For all forms of strabismus, maps of projections from different eyes to the primary visual cortex V1 were constructed for the same three dot-objects located at different distances from the nose in the sagittal plane of space. Their positions were chosen on purpose, because they are located in the spatial loci to which binocular neurons of known groups are tuned: FAR, NEAR or binocular neurons tuned to objects in the plane of fixation (Poggio, Fischer, 1977).

Figure 1 shows maps of projections of these objects in the visual cortex in the case of normal alignment of eyes (a) and in a case with convergent strabismus (b) caused by an anatomical defect (for example, the internal eye muscle of the right eye is shortened or the asymmetric attachment of the muscles is incorrectly attached to the eyeball).



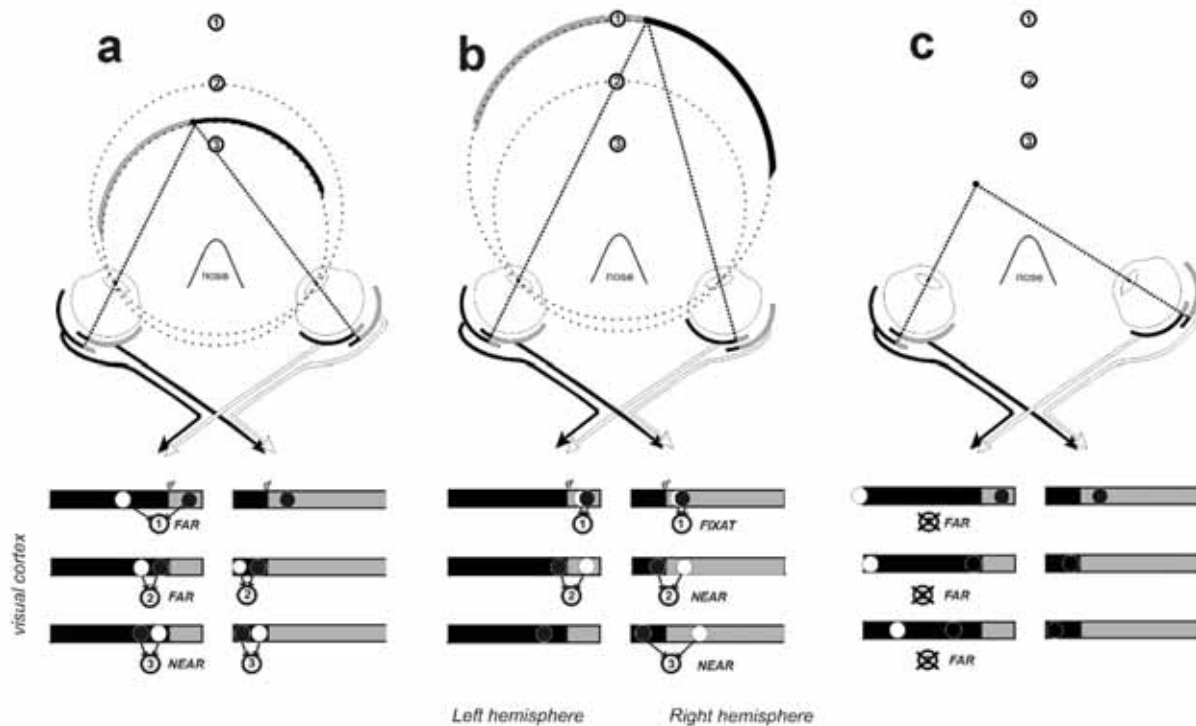
**Figure 1.** Maps of projections of 3 dot-objects to the primary visual cortex in the case of normal eye alignment (a) and convergent strabismus due to the anatomical defect (b). Rectangular diagrams are parts of cortical area where the horizontal meridian of the visual field is represented. FAR, NEAR, FIXAT—binocular cells tuned to the objects.

In the norm (Fig. 1a), object 2 is located at the bifixation point on the frontal plane, object 1 is located behind this plane, object 3 is in front of it. It can be seen that the minimum distance in the cortex between projections through the left and right eyes is observed in object 2 located at the bifixation point. At the same time, the distance between projections for objects located closer and further than the bifixation point is increased. In the zones of these projections, there are monocular neurons that receive inputs from different eyes, and as a result of the convergence of axons (outputs) of these neurons, binocular neurons are formed, tuned to the position of objects 1–3. Such binocular neurons are selective to disparities in the range from  $-1$  to  $+1$  deg (Anzai, De Angelis, 2010). The length of intra-areal neuronal connections (4–5 mm) limits the formation of BNs that are selective to higher disparity values.

In the case of strabismus caused by a congenital anatomical defect of the right eye (fig. 1b) the optical axes of the eyes intersect not at the point of gaze fixation, but in a different position. Therefore, the projections of objects through different eyes to the cortical area do not correspond to each other. The distance between the monocular projections of objects 1 and 2 in the cortex of the left hemisphere is large and exceeds the length of neuronal axons necessary for the formation of binocular neurons tuned to the position of these objects. It is quite possible to form a neuron tuned to object 3. This may indicate the presence of stereo vision in a sepa-

rate sector of space. However, it is obvious for dot- objects only. Strabismus of this origin is corrected in the clinic by surgery of the eye muscles. In addition, if the operation is carried out during the period of plasticity of neural connections (up to 6 years), disparate neurons are formed and children have stereoscopic vision.

Fig. 2 shows diagrams of projections into the cortex in children with convergent (a, c) and divergent (b) strabismus caused by differences in the focal lengths of the optical apparatus of the eyes (optical defects). In these diagrams, the right eye is deviated. In children with differences in eye refraction at the age of the onset of the formation of stereoscopic vision (from 3–4 months of age) and lasting up to 3 years (Daw, 2006). There is a gradual deviation of one eye, aimed at combining the fixation points of both eyes, which is necessary for the formation of an ordered system of binocular neurons.



**Figure 2.** Maps of 3 dot-objects projections to the primary visual cortex in the case of accommodative strabismus. (a)- convergent strabismus with small angles of eye deviation, (c)- with large angle, (b)–divergent strabismus. For most cases of even mild strabismus, only a hypothetical horopter can be drawn.

In this case, the eye and, accordingly, the fixation point of the gaze deviates towards the shorter focus eye, which leads to convergent or divergent strabismus. This is confirmed by clinical observations: farsightedness of the squinting eye (i.e., long-focus eye) is usually combined with a convergent type of strabismus, and myopia with a divergent one. The angle of eye deviation depends on the differences in the focal lengths of the optical apparatus of the eyes. With strabismus, the width of the binocular sector of the visual fields also changes—the sector narrows with convergent strabismus and expands with divergent.

We see by comparison of the diagrams that the frontal plane of space, on which the gaze bifixation point is located, approaches to the observer in the case of convergent strabismus and moves away with divergent strabismus. Such changes in the distances to the fixation plane

and a shift in the fixation point cause a different position of the monocular projections of objects in the primary cortex, however, the conditions for the formation of disparate neurons of all types are preserved (Fig. 2a, b). Under these changed conditions, other types of binocular neurons carry out tuning to objects position. For example, in the case of normal alignment of eyes, the spatial position of object 2 is estimated by a neuron tuned to the fixation plane, in the case of convergent strabismus it is evaluated by neuron *FAR*, and in the case of divergent strabismus by a neuron *NEAR*. This indicates that the information about the position of objects in 3D space changes with accommodative strabismus with a small angle of eye deviation.

At large angles of deviation of the eye (Fig. 2c), BN cannot form due to the increased distance between projections from different eyes.

In the clinic, for the treatment of such forms of strabismus, surgery is not used, the position of the squinting eye returns to normal when the focal lengths of the eyes are adjusted using glasses, contact lenses.

Thus, analysis of the location of monocular projections of dot-objects in the retinotopically ordered primary cortex showed that disparate neurons necessary for stereoscopic vision cannot be formed in strabismus. Only at small angles of strabismus caused by accommodative disturbances, such binocular neurons can be formed. However, the assessment of the position of objects in space by these neurons depends on the position of the fixation point relative to the observer, the position of which changes with convergent and divergent strabismus.

## Discussion

The study showed how the correspondence of information coming from different eyes is important for the structures of stereoscopic vision. The map of 3D space is disturbed even in the primary visual cortex with strabismus. Only at small angles of deviation of the eye during strabismus caused by accommodation disorders, in this visual area of the lowest level of the cortical hierarchy, conditions are created for the formation of all known types of disparate neurons. The preservation of stereo vision in children with “purely accommodative strabismus” at small angles of the eye deviation is noted in the guide to strabology (Жанро, Дюкре, 2022). At large angles of the eye deviation, as well as with other forms of strabismus, stereo vision was not detected. It is clear that these clinical data represent the result of the functioning of all visual areas, and not just the primary cortex.

The disparate neurons of the primary cortex are tuned to the position of objects relative to the point of bifixation. They respond only to the local correspondence between the two images. They responses to random-point stereograms and other specific stimuli that do not cause depth perception (Erkelens, Collewijn, 1985; Masson et al., 1997; Cumming, Parker, 1999, 2000; Miles, 1998; Busettini et al., 2001; Erkelens, 2001). That is, this cortex does not yet solve the problem of correspondence of features of two images, which is necessary for the perception of objects depth. However, such stimuli caused the monkeys to make vergent eye movements, which is important for changing the distance to the fixation plane.

The data obtained when comparing the assessment of the spatial location of the same dot-objects (1, 2, 3) by disparate neurons in accommodative strabismus (Fig. 2ab) and in the norm (Fig. 1a), made it possible to show changes in the types of neurons that were tuned to these objects depending on the eyes alignment—from strabismus to normal alignment. Similar rearrangements in the activation of cells in the primary cortex should occur in children in the treatment of strabismus caused by different refraction of the eyes. In addition to the redistri-

bution of activity in the family of disparate neurons, there are also changes in the activity of the oculomotor structures, which receive signals from the binocular neurons *FAR* or *NEAR*: vergent eye movements change the distance to the fixation plane and versional movements change the location of the bifixation point in this plane. Consequently, in the process of strabismus treatment, quite complex changes occur in many parts of the nervous system. Apparently, this explains the long duration of the process of stereofunction recovery in comparison with the period of their formation and the critical period of their disruption (Daw, 2006).

Note that the map of projections of objects into the cortex with paralytic strabismus wasn't built, because this form of strabismus occurs in neurological disorders (paresis, paralysis of visual and oculomotor structures of the brain). The surgical rotation of the eye does not eliminate pathology in innervation and provides only a cosmetic effect.

Each sensory system formed its representation of the surrounding space, and they are coordinated with the participation of motor control and reinforcement. The map of the visual space in the primary cortex are retinotopic, but not spatiotopic (Gardner et al., 2008). This retinotopic organisation, according to our results, is disturbed by strabismus. In this regard, the information sent by neurons of this area to subsequent visual and oculomotor structures during strabismus does not correspond to information about the spatial position of objects in other "healthy" for this patients sensory systems (auditory, tactile). Given this, the process of restoring stereo vision must be more effective in the plastic period of development of binocular visual neural systems and coordination with other brain structures.

*This study was supported by the State Program 47 GP "Scientific and Technological, Development of the Russian Federation " (2019–2030), theme 0134–2019–0006.*

## References

1. Жанро Н., Дюкре В. Руководство по страбологии. Клинические и терапевтические аспекты; пер. с фр. под ред. Рабичева И. Э. М: Практическая медицина, 2022. 232с.
2. Сергиевский Л. И. Содружественное косоглазие и гетерофории. М: МЕДГИЗ. 1951
3. Рожкова Г. И., Матвеев С. Г. Зрение детей: проблемы оценки и функциональной коррекции. М.: Наука, 2007. 315с.
4. Alekseenko S. V. Neuronal representation of 3-D space in the primary visual cortex and control of eye movements. *Perception*. 2015. 44(8–9): 995–1006.
5. Anzai A., De Angelis G. C. Neural computations underlying depth perception. *Curr Opin Neurobiol*. 2010 June; 20(3): 367–375. doi:10.1016/j.conb.2010.04.006.
6. Atkinson J. Human visual development over the first 6 months of life. A review and a hypothesis *Hum Neurobiol*. 1984. 3(2):61–74. PMID: 6378843
7. Barlow H. B., Blakemore C., Pettigrew J. D. The neural mechanisms of binocular depth discrimination. *J. Physiol*. 1967. 193: 327–342.
8. Braddick O., Atkinson J., Julesz B., Kropfl W., Bodis-Wollner I., Raab E. Cortical binocularity in infants. *Nature*. 1980. 288(5789):363–5. doi: 10.1038/288363a0. PMID: 7432532
9. Braddick O., Atkinson J. Development of human visual function. *Vision Res*. 2011. 51(13):1588–609. doi: 10.1016/j.visres. PMID: 21356229
10. Cumming B. G., Parker A. J. Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci*. 1999. 19: 5602–5618.
11. Cumming B. G., Parker A. J. Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *J. Neurosci*. 2000. 20: 4758–4767.
12. Daw N. W. Visual development. 2006. Springer, 268 p.
13. Erkelens C. J. Organisation of signals involved in binocular perception and vergence control.

- Vision Res. 2001.41: 3497–3503.
14. Erkelens C.J., Collewijn H. Eye movements and stereopsis during dichoptic viewing of moving random-dot stereograms. *Vision Res.* 1985. 25(11): 1689–1700.
  15. Fox R., Aslin R.N., Shea S.L., Dumais S.T. Stereopsis in human infants. *Science.* 1980. 207(4428):323–4. doi: 10.1126/science.7350666
  16. Freeman R.D., Ohzawa I. On the neurophysiological organization of binocular vision. *Vision Res.* 1990. 30: 1661–1676.
  17. Gardner J.L., Merriam E.P., Movshon J.A., Heeger D.J. Maps of Visual Space in Human Occipital Cortex Are Retinotopic, Not Spatiotopic. *J. Neurosci.* 2008. 28(15): 3988–3999. doi: 10.1523/JNEUROSCI.5476–07.2008
  18. Hubel D.H., Wiesel T.N. *Brain and visual perception.* New York: Oxford UP. 2005.
  19. Miles F.A. The neural processing of 3-D visual information: evidence from eye movements. *Eur. J. Neurosci.* 1998. 10: 811–822.
  20. Nikara T., Bishop P.O., Pettigrew J.D. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp. Brain Res.* 1968. 6:353–372.
  21. Parker A., Smith J.E.T., Krug K. Neural architectures for stereo vision. *Philos Trans R Soc Lond B Biol Sci.* 2016. 371(1697): 20150261. doi: 10.1098/rstb.2015.0261
  22. Petrig B., Julesz B., Kropfl W., Baumgartner G., Anliker M. Development of stereopsis and cortical binocularity in human infants: electrophysiological evidence *Science.* 1981.213(4514):1402–5. doi:10.1126/science.7268443. PMID: 7268443 DOI: 10.1126/science.7268443
  23. Poggio G.F., Fischer B. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol.* 1977. 40:1392–1405.
  24. Prince S.J., Cumming B.G., Parker A.J. Range and mechanism of encoding of horizontal disparity in macaque V1. *J. Neurophysiol.* 2002. 87: 209–221.
  25. Roe A.R., Parker A.J., Born R.T., DeAngelis G.C. Disparity channels in early vision *J. Neurosci.* 2007. 27: 11820–11831.

## **Chapter 2. Comparative analysis of neural network and human attention patterns in image classification**

*Antipova A.*

*Lomonosov Moscow State University, Moscow, Russia*

### **Introduction**

People are very good at creating or learning new concepts. With only a few examples, we can identify common content for a new concept. The central problem of cognitive science is how this process of conceptualization, and its underlying representations should be modeled. A related process is categorization, the ability to separate an object from its background and assign it to some class (Bruner, 1977). Research on this cognitive ability can be traced back to the works of Plato and Aristotle, but there is currently no theory capable of satisfactorily describing the flexibility of categorical organization (Boyarskaya, 2011). Nevertheless, the ability to effectively categorize objects is one of the most important cognitive abilities for successful interaction with the world, and therefore this ability is also actively studied in research on artificial intelligence. Thus, Velichkovsky points out that neural networks demonstrate categorization of stimulus situations, the ability to correctly “recognize” slightly modified versions of previously learned concepts (Velichkovsky, 2006). However, a comparison of human and neural network categorization abilities has not yet been carried out. Even though the principle of creating neural networks is biologically mimic and, therefore, we expect similar results when performing cognitive tasks, it seems important to practically compare the strategies of artificial intelligence and humans, since this may further limit the development of AI.

In the “strong-weak AI” paradigm, one of the main difficulties to overcome seems to be the specialized nature of modern forms of artificial intelligence, the ability to solve a single task, such as face recognition or playing chess. It seems that universal artificial intelligence (AGI) will be as multitasking as human intelligence. However, it is still not clear exactly how this can be modeled. One approach to solving the problem is to look for conceptual structures. A conceptual structure is the mental carrier of a concept, responsible for its assimilation, functioning, and change within individual mental experience. Concepts are the point of convergence of general patterns of conceptual thinking and the uniqueness of individual mental experience of a particular person. The conceptual structure is not only a mechanism of information processing, but also a means of its storage. This aspect of the organization of a concept can be described in terms of “semantic networks” and “semantic spaces” considering, however, the fact that the semantic network of each concept is a dynamic formation with constantly changing composition of semantic elements and moving boundaries. Thanks to conceptual structures a special “represented reality” is formed, which, according to J. Piaget, “becomes even more than reality, because a whole world of what can be constructed is discovered and because thinking becomes free in relation to the real world” (Piaget, 1968).

Conceptual structure consists of three mechanisms: semantization, categorization and conceptualization. Semantization is the ability to recognize the meaning of a word or expression and conceptualization is the ability to create certain ontological representations of some array of empirical data. These mechanisms appear to be the most difficult for modeling, although they include the ability to categorize, which is why we propose to begin modeling conceptual structures precisely with it. It is categorization that underlies such tasks as object recognition, object tracking, etc. Modern researchers of artificial intelligence contrast such processes as

intelligence and creativity, but the development of conceptual structures will remove this barrier, as creativity and intelligence are characterized by the number of conceptual connections and the speed of processing conceptual information.

Modeling the ability to efficiently operate on existing concepts and create new ones is one of the main tasks for researchers of intelligent agents (Zwarts, 2015). It is assumed that such a conceptual agent can create and modify conceptual frameworks to interact effectively with the world around it. The theory of concept spaces proposed by Peter Gardenfors in 2000 (Gardenfors, 2000) seems to us a great solution to the problem of structural storage of concepts, but the question of how to create concepts from unlabeled data remains open. A 2017 study shows that InfoGAN latent spaces can serve as domains in the concept space framework (Bechberger & Kühnberger, 2017).

Also, a common problem for the cognitive and computer sciences is the problem of mental representation: in what form a concept or category is stored in the brain or conceptual space. A mental representation is a common approach to explain and describe the nature of ideas and concepts in current philosophy, particularly in branches of metaphysics like philosophy of mind and ontology. A promising direction in this question is the study of the phenomenon of learning to draw (Pelayo & Fonseca, 2013). In the cognitive studies of drawing, the “beginner’s paradox” is well known. A novice artist draws not at all what he perceives visually, but rather seeks to display the general concept of the subject, which he uses in identifying things, while some children with autism spectrum disorders can draw realistic pictures without previous drawing training. It is assumed that in autistics some ability to generalize and conceptualize information is impaired, so sensory stimuli undergo minimal processing. Although drawing is not a direct reflection of those figurative processes that unfold in the psyche of the subject, it nevertheless can be considered as an indirect indicator of verbal-image translation in the structure of corresponding concepts. It is extremely interesting to search for an answer to the question “Why can’t I draw the world the way I see it?”.

However, for the time being practical research of conceptualization mechanisms was limited to psychological experiments within the framework of diagnostics of cognitive abilities of a patient. For example, the technique of formation of artificial concepts offered by German psychologist N. Ah and modified by Soviet psychologist L. S. Vygotsky for an assessment of “level of generalization” in preschool children (Vygotsky & Sakharov, 1998). At the moment, research on conceptual structures is closely related to the search for neural correlates of semantic spaces (Pulvermuller, 2013) but working with the human brain is a complicated process, so in our opinion, modeling conceptual networks is a more efficient job.

Thus, the direct study of conceptualization mechanisms is not yet available to us, but to identify the possibility of studying conceptualization processes at the level of universal abilities of intellectual systems, we propose to compare categorization ability of a human and a neural network. This study is unique in terms of considering categorization and conceptualization within a universal approach to the study of cognitive processes. We aim to consider cognitive and computer science problems on a transdisciplinary level.

## **Theoretical background**

To compare categorization ability, we propose to compare the attention patterns in an image classification task in a neural network and a human. Note that the search for attention patterns in the classification task in a neural network is an important task for the interpretability of the model prediction results. Interpretability of the model becomes necessary because of such

aspects as increasing confidence in the results when solving critical problems, increasing safety and the possibility of challenging the results of calculations. Such a task is of particular interest for such neural network architecture as transformer. A feature of the architecture is the use of the attention mechanism and high efficiency under conditions of parallelization. The Transformer model was first proposed in “Attention is All You Need” by Google developers in 2017 (Vaswani, 2017).

Neural network models are the basis for such a branch of the philosophy of artificial intelligence as connectionism and the main element in the structural approach to modeling the features of natural intelligence. The latest research shows that the architecture of the transformer is equivalent to the mathematical model of the EC-hippocampus. This “memory model” turns out to be equivalent to the self-attention module (Whittington & Warren, 2022). Also in a 2017 article, researchers found that human-selected ROIs on an eye tracker contain the most features for successful image classification by ML models (Zhou et al., 2017). Such studies show that neural networks, and the transformer architecture particularly, are most suitable for interdisciplinary studies of the universality of cognitive processes.

The subjects are offered a solution to an implicit image classification problem, which is not a standard eye-tracking problem. After presenting the image, the subject must assign it to one of the four proposed categories, in average two of which are high-level and two are low-level. We did not analyze the obtained data on the selected class in the experiment, since we specially selected images that are difficult to classify unambiguously, and there is no correct answer in this problem.

As a result, we obtain averaged heatmaps for each image with highlighted areas that a person or Vision Transformer (ViT) paid attention to when solving the classification problem.

## Methods

**Participants.** The experiment involved 25 subjects aged 18 to 29 years with normal or corrected to normal vision without mental disorders.

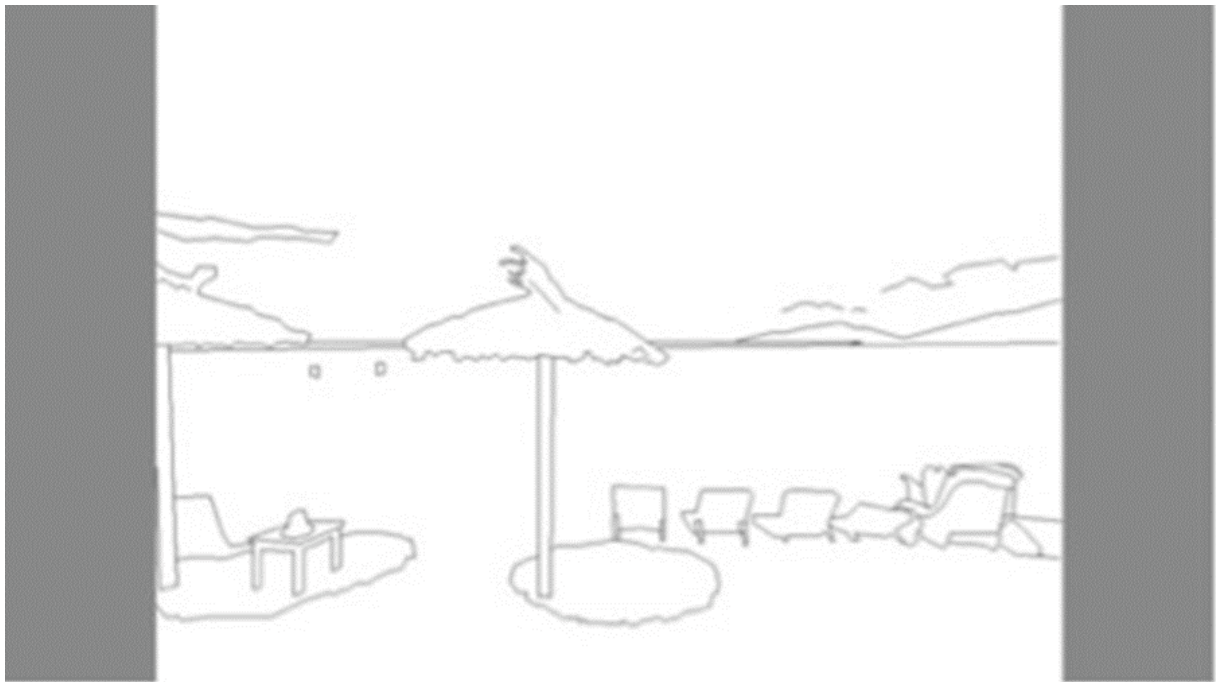
**Equipment.** The study was conducted using the GazePoint GP3 eyetracker, with a registration frequency of 60 Hz and a 9-point calibration before the start of the experiment. The design of the experiment was made on the Neurobureau platform.

**Dataset.** We used 51 images from the open eye-tracking dataset (Borji & Itti, 2015). The selected images belong to 7 different categories: “art”, “object”, “objects”, “landscape”, “poor image quality”, “abstract”, “drawings”. The choice of images from these categories, except for the “object” category, is due to the fact that it is difficult to classify them unambiguously. We believe that the absence of an obvious class to define equalizes the conditions for a human and a neural network trained to classify only 1000 classes. The choice of a relatively small number of images is since a larger number implies a greater cognitive load on the subjects and a decrease in the accuracy of the resulting heatmaps.

**Model.** Vision transformer (ViT) is an image classification model that uses a transformer-like architecture. The model uses self-attention—a kind of attention mechanism, the task of which is to identify patterns only between input data, which brings the model closer to the context in which the person is when solving the problem. The model is pretrained for 1000 classes, which mainly include animals, plants, and simple objects.



*Figure 1. Sample image from the CAT2000 dataset.*



*Figure 2. Sample image from the CAT2000 dataset.*

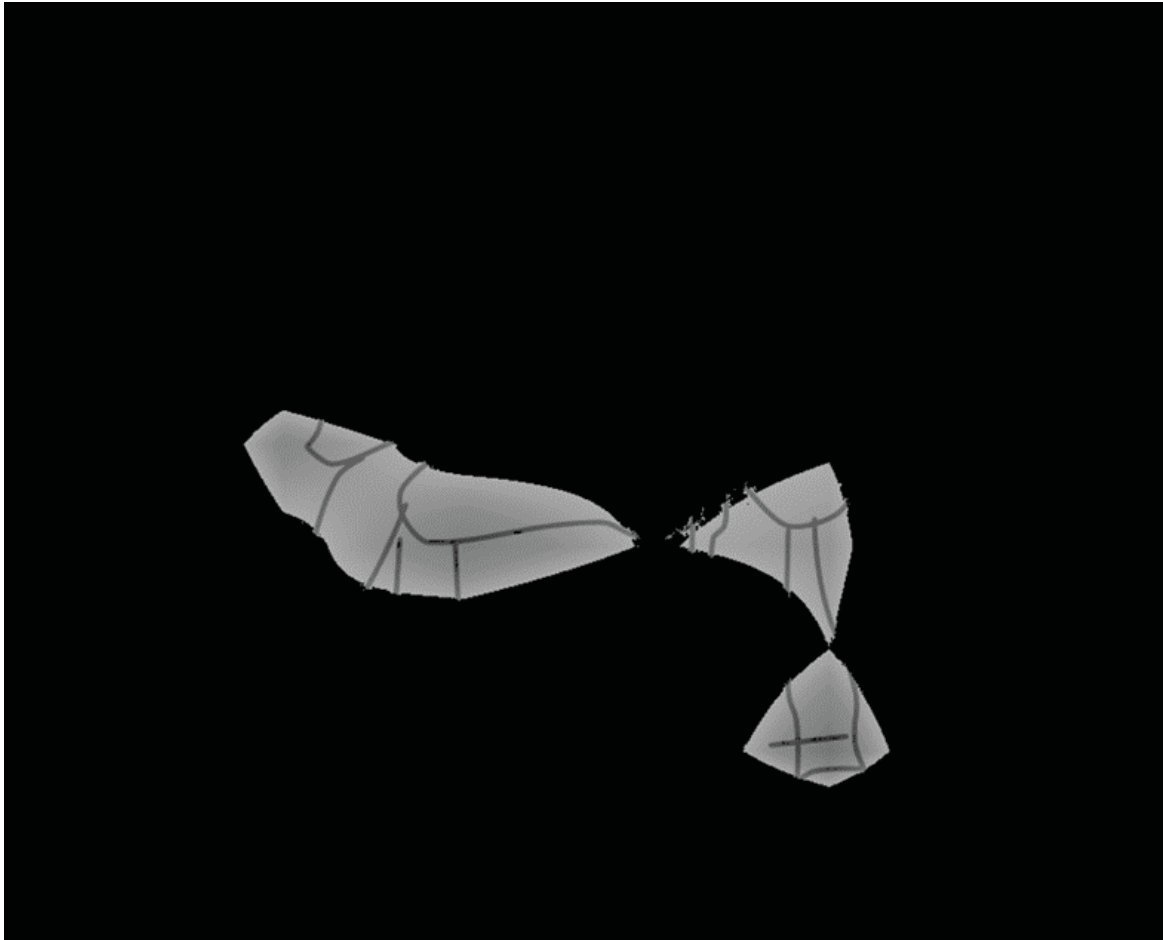
**Hypotheses.** We assume that the patterns of attention will be different for all categories, except for the “object” and, possibly, “drawings” categories, since in these categories, in our opinion, it is easy to separate the object from the background.

**Experiment design.** Stimuli from the CAT2000 dataset were presented to the subjects for 5 seconds, after which it was necessary to assign the image to one of the 4 proposed categories. For example, for the following picture such categories were selected: musician, man, male, guitarist. There was no time limit for category selection.

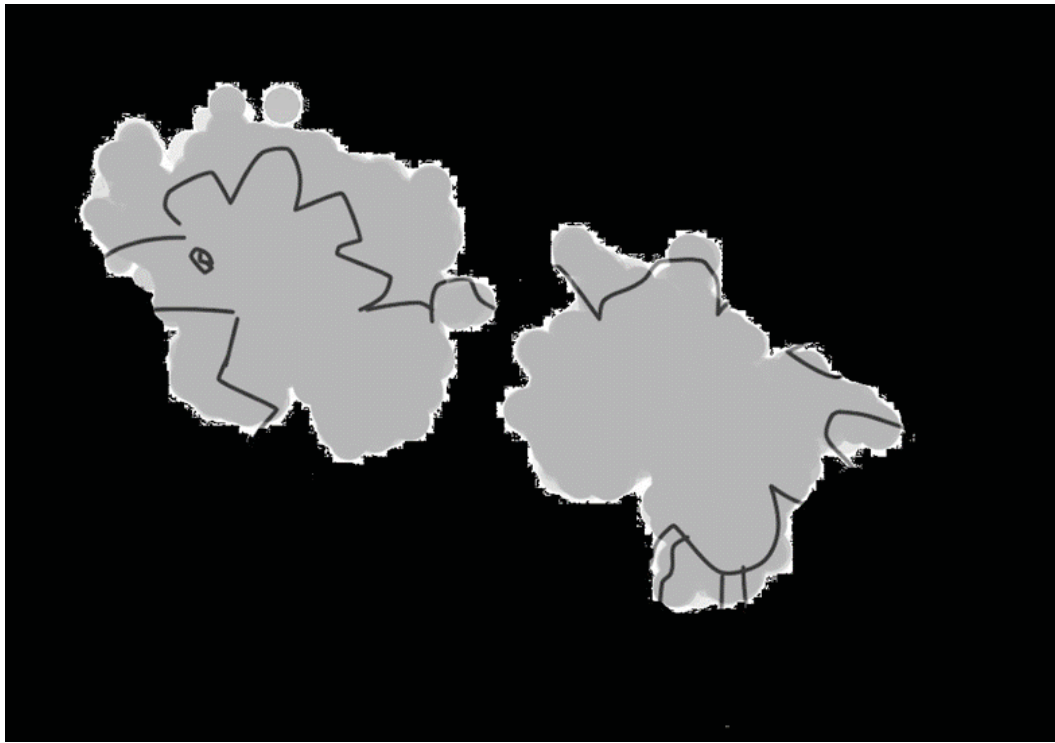


*Figure 3. Sample image from the CAT2000 dataset.*

**Method of analysis.** As a result, we got average attention heatmaps for each image. To compare images, we used a pixel-by-pixel comparison of normalized ViT and human attention heatmaps with color segmented portions of the highest attention in the image.



*Figure 4. ViT attention heatmap segmentation.*

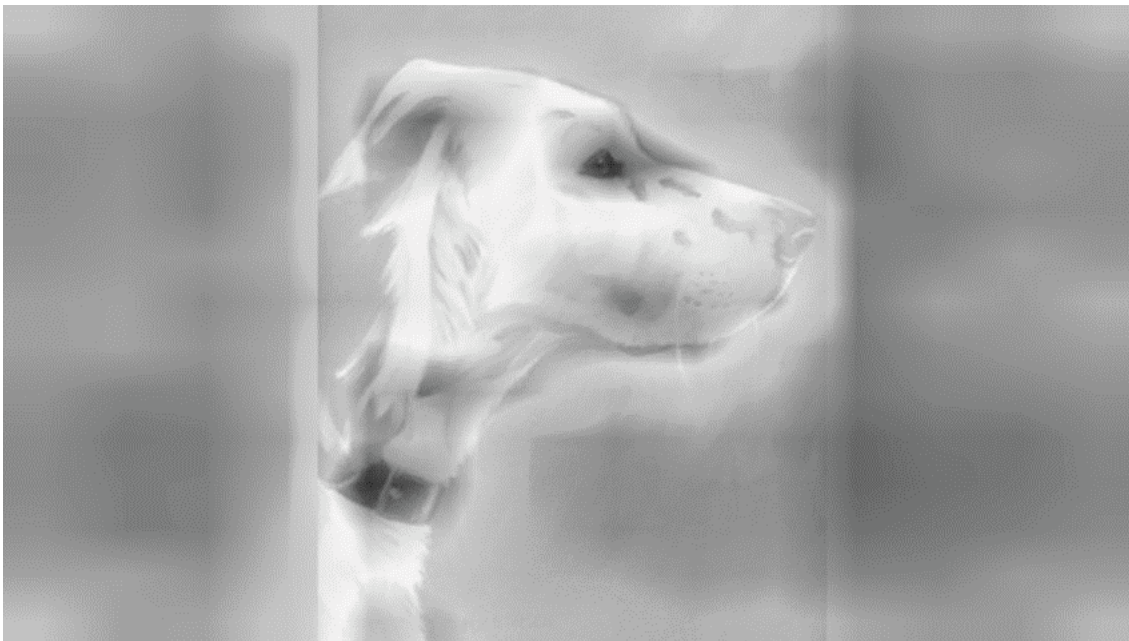


*Figure 5. Human attention heatmap segmentation.*

## Results

In the used dataset, 23 attention heatmaps turned out to be similar in terms of attention patterns and 28 images differed in these patterns. For a group of images with similar heatmaps, the following features can be distinguished:

1. The object/objects can be clearly distinguished in the image.
2. Presence of contrast zones.

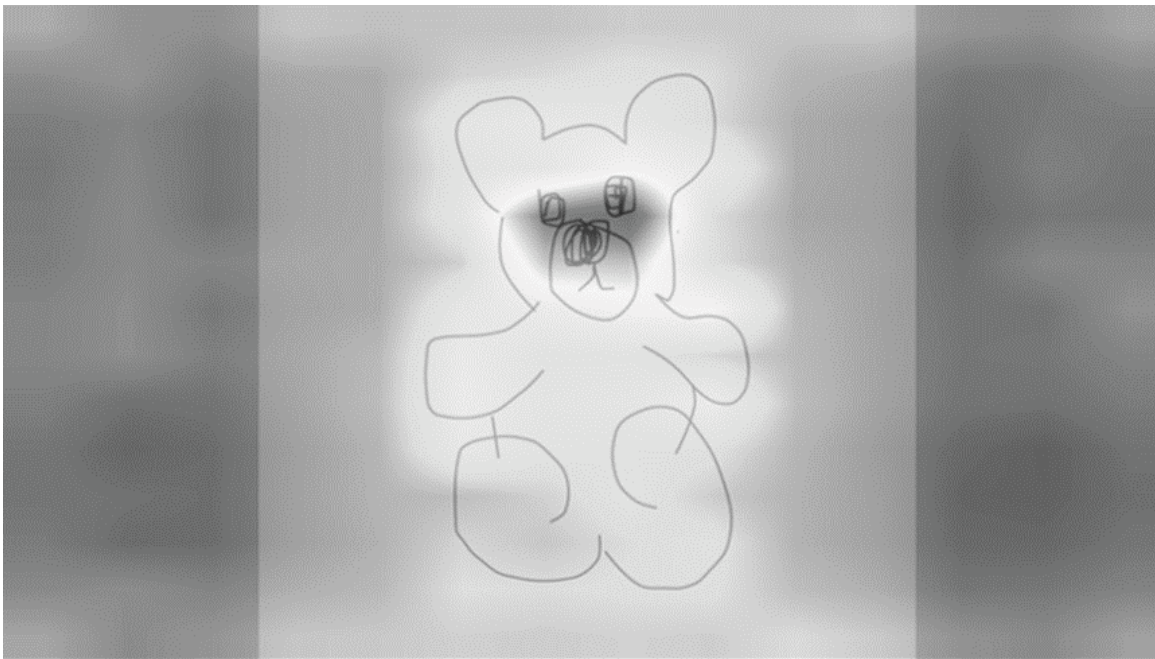


*Figure 6. ViT attention heatmap of the dog image.*

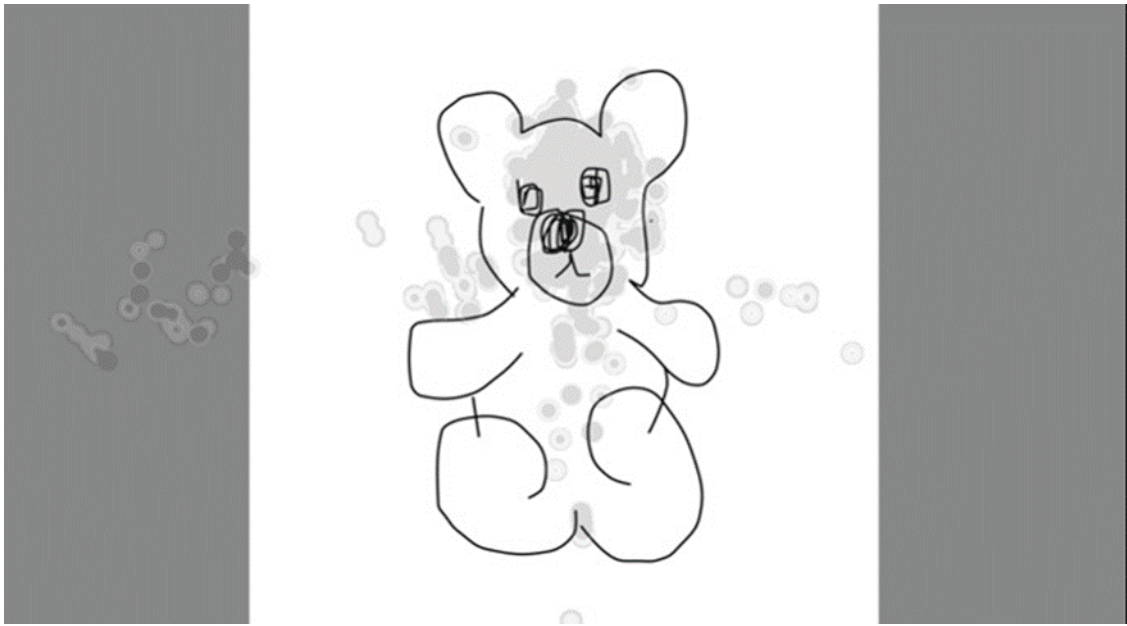


*Figure 7. Human attention heatmap of the dog image.*

It is interesting that the neural network paid attention to the same parts of the image from the “drawing” category as a person. Perhaps such representations of concepts in the theory of drawing can serve as an example of representations of concepts in the conceptual space.



*Figure 8. ViT attention heatmap of the bear image.*

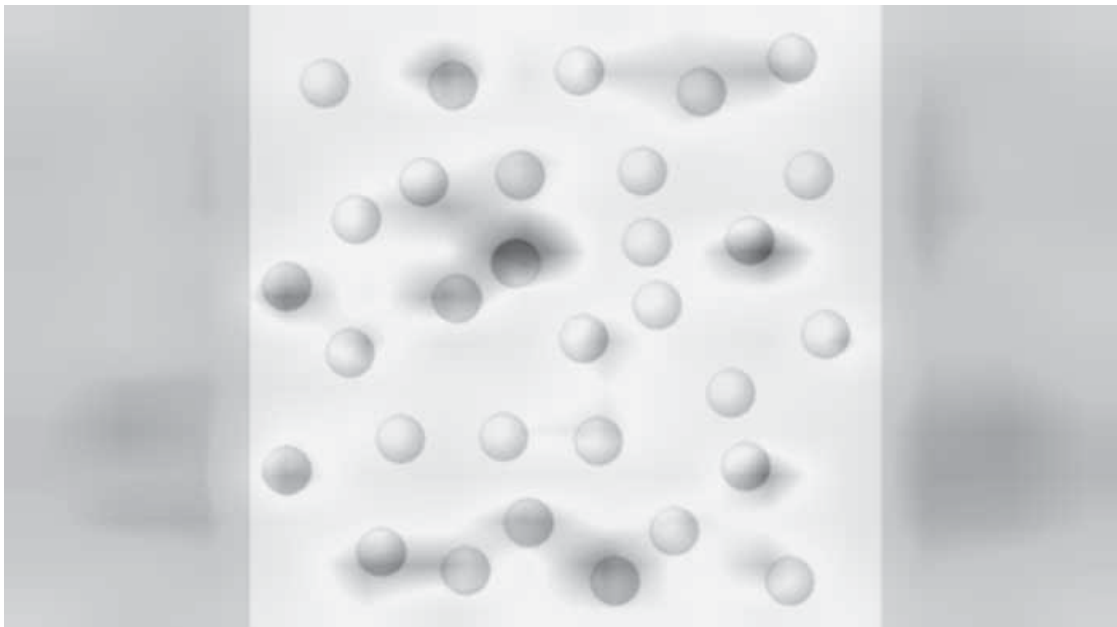


*Figure 9. Human attention heatmap of the bear image.*

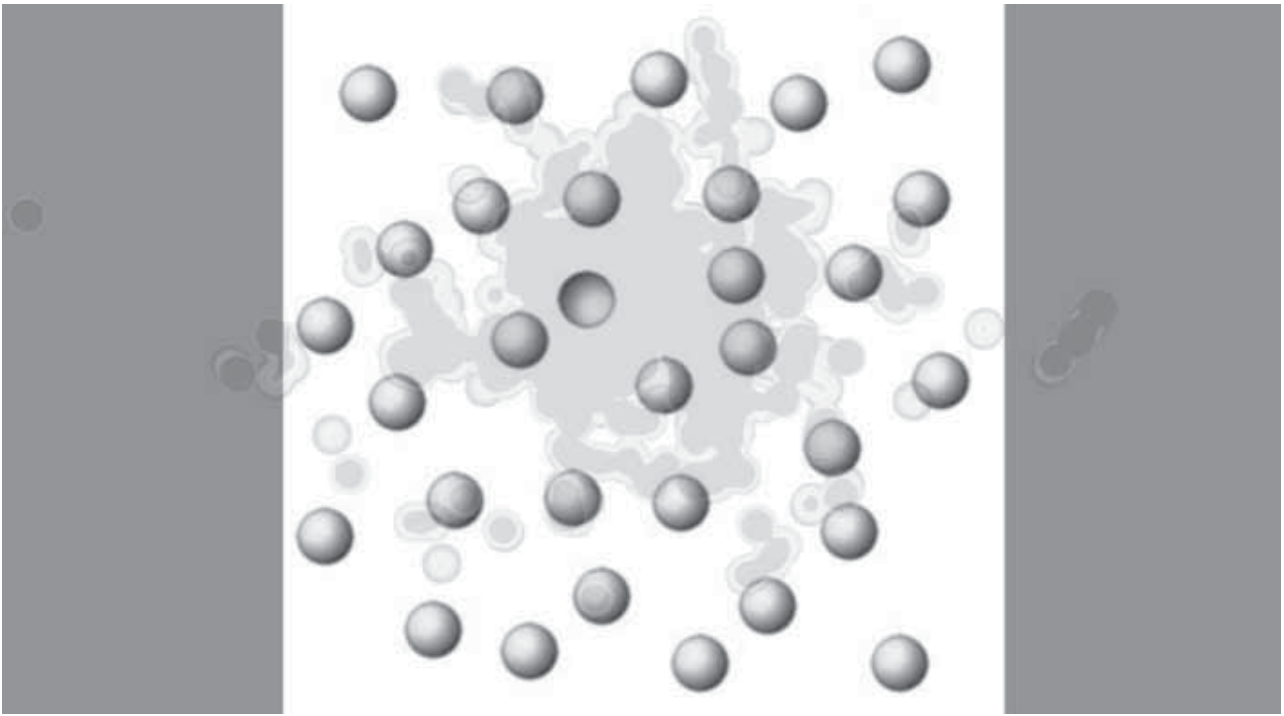
For different heatmaps, the following features can be distinguished:

1. Attention patterns are similar if the image is not uniform. By image uniformity, we mean the degree of uniformity in the distribution of similar objects throughout the image. If the image is homogeneous:

- The person looks to the center.
- The neural network pays attention to the most contrasting parts of the image.



*Figure 10. ViT attention heatmap of the balls image.*



*Figure 11. Human attention heatmap of the balls image.*

The difference in strategies can be explained by the peculiarities of human holistic perception: if there is nothing to “catch” on in the image, then the easiest strategy is to focus the gaze on the center of the image. So, with an embracing, or, in other terminology, ambient (Velichkovsky, 2006), way of perception fixation points accumulate in the central area of the image: “the whole is greater than the sum of its parts”. Due to the absence of such an effect in modern neural networks, it is necessary to “scan” the image in search of contrasting features for prediction.

2. The neural network ignores social stimuli that are important for a person: letters and faces. For successful social engagement, attention modulation to socially relevant stimuli is a crucial component. This qualitatively affects the final patterns of attention for a person: in most cases, in the presence of inscriptions or faces, attention is focused precisely on these parts of the image. Perhaps the absence of such bias is a useful feature, but for the development of social artificial intelligence the need to respond reactively to social stimuli becomes apparent. Characteristically, ignoring social stimuli is a diagnostic marker of autistic diseases. Note that now, the use of literature on psychological deviations in the study of neural networks is becoming increasingly popular (Mizutani, 2022). The most well-known phenomenon of this type is savantism—the presence of outstanding abilities in one or more areas of knowledge, contrasting with a generally limited personality. A person with the savant syndrome can repeat several pages of a text he has heard only once, accurately state the result of multiplication of multi-digit numbers or say on what day of the week May 3 of any year falls. Savants’ abilities are most often manifested in music, mathematics, drawing, calendar calculations, etc. (Bijakji et al., 2021).



*Figure 12. ViT attention heatmap of the people image.*



*Figure 13. Human attention heatmap of the people image.*

Low-level processes are fundamentally inaccessible to introspection, so savants cannot explain the source of their abilities. Alan Snyder suggests that such processes are deliberately inhibited to consciousness, and that an automation strategy speeds up decision-making, especially when providing partial information about the world. (Snyder, 2009). It is worth noting that savant skills are not related to the capacity for creativity; they do not depend on active learning, but on easy reading of raw information. Skills do not improve over time and disappear with the treatment of the disease. However, the cognitive strategy of savants has some advantages for problem-solving classes that involve breaking down cognitive distortions (sa-

vants, by concentrating on details, have no bias in problem solving) (Snyder et al., 2009). Thus, savants have privileged access to low-level processes of the nervous system. The criterion of high/low level in perceptual processes is the vector of information processing and it's mediated by experience. Processes affecting perception directly through the senses are called ascending processes, and processes affecting perception through knowledge, experience, and context are known as conceptually guided processes (Chistopolskaya et al., 2019).

We believe that the development of modern neural networks is at a similar savant stage. Neural networks have unlimited access to “sensory” information (all pixels in an image, frequencies in a piece of music, etc.) without top-down processes. Characteristically, neural networks are called imitators in the same fields in which savants are declared geniuses: mathematics, drawing, music, modeling, etc. Thus, it is established that conceptualization is a necessary condition for the development of artificial intelligence, and it is also possible to model it, given the fact that the ability to categorize in neural networks and humans is similar.

The limitations of the experiment include the presence of a small number of stimuli from different categories, although the resulting heatmaps are stable.

## Conclusions

As a result of the study, the features of the categorization processes in humans and the ViT neural network were established. The obtained results suggest that the main mechanisms (separation of the object from the background, paying attention to the contrasting parts of the material) of categorization can be considered universal, while the difference in strategies is associated with the peculiarity of the holistic perception of the image and the importance of social stimuli in human life. Based on the results, we can also assume the similarity of the internal architecture of the structures responsible for the categorization process. The results of this study can be used as a rationale for the universality of conceptualization processes.

It is important to compare heatmaps of different transformer architectures and create a model that recognizes drawn objects without learning from drawings. It is also extremely interesting to develop the design of the “Conceptual Synthesis” experiment for a neural network. The “Conceptual Synthesis” technique is aimed at identifying conceptual abilities (the ability to construct a certain set of semantic contexts based on three words that are not related in meaning) (Holodnaya, 1988). In the future, we plan to continue to explore visual patterns, as we believe that access to mental representations is easier when modeling cognitive phenomena in the form of a neural network.

*The article was prepared with the financial support of the grant of the Moscow Center for the Study of Consciousness No. 1/2022.*

## References

1. Биджакджи М., Кексал М., Балоглу М. (2021) Клинический случай человека с синдромом саванта из Турции: когнитивные функции и календарный расчет // Клиническая и специальная психология. Том 10. № 1. С. 1–14.
2. Боярская Е. Категоризация как базовая когнитивная процедура // Вестник Балтийского федерального университета им. И. Канта. Серия: Филология, педагогика, психология. 2011.
3. Брунер Дж. Психология познания. За пределами непосредственной информации. Пер. с англ. – М.: Прогресс, 1977. – 413 с.
4. Величковский Б. – Когнитивная наука: Основы психологии познания: в 2 т. – Т. 1, 2006.

5. Выготский Л., Сахаров Л. Исследование образования понятий: методика двойной стимуляции // Хрестоматия по общей психологии. Выпуск III. Субъект познания. – М., 1998
6. Флейвелл Дж. Генетическая психология Жана Пиаже. М.: Просвещение, 1967.
7. Холодная М. Структурный подход в психологическом исследовании мышления // Проблемы философии: основные принципы построения научных теорий. Вып. 77. Киев: Высшая школа, 1988. С. 94–102.
8. Чистопольская А. В., Лазарева Н. Ю., Маркина П. Н., Владимиров И. Ю., Представление о высокоуровневых и низкоуровневых процессах в когнитивной психологии. Теория изменения репрезентации С. Ольссона с позиции уровня подхода // Вестник ЯрГУ. Серия Гуманитарные науки. 2019
9. Bechberger L., Kühnberger K. Towards Grounding Conceptual Spaces in Neural Representations, 2017.
10. Borji A., Itti L. Cat2000: A large scale fixation dataset for boosting saliency research. arXiv preprint arXiv:1505.03581, 2015.
11. Gärdenfors P. Conceptual Spaces: The Geometry of Thought. Cambridge, Massachusetts: MIT Press, 2000.
12. Mizutani R., Noguchi S., Saiga R., Yamashita Y., Miyashita M., Arai M., Itokawa M. Schizophrenia-Mimicking Layers Outperform Conventional Neural Network Layers. Front Neurobot, 2022.
13. Pelayo R, Fonseca T. DRAWING AS A COGNITIVE STRATEGY: PERCEPTION AND CREATIVITY, 2013.
14. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. Attention Is All You Need. arXiv: Computation and Language, 2017.
15. Pulvermuller, F. Semantic embodiment, disembodiment, and misembodiment: in search for meaning in modules and neuron circuits. Brain Lang, 2013.
16. Snyder A., Mulcahy E., Taylor J., Mitchell D., Sachdev P. & Gandevia S. C. Savant-like skills exposed in normal people by suppressing the left frontotemporal lobe. J. Integr. Neurosci. 2, 2003, pp. 149–158.
17. Szymanski L., McCane B., Atkinson C., Conceptual complexity of neural networks. Neurocomputing, Volume 469, 2022, pp. 52–64.
18. Whittington, Warren. RELATING TRANSFORMERS TO MODELS AND NEURAL REPRESENTATIONS OF THE HIPPOCAMPAL FORMATION, 2022.
19. Zhou, Gao, Wang, Yu. Eye tracking data guided feature selection for image classification, 2017.
20. Zwarts J. Conceptual Spaces, Features, and Word Meanings: The Case of Dutch Shirts. In: Zenker, F., Gärdenfors, P. (eds) Applications of Conceptual Spaces. Synthese Library, vol 359. Springer, Cham, 2015.

## Chapter 3. Stimulus display with extra wide color gamut for color discrimination thresholds assessment by strict substitution method

*Belokopytov A.*

*Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia*

### Abstract

*Introduction.* Assessment of human color discrimination thresholds usually requires costly sophisticated apparatus with complex optical system and unusual visual conditions for observer (MacAdam, 1942). The amount of the experimental data is limited due to these obstacles especially outside Adobe RGB color space. Measurements in these areas of human color space are needed to develop color appearance models for novelty imaging devices.

*The goal.* To develop simple device for assessment of human color discrimination thresholds in free-view conditions.

*Methods.* The device uses “strict substitution” method [2] for visual stimuli presentation which allows to significantly simplify optical system. The device consists of stimulus box and Arduino based control box. Stimulus box contains plate with RGB LED strips, two diffuse plastic plates, set of exchangeable apertures, digital temperature sensor, and a fan. Control box has auxiliary LCD display and three rotary encoders. LED brightness of each color channel is regulated with 16-bit PWM. Device was calibrated with X-Rite i1 pro spectrophotometer. Control box software allows to set reference and test stimuli color in xyY CIE color coordinates, and provides replacement of stimuli with frequency of up to 100 Hz.

*Results.* The device has the following features: extra wide color gamut (about 120 % of Adobe RGB), brightness of an equiluminance plane of up to 80 cd/m<sup>2</sup>, stimuli of up to 4 cm diameter, free-view conditions. The device characterization (according to [3]) was conducted, the results were satisfactory. Some cross-talk between color channels was found (less than 10 %), the reasons are under discussion. Thermal stability measurements showed that device should be used in a laboratory with air conditioning. Pilot measurements of color discrimination ellipses gave reasonable results.

*Conclusion.* The device can be used to access color discrimination thresholds.

### Introduction

In 2012 the new standard for imaging devices (mainly television devices) was adopted by ITU-R (BT.2020, 2012) and later it was extended (BT.2100, 2018). However, there are still no commercially available devices that meet one of these standards. The standards have set substantially expanded requirements for the device’s color gamut. Rec. 2020 color space covers 75.8 % of the CIE1931 color space (i.e. human color gamut), whereas Adobe RGB color space covers 52.1 %. The best commercially available displays cover no more than the Adobe RGB color space (for example, NEC PA311D, Eizo CG319X–99 %).

Why there are no imaging devices with more wide gamut? Part of the answer is too few experimental data on the human color discrimination thresholds outside the Adobe RGB color space and lack of color appearance models in this color space area.

Assessment of human color discrimination thresholds usually requires costly sophisticated apparatus with complex optical system and unusual visual conditions for observer. For exam-

ple, MacAdam apparatus clearly is very complex (Fig. 1).

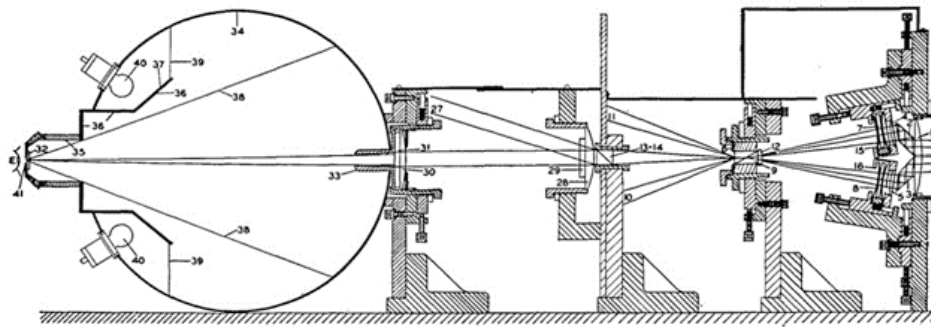


FIG. 2. Vertical cross section of chromaticity discrimination apparatus.

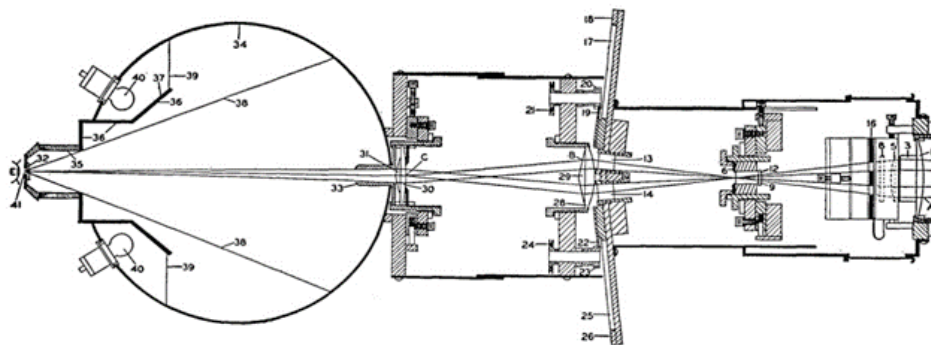


FIG. 3. Horizontal cross section of chromaticity discrimination apparatus.

**Figure 1.** MacAdam apparatus (MacAdam, 1942)

Modern opto- and microelectronics allows to develop simple device for assessment of human color discrimination thresholds. Now high brightness color LEDs and RGB LED strips are widely available. To maximally simplify the optical system of the device it was convenient to use “strict substitution” method (Wysecki, Stiles, 2000, pp. 278–293). In this method test and reference color stimuli interchange in time but have the same physical location and stimulate the same retinal area. Interchange of stimuli is controlled electronically by the microcontroller program.

Traditional technology inevitably led to unusual visual conditions for observer—ocular view which limited investigations to monocular mode and “Maxwellian viewing” (Burns, Webb, 2010) which required special training of the observers. In contrast, now it was not hard to develop device for “free viewing” (“Newtonian viewing”) visual conditions that is much more convenient for observer and opens new possibilities.

## Methods

All spectral measurements (emission/transmittance/reflection) were done with X-Rite I1 Pro rev. D spectrophotometer in hi-res mode (wavelength resolution 3.33 nm) with the help of the spotread utility from argyll cms software package (Argyll, 2021). Assessment of the LED’s temperature was done with 1) digital thermometer DS18B20 and 2) Fluke TiS60 Thermal Imager.

The device consists of the “Control Box” with electronics and “Stimulus Box” with LEDs and fans (Fig. 2).



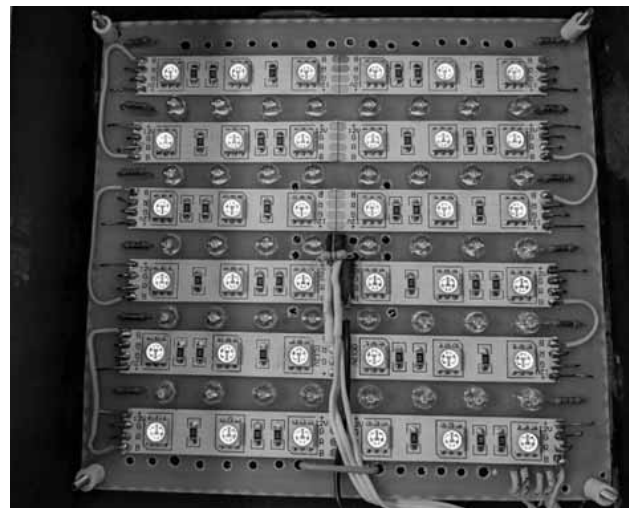
**Figure 2.** *General view of the device*

Off-the-shelf electronic components are used in the device: control box–Arduino Leonardo microcontroller, PSU MeanWell 12V 2A, three EC11 rotary encoders, LCD display 20×4; stimulus box–decorative RGB LED strips, digital thermometer DS18B20, fans Evercool EC6015M12B (60×60×15 mm). Microcontroller provides at least 3 hardware PWM channels with 16-bit resolution each. LED strips were annealed for 100 hours at full brightness. Additionally, the control box has RJ45 port to connect optional module with three encoders (3 m cable) which operate parallel to the encoders mounted in the control box.

Stimulus box was made of binding board, has dimensions 135×135×50 mm. Inside the box there are two diffuse plastic plates (one mounted in the cover), PCB with LED strips, digital thermometer (Fig. 3). Fans are mounted at the back side of the box, there are holes in the box and the PCB for the air flow. Stimulus box is accompanied with set of exchangeable diaphragms with round holes of various diameters (maximum 4 cm).



**A**



**B**

**Figure 3.** *Inside stimulus box. A. Diffuse plastic plate over PCB with LEDs. B. PCB with LED strips, digital thermometer is mounted in the center of the PCB.*

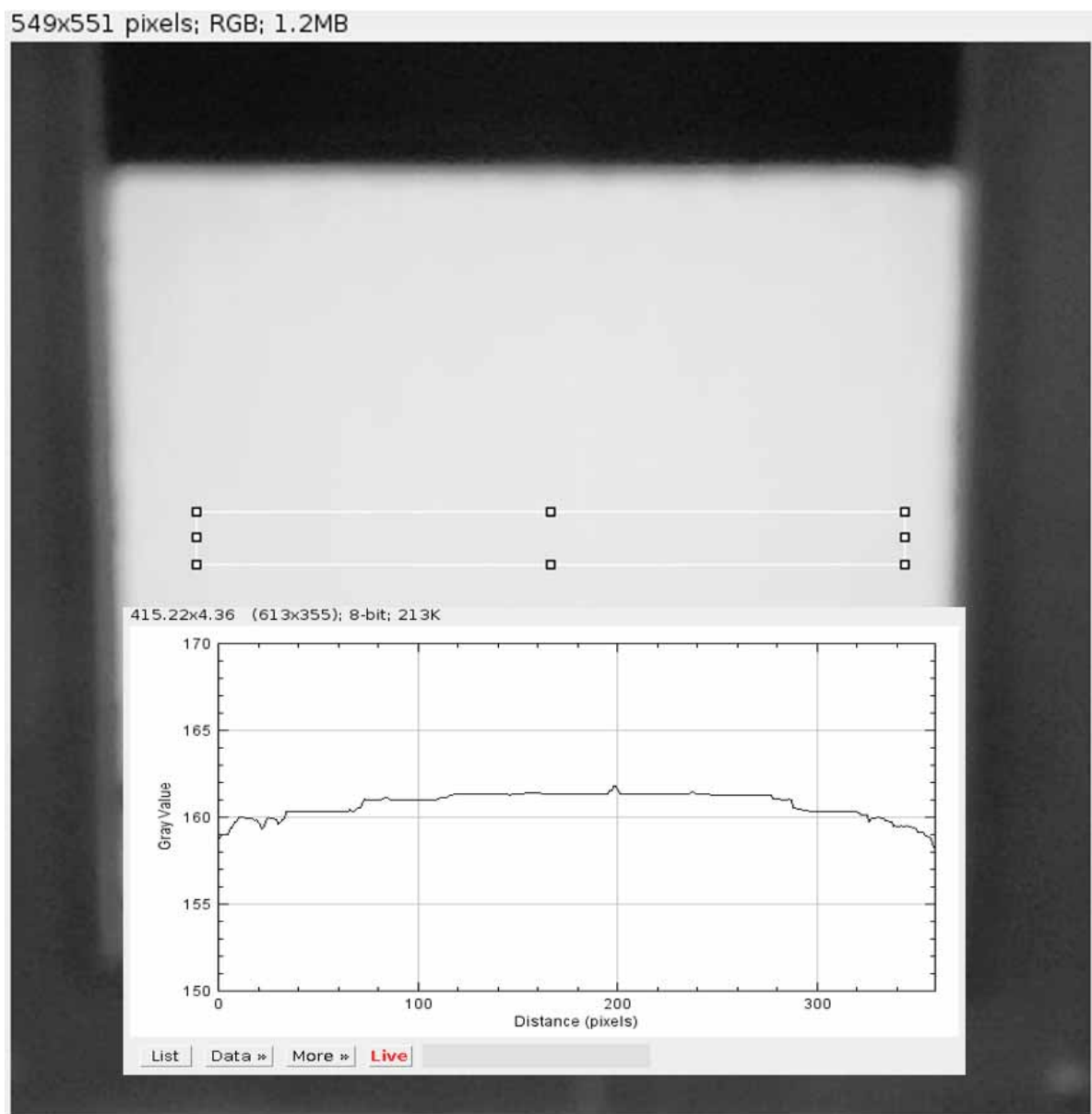
Control box software allows to set color center in CIE xyY color coordinates, to measure JND in polar coordinates, to replace stimuli with a minimum interval of 10ms. Experimenter sets desired parameters with 3 knobs (encoders), and observer operates only 2 knobs. Further technical details one can find at <https://github.com/abelokopytov/sd>.

Optical pulses measurements were done with shunted Vishay BPW34 photodiode and oscilloscope. Stimulus display characterization as new imaging device was conducted according to (Brainard et al., 2002).

## Results

### *Spatial homogeneity (field uniformity)*

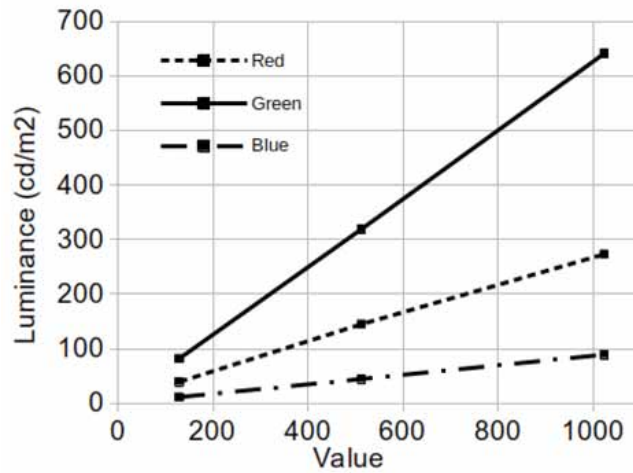
Field uniformity was assessed by analyzing luminance profile of the stimulus box window photographs with the help of the ImageJ software (Fig. 4). The conclusion is that two thirds of the window can be used for measurements.



**Figure 4.** Photograph of the stimulus box window with marked ROI (region of interest) superimposed with luminance profile of the ROI

### Luminance linearity

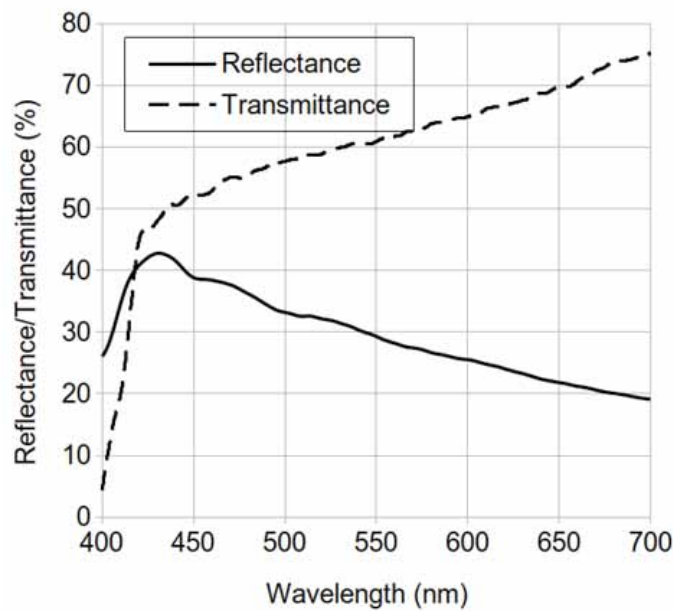
The evaluation of the luminance linearity led to the conclusion that the device demonstrates good linearity (Fig. 5). Measurements were conducted in 10-bit PWM mode but there is confidence that the results will be the same for 16-bit mode.



**Figure 5.** Color Red, Green and Blue channels luminance dependence of the PWM value.

### Diffuse plate spectral properties

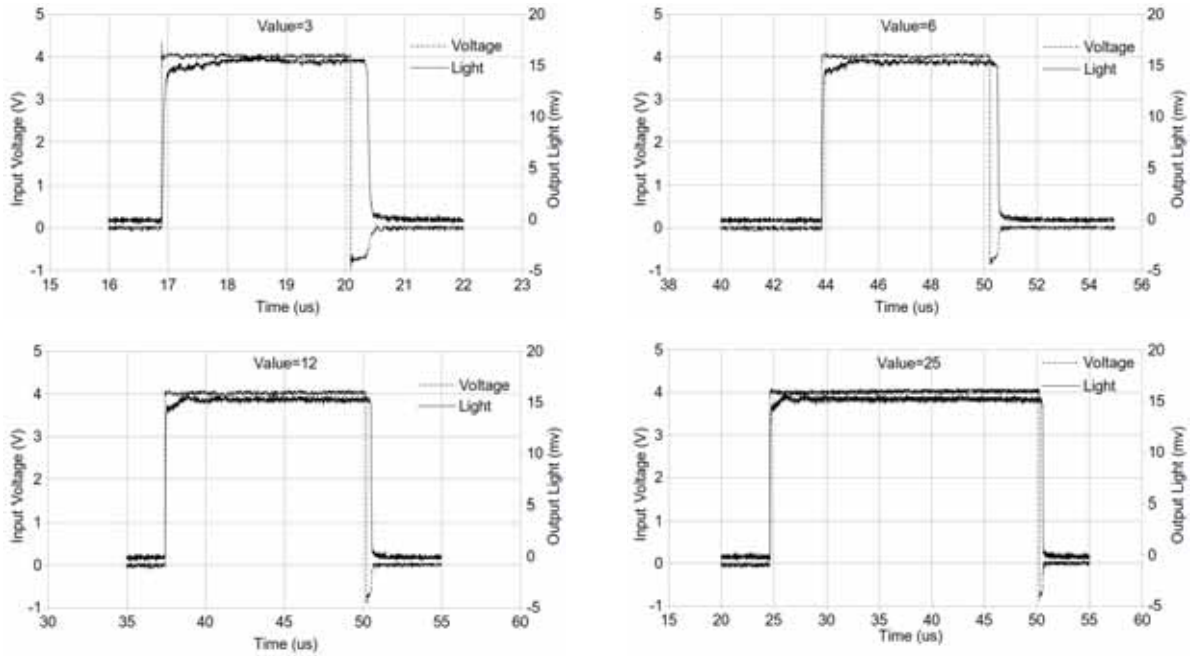
Spectral reflectance and transmittance of the diffuse plate are presented in Fig. 6.



**Figure 6.** Diffuse plate spectral reflectance/transmittance

### Light pulses assessment

It was important to assess quality of linearity at low PWM levels to define limits of applicability of the device at low light levels (when luminance of the stimulus box is less than 0.1 cd/m² and cannot be reliably measured with the spectrophotometer used). To accomplish this task light pulses were measured directly (in 10-bit PWM mode) with the help of photodiode and oscilloscope (Fig. 7).

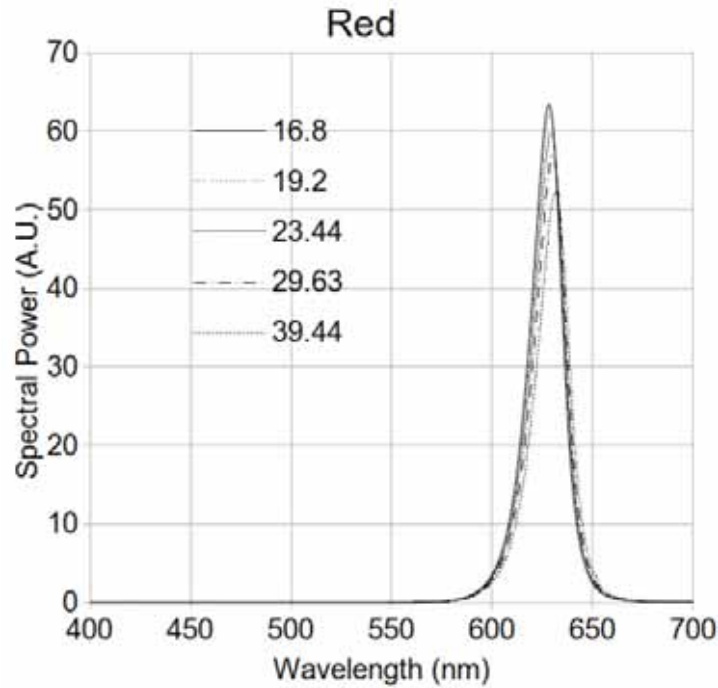


**Figure 7.** Output light and input electrical pulses at various PWM values.

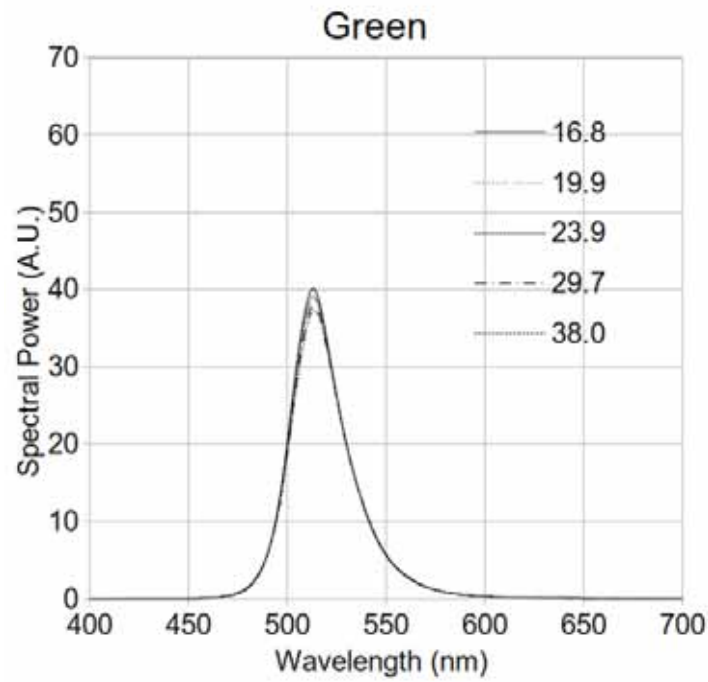
It was reasonable to set the low limit of the PWM value to about 6 in 10-bit PWM mode, i.e. to  $6 \cdot 64 \approx 400$  in 16-bit PWM mode.

#### *Stimulus display spectral characteristics*

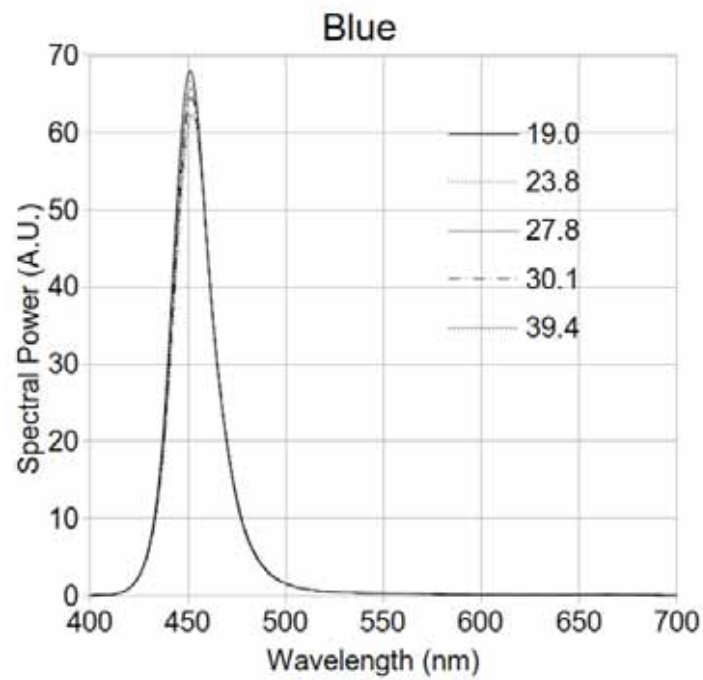
The spectral power diagrams were measured for Red (Fig. 8), Green (Fig. 9) and Blue (Fig. 10) channels at various temperatures (temperature was assessed with digital thermometer inside the stimulus box).



**Figure 8.** Spectral power distribution of the Red channel at various temperatures indicated in the legend.



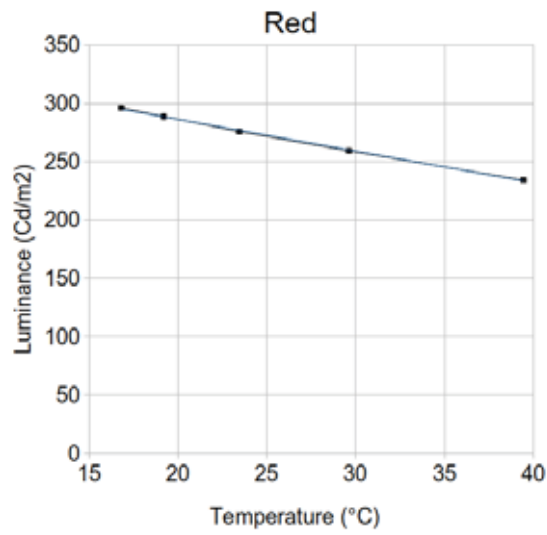
**Figure 9.** Spectral power distribution of the Green channel at various temperatures indicated in the legend.



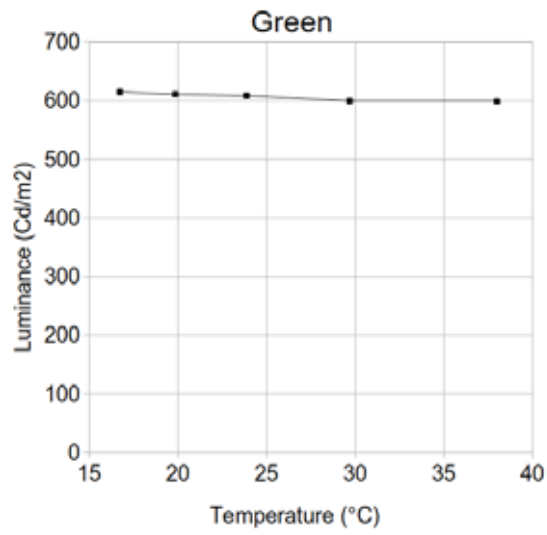
**Figure 10.** Spectral power distribution of the Blue channel at various temperatures indicated in the legend.

#### *Luminance dependence on the temperature*

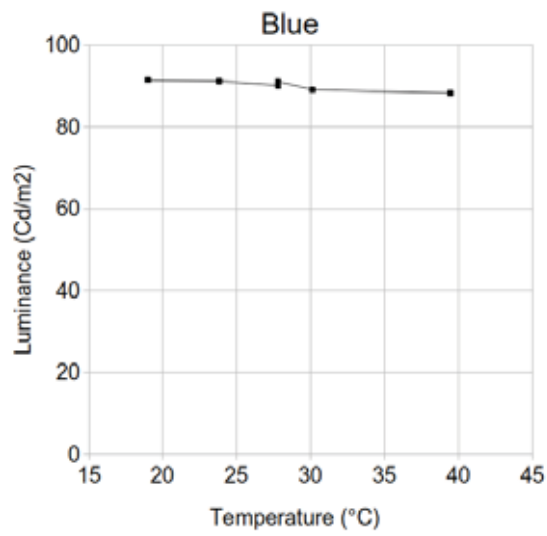
It was found that time lag to stabilize temperature and emission of the stimulus box is about 2 minutes. Luminance dependence on the temperature for Red, Green and Blue channels are presented in Fig. 11–13 accordingly.



*Figure 11. Luminance of the Red channel at various temperatures.*



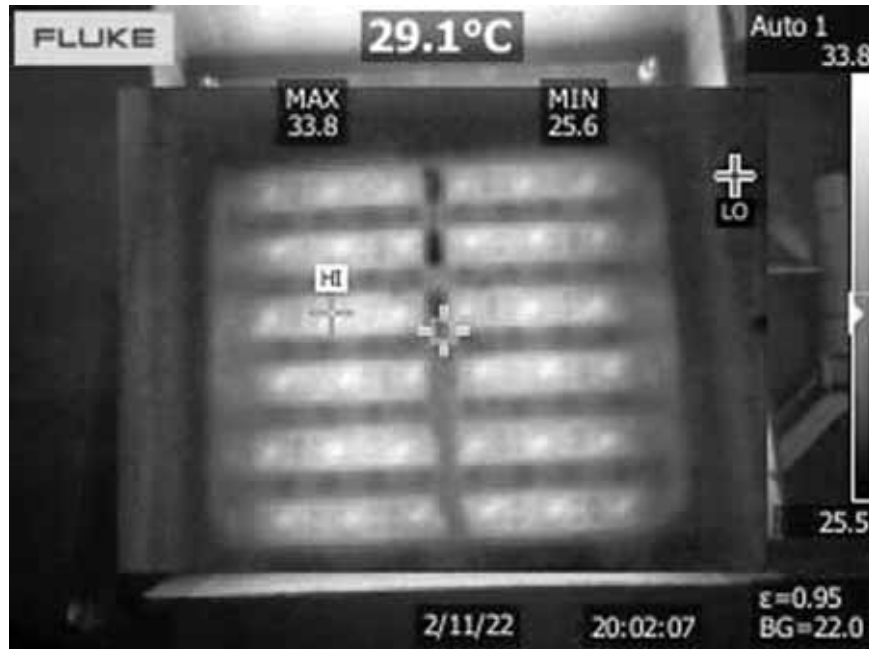
*Figure 12. Luminance of the Green channel at various temperatures.*



*Figure 13. Luminance of the Blue channel at various temperatures.*

### *Assessment of the LED temperatures*

It was of interest to compare digital thermometer readings with other device. Measurements were done with the following RGB values: R=21989; G=9478; B=65535 (luminance 273 cd/m<sup>2</sup>). The data on LEDs temperatures obtained by Fluke thermal imager are shown in Fig. 14 and comparison with digital thermometer readings and ambient temperature are presented in the Table 1.



**Figure 14.** Data obtained by Fluke thermal imager

**Table 1.** Ambient and LEDs temperatures.

Source	Temperature (°C)
Ambient	22.0
Digital thermometer inside stimulus box	26.7
Thermal Imager maximum	33.8

### *Working temperature of the device.*

In air conditioned laboratory with ambient temperature 22 °C digital thermometer reading at zero luminance is about 24 °C due to fans self-heating (though airflow is outside the stimulus box). At the equiluminance plane with  $Y = 80 \text{ Cd/m}^2$  its readings are about 26 °C, so temperature difference compared to zero luminance is about 2 °C.

### *Channels cross-talk*

It was found that channels luminance does not obey additivity law. It means that

$$LW \neq LR + LG + LB$$

Here LW denotes luminance of “white” (i.e. R=G=B=value), LR, LG, LB denote luminances of Red, Green and Blue channels at the given PWM value accordingly.

Relevant data is presented in the Table 2.

**Table 2.** Luminance additivity check results.

PWM Value	LW	LR	LG	LB	Error (%)
128	121.4	37.9	78.6	11.3	5.3
512	476.2	141.1	307.2	44.0	3.4
1023	942.1	265.9	615.5	88.1	2.9

PSU voltage was stable in all conditions.

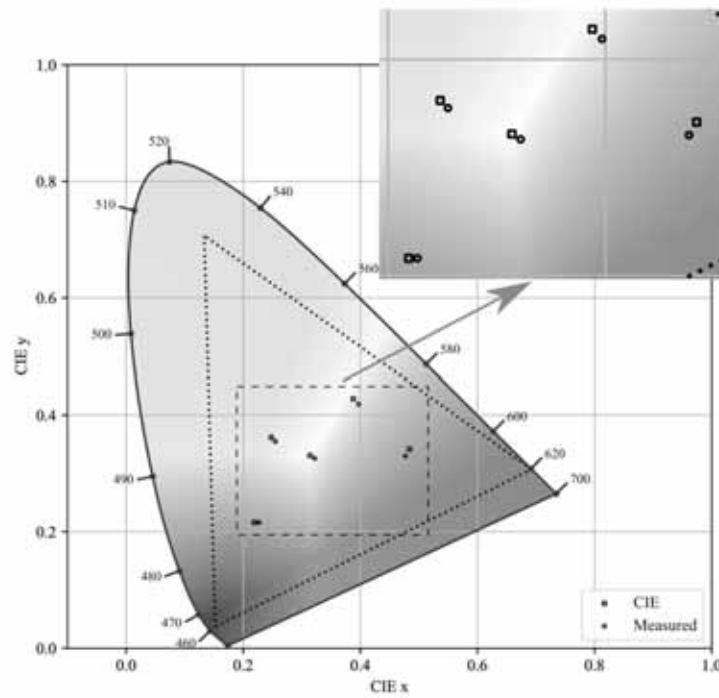
#### *Device accuracy*

The device was calibrated (CIE xyY color coordinates of the primaries were measured with spectrophotometer at ambient temperature 22 °C and parameters in the control box software were adjusted accordingly) to present equiluminance plane  $Y = \text{const}$ . Maximum Y achieved was more than 80 cd/m<sup>2</sup>.

Device accuracy was assessed in the 5 CIE color centers (Robertson, 1978). Data on luminance accuracy is presented in Table 3, and data in Fig. 12 allows to visually evaluate xy accuracy.

Table 3. Target and measured device luminance (Y) in the 5 CIE color centers.

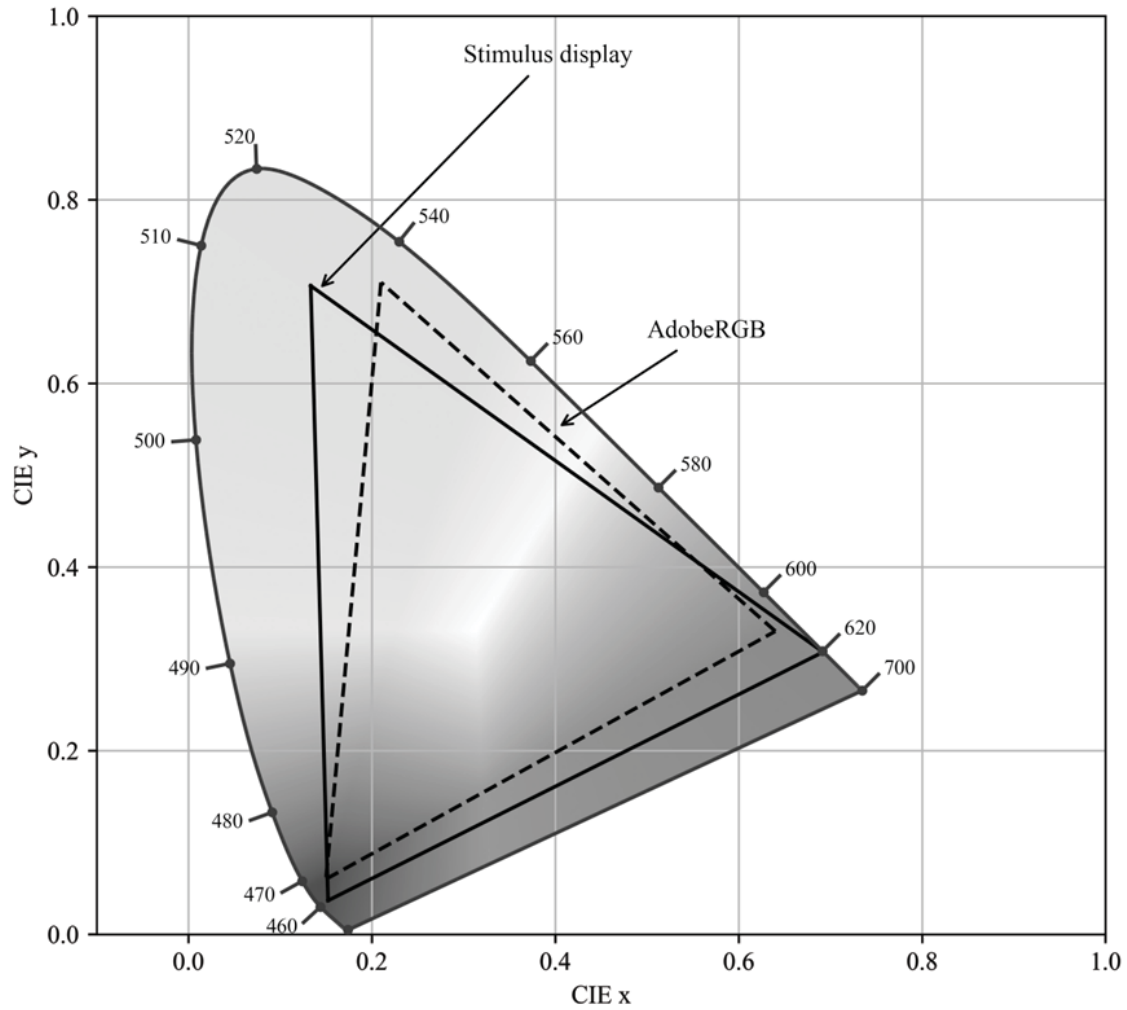
CIE color center	CIE Y (cd/m2)	Y measured (cd/m2)	Error (%)
Gray	30.0	31.07	3.6
Red	14.1	15.71	11.4
Yellow	69.3	71.36	3.0
Green	24.0	24.88	3.7
Blue	8.8	9.62	9.3



**Figure 15.** CIE1931 chromaticity diagram with target (square markers) and measured (circle markers) CIE xy color coordinates in the 5 CIE color centers

### *Stimulus display gamut*

The developed stimulus display covers 120 % of the Adobe RGB gamut (color space) except some area (Fig. 16).



**Figure 16.** CIE1931 chromaticity diagram with gamuts of the developed stimulus display (solid triangle) and Adobe RGB (dashed triangle).

### **Discussion**

Spectral measurements of the diffuse plate used showed that it has rather big reflectance coefficient (Fig. 4) hence follows that color threshold measurements should be conducted in the dimmed room. Illumination of the plate mounted in the cover of the stimulus box should not exceed 1 lux.

The device by design has high dynamic range: PWM value can vary from 0 to 65535 (16-bit mode). Luminance linearity analysis increased low limit of the PWM value from 0 to 400. And diffuse plate reflection also somewhat decreases this range.

Stimulus display emission is temperature unstable. For this reason, the device should be used in the air-conditioned laboratory at the recommended temperature 22 °C. The effect of cross-talk between channels may be explained by the equal cross sections of the common wire and channel wires in the LED strips used but this caveat needs further investigations.

## Conclusion

The device developed has the following useful features:

1. Extra wide gamut of the stimuli allows to measure color thresholds outside the Adobe RGB color space.
2. High brightness—equiluminance plane  $Y=80$  cd/m<sup>2</sup> achieved; maximum brightness of 940 cd/m<sup>2</sup> achieved (for “white”).
3. High dynamic range—16-bit PWM value range from 400 to 65535.
4. Free viewing is very convenient for the observer and allows to use binocular or monocular modes and explore peripheral color vision.
5. Ease of use—three knobs for experimenter and two knobs for the observer. Experimenter set directly CIE xyY color coordinates of the reference stimulus.
6. Low cost—cheap electronic components and materials.

On the other hand, the device has the following drawbacks:

1. Temperature instability. The device should be used in the air-conditioned laboratory.
2. The laboratory should be dimmed.
3. Channels cross-talk and other factors limits the device accuracy.

Undoubtedly the construction of the stimulus box can be improved. There are wide possibilities to increase maximum luminance and expand color gamut (adding appropriate color LEDs).

Pilot experimental results of color threshold assessment by a stimulus display developed are described in another chapter of this book.

## References

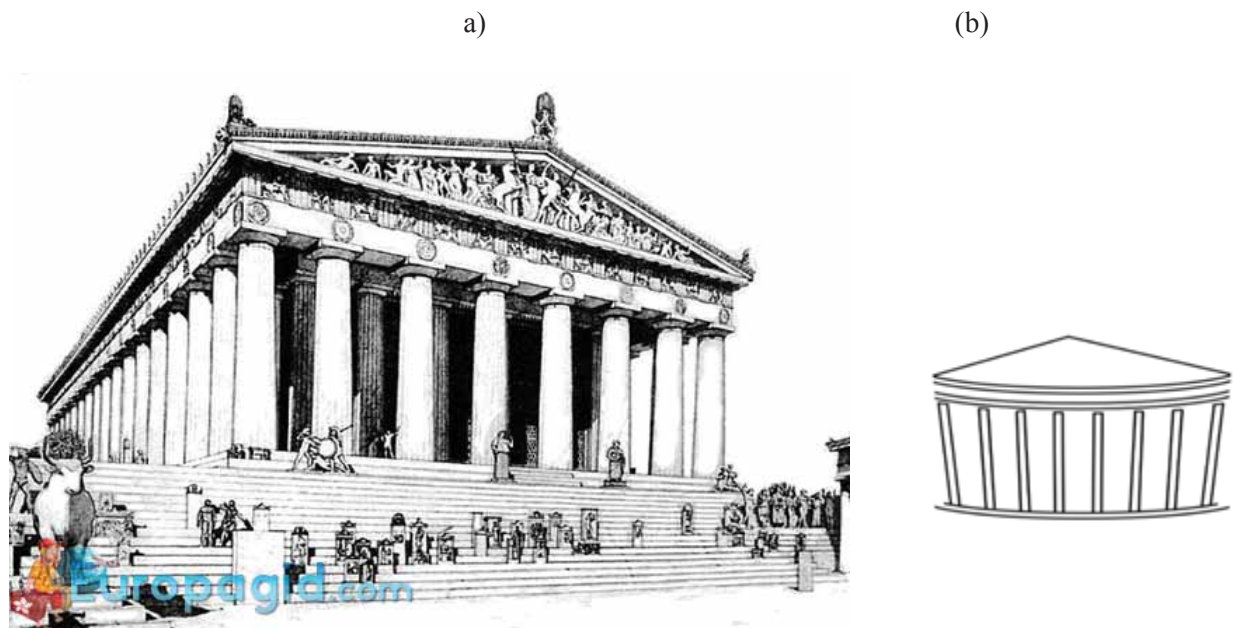
1. Argyll, 2021. <https://www.argyllcms.com/>. Retrieved 2022–12–08.
2. Brainard, D. H., Pelli, D. G. and Robson, T. Display Characterization. In: J. P. Hornak (ed.), *Encyclopedia of Imaging Science and Technology*. 2002.
3. (BT.2020) BT.2020: Parameter values for ultra-high-definition television systems for production and international programme exchange. International Telecommunication Union. 2012. <https://www.itu.int/rec/R-REC-BT.2020-0-201208-S/en>: Retrieved 2022–12–05.
4. (BT.2100) BT.2100: Image parameter values for high dynamic range television for use in production and international programme exchange. International Telecommunication Union. 2018. <https://www.itu.int/rec/R-REC-BT.2100/en>: Retrieved 2022–12–05.
5. Burns S., Webb R. Optical generation of the visual stimulus. In: Bass M. (ed.), *Handbook of Optics*. V. III. Ch. 5. 3<sup>rd</sup> edition. McGraw Hill. 2010.
6. MacAdam D. L. Visual sensitivities to color differences in daylight. *JOSA*. 1942; 32(5); 247–274. <https://doi.org/10.1364/JOSA.32.000247>.
7. Robertson A. CIE Guidelines for coordinated research on colour difference evaluation. *Color Research & Application*. 1978. 3(3):149–151.
8. Wyszecky G., Stiles W. S. *Color science. Concepts and methods, quantitative data and formulae*. John Wiley and Sons. 2000. 976 p

## Chapter 4. Perception of Classical Architecture: Psychophysical Researches and Aesthetics

*Bondarko V.M., Danilova M. V., Solnushkin S. D., Chikhman V.N., Shelepin Yu.E.*

*Pavlov Institute of Physiology Russian Academy of Sciences, St. Petersburg, Russia*

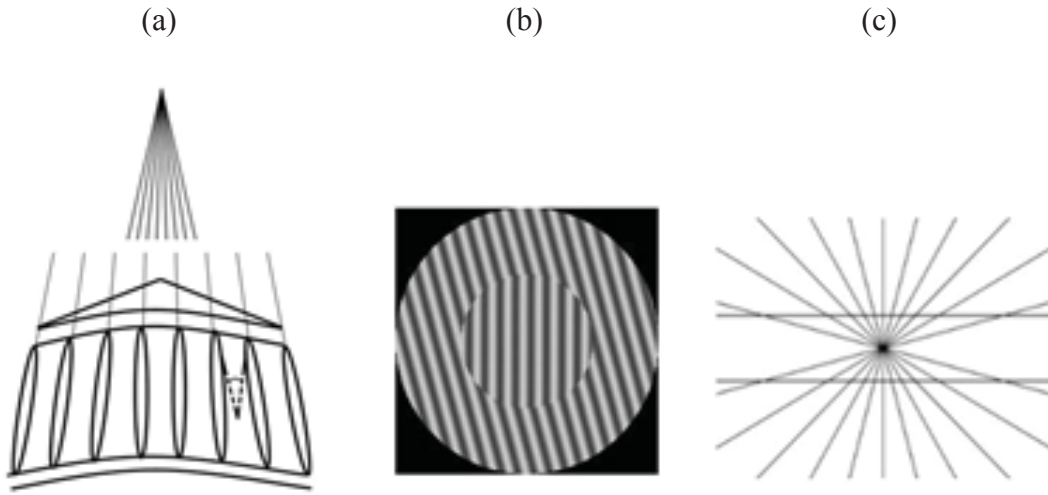
The Parthenon is an ancient temple of Athena, which was built in 447–438 BC. It is the most important surviving building of Classical Greece. In the architecture of the Parthenon, the upper parts of columns taper upward and the two-dimensional projections of the building have many convex horizontal and vertical lines, although the temple is perceived to be perfectly straight. The stylobate (floor) rises towards the center, while the corner columns tilt towards the center and have a barrel shape. Frith (2007) demonstrates a hypothetical two-dimensional projection of the Parthenon taken from a paper by Pennethorne (1844); it lacks the correction in the curvature and the slopes.



**Figure 1.** Image of the Parthenon (a) and its hypothetical two-dimensional projection (b), if correction was not made.

These corrections were not made by accident. To make the temple appear higher, the laws of perspective were used and the columns tapered upward. This, in turn, led to an additional effect. The tilt illusion arises, where the columns at the top appear to diverge. To prevent this the columns of the temple were tilted inwards. But this caused the genesis of the Wundt-Hering illusion (or the fan illusion) in which the apparent concavity of the stylobate (floor) can be observed. To avoid this, the stylobate line was made convex.

Fig. 2a shows an exaggerated temple diagram with the tilt and the Wundt-Hering illusions. The columns are tilted inwards and form a fan. The tangents to the columns form sharp angles (dashed lines). In the tilt illusion (Fig. 2b) a vertically oriented grating in the center of the circle seems to be tilted clockwise due to the surroundings. A similar illusion is observed when lines are used as images (look at the sharp angle at the top of the columns in Fig. 2a). In the Wundt-Hering illusion (Fig. 2c) the horizontal straight lines are visible as convex (top line) or concave (bottom line) when a fan image is superposed on them.



**Figure. 2.** *The temple diagram (a), the tilt illusion with vertical grating in the center (b) and the Wundt-Hering illusion (c).*

We considered the architecture of the Parthenon from the point of view of the tilt and the Wundt-Hering illusions. Psychophysical experiments were carried out to estimate the values of these illusions. Moreover, curvature of some elements without context were estimated. The Parthenon is the canon for classical architecture, which is strictly followed by all buildings of this style (Vitruvius, about 13 BC; Palladio, 1570; Puppi, 1989).

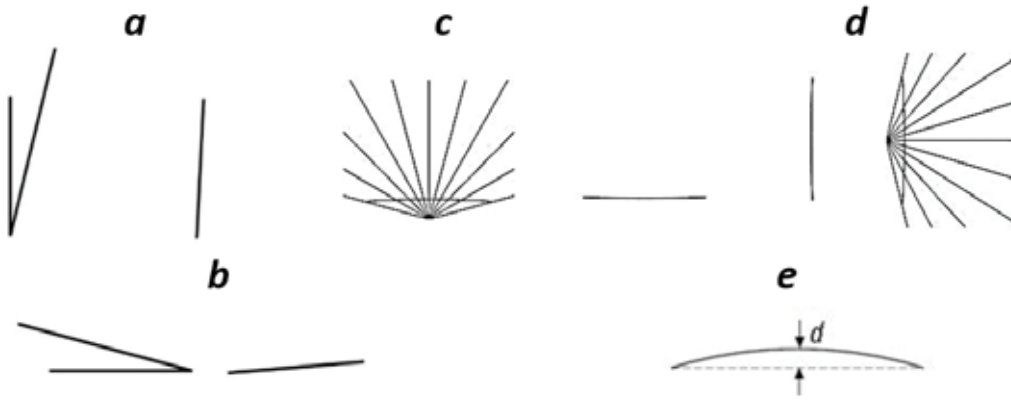
## Methods

**Observers.** Three observers with normal or corrected-to-normal visual acuity participated in both experiments. The observers were naïve to the purpose of the study.

**Apparatus.** The stimuli were synthesized on a Mitsubishi Diamond Plus 230SB monitor with diagonal size 22 in. and screen resolution 1600×1200 pixels and a refresh rate of 100 Hz. The software was developed in Delphi and Python. Dark lines (5 cd/m<sup>2</sup>) were drawn on a bright background (40 cd/m<sup>2</sup>), the lines were 0.8 arcmin thick. A 130×100 cm<sup>2</sup> mask with a circular aperture 28 cm in diameter was superimposed on the monitor screen.

**Stimuli.** When we studied the tilt illusion our stimuli consisted of target (T) lines, target lines joined with inducing (I) lines (Fig. 3, a, b) or inducing lines (distractors) were located at different distances from the T lines. The test lines with varied orientation were at the other visual hemifield and compared with T lines. The lengths of the T lines were 1.2 times smaller than the lengths of the I lines and were equal to 2.5 deg. The orientation of the T lines was fixed at 87° or −3° to the horizontal. The orientation of the I lines differed by ±10–30° or ±15–75° from the orientation of the T lines. Fig. 3a, b shows an examples of these stimuli.

In the Hering illusion we used a fan with straight, slightly concave or convex lines. The fan was oriented horizontally or vertically (c, d). The angular size of the horizontal fan was 6.7×5.1 deg, the distance from the center of the fan to the lines was 0.63 deg, the length of straight lines was 4.7 deg. The concave and convex lines were arcs of a circle with a large radius. The distance from the 6.7 deg horizontal line to the vertices of the arcs was 3.8 arcmin. Distance *d* (Fig. 3e) was taken as measure of curvature. The same concave and convex lines were used to estimate curvature of lines without context also.

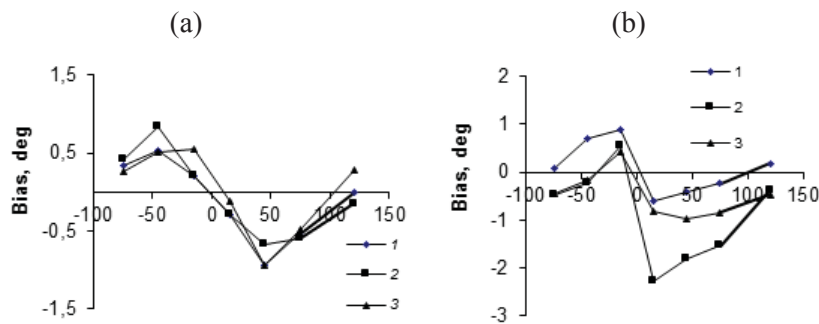


**Figure 3.** Stimuli that were used in experiments.

**The observer's task** was to discriminate the orientation of the target lines or the curvature of the lines. The screen was viewed binocularly at the distance of 115 cm. The stimuli time presentation was 1 s.

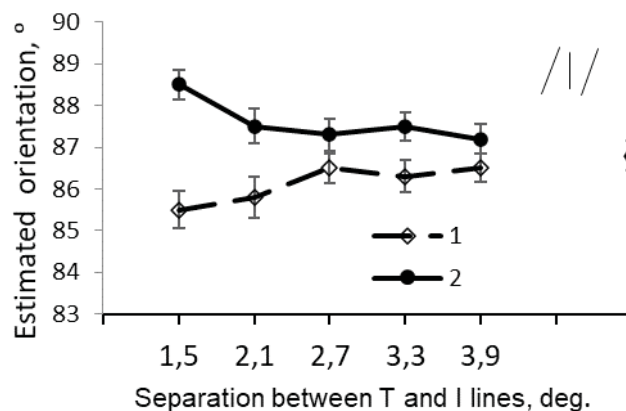
## Results

### *Orientation estimation in the tilt illusion*



**Figure 4.** Difference between reference and additional lines orientation (deg).

Figure 4 shows the tilt illusion for abutting lines for the three observers in  $-87^\circ$  (a) and  $-3^\circ$  (b) referent lines orientation respectively. X- axes are differences between reference and additional lines orientation (deg). The right-hand disconnected points correspond to the biases for isolated lines. Subjects see the lines as tilted away from the inducing lines as was shown in several investigations (for example, see Blakemore, Carpenter, and Georgeson, 1970). Curves 1, 2 and 3 are data of different observers.



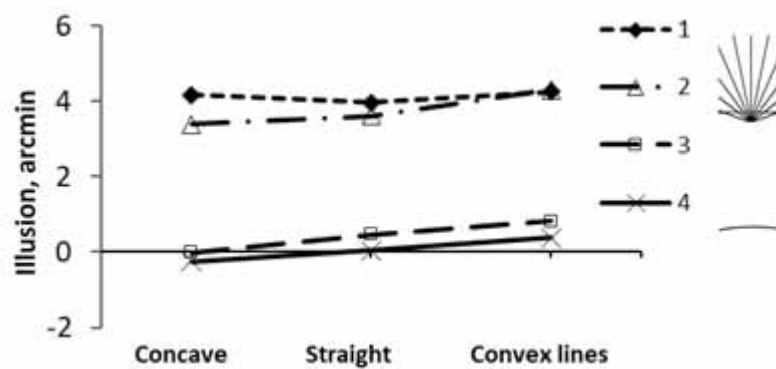
**Figure 5.** Lines orientation estimations.

Figure 5 shows the apparent orientation of the lines when the distractors (I lines) are rotated clockwise or counterclockwise (curves 1 and 2) by  $15^\circ$  from referent lines with orientation  $87^\circ$  (T lines) in depends on distances between T and I lines. The data of three observers are averaged. Vertical lines are SEM that represents intersubjective variability. The right-hand disconnected point corresponds to the estimations for the isolated lines.

**Thus, without correction for the slopes of the columns, they may appear to diverge at the top.**

### *Curvature assessment*

Figure 6 shows the magnitudes of the averaged Wundt-Hering illusions, computed as the difference between the mean values of the resulting normal distributions and the physical curvature of the reference stimuli, separately for concave, straight, and convex reference horizontal (1) or vertical (2) lines. Illusions do not depend on curvature referent lines and similar for vertical and horizontal orientation. In the case of isolated lines (3,4), there are no distortions in the estimates (there is no illusion).



**Figure 6.** Estimates of the curvature for lines with different curvature and orientation in context (1,2) and without context (3,4) are drawn. Magnitude of the estimations for horizontal (1,3) and vertical (2,4) lines, respectively.

The stylobate (floor) line of the Parthenon corresponds to our concave line in our experiments. The curvature of the concave line is overestimated by 4.27 arcmin (Fig. 6). When the curvature of the concave line is  $-3.8$  arcmin, it is perceived as a straight line (possibly somewhat convex), since its deviation from a straight line with a length of  $6.7$  deg at the center of the arc is only  $0.47$  arcmin ( $=4.27-3.8$ ). The initial curvature of the stylobate line is somewhat less in absolute value than the one we used in our experiments.

**Therefore, the line of stylobate can be perceived as a straight line.**

**But curvature of the columns is estimated correctly. The convex shape of the columns may be due to the aesthetics of perception.**

### *Aesthetics of perception*

We considered Berlyne's theory (Berlyne 1974), of the relationship between beauty and complexity of images. The dependence of the beauty score on complexity can be linear or bell-shaped (with increasing complexity the beauty score increases and then decreases).

Early we showed that the complexity estimations depend on the alphabet of images (Chikhman et al., 2012). For contour images, the complexity estimates are correlated with the number of turns in the images. Experimentally obtained estimates based on the data of 35 observers are

shown in the figure 7.

For Chinese characters, the best approximation in the model was obtained by taking into account the product of the square of the median of the amplitude spectrum and area of the image, when images had the equal size (compare the third and the fourth rows of Fig.8).



**Figure 7.** Contour images arranged in order of increasing complexity. Complexity increases from the left to right and then from the upper row to the lower row.

中行年币公念发纪熊金周最猫隆重  
 中周币重公金纪年行最猫念发隆熊  
 中币公年行周纪发念金重猫最熊隆  
 中公币行年纪周金发念重猫最熊隆

**Figure 8.** Hieroglyphs arranged in the order of increasing complexity. Complexity increases from left to right. Upper row is complexity that calculated as the stimulus area. In the second line complexity was calculated as the median of spatial frequency distribution. In the third line complexity was estimated as the squared median multiplied by the stimulus area or median multiplied by perimeter image. In the fourth line experimental obtained estimations of complexity are shown, that were calculated as the sum of positions assigned by 19 observers.

High correlation between experimental data and model calculations, where the product of the square of the median of the amplitude spectrum and area of the image was taken into account, means that complex images have more details of various shapes, orientations and sizes. Projections of buildings are closer to the alphabet of hieroglyphs. A confirming example is the hieroglyph responsible for the concept of “house” (Fig. 9b).



**Figure 9.** *Classical Stock Exchange building in St. Petersburg (a) and the Chinese character responsible for the concept of «house» (b). Note, that the cornices of the Stock Exchange building are straight, but appear to sag slightly due to the Wundt-Hering illusion.*

Classical buildings have more divergent details than the modern many-storeyed buildings of typical construction, but less than eclectic buildings. They seem to be more aesthetic. This confirms the hypothesis that beauty is related to the complexity of images. The dependence of the beauty score on complexity is rather bell-shaped (with increasing complexity the beauty score increases and then decreases).

## Conclusion

The obtained data were matched with the Parthenon's architecture. We observed an illusion of repulsion in the tilt illusion with a small difference between the orientation of lines. The angle between the lines seemed to be enlarged. This illusion can arise due to the thinning of the upper part of the columns in the temple (Fig. 2a). As a consequence, it will seem that the Parthenon columns are deviated from each other at the top part similarly to a fan. To avoid the influence of the tilt illusion, the temples' columns were inclined inward. This procedure led to the genesis of the Wundt-Hering illusion (or fan illusion). We obtained our experimental data using concave, convex and straight lines in the modified Wundt-Hering illusion and showed that the curvature of these lines was distorted. The curvature corresponding to the temple's lines decreased.

Thus, distortions in the perception of the curvature of lines was compensated in the architecture of the Parthenon. As a result, all lines in the Parthenon are perceived as straight ones.

The influence of the context on the curvature estimation was revealed. The curvature is adequately assessed in the absence of a fan image, that is, from relatively small distances, the convexity of the columns should be noticeable. Such elements in architecture may be related to the aesthetics of perception.

*This study was supported by the State Program GP-47 "Scientific and Technological Development of the Russian Federation " (2019–2030), theme 0134–2019–0006.*

## References

1. Berlyne D.E. Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation. New York: Wiley, 1974. 348 p.
2. Blakemore C., Carpenter R. H.S., Georgeson M. A. Lateral inhibition between orientation detectors in the human visual system. *Nature*. 1970. 228(5266): 37–39.

3. Bondarko V. M., Solnushkin S. D., Chikhman V. N. Curvature estimation and architecture of the Parthenon. *J. Optical Technology*. 2021. 88(6): 56–62.
4. Chikhman V. N., Bondarko V. M., Danilova M. V., Goluzina A. G., Shelepin Y. E. Complexity of images: Experimental and computational estimates compared. *Perception*. 2012. 41: 631–647.
5. Frith C. D. (2007) *Making Up the Mind: How the Brain Creates Our Mental World*. Malden: Blackwell Publ., 2007. 234 p.
6. Palladio A. *Quattro Libri dell'Architettura*. 1570.
7. Pennethorne, J. *The Elements and Mathematical Principles of the Greek Architects and Artists Recovered by an Analysis and Study of the Remaining Works of Architecture*. London: William Clowes and Sons, 1844. 21–49 p.
8. Puppi L. *Andrea Palladio: the complete works*. N.J.: Electa/Rizzoli, 1989. 311 p.
9. Vitruvii M. *Ten books on architecture*. M.: KomKniga, 2005. (In Russian).

## **Chapter 5. Cross-modal interaction in the integral assessment of image and sound quality**

*Glasman K., Grinenko E., Shelepin Yu.*

*St.-Petersburg Electrotechnical University LETI, St.Petersburg, Russia*

*Pavlov Institute of Physiology Russian Academy of Sciences, St.Petersburg, Russia*

### **1. Audiovisual information perception and quality assessment**

#### **1.1. Audiovisual information**

The modern world is a world of media, or mass communication systems. Television, film, radio, newspapers, and the Internet have a serious informational impact on society, its state and development (Моль, 1973). The amount of information assimilated by a person depends on the number of perception organs involved in the information receiving process. The success of film and television is to some extent related to the fact that hearing and vision organs, which gives a person the lion's share of information, have been involved in the process of perception. It is television that has become the most popular way of mastering information. The information society—a society controlled by information, was formed with invention of television. The possibilities of this technical phenomenon allow to transmit the maximum amount of information in the form which is extremely easy assimilated and, therefore, has the greatest impact (Левченко, 2005).

The progress in the creation of effective management, communication and media communication systems urges the consideration of human parameters as the most important link in such systems. The semantic side of the information should be considered when this information serves as the basis for certain practical actions. However, film, television, and other media systems are not only media. Film and television are also fields of culture and art. Therefore, the methods of the theory of information should also be extended to the sphere of aesthetic perception and objects of art (Моль, 1966).

Semantic information allows it's accurate representation, it is translatable into other languages, and the rules for handling it and the symbols that appear in it are generally accepted. Aesthetic information is personalized information, it is almost entirely unpredictable, it is untranslatable, because there is no other language to convey this information.

Messages containing purely semantic or purely aesthetic information are only extreme cases. Every real message is always a mix of both semantic and aesthetic information. When we listen to the latest news on the radio, some messages carry semantic information: for example, the weather forecast, stock prices if we are stock holders; administrative orders if they apply to us. Such messages determine the reactions of the majority, or at least of a logically definable group of listeners. The rest of the messages are essentially useless to us; they inform us, cause us resentment or satisfaction, but in no way determine our actions either in the present or in the future.

Aesthetic information disseminated by the media does not so much reflect the real facts as it evokes certain states of mind, reactions and emotions in the audience. Information conveyed in this way becomes more stable for the audience than facts and arguments presented in the form of semantic information. The aesthetic information object is based on what different people believe in. Often these objects are not based on any facts or evidence, but are created

by collective experience. At the same time, as a rule, they are not clearly recognized by the individual himself (Поцелуев, 1999).

Every real message in film and television contains both semantic and aesthetic information, but these types of information exist separately in the message. Some messages can contain mainly semantic or aesthetic information.

In artistic works, aesthetic information is of the most importance. For example, the role of aesthetic information in musical works is very large – this arises from orchestration variations, differences in the instruments construction, interpretation features, etc. Specific features of different types of art are reflected in aesthetic information.

## **1.2. Cross-sensory interaction in audiovisual information perception**

Perception is the holistic image formation of an object resulting from the impact of the objective world on the human senses. Perception reflects external aspects of objects. There are two main forms visual and auditory perception. Perception is formed from sensations that human obtain while interacting with objects in the environment. In this interaction, the contours of the object determine, for example, the trajectory of the hand touching the object, and the trajectory of the hand movement, in turn, determines the structure of the visual image. Perception belongs to the individual and it is a model of external objects and our body. Representations are built from perception (Грицанов, 1998).

It can be said that perception is the first moment when a person meets the world, in which the foundations of future cognitive strategies and intellectual intuitions are laid. In the modern world, there is an increase in visual communications; visually active presentations (computer graphics, hypertexts of electronic manuals (Кухта, 2004)) replace the textual presentations.

Information perception is characterized by a polysensory nature. Different sensory systems participate in perception: hearing, vision, skin-tactile sense, chemoreception (smell, taste), thermoreception (sense of heat – cold). Each of these sensory systems, or analyzers of information of the external world consists of the peripheral (receptor), conductive (sensitive nerve) and central part, i.e. the corresponding areas of the brain, where information of the external world, perceived by receptors, is converted (decoded) into visual, auditory, skin-tactile, olfactory, thermal sensations and representations. The polysensory nature of perception ensures the human possibility to receive practically all biologically and socially significant types of external world information (Морозов, 1998; Шмаков, 1954).

The information transmitted in film and television systems are multimedia messages containing two main components image and sound. These components are transmitted through different human sensory channels, but they are synthesized during the perception of multimedia messages, which corresponds to the principle of perception integrity. Scientists from different fields have studied the intersensory perception phenomenon existence for many years: psychoacoustics, psychology, physiology, neurology, philosophy, musicology, and ecology. Currently, this problem has begun to be considered by specialists in the field of human-computer interaction.

A qualitative characteristic of perception is modality. Modality describes the information perceived by particular sensory system. The relevant modalities for modern film and television systems includes visual and auditory modalities. Audiovisual perception is characterized by the fact that information received within one sensory modality influences the perception of information received by another sensory modality (Storms, 1998).

Cross-modal audiovisual interaction has been studied for many years. There are many phenomena confirming that cross-modal interaction exists. For example, it has been proven that viewers associate an increase in the audio tone pitch with an increase in the brightness of the image (Marks, 1989; Martinussen, 2007; Masry, Hemami, Rohaly, Osberger, 2001; Stein, Meredith, 1993). The phenomenon of synesthesia is known as a phenomenon of perception, when the receptors of one sense organ are stimulated, along with sensations specific to it, sensations corresponding to another sense organ arise, in other words, signals emanating from different sense organs are mixed. A person not only hears sounds but also sees them, not only touches an object but also feels its taste (Neuman, 1990). Ventriloquism is a type of audiovisual cross-modal interaction (Bermant, Welch, 1976; Howard, Templeton, 1966; Ragot, Cave, Fano, 1988). The effect is to create the illusion that we are hearing the puppet's speech on the screen, when in fact we are hearing the voice of the ventriloquist. Even though we are aware that the doll cannot speak, it seems that the sounds come from her mouth. This effect is an illustration of the spatial connection that occurs during the perception of auditory and visual information. The dominance of one modality over another has been described in several works (more often the authors describe dominance of visual information over auditory ones) (Burrows, Solomon, 1975; Colavita, 1974; Jones, Kabanoff, 1975; Knudsen, Brainard, 1995). The influence of sound on eye's light and color sensitivity (Kravkov, 1936) and the influence of visual images on hearing sensitivity are known (Gregg, Brogden, 1952).

The described phenomena and effects confirm the need for a new approach to the development of integral quality assessment methods in television, which should take into account the phenomenon of intersensory interaction in the quality perception of audiovisual programs.

### **1.3. From content delivery to quality of experience delivery**

Digital television is a system where new information technologies are used in the production, distribution, and presentation of television programs. Yesterday television was a means of mass information, agitation, culture, and art, i.e. a social and political institution, today it also become a component of the information environment as an ecosystem where the passive audience is replaced by a multitude of actors who are both creators and consumers of information and media content (Воробьев, Степанов, 2011).

The convergence of high-definition television, film, and data networks is making a new service real, "quality of experience delivery", a service that replaces traditional "content delivery". Viewers of ultra-high-definition digital television already talk not only about the images and sound quality itselfs, but also about the emotional impressions left behind by the images and sounds they watch and listen to. This means that the television viewer will soon want to receive, anywhere and anytime, not just a television program, film, or audiovisual material, but a personalized visual and auditory experience, impressions, emotions (Гласман, 2006).

A similar trend can be seen in other areas of technology related to the audiovisual information transmission. Not only the digital indicators of mobile phones (dimensions, screen resolution, etc.) determine the popularity of certain devices. One of the main success factors of Apple Company is that people who use Apple devices every year get new sensations and impressions, often without purchasing new devices, but only by installing new versions of software (Bajarin, 2012).

Business models in the telecommunications industry have changed fundamentally in recent years. Telephony was a service that provided reliable communication anywhere and anytime and saved time and money in recent years. Today young people use telecommunications to

build their own social and entertainment ecosystem. The main purpose of the telecommunications system nowadays is to deliver personalized emotions and experiences (Bau, 2011).

“Quality experience delivery” should be the main task for the media industry. The study of viewers’ perception and the development of quality of experience evaluating methods is a significant task. And it serves as the basis for the technical requirements development for future media industry systems.

Assessment of qualitative parameters of media systems is reasonable to consider in the context of the structure of the information ecosystem. Society has entered the era of digital communications due to the rapid development of radio, television, communications and information technologies. All these technologies are associated with the creation, transmission, distribution, and processing of information obtained through communication channels. These instruments are components of the information environment, where there are many subjects—producers, and consumers of content simultaneously. Such an environment can be considered as an information ecosystem—a dynamic self-organizing system that has a network structure and consists of modules (Kilkki, 2008). The modules exchange information flows. There are producers and consumers of information in the system, and media technologies, that are used to produce and distribute information.

Existing network media technologies, that offer a huge variety of services, well demonstrate that user perception and evaluation are more important than any technical mechanisms used within the network and technical parameters. An important question is how different technical parameters affect end users and market players.

The quality of service traditionally applied only in a technical context in the evaluation of data networks. In rapidly network technologies development, which modern television systems are also based on, the stack of TCP/IP protocols has become the main protocol of media data transmission. The term “quality of service” is defined as a set of parameters that can be used to describe the network performance that are necessary to meet the requirements of various applications. Quality of service is a purely technical concept that describes the interaction between applications and network services. The usage of definition of “quality of service” is limited by the technical domain, means that another concept that describes and manages the interaction between users and applications and describes the degree of user satisfaction should be introduced.

Three modules can be defined in the structure of the information ecosystem: the user, the application and the network, the relationships between which are described using quality of service and user satisfaction (Moorsel, 2001; Muhammad, Chiavelli, Soldani, Li, 2006; Reichl, 2007). First of all, it is necessary to define and apply terminology that corresponds to the nature of each module. It is equally misleading to use technical terms to describe a human module and to use human concepts in a technical module. Therefore, it is necessary to create an algorithm or model that effectively converts technical parameters into measures of human perception.

The proposed information ecosystem structure can be used to define the key term—user satisfaction. Many definitions of this term have been proposed in the literature. The International Telecommunication Union ITU offers the following definition: the acceptability of an application or service as a degree of subjective perception by the end user (ITU-R WP6C, 2012). Work (Muhammad, Chiavelli, Soldani, Li, 2006) explains the difference between the terms user satisfaction and quality of service as follows: the main goal of the network and services

should be to achieve high user ratings, while network quality is the main structural element for the effective achievement of this goal. In (Li-yuan, Wen-an, Jun-de, 2006), user satisfaction is defined as a function of two aspects: monitoring/controlling the user's perception in real time; controlling and matching the service based on user satisfaction with the assurance that the quality of service will meet the users' requirements. As defined in (Lopez, Gonzalez, Bellido, Alonso, 2006), user satisfaction is defined as an extension of traditional quality of service.

Article (Moorsel, 2001) provides a more general definition of satisfaction, starting with a fundamental understanding of keywords. However, to describe specific role of a person, for example, as a user or a customer, the author suggests using clarifying definitions: the degree of user satisfaction and the degree of customer satisfaction. In all cases, the impressions and experiences acquired correspond to the role played by the person. The user satisfaction metric is usually measured at the final playback device and can conceptually be considered as the "residual" quality after the distortions placed at the stage of content production and content delivery over networks to the final device decoder. The assessment of perceived quality of service is the subject of both objective and subjective measurements.

Each module of the information ecosystem has its own terminology and language that cannot be directly applied for other modules. The definition of terms is of particular importance in customer and user modules, since an incorrectly defined term can affect the assessment of the subjective perception and degree of satisfaction of the person / customer and person / user. In work (Гласман, 1998), it was shown that quality scores in terms of the category scale of different languages turn out to be different on the interval scale. It is important to define the basic terms related to the assessment of user satisfaction degree in each language spoken by people and subjective measurements are made.

The term "Quality of Experience" is used as a summarized term to describe the degree of user satisfaction with certain services in the ITU documents published in English (ITU-R WP6C, 2012). It should be taken into account that a term that defines the degree of user satisfaction and is called "Quality of Experience" exists and is used in psychology and philosophy (Herman, 1990; Csikszentmihalyi, 1992), where it often means "quality of perception". In medical terminology, there is also the term "Quality of Experience". It is a subjective assessment of patient satisfaction with the outcome of medical care. In the context of television systems and telecommunications networks, the term "user satisfaction" is a subjective assessment that describes, from the user's perspective, the integral quality and quantity of services provided.

Summing up, we come to the conclusion that as a generalized term for describing the degree of user satisfaction with certain services of the media industry, it is reasonable to use the term called "Quality of Experience" which is understood as a subjective assessment that describes the degree of user satisfaction with audiovisual services, depending on a number of factors: goals pursued by the user while viewing; user expectations based on previous experience with similar services; the psychological state of the viewer; network parameters such as bandwidth, latency, jitter, and packet loss; type of device used for viewing; and finally, the type of content (news, clips, sports, etc.).

#### **1.4. Standardized methods for image and sound quality assessment**

The International Telecommunication Union (ITU) has developed a number of recommendations related to television image quality assessment: ITU-R BT.500-13, ITU-R BT.710-4, ITU-R BT.1128, ITU R BT.1129, ITU-R BT.1788, ITU R BT.1082, ITU-R BT.2021 (ITU-R

REPORT BT.1082–1, 1990; Rec ITU-R BT.1129–2, 1998; Rec ITU-R BT.1788–0, 2007; Rec ITU-R BT.2021, 2012; Rec. ITU-R BT.1128–2, 1997; Rec. ITU-R BT.500–13, 2012; Rec. ITU-R BT.710–4, 1998).

The main standard for subjective assessments in television is Recommendation ITU-R BT.500–13 (Rec. ITU-R BT.500–13, 2012). This Recommendation, first published in 1974 (and subsequently updated several times), describes a number of procedures for human evaluation of image quality. The Recommendation defines two main classes of subjective evaluations: evaluations that determine the quality of the system under optimal operating conditions (such evaluations are usually called quality assessments); evaluations that determine the ability of the system to maintain quality under suboptimal operating conditions due to signal transmission and distribution (such evaluations are usually called impairment assessments). Subjective assessments are carried out in accordance with a certain methodology, which defines the choice of test sequences, the choice of participants, assessments scales, observation conditions, measurement procedures and processing of results.

The Recommendation describes several methods for subjective assessments of perceived quality. The main ones are two: the Double Stimuli Impairment Scale (DSIS) method and the Double Stimulus Continuous Quality Scale (DSCQS) method.

The methodological basis for the development of technologies and methods of stereoscopic image quality assessment is provided by Recommendation ITU-R BT.500–13 (Rec. ITU-R BT.500–13, 2012). The improvement of this methodology in order to develop methods for assessing the quality of stereoscopic images is given in Recommendation ITU-R BT.1438 (Rec. ITU-R BT.1438, 2000). The subjective evaluation factors of stereoscopic television images, evaluation methods, viewing conditions, selection of participants and test materials containing still and moving images are considered in this recommendation.

Recommendations regulating the choice of test sequences, the choice of participants, assessment scales, observation conditions, measurement procedures and processing of results also exist for subjective sound quality assessments. The recommendation ITU-R BS.1284 (Rec. ITU-R BS.1284–1, 2003) should be considered as the main one. The exchange of sound materials and its types are constantly increasing. The ability to control and maintain the quality of sound materials becomes increasingly important. These considerations led to the development of technical regulation EBU3286 (1997) (EBU Document Tech. 3286, 1997). Recommendation ITU-T P. 800 defines subjective quality assessment methods that can be used to assess the quality distortions caused by various telephone network equipment (Rec. ITU-T P. 800, 1996). Recommendation ITU-T P. 805 (Rec. ITU-T P. 805, 2007) describes methodologies of sound quality assessments in dialogue (conversation) environment. Methodologies for subjective evaluations to detect subtle distortion in audio systems, including multichannel systems, are presented in Recommendation ITU-R BS.1116–1 (Rec. ITU-R BS.1116–1, 1997). Recommendation ITU-T P. 800.1 (Rec. ITU-T P. 800.1, 2006) discusses the terminology that can be used to describe the speech quality based on the mean opinion score of speech intelligibility (MOS). Recommendation ITU-T P. 800.2 (Rec. ITU-T P. 800.2, 2013) is a supplement. It discusses more general types of mean opinion scores for a group of participants, and describes the minimum information that should be obtained for a correct interpretation of the results.

The widespread multimedia services development led to the creation of a special study group “Audiovisual Quality in Multimedia Services”. As a part of the International Telecommunication Union, this group created one of the most important documents regulating quality

assessment methods in multimedia applications, Recommendation ITU-T P. 910 (Rec. ITU-T P. 910, 1999). The Recommendation ITU-T P. 910 (249) defines methods for subjective quality assessment of digital video images encoded with different bit rates in multimedia applications such as video telephony, videoconferencing and telemedicine. Recommendation ITU-T P. 911 (Rec. ITU-T P. 911, 1998) was developed as an extension of Recommendation ITU-T P. 910. It introduces two groups of subjective tests: optimization and qualification tests. Optimization tests are subjective tests that are carried out during the development and standardization of new signal processing algorithms or new systems. The purpose of these experiments is to evaluate the qualitative parameters of new systems or new processing algorithms. Qualification tests are subjective tests that are conducted to compare systems or equipment parameters available on the market. These tests should be carried out in an environment close to the actual operating conditions of the systems. An important part of the Recommendation ITU-R. P. 911 (Rec. ITU-T P. 911, 1998) is Appendix C, which describes the relationship between sound, image and audiovisual quality scores.

However, despite the fact that there are many standardized methods, the problem of assessing image and sound quality is still actual and far from being finally solved. The introduction of new services such as high-definition television, mobile television, and the integration of television with data networks require changes in the subjective assessment methods and the development of new methods for subjective quality assessment.

### **1.5. Quality assessment of television materials: current challenges**

A number of works have been done on integral quality assessment or quality of experience assessment in television broadcasting and related fields in recent years. Integral quality assessment in television systems, that includes evaluation of both image and sound quality, is conducted in a wide variety of television applications. For example, in (Neumann, Crigler, Bove, 1991), the influence of the sound quality of television receivers on the viewer's perception of television programs was studied. An interesting result was obtained a video clip with higher sound quality was always rated as more interesting. An additional unexpected result was noted that the image received a higher score with high sound quality accompany.

Integral quality assessment is important in virtual reality systems and other multimedia systems. In (Larsson, Västfjäll, Kleiner, 2001), an ecological approach to the multimodal perception of a virtual environment (virtual reality) was developed, according to which different perception modalities should interact in the process of forming an impression of complex objects. Typically, virtual reality systems are designed for visual observation. The results of the work showed that sound in virtual reality systems can be used not only as a channel for transmitting information, but also as a source of improving the sense of reality in virtual environment systems. The use of sound to create or enhance the sense of presence in virtual reality systems was also studied in (Turner, McGregor, 2007). The work (Martinussen, 2007) investigated the effect of using animation with stereo sound on the level of presence in multi-dimensional virtual reality systems.

The influence of various additional stimulus, such as light, wind, vibrations during the demonstration of audiovisual programs, i.e. increasing the number of multimedia programs modalities, is studied in (Waltl, Timmerer, Hellwagner, 2009). In (Winkler, Symmetricom, San Jose, Mohandas, 2008) it is proposed to normalize the level of quality of service and guarantee the necessary bandwidth in multimedia conferencing systems. The proposed approach is discussed using the results of sound quality assessment experiments. Similar studies

are described in (Bouch, Sasse, 2000). The works (Bench, Sasse, H. de Meer, 2000; Wilson, Sasse, 2000) study the users' requirements for the quality of service in multimedia data transmission networks. Various tasks of quality of experience assessment in multimedia systems are investigated in (Bevan, 2009; Ganglbauer, Schrammel, Deutsch, Tscheligi, 2009; Roto, Obrist, Väänänen-Vainio-Mattila, 2009; Väänänen-Vainio-Mattila, Segerstahl, 2009; Zaman, 2009). The fundamentally new Dolby AC-4 and DTS -UHD multichannel sound systems providing high sound quality developed by Dolby Labs and DTS for ultra-high definition video files. New systems allow to get full-fledged 3D sound, which leads to the ability to accurately localize apparent sound sources on a virtual sound stage, in whatever direction they move across the screen. One of the goals of creating new systems is a new level of sound and image integration (Быструшкин, Степаненко, 2014).

Cross-modal interaction in the process of audiovisual perception allows to optimize the parameters of television systems in order, for example, to reduce the data bit-rate. The paper (Lee, De Simone, Ebrahimi, 2009) investigates the effect of the focus of audiovisual attention on the perceived quality of audiovisual sequences of standard definition. It is shown that the sound source attracts visual attention, so image distortions in areas far away from the sound source are less noticeable. This fact can be taken into account when encoding video signals.

In (Hulusic, Aranha, Chalmers, 2008) the influence of cross-modal interaction between image and sound perception on the threshold parameters of the human visual system was investigated. The results found made it possible to achieve a reduction in the computing power required to process dynamic images, due to image coarser quantization in areas that were far from the center of attention, which associated with the sound source. At the same time, the overall video quality was not reduced. Similar problems are solved in (Bonneel, SinVri, Vijuid-Deltiuni, DrettakLs, 2010). The goal of (Vleeschouwer, Delmot, Marichal, Macq, 1997) is the adaptive distribution of the data bit-rate between different areas of the image according to their semantic significance. The proposed method shows its effectiveness at low and very low encoded data bit rates.

Research in the field of quality of experience assessment is very important for creating efficient mobile TV systems, that operate in conditions of low and variable bandwidth of communication channels. The main factors affecting the quality of experience in mobile TV systems are investigated in (Knoche, Sasse, 2008). The development of audiovisual quality assessment methods on mobile devices operating in 3G networks is carried out in (Ries, Puglia, Tebaldi, Nemethova, Rup, 2005). It is noted that the quality of music and voice playback affects the perceived image quality.

The evaluation of video quality in mobile television is carried out in (Jumisko-Pyykkö, Häkinen, 2005). The influence of basic parameters as data bit-rate, frame rate and screen size of a mobile device is investigated. It is shown that the H.264 codec provides the highest quality when the bandwidth and data bit-rate is tightly limited. A combination of video and audio streams is also investigated for low bit-rate streams. It is noted that the ratio of sound and image bitrates of 32/128 Kbps is the most relevant for the viewer, although research of the impact of program genre is required.

Low bit-rate audio coding for mobile multimedia applications is studied in (Kontola, Makinen, Taleb, Bruhn, Bessette, Salami, 2006). The work (Utriainen, Jumisko-Pyykkö, 2010) investigates the influence of image and sound coding parameters (frame rate, bit-rates) in mobile stereoscopic television. The influence of these parameters in the context of mobile devices use is considered in (Jumisko-Pyykkö, Utriainen, 2010). An ecological approach to as-

sessing the quality of service in mobile television is considered in (Thurstone, 1931), and user requirements for network quality of service are discussed in (Bouch, Sasse, 1999). The aim of work (Winklerand, Faller, 2005) is to evaluate audiovisual quality in mobile applications using the AVC/H.264 codec for image and AAC for audio, and to study the interaction between image and sound in terms of perceived audiovisual quality. The effect of text readability on the perceived video quality is explored in (Knoche, McCarthy, Sasse, 2006). The results can be used to obtain quantitative estimates of image quality improvement when delivering images and text separately in mobile television.

## **1.6. Quality assessment methods and models**

In the field of integral quality assessment different approaches are used to solve the tasks and different models and methods are proposed. One of multimedia quality assessment models is discussed in (Hands, 204). It is noted that image and sound quality are integrated in the human nervous system using the multiplicative law. For “talking head” type of content, both modalities have a noticeable effect, with the weight of sound quality even larger than the weight of image quality. For dynamic scenes, image quality is more important and has more weight. Various methods for objective image and integral quality assessments are considered in (Barber, Laws, 1994). The poles are the signal-to-noise ratio, on the one hand, and hybrid metrics, on the other. It is noted that metrics based on the data bit- rate are the most promising for practical applications.

The influence of context, i.e. circumstances in which mobile devices are used to quality of the experience assessments in mobile television was studied at the work (Jumisko-Pyykkö, Hannuksela, 2008). In each task in which a mobile device was used, the variable parameter affecting quality was the bit-rate of transmitting data. In (Wilson, Sasse, 2000), a combined method, which uses task characteristics, user satisfaction, and user cost, is proposed to estimate the level of multimedia quality required in various network multimedia applications. A similar approach is developed in (Wilson, Sasse, 2000, 2001) for multimedia teleconferencing systems.

The paper (Jumisko-Pyykko, Hakkinen, 2007) shows that the main factors to be taken into account when assessing audiovisual quality are image quality, genre of the program, sound quality, and cross-modal audiovisual interaction. The importance of integrated audiovisual quality assessment is noted in (Winkler, Faller, 2005). This area has not been sufficiently developed, although there are many standards that describe methods of separate assessments of image and sound quality. To take into account audiovisual interaction is relevant for applications that use low data bit-rates. The optimal ratio of audio and video data bit-rates, that provides the highest integral quality is proposed in the paper. However, the proposed ratio is a particular case and cannot be extended to different values of the total bit-rate and various types of content of audiovisual programs. A similar study was carried out for mobile television systems (Jumisko-Pyykko, Hakkinen, 2006). The paper presents the results of extensive research, but the conclusions are also made for particular case.

The paper (Yamagishi, Tominaga, Hayashi, Takahashi, 2007) discusses a model of objective quality assessment, that makes it possible to predict the subjective image quality based on the physical parameters of a videotelephone network system (frame rate, bit rate, packet loss rate). The paper does not describe the cross-modal interaction of image and sound scores. In (Watson, Sasse, 1998), it is noted that the existing standardized methods for measuring the image and sound quality do not correspond to the tasks of new multimedia applications. Many

works are known to assess the quality of service at the network level of multimedia applications, but there is a need for multidimensional assessments that take into account the subjective quality assessment of the user, that alone can determine the relevance of new multimedia applications implementation. The need for a multimodal assessment is proved in (Strohmeier, Jumisko-Pyykkö, Kunze, 2010). A model that allows to predict subjective image quality assessment based on the results of objective measurements (bit rate, image size) in mobile television is proposed in (De Koning, Veldhoven, Knoche, Kooij, 2007).

A particular structure in solving the complex problem of quality of experience assessment is introduced in the work (Wikstrand, 2003). The author proposes to solve the problem based on the task decomposition, for the first time introducing a model of three levels: the content level of audiovisual programs; the media level; the network level (Gulliver, Ghinea, 2006). The content level of audiovisual program is related to the transmission of information from content creation to the viewer and its impact on the audience. The media level (it could be called the transport layer) covers issues related to encoding, data compression for transmission through networks and playback on the viewer's device. The network level covers information delivery and exchange over networks. Technical parameters that affect the quality of service at the network level include packet loss, jitter, delays.

A number of papers address quality measurement and quality management issues at three levels. At the network level, bit error values, segment loss, segment order (Ghinea, Thomas, 2000), delay and jitter (Claypool, Tanner, 1999; Ghinea, Thomas, 2000; Procter, Hartswood, McKinlay, Gallacher, 1999) were changed to study the quality of service. All of the listed technical parameters, including channel bandwidth allocation (Wang, Claypool, Zuo, 2001), were used to measure the network performance.

At the media or transport level, the effects on the quality of frame rate (Apteker, Fisher, Kisimov, Neishlos, 1995; Ghinea, Thomas, 1998; Kawalek, 1995; Kies, Williges, Rosson, 1997), subtitling (Gulliver, Ghinea, 2003), animation method (Wikstrand, Eriksson, 2002), image spatial resolution (Kies, Williges, Rosson, 1997), audio and video stream synchronization (Steinmetz, 1996), and video compression parameters (Masry, Hemami, Rohaly, Osberger, 2001; Winkler, 2001) have been investigated.

At the content level, or user level, there are a number of parameters and factors that affect the quality of the experience. In (Rimell, Hollier, 1999) and (Ghinea, Thomas, 1998), the type of content was a variable parameter. In the first study two types of content were used: the bank annual report and the staged sexual harassment. The second study used fragments from a news program, a music video, and a rugby match. Level of understanding, level of interest, comprehension, and memorization were analyzed as parameters for measuring quality. Papers (Ghinea, Thomas, 1998; Gulliver, Ghinea, 2003; Masry, Hemami, Rohaly, Osberger, 2001; Procter, Hartswood, McKinlay, Gallacher, 1999; Rimell, Hollier, 1999; Steinmetz, 1996) investigated the effect of content type changing on quality of experience. The authors of the works used different parameters to measure the quality of audiovisual programs and different ways to manage it.

The parameters that are most important and affect the quality of experience at the content level may vary according to the function and content type. For games, content level quality parameters include the degree of enjoyment and pleasure, as well as the ability to win the game and get a high score. Especially for conversational-type programs parameters that take into account how well the meaning of the information is conveyed should be included (Veinott, Olson, Fu, 1999). Lip-reading ability and comprehension (Barber, Laws, 1994), speech intel-

ligibility, facial and emotion recognition, and comprehension of presented material, for example, may be important (Barber, Laws, 1994; Kies, Williges, Rosson, 1997; Masry, Hemami, Rohaly, Osberger, 2001).

Quality of experience research can take place at one level or at several levels at once.

The accuracy and quality of television programs and audiovisual materials can be measured through objective measurements or subjective assessment.

### **1.7. Objective quality measurement methods**

Objective measurements that determine the accuracy of playback and quality were carried out in (Verscheure, Frossard, Hamdi, 1998; Wang, Claypool, Zuo, 2001; Winkler, 2001). (Verscheure, Frossard, 1998; Winkler, 2001) used the signal-to-noise ratio as an objective parameter characterizing the fidelity of playback. The signal-to-noise ratio is a mathematical indicator that does not take into account characteristics of human perception.

(Verscheure, Frossard, Hamdi, 1998) also used motion picture quality metrics as a model to predict how users would rate material. The frame rate was used as an objective parameter for accuracy assessing in (ETSI TS133 220 (3GPP SA 3); Wang, Claypool, Zuo, 2001). In (Winkler, 2001), the author used the Perceptual Distortion Metric<sup>3</sup> (PDM) as a potential model to predict how users will evaluate the quality of streaming video. The author proposes an improved version of PDM that takes into account the visual appeal of streaming video. External attractiveness of presented video was calculated as a combination of color reproduction and sharpness.

Objective and measurable indicators affecting the quality can include: signal quantization step, frame rate, data bit-rate, spatial resolution. In different works the influence of changes in these parameters was investigated: in (Apteker, Fisher, Kisimov, Neishlos, 1995; Barber, Laws, 1994; Kies, Williges, Rosson, 1997; Masry, Hemami, Rohaly, Osberger, 2001; Wilson, Sasse, 2000) the effect of frequency and frame loss changes was studied; (Masry, Hemami, Rohaly, Osberger, 2001), (Verscheure, Frossard, Hamdi, 1998) and (Winkler, 2001) investigated bit-rate changes during video encoding. In (Barber, Laws, 1994) and (Kies, Williges, Rosson, 1997), image resolution and physical size were varied. In (266) the effect of image and sound desynchronization on the overall quality of experience was studied. Different codecs were tested in (Masry, Hemami, Rohaly, Osberger, 2001) and (Winkler, 2001). (Moore, Foley, Mitra, 2001) simulated errors and artifacts that are typical for MPEG-2 compression format.

In the articles (Wilson, Sasse, 2000), by measuring physiological parameters: heart rate and galvanic skin response, the stress that occurs when watching audiovisual programs with low image quality was studied.

### **1.8. Subjective quality measurement methods**

#### **1.8.1. Basic principles**

The purpose of subjective assessments is to evaluate subjective, perceived, or experienced quality. Subjective assessments are characterized by an assessment of stimulus quality based on the process of human perception. Subjective assessments are usually carried out during the development and optimization of television system parameters. They are necessary when developing objective quality metrics.

Subjective measurements of playback accuracy and streaming video quality have been made

in (Fong, Hui, 2001; Kies, Williges, Rosson, 1997; Moore, Foley, Mitra, 2001; Rimell, Hollier, 1999; Steinmetz, 1996; Winkler, 2001). In (Moore, Foley, Mitra, 2001) and (Steinmetz, 1996), the probability of error detection and the degree of annoying stimulus were used as parameters describing the accuracy of playback. The detected errors were assigned different levels of annoying stimulus depending on their type, size, duration, and location.

In (Fong, Hui, 2001), viewers were asked to rate the quality of displayed content using a scale from 0 to 100 %. In the study described in (Kies, Williges, Rosson, 1997), participants were asked to compare the quality of streaming video with the original undistorted video. The authors asked to rate the quality using the following statements: “the video was as good as a live lecture in the same room”, “the video was as good as watching it with a videocassette recorder (VCR) or regular TV”, using a seven-point Likert-style scale. In (Procter, Hartswood, McKinlay, Gallacher, 1999) the same scale was used to evaluate streaming video quality.

In (Winkler, 2001) Double Stimulus Continuous Quality method was used to determine which of the two clips viewers preferred more. In this method, two clips were used, one of which is the original, undistorted clip, the second is a distorted clip. Two versions were presented to the viewer at the same time. Participants were asked to rate the quality of each clip on continuous scales with two marks “poor” and “excellent”.

A huge number of subjective quality assessment methods can be grouped into three classes: psychophysical methods, descriptive methods, and user-oriented methods (Jumisko-Pyykkö, Strohmeier, 2008).

### **1.8.2. Psychophysical methods**

Experiments using psychophysical methods are highly regulated and formalized. They are usually carried out in accordance with the recommendations of international organizations for standardization. Quantitative subjective image quality assessment has been carried out for a long time based on the Recommendation ITU-R BT.500–13 “Methodologies for the subjective assessment of the quality of television images” (Rec. ITU-R BT.500–13, 2012). The document was improved several times and is currently the main document regulating the experiments in the field of image quality evaluation. The documents regulating audio quality evaluation experiments are known (EBU Document Tech. 3286–1997, 1997; EBU Technical Recommendation R90–2000, 2000; 233. Rec. ITU-R BS.1284–1, 2003). In addition to the Recommendations ITU-R BT.500–13 (Rec. ITU-R BT.500–13, 2012) and ITU-R BT.1438 (Rec. ITU-R BT.1438, 2000), regulating the methods of subjective quality assessment of mono- and stereo television images respectively, the Recommendations ITU-R P. 910 (Rec. ITU-T P. 910, 1999) and ITU-R P. 911 (Rec. ITU-T P. 911, 1998) offer a number of quality assessment methods of video and audio-visual quality of multimedia systems.

### **1.8.3. Descriptive methods**

Works carried out using descriptive methods allow us to find parameters that describe quality from the user’s point of view. The descriptive methods are used when subjective quality cannot be considered as a value that depends on some set of defined parameters. In this case, the definition of system quality is vague, dependent on the personal knowledge of the user and interpreted individually by each expert (Reiter, Köhler, 2005). Human behavior in the process of evaluating subjective quality can be determined by testing, and testing methods should be open-ended in order to understand the underlying attributes and properties of quality (Jumisko-Pyykkö, Hakkinen, Nyman, 2007). The interview method can be used to investi-

gate individual quality parameters that are important for each user, in addition to quantitative scores given by experts on an absolute category scale. In (Jumisko-Pyykko, Reiter, Weigel, 2007), the results of a paired comparison test, where interviews were used to investigate the differences between each pair of stimuli, are presented. The approach proposed in (Williams, Arnold, 1984) is used in many areas related to sensory assessments. This approach allows participants of experiment to use their own vocabulary to describe the product being tested.

#### **1.8.4. User-oriented methods**

Different works in the field of user-centered quality assessment extend the boundaries of the usage of standardized methods, allowing them to be applied to real systems or services. User-centered methods rely on the introduction of the so-called “user experience” (Jumisko-Pyykkö, Weitzel, Strohmeier, 2008). Currently, there is no generally accepted definition of the term of user experience among researchers focus on human-computer interaction. However, there are many interpretations of this term (Forlizzi, Battarbee, 2004). In (Hassenzahl, Tractinsky, 2006) the following interpretation of the term of user experience is proposed: user experience is the cumulative result of the internal state of the user, the parameters of the developed system and the context, or conditions in which human-system interaction occurs. This definition is supplemented in (Roto, 2006), the authors developed a model describing user perception in the conditions of viewing audiovisual materials on mobile devices. The model consists of three modules: user, system and context. The work also summaries the definitions, specifies the vocabulary and the scope of application for each of them.

The “user” module is defined as the person who controls and manages the system, he can be defined based on the characteristics of his needs, motivation, experience, mental state and resources.

“System” is defined as an organized set of components, that is necessary for the operation or usability of the product under study. From a user’s perspective, a system in an environment such as mobile television might be a device, a browser or player, a connection (connection, access to resources) that provides access to a content. Content in these works refers to any type of moving image or video. Short clips, news, weather, sports or music videos are the most popular content for viewing on mobile terminals. A large part of the content viewed on mobile terminals is so-called user-generated content. The concept of service, including the commercial and service model, often overlap with the term “systems” from a user perspective. Service parameters are part of the system definition in the mobile television environment.

“Context” refers to the search and viewing conditions of audiovisual content. The main components of context are physical, temporal, social, and problem-oriented factors. Physical factors include visible features of a situation or physically felt conditions, including location. For example, the most common places to watch mobile television are on the go, at home, and at work (O’Hara, Mitchell, Vorbau, 2007; Oksman, Noppari, Tammela, Miikinen, Ollikainen, 2007). Temporal context characterizes the time available to complete a task. Actions in the temporal dimension can be classified into fast, normal, and pending (Tamminen, Oulasvirta, Toiskallio, Kankainen, 2004). For example, the duration of watching mobile TV programs ranges from 10 to 15 minutes on average. Social context is a description of other people present, their characteristics, visible roles, and interpersonal interactions. Social context includes an assessment of other people’s influence on the user and the user’s contribution to solving social tasks (Belk, 1975). Social context refers to the way of viewing—by one or more users at the same time. Problem-oriented context describes multitasking and the possible interrup-

tion of task-related activities (Hiltunen, Laukka, Luomala, 2002; Roto, 2006), such as mobile TV viewing. It also characterizes the content viewing goals (entertainment or educational purposes). It should be noted that, for example, in mobile communication the physical and social environments are heterogeneous and can change during use, e.g., from individual use to use by a group of people, from purposeful to unplanned actions (Tamminen, Oulasvirta, Toiskallio, Kankainen, 2004).

Various standardization organizations actively develop documents regulating the quality of experience. The ETSI HF STF 354 (Specialist Task Force 354 within the Human Factors Group of ETSI) is developing a set of required recommendations for real-time multimedia services that aim to provide a good level of quality of experience. The goal is to provide objective and subjective measurements of quality of experience parameters for specific communication situations and levels of quality of service. The recommendations are formulated in ETSI documents EG 202 534 and ETSI TR102 535 (ETSI EG 202 534, 2007; ETSI TR102 535, 2007).

In contrast to psychophysical approaches, user-centered methods extend the boundaries of subjective quality assessment, based on the consideration of human behavior. For example, the approach of Gini and Thomas (Ghinea, Thomas, 1998) to assess the quality of experience takes into account the user's overall satisfaction with audiovisual content and the user's ability to understand the information contained (semantic information) in the content. The approach allows to describe user experience taking into account two aspects. Quality is defined at the media level (transport layer) independently of the content, i.e. users evaluate objective parameters of the quality of service. Also, quality is evaluated at the content level, i.e. the level of user satisfaction is determined.

### **1.9. Quality of experience assessment**

The challenge for current quality assessment researches is to take into account all aspects of creating a holistic approach to quality assessment that will define the quality of experience in terms of user characteristics, system and context of use. The development of new user-oriented quality assessment methods should be carried out using some basic principles (Jumisko-Pyykkö, Strohmeier, 2008).

First, quality perception is an active process involving different levels of human information processing. This process includes information processing at a low sensory level and information processing at a high cognitive level, including personal knowledge, attitudes, expectations, and emotions (Neisser, 1976). Each sensory modality has specific parameters that depend on the stimulus physical dimensions. At the cognitive level stimuli are interpreted by the individual, their particular meaning, importance and goals are determined. For example, individual emotions, knowledge, expectations, and representational patterns, in reality, influence the value that each sensory modality has and govern a person's situational behavior and active interpretation of quality (Jumisko-Pyykko, Hakkinen, Nyman, 2007). As it was noted earlier, integral multimodal perception is created when information from the visual and auditory channels is combined. Different modalities can interact and influence the perception of information received by another perceptual channel. Multimodal quality appears to be more than just a sum of two different perceptual channels (Hands, 2004).

Second, the principle of the integrity of user experience, based on the introduction of a single complex consisting of user, system, and context, must be considered. It allows critical components of the system to be discovered (Hassenzahl, Tractinsky, 2006; Roto, 2006).

Third, quality assessment experiments are part of a design process that aims to meet user needs. User-centered design is a process that reflects knowledge about users and their involvement in the design process as a key stakeholder (Preece, Rogers, Sharp, Benyon, Holland, Carey, 1994). The process is characterized by active user participation to understand user needs, iterative design and evaluation, and a multidisciplinary approach. The resulting benefits of user-centered design are characterized by increased end-user satisfaction and reduced system development costs.

The quality of experience assessments should be an important element in the new multimedia systems design process. In recent years, a large number of works dedicated to the quality of experience evaluation have been carried out. However, there is no universally accepted definition of the quality of experience, and there are no well-developed and standardized methods for its evaluation. An analysis of works has shown that the quality of experience depends on a large number of factors (the goals pursued by the user while watching; user expectations based on previous experience from similar services; the psychological state of the viewer; network parameters such as bandwidth, latency, jitter and packet loss; the type of device used for viewing; and finally, the type of content (news, clips, sports, etc.) and the unfolding scenario), the task of formalizing the influence of which is relevant and unresolved at the present stage. The most difficult task is to describe the context (conditions of viewing; psychological state and emotions of the viewer; gender and age of the viewer; viewer's viewing goals for a particular program, etc.) to be able to apply the results in practice.

Existing standardized psychophysical quality assessment methods require a high degree of regulation and formalization in order to ensure reproducibility of results; they clearly regulate the choice of test sequences, the choice of experts, assessment scales, viewing conditions, measurement procedures, and results processing. To assess the quality of experience, various works have proposed descriptive methods that use viewer interviews, or user-oriented methods, which, extend the boundaries of the usage of standardized methods, allowing them to be applied to real systems or services. User-centered methods rely on the introduction of the term so-called "user experience". The transformation from existing subjective assessment methods to the development of new methods for assessing the quality of experience cannot be done all at once. Two steps should be done. At the first step, it is necessary to study the integral quality, as an assessment of perception of a continuous set of two components of an audiovisual program (image and sound). Integral quality is the most important component of quality of experience. It is important to investigate interaction and relationship between the image and sound quality perception. There are many phenomena that confirm the existence of this interdependence. The described phenomena and effects confirm the need for a new approach to the development of integral quality assessment methods in television, which should take into account the phenomenon of intersensory interaction in the perception of audiovisual programs. In the second step, the influence of context must be taken into account.

### **1.10. Conclusions to section 1**

1. Television broadcasting today is undergoing a radical transformation associated with the use of digital and information technology and the transition from hardware solutions to software solutions. Television is becoming a component of the information environment as an ecosystem, in which a passive audience is replaced by a multitude of actors who are both creators and consumers of information and media content.
2. Television is no longer the only audiovisual playback device. TV programs can be

displayed on the screen of a desktop computer, laptop, handheld computer or mobile phone. The market for mobile communication services is huge, and the new mobile television has all the opportunities for formation, development and growth.

3. The convergence of high-definition television, cinema and data transmission networks, which symbolizes the transition to a post-industrial: or information society, and which will certainly happen soon, makes a new service real – “quality of experience delivery,” which begins to replace the usual “content delivery. The television viewer will soon want to receive, anywhere and anytime, not just a television program, film or audiovisual material, but a personalized visual and auditory experience, impressions, emotions.
4. Digital television broadcasting development is directly related to the video compression methods and technical parameters improvement, or compression of digital video data streams and audio compression, or compression of digital audio data streams. Compression allows the most expensive resource of communications and information systems – the bandwidth of communication channels – to be used efficiently.
5. Application of compression requires solving an optimization problem aimed at achieving the maximum level of displayed image and sound quality while meeting the technical limitations associated with the bandwidth of communication channels. Evaluation of image and sound quality should be based on taking into account the characteristics of subjective perception of distortions and artifacts caused by codecs of image and sound signals compression.
6. Determination of optimal video and audio codec parameters is usually done separately for image and sound, however, image and sound quality evaluations in audiovisual programs are not independent. A new approach is needed to optimize codec parameters based on integral quality assessment in television and taking into account the phenomenon of inter-sensory interaction during the perception of audiovisual programs. This approach corresponds to the principle of the integrity of perception as a process of forming a holistic image of the object, that arises as a result of the impact of the objective world on the senses.
7. It is reasonable to consider quality assessment of audiovisual systems with compression in the context of information ecosystem structure – a dynamic self-organizing system with a network structure. Quality assessment in such systems with a network structure has two aspects: quality of service and quality of experience. Quality of service is defined as a set of parameters that can be used to describe the network performance that is necessary to meet the requirements of different applications. Quality of experience is a subjective assessment that characterizes the user’s satisfaction with audiovisual services and depends on a number of factors: the user’s viewing goals; user expectations based on previous experience with similar services; network parameters such as bandwidth, latency, jitter and packet loss; the type of device used for viewing; and finally, the type of content (news, clips, sports, etc.) and the unfolding scenario.
8. Evaluating the quality of experience as a subjective value that characterizes the degree of user satisfaction of audiovisual services generates the need to consider the semantic side of information, especially when this information should serve as the basis for certain practical actions. But television and other media systems are not only media, they are also areas of culture and art. Therefore, the quality of experience evaluation should be extended to the sphere of aesthetic perception and objects of art as well.

9. Assessing the degree of viewer (and listener) satisfaction with the quality of image and sound displayed in digital television systems with a data compression, is an important and relevant task for theory and practice. The relevance of the task of studying the effect of compression parameters on the quality of the displayed image and sound increases with an increase in the degree of compression, high values of which are achieved by reducing psychophysical redundancy and introducing distortions and artifacts. Such studies are especially important in mobile television systems, which use the limited broadband communication channels, and the target compressed data bit rate drops to megabits per second and even to lower values.

## **2. Integral assessment of image and sound quality, taking into account the cross-modal interaction of the components of an audiovisual program**

### **2.1. Methodology for integral quality assessment**

A technique that ensures the consistency and reproducibility of results is of paramount importance when conducting psychophysical experiments. The review of standard and possible alternative methods, performed in Section 1, showed that the existing methods do not meet the objectives of this work, therefore, a new method for conducting psychophysical experiments should be developed, which can be used to study cross-modal interaction when assessing the integral quality in digital television systems with data compression.

A number of standards can be identified as prototypes, which contain recommendations for conducting subjective experiments in television. Recommendation ITU-R BT.500–13 (Rec. ITU-R BT.500–13, 2012) regulates subjective experiments to evaluate the image quality of television systems. Recommendation ITU-R BS.1284–1 (Rec. ITU-R BS.1284–1, 2003) regulates the conduct of subjective experiments to evaluate sound quality, without regard to image quality. Recommendation ITU-T P. 911 (Rec. ITU-T P. 911, 1998) regulates the conduct of subjective experiments to evaluate the audiovisual quality of multimedia programs. However, the existing standards: Recommendation ITU-R BT.500–13 (Rec. ITU-R BT.500–13, 2012), Recommendation ITU-R BS.1284–1 (Rec. ITU-R BS.1284–1, 2003) regulate the conduct of subjective experiments to evaluate image quality without taking into account sound quality or sound quality without taking into account image quality. Therefore, the direct use of these documents for multimodal quality assessment is not possible.

The methodology for conducting a quality assessment experiment should be closely related to the purpose of the system. As discussed in Section 1, performance evaluation of audiovisual and television systems is useful to consider in the context of the structure of the information ecosystem, where quality evaluation has two aspects: quality of service and quality of experience. The evaluation of the integral quality as the “impression quality”, which is a subjective value that reflects the degree of user satisfaction with an audiovisual program of different genres with an unchanged viewing context, should take into account the informational component of the audiovisual program. Sound and image are semantically linked in television programs, and in order to take into account their mutual interaction, the semantic component must play an important role.

The techniques described in ITU-T P. 911 (Rec. ITU-T P. 911, 1998) are suitable for assessing quality depending on technical parameters, such as link bandwidth, delay, jitter and packet loss, image resolution, etc., without taking into account changes in the amount of semantic and aesthetic information contained in an audiovisual program under the influence of various technical parameters. In this section, when planning an experiment, a special place is given

to the development of a methodology for integral quality assessment, taking into account the informational approach based on the existing standardized method using an impairment scale with two stimuli.

To conduct experiments on assessing the integral quality of an audiovisual program, it is proposed to use an improved quality assessment method with a double stimulus and fixing the results on an impairment scale.

The method proposed in this work was based on the method using the two-stimulus impairment scale (DSIS) (Rec. ITU-R BT.500–13, 2012). The order of sequence presentation, time intervals and other conditions of the experiment were in accordance with Recommendation ITU-R BT.500–13 (Rec. ITU-R BT.500–13, 2012). A group of observers were sequentially presented with pairs of audiovisual sequences; each pair includes a reference uncompressed sequence and a test sequence with the same content, but distorted by video and audio compression artifacts. The experts were asked to compare the integral quality of the test sequence with the quality of the reference sequence and rate it on a five-point impairment scale.

A total of 25 pairs of sequences were randomly presented with different values of video and audio data bitrates corresponding to a wide range of changes in audiovisual quality (from the reference level corresponding to the highest rating to the level of minimum ratings). After the end of the experiment, for each combination of parameters, the score averaged over all experts was calculated.

The recommendations do not strictly regulate the approach to the preparation of instructions for experts or observers, however, instructions are an important part of the experiment.

Before the beginning of each session of experiments, the observers got acquainted in detail with the goals and objectives of the experiment, the manifestations of video and audio compression distortions (with demonstration of examples of distortions), the order of presentation of materials, and the rating scale. The purpose of this experiment was to determine the characteristics of the integrated perception of audiovisual programs when assessing distortions (artifacts) that occur during compression, and to determine the acceptable range of introduced artifacts.

The introduction of audio and video data compression is always associated with the appearance of noise and the appearance of image and sound distortions, the visual and auditory manifestation of which increases with increasing compression ratio. The following distortions are possible.

## **Image**

1. The most noticeable artifact is the block structure. The block is a simple geometric figure and therefore is clearly visible in the image; blocks have clear boundaries—jumps in brightness and color.
2. Image blur—reducing the sharpness, definition of the image; blurring or complete disappearance of small details.
3. The appearance of edging on the borders of the image.
4. Destruction of boundaries—breaks in the contours of objects.
5. Jitter—most noticeable on stationary or slow moving objects.
6. Intermittent movement.

## Sound

1. The appearance of extraneous sounds—various overtones, hissing, ringing, the appearance of background noise.
2. Violation of transparency and naturalness of sound.
3. Emergence of the echo effect.
4. Violation of the spatial impression—the stereo panorama becomes flatter, the sound planes disappear.

Another artifact of television program distortion is the appearance of a delay between sound and image, caused by the difference in signal processing time in the decoding and decompression paths in the receivers. In this work, the influence of this artifact is not studied.

The assessment of the integral quality should be based on the overall, integral impression of viewing the audiovisual sequence. To build a scale of subjective assessments and explain to experts from what positions they should approach the assessment of the perception of integral quality, this paper proposes to use a new, informational approach (Белозерцев, Гриненко, 2009; Гриненко, Поддубная, Гласман, Белозерцев, 2013; Перегудов, Гласман, Белозерцев, Гриненко, 2008; Перегудов, Гласман, Белозерцев, Гриненко, 2009; Grinenko, Poddubnaya, Glasman, Belozertsev, 2011; Peregudov, Glasman, Belozertsev, Grinenko, 2008).

An audiovisual program contains a certain information. In news and other similar stories, this is mainly semantic information. Bringing this information to the viewer leads to a decrease in the degree of uncertainty (an increase in the amount of knowledge) about any fact, event, phenomenon. Artistic, musical and other similar programs contain mainly aesthetic information. Receiving aesthetic information gives the viewer aesthetic pleasure.

The integral perception of audiovisual material depends on many factors: image and sound parameters, program genre, plot content, etc. Compression distortion leads to a reduction in the amount of information delivered to the viewer. The impairment of the quality of an audiovisual work due to distortions can be integrally assessed by the reduction in the amount of information (the completeness of the elimination of uncertainty).

If one of the components of the test program (image or sound) has a reference quality, but a large proportion of the information load is concentrated in the second, distorted by compression, component, then a significant overall impairment can be observed. For example, image distortion in view programs leads to a significant deterioration in the overall impression even with undistorted sound.

On the other hand, if one component is distorted by compression, but most of the information is contained in the second component, then the overall degradation may be absent or minimal. An example is the distortion of the announcer's image in a news program does not lead to a serious deterioration in the overall impression (Гриненко, 2011; Перегудов, Гласман, Белозерцев, Гриненко, 2009; Peregudov, Glasman, Belozertsev, Grinenko, 2008).

After viewing a pair of audiovisual sequences, the expert-observer was asked to assess the degree of degradation of the second of the presented video sequences relative to the first one (reference). The score had to reflect the reduction in the amount of semantic and/or aesthetic information in the test sequence compared to the reference one. Distortions are most noticeable and affect the perception of the program as a whole, if the most important component of

the program (video or sound) is distorted. The experiment is designed in such a way that the degree of compression of audio and video data is different in each experiment. The scale that is used for expert assessments is developed in the next section.

## **2.2. Distortion Rating Scale**

The main advantage of psychophysical experiments is that the measured quality of television systems directly reflects the reaction of viewers. The data obtained can be used in the development of new systems, improvement and modernization of existing ones. The quality that is formed at the output of a television system depends on a number of technical parameters or objective indicators. Therefore, psychophysical experiments are often carried out to build mathematical models that characterize the influence of various technical parameters on subjective quality.

As shown in Section 1, there are various techniques for conducting psychophysical experiments and measuring quality. In the general case, the measurement process is a procedure by which the measured object is compared with some reference and a numerical expression is obtained on a certain scale (Ядов, 1987).

In order to assign numbers to the properties of the measured object according to certain rules, it is necessary to identify its content structure and find a correspondence between it and the measurement tool. The most important aspect in conducting subjective experiments is the choice of measurement standard, that is, the choice of the scale and the determination of its properties. Often, qualitative characteristics do not have established measurement standards. They have to be developed based on the nature of the measured object and according to the research hypotheses.

The search for a measurement standard can be divided into four stages. At the first stage, a qualitative classification of the object is carried out. At the second stage, it is established whether the found properties have a discrete or continuous nature, whether they can be represented in the form of various successive states of the quality being measured. At the third stage, empirical indicators are established, or outwardly well-distinguishable indicators of the properties of the object. At the fourth stage, it is checked whether all the properties of the object according to the accepted indicators fit into the ranged series. Relationships between scale items should reflect the relationship of the object's properties in terms of "equal", "greater than", "less than". The validity of the scale lies in the fact that it purposefully measures a well-defined property or feature, without mixing it with others.

The accuracy and correctness of the measurement depend on: the degree of stability of the measured object, the sensitivity of the measurement standard, the absence of systematic measurement errors, and, finally, the stability of the measurement itself.

There are different types of scales. One of the most important properties of the scale is related to the ability of the scale to meet the requirements of various operations with numbers. The following types of scales can be distinguished (Ядов, 1987).

The simplest scale is the nominal or naming scale, which establishes relations of equality between properties of the same class. The points of the scale are standards for the qualitative classification of properties. This scale allows you to find the distribution frequencies by scale points using percentages or in natural units, as well as the average trend in modal frequency. The most powerful method of quantitative analysis in this case is to establish the relationship between the rows of properties that are randomly located.

An ordinal scale, or an ordered nominal scale, is a completely ordered scale of names that establishes relations of equality between phenomena and relations of sequence in terms of more or less. Rank scales are a commonly used variation of this type of scale. Using these scales, you can order the measured objects from the most to the least important, significant, preferred. The intervals in the scale are not equal, the numbers on the scale indicate only the order of the features. All operations are carried out with ranks, and not with a quantitative expression of properties in each point of the scale. Numbers that reflect the order of the properties of measured objects are amenable to monotonous transformations. They can be replaced by others, but with the preservation of the previous order, so that the relationship between the ranks remains unchanged. For further processing of the results obtained using an ordered scale, in addition to modal indicators, you can find the average trend using the median, which divides the ranked series in half. The median is used to detect thresholds on the scale: to the right and to the left of it there are signs that gravitate towards opposite poles. The strongest indicator that can be found is the rank correlation coefficient, which indicates the presence or absence of functional relationships in two rows of features.

The interval scale refers to the type of metric scales, which, unlike the first two types of nominal scales, not only establish the relationship between the points of the scale in terms of more–less, but also allow you to determine the size of the interval. An interval scale is a fully ordered scale with measured intervals between scale items. The reading in the scale can start from an arbitrary point. The reference point on the scale of intervals, as well as the value of the dimensional quantity, can be changed. The numbers on such scales remain unchanged after performing linear operations. Instead of a rank coefficient, you can use a more sensitive pair correlation coefficient and multiple correlation coefficients, which allow you to predict changes in one variable depending on changes in another or a number of other variables.

The scale that allows you to do all the operations with numbers is the ratio scale. It is a metric scale, similar to the interval scale, but with one advantage: the reading in the ratio scale does not start from an arbitrary point, but from an experimentally established zero. On this scale, you can determine how much each item exceeds or is below the other.

The modified method with a double stimulus proposed in this paper involves the use of an impairment scale to fix integral quality scores. The scales proposed in the Recommendations developed by the International Telecommunication Union (ITU), including the impairment scale, are ordinal scales. The numbers on such scales are ranks that allow you to sort the measured quality according to external properties and features that correspond to the goals and objectives of the experiment. In order to be able to use the results obtained in the course of subjective experiments for the mathematical description of cross-modal interaction in the process of forming the viewer's impression of the watched audiovisual programs of various content, it is necessary to use an impairment scale to measure the quality, which has the properties of an interval scale or, more preferably, a ratio scale.

Thurstone's method (Thurstone, Chave, 1929; Thurstone, 1925; 1931) was used to construct an impairment scale that has the properties of interval or ratio scales. The scale of seemingly equal intervals was first constructed by Thurstone (Thurstone, Chave, 1929). With the help of it, the author investigated and measured the attitude of society towards the church as a social institution.

The technique for constructing the Thurstone scale includes the following stages. At the first stage, a certain set of judgments is proposed, each of which reflects the attitude towards the measured object. At the next stage, in the process of psychophysical examination, a scale val-

ue is determined for each judgment and the final selection of judgments for the scale is made.

To assess the distortion of the image and sound that occurs during the production, transmission and playback of audiovisual information, in comparison with some “ideal”, in which there are no distortions of the image and sound, 11 judgments were proposed—categories ranging from “ Impairment are extremely annoying” to “Impairments are imperceptible “ (Grinenko, Glasman, Belozertsev, 2014):

- C1 Искажения незаметны (Impairments are imperceptible)
- C2 *Искажения практически незаметны (Impairments are practically imperceptible)*
- C3 Искажения немного заметны, но не мешают (Impairments are slightly perceptible, but do not interfere)
- C4 *Искажения заметны, но почти не мешают (Impairments are perceptible, but almost do not interfere)*
- C5 Искажения слегка мешают (Impairments are slightly interfere)
- C6 *Искажения заметно мешают (Impairments noticeably interfere)*
- C7 Искажения мешают и раздражают (Impairments interfere and are annoying)
- C8 *Искажения довольно сильно раздражают (Impairments are quite annoying)*
- C9 Искажения очень сильно раздражают (Impairments are very annoying)
- C10. *Искажения исключительно сильно раздражают (Impairments are exclusively annoying)*
- C11. Искажения предельно раздражают (Impairments are extremely annoying)

The experts were offered a special form on which a vertical line was drawn to assess the impairments of the image and sound—a graphical scale. There are 6 marks on the scale at equal intervals from each other, denoted by the letters А, Б, В, Г, Д, Е (Cyrillic letters). The scale has two reference points corresponding to the marks with the letters А and Е. The upper point with the letter А corresponds to the assessment of an audiovisual sequence in which there are no distortions and impairments of the image and sound. It’s called “No Impairments”. The bottom point with the letter Е corresponds to an audio-visual sequence whose level of impairments is so great that it is extremely annoying to the viewer/listener and in fact makes the perception of the image and sound absolutely impossible. It is called “Impairments are absolutely unacceptable”.

Point Е of the scale is the zero point of the scale. The interval between points Д (Е) and Е determines the scale unit (the point Д of the scale has the value of one scale unit). The distance between adjacent marks of the scale is equal to one. Point А of the scale thus has a value of five scale units. This is the highest rating, because there can be no higher quality than the quality of the image and sound, which are free from distortions and impairments. Thus, the marks of the scale correspond to the following numbers of the ratio scale: Е=0, Д=1, Г=2, В=3, Б=4, А=5. However, these numbers were not indicated on the form so that the experts would not have any associations between the numbers on the scale and the numbers of judgments. Experts should be guided only by the semantic meaning of the judgments used in Russian and their own idea of where on the scale each judgment should be located. During statistical processing of the results, marks А-Е of the scale are replaced by numbers 5–0 (Grinenko, Glasman, Belozertsev, 2014).

The purpose of the experiment is to determine those judgments that will be located by experts exactly (or almost exactly) on scale marks that are at equal intervals from each other. If such judgments are found, they will define a quality scale that has the properties of an interval scale. If among the equal-interval judgments a judgment is found that is located at the zero point E (F) of the scale, then the selected subset of judgments will determine the quality scale that has the properties of a ratio scale.

117 students of St. Petersburg University of Film and Television participated as experts in psychophysical experiments. The task of the experts is to place on the scale all the given judgments relative to each other and to the reference points (“No Impairments”/“Impairments are absolutely unacceptable”). For convenience, each category is assigned a number. The number of each judgment should be put in correspondence with one of the marks on the scale, which, in the opinion of the expert, best characterizes the position of the judgment on the scale.

The experts were not limited in time, they could make as many corrections as they wanted, returning to the previous ones and moving on to subsequent judgments. The instructions that the experts were to follow explained that the number of the judgment was only an identifier, or the name of the judgment, that the judges should not distribute all the judgments evenly on the scale. Different judgments can be located at the same mark on the scale. When making a decision, experts should be guided only by the semantic meaning of the judgments used in Russian.

The results of the experiment, summarizing the opinions of 117 experts on the position of judgments on the scale, are shown in Table 1.

**Table 1.** *Distribution of expert opinions on the position of judgments on the scale*

Judgments		C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10	C-11
Marks on scale												
A	5	109	55	7	0	0	0	0	0	0	0	0
Б	4	8	61	102	65	18	1	0	0	0	0	0
В	3	0	1	7	47	81	44	11	0	0	0	0
Г	2	0	0	1	5	17	61	73	39	6	0	0
Д	1	0	0	0	0	1	11	31	63	76	34	3
Е	0	0	0	0	0	0	0	2	15	35	83	114

The rows of the table show the number of experts who put judgments C1-C11 on the divisions of the scale A-E, which correspond to the numbers 5–0. Each column thus contains the distribution of expert opinions on the position of judgments C1-C11 on the scale. For example, as can be seen from the table, 65 experts out of 117 placed judgment C4 (the judgment «Искажения заметны, но почти не мешают» (Impairments are perceptible, but almost do not interfere) on mark Б=4, 47 experts – on mark В=3 and 5 experts – on mark Г=2.

To achieve the goal of the experiment and determine those judgments that experts place exactly (or almost exactly) on the scale marks, statistical processing is necessary, for example, interval estimation, aimed at constructing an interval in which the true position of the judgment-category on the scale is with some probability. The use for these purposes of such widely used statistical methods as, for example, the Student’s criterion, in this case is impossible, since there is no reason to believe that the original data are distributed according to the normal

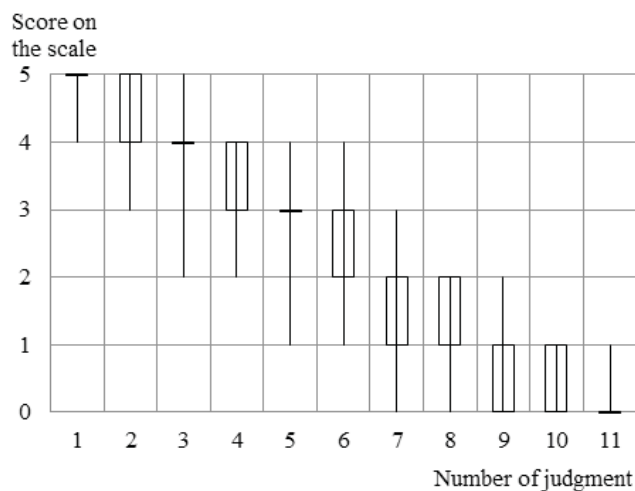
law. It is advisable to use order criteria (12, 15, 37, 39, 42, 67), which are nonparametric and do not depend on the distribution. The interval estimation of the distribution of expert estimates is performed by calculating the median and deviations from the median point for each of the 11 judgments. As can be seen from the data in Table 1, some distributions are highly asymmetric, which is an additional argument in favor of using the median as an average value (Вентцель, 1969).

When processing the results, that is, after the experts have distributed all the judgments on the scale, the marks corresponding to the letters from A to E are assigned numbers from 5 to 0. The median divides the data set into two halves, which means that 50 % of the experts give a score less than this value, and 50 % is greater (so the median can be called the 50th percentile). In addition to the median, there are the 25th and 75th percentiles. The 25th percentile is also called the first quartile of Q1, the median or 50th percentile is also the second quartile of Q2, the 75th percentile is the third Q3, and the 100th percentile, respectively, is the fourth quartile of Q4. To describe deviations, that is, to determine the spread for each judgment, quartiles and the interquartile range are used, which determines the central part of the distribution and is equal to the difference between the third and first quartiles. Within the interquartile range, i.e. around the median are concentrated estimates of fifty percent of experts. Table 2 shows the median values, the maximum and minimum values, the value of the 1st and 3rd quartiles, as well as the interquartile range for 11 judgment categories.

**Table 2.** Indicators of the scatter of expert opinions on the position of judgments on the scale

Judgment	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Median	5	4	4	4	3	2	2	1	1	0	0
1st quartile	5	4	4	3	3	2	1	1	0	0	0
3rd quartile	5	5	4	4	3	3	2	2	1	1	0
Maximum	5	5	5	4	4	4	3	2	2	1	1
Minimum	4	3	2	2	1	1	0	0	0	0	0
Range	0	1	0	1	0	1	1	1	1	1	0

The values obtained give a good description of the center, dispersion and shape of the distribution. The data obtained in the course of experimental studies are presented in Figure 1 in the form of a box diagram.



**Figure 1 -** Indicators of the scatter of expert opinions on the position of judgments on the scale

On the box diagram, the distribution of experimental data is represented as a rectangle with “whiskers”. The projections of the side edges of the rectangle correspond to the width of the interquartile range, i.e. the range within which 50 % of all data for each judgment is located, and the “antennae” correspond to the largest and smallest value. The position of the median corresponds to the horizontal line. Thus, with the help of a box diagram, the position of the median, the degree of scatter of values (interquartile range and amplitude of values) and the asymmetry indicated by the median line, which is not exactly in the center of the rectangle, are clearly shown.

As can be seen from Table 2 and Figure 1, the interquartile range of distributions of opinions for some judgments is equal to zero (C1, C3, C5, C11). This indicates a high agreement of opinions, in which the values of the first, second (median) and third quartiles are equal to the same value, coinciding with some division of the scale. Zero interquartile ranges are practically possible only for discrete random variables, which are the results of this experiment. However, the interquartile range is only half of the responses. Does the zero value of the interquartile range give reason to believe that the true value of the median in this case coincides with some mark on the scale? This issue is of particular importance for judgments C1 and C11, for which the distribution of opinions is completely asymmetric.

Consider the distribution of opinions for judgment C1 (“Impairments are imperceptible”). 109 experts placed judgment C1 on mark A of the scale ( $A=5$ ), 8 – on division Б ( $B=4$ ) (the total number of experts is 117). The number of experts who placed judgment C1 on mark A can generally be considered as a random variable equal to the sum of random variables  $X_i$  ( $i=1, 2, \dots, n$ ), where  $n=117$  is the number of experts, and  $X_i$  is a random variable, taking values 1 and 0 with probabilities  $p_i$  and  $(1-p_i)$  respectively. Here  $p_i$  is the probability that expert  $i$  places judgment C1 on scale division A,  $(1-p_i)$  is the probability that expert  $i$  places judgment C1 on some other and lower mark of the scale (Б-Е), which has a smaller value than mark A. If we assume that the probabilities  $p_i$  are the same for all experts, then the sum  $X_i$  is a random variable distributed according to the binomial law (Ван дер Варден, 1960; Вентцель, Овчаров, 2000; Гришин, 1975; Джонсон, Лион, 1980).

Let's check the null hypothesis  $H_0$ , according to which the probability  $p=1/2$  and the true value of the sum  $X_i$  (the number of experts who must place the category C1 on the division A of the scale) is equal to  $n/2$ . If the hypothesis  $H_0$  is true, the experts choose whether to place C1 on mark A or on any mark below, with equal probability. The deviation of the number of experts who placed category C1 on division A of the scale from  $n/2$  is random.

The alternative hypothesis  $H_1$  is that the number  $k$ , equal to the number of experts who placed judgment C1 on mark A of the scale, exceeds the value  $n/2$  is not random, the probability  $p>1/2$ . Placing judgment C1 on mark A of the scale is statistically significant.

Hypothesis testing can be performed using a one-sided sign test (Ван дер Варден, 1960; Джонсон, Лион, 1980; Митина, 2008). The hypothesis  $H_0$  is rejected if the number  $k$  exceeds the critical value  $m$  for a given significance level  $\alpha$  (Ван дер Варден, 1960). The critical level  $m$  is 69 for the significance level  $\alpha=0.05$  and 73 for the significance level  $\alpha=0.01$  at  $n=117$  (Ван дер Варден, 1960; Митина, 2008). The number  $k=109$  observed in the experiment exceeds the critical value at both significance levels. The  $H_0$  hypothesis is rejected in favor of the alternative  $H_1$  with a significance level not exceeding 0.01. The statement that judgment C1 (“Impairment are imperceptible”) is placed by experts on the mark of the  $A=5$  scale does not contradict the experimental data and is statistically significant.

Similarly, one can perform a statistical test of the distribution of expert opinions for judgment C11 (“Impairments are extremely annoying”). 114 experts placed judgment C11 on the E=0 division of the scale, 3 – on the Д=1 division (the total number of experts is 117). The number  $k=114$  observed in the experiment exceeds the critical value at the significance level  $\alpha=0.01$  ( $n=117$ ). The statement that judgment C11 (“Impairments are extremely annoying”) is placed by experts on the mark of the E=0 scale does not contradict the experimental data and is statistically significant with a significance level not exceeding 0.01.

The sign test can be used to test the hypothesis that the distribution of a random variable has zero median and is symmetric in this sense. If you need to test the hypothesis that the median is equal to a given value, then you can introduce a new random variable shifted by a given value, and then use the sign test (Ван дер Варден, 1960). For example, judgment C3 was placed by the experts quite closely relative to mark Б of the scale, which was assigned the value 4. 7 experts placed C3 at mark А (А = 5), 102 experts – at mark Б, 7 experts – at mark В, 1 expert – at mark Д. To test the hypothesis that the median is 4 (mark Б), we can shift the distribution to point 4 and discard zero differences. Of the remaining differences, 7 are positive and 8 are negative. For 15 experiments, the critical region is limited by numbers 4 and 11 for a significance level of 0.05 (two-tailed test with a confidence level of 0.95) and by numbers 3 and 12 for a significance level of 0.01 (Ван дер Варден, 1960). Neither the number of positive differences nor the number of negative ones goes beyond the boundaries of the critical region, so the distribution of expert opinions can be considered quite symmetrical and the hypothesis that the median for judgment C3 is 4 (mark Б of the scale) is accepted.

A similar conclusion can be drawn with regard to judgment C5 (81 experts placed C1 at mark Б, 18 experts – at mark В, 17 experts – at mark Г, 1 expert – at mark Д). This gives 18 positive differences and 18 negative ones. The boundaries of the critical region are 12 and 24 at a significance level of 0.05 and 10 and 26 at a level of 0.01. The hypothesis that the median for judgment C5 is 3 (mark Б on the scale) is accepted.

The distributions of expert opinions for other judgments are not symmetrical. The application of the sign criterion requires discarding zero differences. The results of this experiment are discrete values, and the number of zero differences is large, as can be seen from the data in Table 1. However, they are discarded when applying the sign criterion to test hypotheses. It is advisable to use another method of statistical estimation, which involves the construction of a confidence interval, in which the true value of the estimated parameter is found with a given probability – interval estimation (Ван дер Варден, 1960; Гмурман, 2003; Закс, 1976; Кобзарь, 2006).

Two-sided confidence limits for the median in the general case are order statistics with numbers  $(n-m)$  and  $(m+1)$ , where  $n$  is the sample size,  $m$  is a parameter corresponding to the confidence level and setting the boundary of the critical region for the sign criterion (Ван дер Варден, 1960). With a sample size of  $n=117$  (number of experts), the two-sided confidence limits for the median are the numbers of ordinal statistics 47 and 71 for a significance level of 0.05 (confidence probability 0.95) and the numbers 44 and 74 for a significance level of 0.01 (confidence probability 0.99). Medians and limits of the confidence interval for the median at a significance level of  $\alpha=0.01$  for judgments C1, C3, C5, C7, C9, C11 are shown in Table 3.

**Table 3.** Median and limits of the confidence interval for the median at a significance level of  $\alpha=0.01$

Judgment	C1	C3	C5	C7	C9	C11
Order statistics (n-m)	5	4	3	2	1	0
Median	5	4	3	2	1	0
Order statistics (m+1)	5	4	3	2	1	0

From the data in Table 3 it follows that the width of the confidence interval of opinions for judgments C1, C3, C5, C7, C9, C11 is equal to zero. This indicates a high consistency of expert opinions, in which the values of the ordinal statistics with the number (n-m), or the lower limit of the confidence interval, the median and the ordinal statistics with the number (m + 1), or the upper limit of the confidence interval, are equal to the same value, coinciding with some scale mark. Zero confidence intervals are practically possible only for discrete random variables, which are the results of this experiment.

Thus, the hypotheses that the judgments C1, C3, C5, C7, C9, C11 are placed by experts exactly on the scale marks A=5, B=4, B=3, Γ=2, Д=1, E=1, respectively, are accepted with a confidence level of 0.99. The intervals between judgments C1, C3, C5, C7, C9, C11 are the same, i.e. the scale has the properties of an interval scale. Previously, the statistical significance of the hypothesis was confirmed that judgment C11 (“Distortions are extremely annoying”), according to experts, coincides with division E (“Distortions are absolutely unacceptable”), which is the zero point of the scale. Therefore, the impairment scale built on the basis of judgments C1, C3, C5, C7, C9, C11 has the properties of a ratio scale. The equivalents of judgments C1, C3, C5, C7, C9, C11 are the numbers 5, 4, 3, 2, 1, 0, with which all mathematical operations can be performed.

Successful testing of hypotheses about the properties of the proposed impairment scale allows assessing the accuracy of the placement of judgments by experts using the usual statistical procedures (estimation of the mean, root mean square, confidence interval for the mean) recommended in Recommendation ITU-R BT.500–13 (ITU-R BT.500–13, 2012). This will also allow us to evaluate the position on the equal-interval scale of those judgments (C2, C4, C6, C8, C10), the distributions of which do not allow us to draw certain conclusions using the sign criterion and estimating the confidence interval of the median (Table 4).

**Table 4.** Statistical indicators of expert opinions on the position of judgments on the scale (significance level  $\alpha=0.05$ )

Judgement	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Mean value	4,93	4,46	3,98	3,51	2,99	2,30	1,79	1,21	0,75	0,29	0,03
Root-mean-square value	0,25	0,52	0,39	0,58	0,58	0,65	0,62	0,65	0,54	0,46	0,16
Lower boundary of the confidence interval for the mean		4,37	3,91	3,41	2,89	2,18	1,68	1,09	0,65	0,21	
Upper boundary of the confidence interval for the mean		4,56	4,05	3,62	3,10	2,42	1,91	1,32	0,85	0,37	

From the data in Table 4, it follows that the average values of the position of judgments C1 and C11 are very close to the scale marks 5 and 0, respectively. The standard deviation is a quarter or less of the interval between adjacent marks, which seems to be a very small value (178). Confidence intervals for the position of judgments C1 and C11 were not calculated due to the significant asymmetry of the distributions.

The data in the table confirm the hypotheses that judgments C3 and C5 are located at scale marks 4 and 3, respectively (confidence intervals of the mean values of the position of judgments C3 and C5 cover points 4 and 3, respectively, with a significance level of 0.05).

The mean values of the position of judgments C7 and C9 are below the scale marks 2 and 1, respectively, but the deviation does not exceed a quarter of one point, i.e. interval between adjacent marks.

Judgements C2 and C4 can be used to describe impairment levels intermediate between 5, 4 and 4, 3. The confidence interval for the mean position of judgment C2 covers the middle of the interval between 5 and 4 (point 4.5). The confidence interval for the mean position of judgment C4 covers the point 3.5.

Judgements C6, C8 and C10 can be used to describe the level of impairments intermediate between the marks 3, 2, 1 and 0. The mean values of the position of judgments C6, C8 and C10 are slightly below the midpoints of the corresponding intervals, but the standard deviations are relatively small.

Thus, experimental studies have shown that the proposed scale of impairments has the properties of a ratio scale. The accuracy of the correspondence of the position of judgments to the numbers of the scale is very high in the field of judgments describing relatively small impairments: C1 (Искажения незаметны–5,0), C2 (Искажения практически незаметны–4,5), C3 (Искажения немного заметны, но не мешают–4,0), C4 (Искажения заметны, но почти не мешают–3,5), C5 (Искажения слегка мешают–3,0). Accuracy is not as high in the area of judgments describing significant impairments: C6 (Искажения заметно мешают–2,5), C7 (Искажения мешают и раздражают–2,0), C8 (Искажения довольно сильно раздражают–1,5), C9 (Искажения очень сильно раздражают–1,0), C10 (Искажения исключительно сильно раздражают–0,5). However, deviations are equal to values not exceeding a quarter of a point.

A valuable property of the scale is the exact position of judgment C11 at the zero point of the scale. This is very important, since it is the presence of the experimentally established position of zero that turns the scale of intervals into a scale of ratios. The scale can be recommended for use in studies and assessments of the quality level of television broadcasting systems.

### **2.3. Conclusions to section 2**

1. A technique that ensures the consistency and reproducibility of the results is of paramount importance when conducting a psychophysical experiment. The review of standard and possible alternative methods, carried out in Section 1, showed that the existing methods do not meet the objectives of this work, therefore, a new method for conducting psychophysical experiments should be developed that can be used to study cross-modal interaction when assessing the quality of experience in digital television systems with data compression.

2. A number of standards were taken as prototypes, which contain recommendations for conducting subjective experiments in television. However, Recommendation ITU-R BT.500–13, Recommendation ITU-R BS.1284–1 regulate subjective experiments to evaluate image quality without sound quality or sound quality without image quality. The techniques described in ITU-T P. 911 are suitable for estimating quality of service depending on technical parameters, such as link bandwidth, delay, jitter and packet loss, image resolution, etc., without taking into account changes in the number of semantic and aesthetic information contained in the audiovisual program.
3. To build a scale of subjective assessments of the integral quality, a new informational approach was proposed in the work, which takes into account the presence in films and television programs of both semantic information and aesthetic information. The degradation of the quality of experience due to impairments is proposed to be assessed by a decrease in the amount of semantic information and a decrease in the degree of aesthetic enjoyment (aesthetic information).
4. An important aspect in conducting subjective experiments is the choice of a measurement standard, that is, the choice of a scale and the determination of its properties. There are 4 types of scales in accordance with the ability to meet the requirements of more diverse operations with numbers: nominal scale, ordinal scale, interval scale, ratio scale. In accordance with the proposed informational approach, the Double-stimulus impairment assessment (DSIS) methodology was modified to use a new version of the impairment scale, the prototype of which is described in Recommendation ITU-R BT.500. However, the scales proposed in the International Telecommunication Union (ITU) Recommendations, including the impairment scale, are ordinal scales. The numbers on such scales are ranks that only allow you to sort the measured quality according to external properties and features that correspond to the goals and objectives of the experiment.
5. Thurstone's method was used to construct an impairment scale that has the properties of interval or ratio scales. It has been shown that the impairment scale, built on the basis of the following set of judgments: C1-Искажения незаметны (Impairments are imperceptible), C3-Искажения немного заметны, но не мешают (Impairments are slightly perceptible, but do not interfere), C5-Искажения слегка мешают (Impairments are slightly interfere), C7-Искажения мешают и раздражают (Impairments interfere and are annoying), C9- Искажения очень сильно раздражают (Impairments are very annoying), C11-Искажения предельно раздражают (Impairments are extremely annoying), has the properties of a ratio scale. The equivalents of judgments C1, C3, C5, C7, C9, C11 are the numbers 5, 4, 3, 2, 1, 0, with which all mathematical operations can be performed. The scale can be recommended for use in studies of cross-modal interaction in assessing the quality of experience—integral assessment of the quality of the image and sound.

### **3. Development of subjective perception mathematical model that allows to predict an integral quality assessment of compressed multimedia materials**

#### **3.1. Statement of the problem**

The perception of audiovisual programs is characterized by two modalities: visual and auditory, each of which can complement and modify the perception experience created by the other one. (Hands, 2004; Jumisko-Pyykko, Hakkinen, Nyman, 2007; Ries, Puglia, Tebaldi, Nemethova, Rup, 2005; Storms, 2000; Winkler, Faller, 2005). It was noted in Section 1 a cross-modal interaction exists in producing the quality perception of audiovisual programs. It

is desirable to have an empirical model of multimodal quality perception to use the features mentioned above in practice.

The goal of work described in this section is to develop the model of integral quality assessment of compressed television materials taking into account the effect of cross-modal interaction of program components.

Modeling the assessment of image and sound quality in modern multimedia systems is a complex task. Level decomposition can be used as an effective way to solve it. In this paper, it is proposed to use level decomposition when developing a model for integral quality assessment. At the first level, the model describes the relationship between image quality and sound quality and bit rates of compressed video and audio signals. As it can be seen from the studies described in Sections 1 and 2 these functions will depend on the genre of audiovisual program. The quality measures predicted at the first level of the model can be called monomodal ones. The model of the first level should correspond to the laws of subjective perception that describe the influence of the video bit rate on the image quality and the audio bit rate on the sound quality. At the second level, the model describes the relationship of the integral quality as a function of the image quality and sound found at the first level. The integral quality predicted by the model at the second level can be called multimodal.

### **3.2. Development of the first-level subjective perceptual model of audio and video quality**

#### *3.2.1. Experimental procedure*

The model parameters should be found on the basis of experimental data. The first task of the current investigation was to conduct a study of the main factor affecting quality perception of video and audio—the bandwidth. There were conducted two series of experiments in the framework of the first task. The influence of the video and audio bit rate on the picture and audio quality of video sequences was studied separately. Three test audiovisual sequences were used in the experimental study:

- news program fragment: is an example of head and shoulder television contents with a small number of small details and low intensity of motion in the frame; audio accompaniment—the speaker's voice;
- music video clip: a video sequence with medium detail and motion dynamics in the frame accompanied by a story-related musical audio track;
- sports report (ski-rolling competitions) with high detail and dynamics of the frame, accompanied by the commentator's speech and the background noise of sports competitions.

These test sequences were selected according to their audiovisual characteristics. The levels of spatial details and temporal motions varied in the visual domain. Speech and music varied in audio domain.

The original materials for the test sequences were sourced from DV-tapes (video: 720×576@50i, 4:2:0, 25Mbps, stereo audio: 48kHz, 16 bits) and converted to AVI frames (Adobe Premier Pro 1.5). AVI frames were used as the input to produce the originals. The original clips were processed to produce 10–15 second test sequences. Video was encoded with H.264 software MediaCoder v.0.6.1 (Main Profile, Level 5.1) (MediaCoder). Audio was encoded with AAC encoder. In mobile television, bitrates are much lower than DV 25Mbps. That is why decoded

DV sequences were used as the reference ones. The reference sequences compressed using H.264 and AAC encoders were test sequences used in the experimental study.

The test sequences had different values of video and audio bitrates. But the reference and test sequences always had the same values of the other factors that varied in the study: spatial resolution, dimensions of a picture in height and width, black and white or colour image, contrast of the picture, loudness.

Sony monitor (PVM-14L2) was used for video playback. Loudspeakers (M-Audio Studio-ophile BX-5) were used for audio playback. All test sequences were played from the memory of the laboratory installation using an auto selection from the play list.

The experiment on subjective quality assessment was conducted in the research laboratory of the St. Petersburg State University of Film and Television. The order of sequences presentation, time intervals, and other conditions of the experiment were set according to Recommendation ITU-R BT.500-13 (ITU-R BT.500-13, 2012) and Recommendation ITU-R BS.1284-1 (Rec. ITU-R BS.1284-1, 2003). Room illumination was controlled; the image viewing distance was about 4 image heights (4 H). For 320×240 image resolution the height H on the monitor screen was 16 cm. The maximum luminance of the screen was set equal to a value of 80cd/m. The luminance of the surrounds was set equal to 10 % of the maximum luminance of the screen. 10 students from St.Petersburg University of Film and Television took part in these experiments as an experts. Each assessor took part in nine series of experiments. Each experiment was repeated three times.

A double-stimulus impairment scale (DSIS) method was used in the study.

### *3.2.2. Results of experimental study*

A study was conducted to examine the influence of the main factor affecting quality perception of video and audio—the bandwidth. The investigation was to study the influence of video and audio bit rates on the perceived quality of image and sound separately. In the first series, the influence of video bit rate on picture quality of video sequences was studied. Assessors were presented with a video sequence coded with different compression ratios (different bit rates) without sound. The experts were asked to evaluate the image quality of 3 programs: news program (head and shoulder television content static frame), sports competitions (dynamic frame) and the video sequence of the music video. Then, the mean opinion score for the sequence was calculated using the standard double-stimulus impairment scale (DSIS) method. Such a quality could be called a monomodal one. The relationship between the video bit rate (VBR) and the mean opinion score of picture quality ( $Q_v$ ) for news, music video and sport sequences is shown in Figure 2. Experimental data are approximated by exponential curves. The method for calculating the approximation parameters is given below.

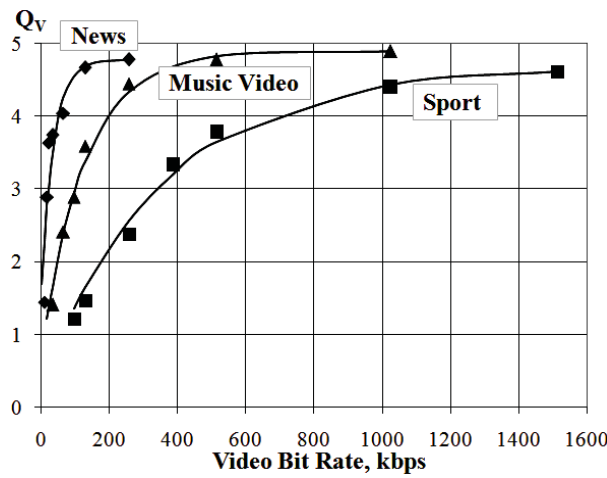
In the second series, the influence of the audio bit rate on the sound quality of video sequences was studied. The assessors were presented with the sound accompaniment of video sequences without pictures. The audio data of the sound accompaniment was compressed/decompressed with different compression ratios (different bit rates). Then, the mean opinion score for the sound accompaniment was calculated using the impairment scale method. The relationship between the audio bit rate (ABR) and the mean opinion score of sound quality ( $Q_a$ ) for news and music video sequences is shown in Figure 3. This is the quality perception of the second modality or the monomodal audio quality. Experimental data are approximated by exponential curves. The method for calculating the approximation parameters is given

below.

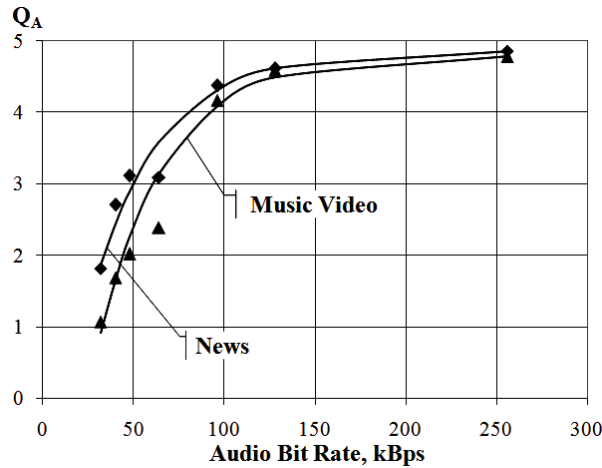
First of all, it should be noted that monomodal picture quality (Figure 2) is content dependent. In the high motion sport sequence, the video bit rate required is much more than in the news sequence at one and the same quality level. The difference between the audio bit rate values required for different sequence is less (Figure 3). Nevertheless, this difference is significant especially at low levels of quality.

### 3.2.3. Definition of parameters of the empirical models of video and audio monomodal quality perception

Analysis of experiments data allows us to determine the characteristics of the subjective perception of monomodal audio and video quality and find the parameters of the approximation of the first level models. Based on the type of relationships shown in Figures 2 and 3, it can be assumed that the experimental data are well approximated by exponential curves (3.1, 3.2).



**Figure 2.** Relationship between the video bit rate and the mean opinion score of picture quality ( $Q_v$ ) for different test sequences.



**Figure 3.** Relationship between the audio bit rate and the mean opinion score of sound quality ( $Q_A$ ) for news and music video.

$$Q^*_V = k_{V1} + k_{V2} \cdot \exp(k_{V3} \cdot VBR) \quad (3.1)$$

$$Q^*_A = k_{A1} + k_{A2} \cdot \exp(k_{A3} \cdot ABR) \quad (3.2)$$

where:

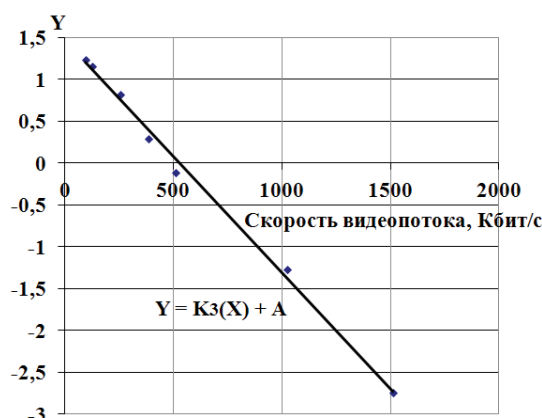
$VBR$  – video bitrate,  $kbps$ ;

$ABR$  – audio bitrate,  $kbps$ ;

$Q^*_V, Q^*_A$  – predicted monomodal video and audio quality levels;

$k_{V1}, k_{V2}, k_{A1}, k_{A2}, k_{V3}, k_{A3}$  – parameters of the empirical models of video and audio monomodal quality perception.

This assumption can be verified by displaying the same relationship on a logarithmic scale using the additional function  $Y(Q) = \ln(Q - k_1)$  and the parameter  $A = \ln(k_2)$ . Figure 4 shows the relationship between the video bit rate and the mean opinion score of picture quality ( $Q_V$ ) for sport sequence on a logarithmic scale. Similar graphs were obtained for other sequences. It can be seen from the figure the straight line is a good approximation, that confirms the possibility of using exponential curves (3.1) and (3.2) as first-level quality assessment models (Белозерцев, Гриненко, 2009; Гриненко, 2011; Перегудов, Белозерцев, Гласман, Гриненко, 2009; 2010).



**Figure 4.** Relationship between the video bit rate and the mean opinion score of picture quality ( $Q_V$ ) for sport sequence on a logarithmic scale

The parameters of the approximation in equations (3.1) and (3.2) represent empirical models of video and audio monomodal quality perception can be found using the method of least squares.

The parameters  $k_{V1}, k_{V2}, k_{V3}, k_{A1}, k_{A2}, k_{A3}$ , are given in Table 5.

**Table 5.** Parameters of the empirical models of video and audio monomodal quality perception

	News	Music	Sport
$k_{A1}$	4,8600	4,7900	4,6860
$k_{A2}$	- 6,9414	- 9,0359	- 12,6847
$k_{A3}$	- 0,0263	- 0,0265	- 0,0390
$k_{V1}$	4,7800	4,8900	4,6800
$k_{V2}$	- 3,1240	- 4,1645	- 4,3397
$k_{V3}$	- 0,0277	- 0,0079	- 0,0028

### 3.3. Development of the second-level subjective perceptual model of multimodal quality

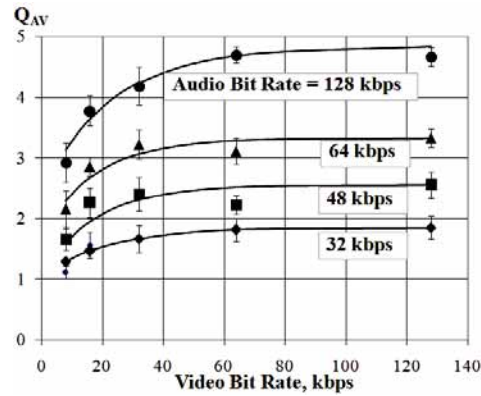
#### 3.3.1. Results of experimental study

Parameters of the subjective perception model describing the effect of audio and video bitrate on the integral (audiovisual) quality of audiovisual programs are based on the experimental study. Viewers were asked to assess the integral (audiovisual) quality of three audiovisual

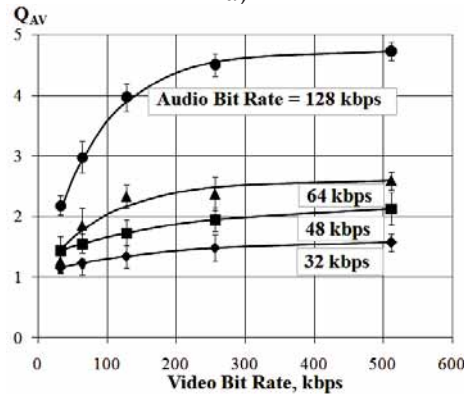
sequence: news program, sports report and music video. These experiments can be called multimodal, they investigate simultaneously two modalities of audiovisual sequence (video/audio). Assessment of integral (audiovisual) quality, expressed in expert scores, can be called multimodal.

A modified double-stimulus impairment scale (DSIS) method was used in the study of multimodal quality perception. A new approach was proposed to define the subjective scale of multimodal audiovisual perceptual quality and to explain the assessors how they should express their judgements. This method is described in details in Section 2.

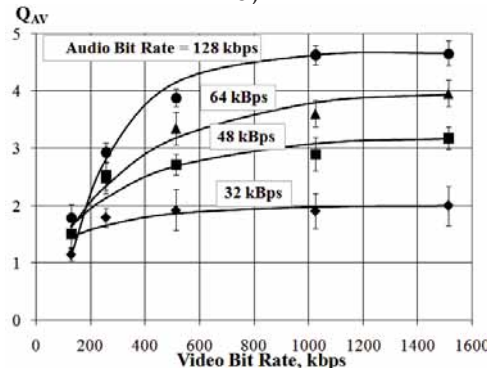
The relationships between the VBR and the mean opinion score of multimodal quality ( $Q_{AV}$ ) for news, music video, and sports sequences are shown in Figure 5. The ABR is the parameter of the curves. Confidence intervals corresponding to 0.95 are shown.



a)



b)



c)

**Figure 5.** Relationship between video bit rate, audio bit rate, and the MOS of multimodal quality ( $Q_{AV}$ ) for different test sequences (a – news; b – music video; c – sport).

A large amount of experimental data was obtained, which allows to conclude that the qual-

ity assessments of the components of an audiovisual program are mutually compensated, when one of the components (video or audio) complements and modify the subjective perception created by the other component. Both modalities (video and audio) affect the multimodal quality judgment of assessors (Figure 5 a, b, c). The importance of visual and audio partial parameters varies according to the content of television materials. For example, video quality has relatively more weight than audio in high motion sports content. The role of audio quality is most important in head-and-shoulder news content. Audio and video content interact in producing the perception of quality. It is desirable to have an empirical model of multimodal quality perception to use the features mentioned above in practice ((Белозерцев, Гриненко, 2009; Гриненко, 2011; Перегудов, Белозерцев, Гласман, Гриненко, 2009; 2010).

### *3.3.2. Definition of parameters of audiovisual multimodal quality perception empirical models*

At the first step it was shown that the exponential model describes well the relationship between video and sound quality indicators (monomodal video and sound quality) as a function of various signal processing parameters of audiovisual programs. At the second level, the model of subjective quality perception should describe the relationship between the integral multimodal quality score  $Q^*_{AV}$  as a function of video and audio quality scores  $Q^*_A$  and  $Q^*_V$ , obtained during the video and audio estimation experiments made at the first level.

Appropriate type of the empirical model should be selected. One of the main requirements for a model is the ability to predict the value of dependent variable with the required accuracy. To approximate data by a curve (in the case of one independent variable) or a surface (in the case of two independent variables), the existence of a functional relationship of a certain type is first assumed (Джонсон, Лион, 1980). Then, parameters of the curves are found to give the best approximation according to some criterion. It cannot be said that the best description of the functional relationship is obtained in this way, but the best parameter estimates are obtained according to the given function choice and criterion. It is possible to find out how good a given relationship is, but it is possible to get a better one by choosing a different function and a different criterion. In this case, it is possible to sort through combinations and data transformations to infinity, which is not appropriate. It is also wrong to assume that the model found will be the best one just because it gives a good approximation, but is totally unrelated to the actual physical or technical links. A valid functional form should be considered first, regardless of whether it was obtained by analytic deduction or by some other prior knowledge of the properties of the variables.

The review of studies carried out in Section 1 showed that still there is no satisfactory understanding of the multimodal quality assessment model properties. A summary of the current studies is made in Appendix C of the Recommendation ITU-R. P. 911 (ITU-R. P. 911, 1998). One of the most important results of this summary is the description of the first effort to establish a relationship between video and audio quality indicators determined in separate modality evaluation experiments and integral audiovisual quality indicators. It has been noted that an integral audiovisual quality MOS(AV) can be predicted by using assessments of video quality MOS(V) and sound quality MOS(A), obtained by separate subjective quality assessments of the two modalities. It was shown that the integral quality score of an audiovisual programme can be found using a multiplicative model:  $MOS(AV) = a + b * MOS(A) * MOS(V)$ , where the value of parameter a is in the range from 1.1 to 1.5 points, the value of parameter b is in the range 0.107 to 0.121. A 9-point category scale was used for experts scores. The recommendation notes that the results presented are preliminary and require further development.

It can be noted that a purely multiplicative model is not suitable for describing all the properties of integral perception, for example, it cannot display the different weights of modalities when evaluating television programs of different genre. However, it must be admitted that the multiplicative model, which includes multiplication of the video and sound quality indicators, is indeed a simple and effective way to describe cross-modal interaction.

It is reasonable to start with a graphical representation of the experimental data because the form of the relationship is unknown, since the issues related to audiovisual quality perception and its evaluation have not been sufficiently studied. Graphical representation helps to guess the form of the functional relationship. As a first approximation, it is advisable to use a relatively simple form of the equation, such as polynomial or exponential. The polynomial function is linear with respect to unknown coefficients, which simplifies the processing of results (Адлер, Маркова, Грановский, 1976).

The existing mutual relationship and compensation of two components, when one of the components (video or audio) complements and modifies the subjective perception created by the other component, can be described mathematically in different ways. The polynomial representation contains a multiplicative term, which allows to take into account the cross-modal interaction.

A trade-off between complexity of a model and correlation with the experimental results can be found if the polynomial approximation of multimodal audiovisual quality is used (3.3):

$$Q_{AV}^* = c_0 + c_{A1} \cdot Q_A^* + c_{V1} \cdot Q_V^* + c_{AV} \cdot Q_A^* \cdot Q_V^* + c_{A2} \cdot Q_A^{*2} + c_{V2} \cdot Q_V^{*2} + c_{AV2} \cdot Q_A^* \cdot Q_V^{*2} + c_{A2V} \cdot Q_A^{*2} \cdot Q_V^* + c_{A3} \cdot Q_A^{*3} + c_{V3} \cdot Q_V^{*3} \quad (3.3)$$

Of course, the number of terms in (3.3) should be minimal to have a simple model. The point is how many terms of the polynomial approximation might be used that the model would have a good correlation with the experimental data (Джонсон, Лион, 1980). The more non-zero coefficients are in the model, so it is more complicated for practical use.

To find the minimal number of terms it is reasonable to consider the relationship between monomodal quality levels ( $Q_A^*$  and  $Q_V^*$ ) and the multimodal quality level ( $Q_{AV}^*$ ) for different numbers of terms and compare these relationships to the experimental data. Let us consider the relationship between monomodal quality ( $Q_A^*$  and  $Q_V^*$ ) and multimodal quality ( $Q_{AV}^*$ ) levels, that can be obtained using the model with a different number of nonzero coefficients (3.3).

The simplest model of multimodal quality perception occurs when only  $c_0$ ,  $c_{A1}$  и  $c_{V1}$  are non-zero coefficients.:

$$Q_{AV}^* = c_0 + c_{A1} \cdot Q_A^* + c_{V1} \cdot Q_V^* \quad (3.4)$$

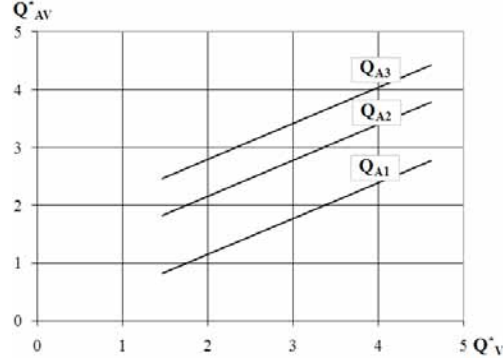
It is a linear model when the monomodal quality levels  $Q_A^*$  and  $Q_V^*$  are summed with weighting factors  $c_{A1}$  and  $c_{V1}$  to get the multimodal quality  $Q_{AV}^*$ . The relationship between  $Q_V^*$  and  $Q_{AV}^*$  is shown in Figure 6 (a). The parameter of the lines in Figure 6 (a) is the level of  $Q_A^*$ . All the lines are straight and parallel. The slope of the lines is determined by  $c_{V1}$  and the point of crossing with the ordinate axis is determined by the sum  $c_0 + c_{A1} \cdot Q_A^*$ .

The model when factor  $c_{AV}$  is also non-zero is shown in Figure 6 (b). In this case the model can be called an incomplete quadratic model (1) or a parametric mode.

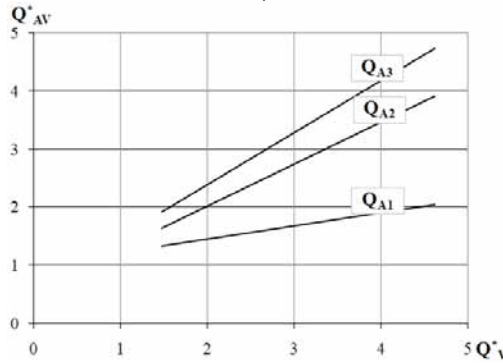
$$Q_{AV}^* = c_0 + c_{A1} \cdot Q_A^* + c_{V1} \cdot Q_V^* + c_{AV} \cdot Q_A^* \cdot Q_V^* \quad (3.5)$$

All the lines are straight but non-parallel. The slope of a line with parameter  $Q_{A1}^*$  is determined by the factor equal to  $c_{V1} + c_{AV} \cdot Q_{A1}^*$ . The slope of lines with parameters  $Q_{A2}^*$  and  $Q_{A3}^*$  is determined by the factor equal to  $c_{V1} + c_{AV} \cdot Q_{A2}^*$  and  $c_{V1} + c_{AV} \cdot Q_{A3}^*$  respectively. The better the quality  $Q_A^*$ , the more the slope of the line. The point of crossing with the ordinate axis is determined by the sum  $c_0 + c_{A1} \cdot Q_A^*$ . The non-zero multiplicative factor  $c_{AV}$  indicates the effect of the interactions between the two modalities. If factors  $c_{A2}, c_{V2}$  are non-zero the lines will not be straight, the model will be a polynomial of the second degree (3.6) (Figure6 (c)).

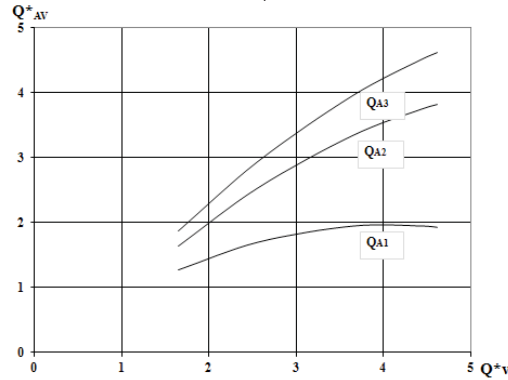
$$Q_{AV}^* = c_0 + c_{A1} \cdot Q_A^* + c_{V1} \cdot Q_V^* + c_{AV} \cdot Q_A^* \cdot Q_V^* + c_{A2} \cdot Q_A^{*2} + c_{V2} \cdot Q_V^{*2} \quad (3.6)$$



a)



b)



c)

**Figure 6.** Relationship among  $Q_V^*$ ,  $Q_A^*$ , and  $Q_{AV}^*$  for different numbers of terms of the polynomial approximation (3.6) : a – non-zero coefficients are  $c_0, c_{A1}$  u  $c_{V1}$ ; b – non-zero coefficients  $c_0, c_{A1}, c_{V1}$  u  $c_{AV}$ ; c – non-zero coefficients  $c_0, c_{A1}, c_{V1}, c_{AV}, c_{A2}$  u  $c_{V2}$

To find the proper number of terms it is needed to compare the relationships between the monomodal quality ( $Q_A^*$  and  $Q_V^*$ ) and multimodal quality ( $Q_{AV}^*$ ) levels calculated with different number of non-zero coefficients in the polynomial approximation (3.3) with experimental data.

The relationship between the monomodal quality levels ( $Q_A$  and  $Q_V$ ) and the multimodal quality level ( $Q_{AV}$ ) that was calculated on the basis of experimental data for the music video sequence is

illustrated in Figure 7.

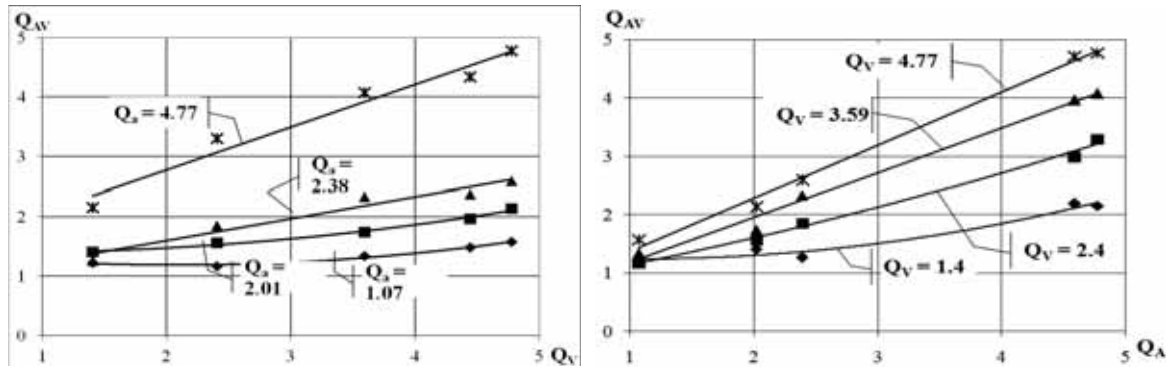


Рисунок 6 – Соотношения между уровнями мономодального качества ( $Q_A$  и  $Q_V$ ) и значением мультимодального качества ( $Q_{AV}$ ) для музыкального клипа.

**Figure 7.** Relationship between the monomodal quality levels ( $Q_A$  and  $Q_V$ ) and the multimodal quality level ( $Q_{AV}$ ) for the music video sequence.

Comparing the relationships shown in Figure 6 and 7, it can be concluded that the form of the curves is quite similar, therefore, it is reasonable to calculate the parameters of the polynomial model using regression analysis methods.

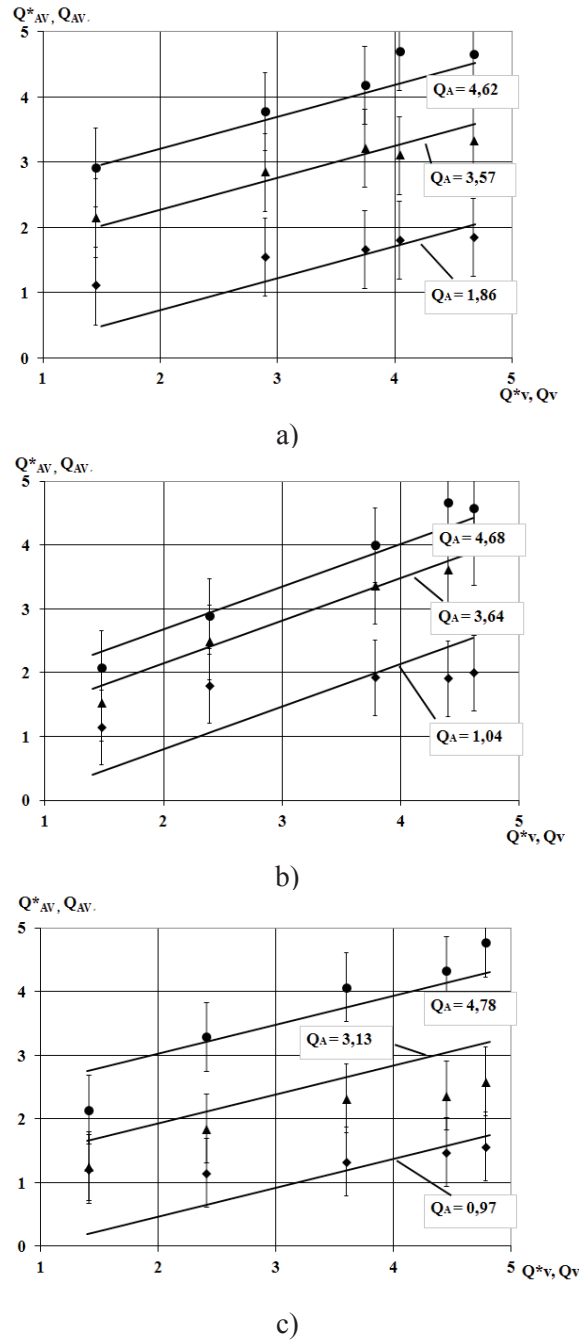
The regression analysis was carried out using the method described in (Адлер, Маркова, Грановский, 1976; Джонсон, Лион, 1980).

The parameter set  $\{c_0, c_{A1}, c_{V1}\}$  of the approximation of the experimental data  $Q_{AV} = f(Q_A, Q_V)$  are shown in Figures 5 by the polynomial  $Q_{AV}^* = c_0 + c_{A1} \cdot Q_A^* + c_{V1} \cdot Q_V^*$  represent the simplest model of multimodal quality perception. They are given in Table 6.

**Table 6.** Parameters of the empirical model of multimodal quality perception

	News	Music	Sport
$c_0$	-1,9222	-1,0604	-1,0429
$c_{A1}$	0,9006	0,6662	0,5128
$c_{V1}$	0,4882	0,4562	0,6627

Figure 8 shows the relationships between monomodal video quality  $Q^*V$  and sound quality  $Q^*A$  and multimodal quality  $Q^*AV$  for nonzero coefficients  $c_0, c_{A1}, c_{V1}$  in polynomial approximation and experimental data for three sequences: news program, music video and sports report. The root mean square error is shown in the figure by the vertical lines. The corresponding sound quality values were chosen as the parameter of the curves. A linear model when the monomodal quality levels  $Q^*A$  and  $Q^*V$  are summed with weighting factors  $c_{A1}$  and  $c_{V1}$  to get the multimodal quality  $Q^*AV$  in general describes the relationships obtained in the experimental studies for all three sequences. But it can be concluded from the Figure 8 that the model does not fully correspond to experiment results and does not take into account the interaction of the two modalities (the lines are parallel for each concluded  $Q_A$ ).



**Figure 8.** Relationship among  $Q^*_V$ ,  $Q^*_A$ , and  $Q^*_{AV}$  for non-zero coefficients  $c_0$ ,  $c_{A1}$  u  $c_{V1}$ : a – news program; b – music video; c – sport

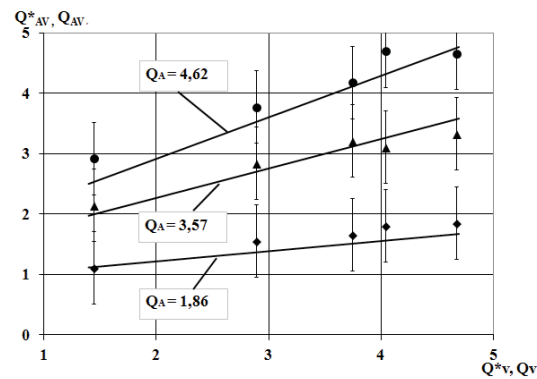
The cross-modal interaction should be taken into account while choosing an appropriate empirical model. This means that the multiplicative coefficient  $c_{AV}$  corresponding to the interaction effect, should be non-zero. In this case, the model will have the form of an incomplete square model and will describe the interaction of two modalities most simply. The parameter set  $\{c_0, c_{A1}, c_{V1}, c_{AV}\}$  of the approximation of the experimental data  $Q_{AV} = f(Q_A, Q_V)$  by the polynomial  $Q^*_{AV} = c_0 + c_{A1} \cdot Q^*_A + c_{V1} \cdot Q^*_V + c_{AV} \cdot Q^*_A \cdot Q^*_V$  represent the multimodal quality perception. They are given in Table 7.

**Table 7.** Parameters of the empirical model of multimodal quality perception

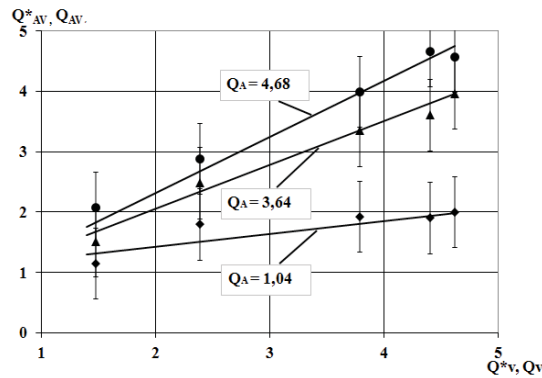
	News	Music	Sport
$c_0$	0,4096	0,8575	1,1739

	News	Music	Sport
$c_{A1}$	0,2459	0,0505	-0,1507
$c_{V1}$	-0,1786	-0,1227	0,0070
$c_{AV}$	0,1873	0,1853	0,1963

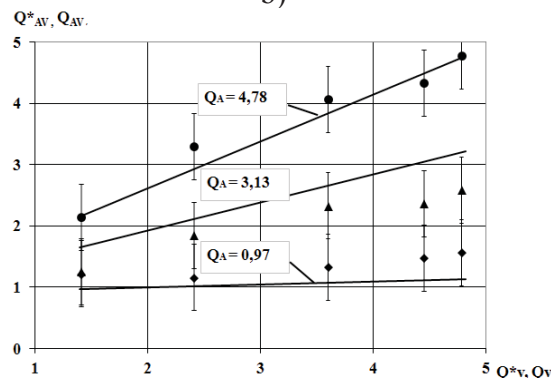
It can be seen from the table that the values of the factors  $c_{A1}$ ,  $c_{V1}$ , reflecting linear effects, vary greatly depending on the genre of audiovisual program, while the multiplicative term, reflecting the effects of interaction, practically remain at the same level for the three sequences, indicating their influence. Figure 9 shows the relationships between monomodal video quality  $Q_V^*$  and sound quality  $Q_A^*$  and multimodal quality  $Q_{AV}^*$  for nonzero coefficients  $c_0$ ,  $c_{A1}$ ,  $c_{V1}$  and  $c_{AV}$  in polynomial approximation and experimental data for three sequences: news program, music video and sports report. The root mean square error is shown in the figure by the vertical lines. The corresponding sound quality values were chosen as the parameter of the curves.



a)



b)



c)

**Figure 9.** Relationship among  $Q_V^*$ ,  $Q_A^*$ , and  $Q_{AV}^*$  for non-zero coefficients  $c_0$ ,  $c_{A1}$ ,  $c_{V1}$  u  $c_{AV}$ : a – news program; b – music video; c – sport

All the lines are straight but non-parallel. The slope of a line with parameter  $Q_{Ai}^*$  is determined by the factor equal to  $c_{V1} + c_{AV} \cdot Q_{Ai}^*$ . The better the quality  $Q_A^*$ , the more the slope of the line.

The incomplete quadratic model better describes the experimental data, reflecting the properties of the mutual compensation of the quality of modalities. However, it seems from the relationships between the levels of monomodal quality ( $Q_A$  и  $Q_V$ ) and the value of multimodal quality  $Q_{AV}$ , obtained from the experimental data presented in Figure 7, that some lines are straight and some are not, which means that factors  $c_{A2}$  and  $c_{V2}$  must be taken into account.

If factors  $c_{A2}$ ,  $c_{V2}$  are non-zero the lines will not be straight, the model will be a polynomic of the second degree. The parameter set  $\{c_0, c_{A1}, c_{V1}, c_{AV}, c_{A2}, c_{V2}\}$  of the approximation of the experimental data  $Q_{AV} = f(Q_A, Q_V)$  are given in Table 8.

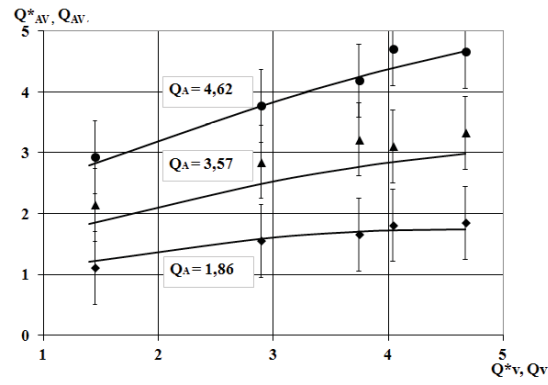
**Table 8.** Parameters of the empirical model of multimodal quality perception

	News	Music	Sport
$c_0$	-0,18906	1,0733	0,2786
$c_{A1}$	-0,3987	-0,8662	-0,2864
$c_{V1}$	1,8359	0,4411	0,7666
$c_{A2}$	0,0944	0,1571	0,0233
$c_{V2}$	-0,2887	-0,0868	-0,1198
$c_{AV}$	0,1873	0,1858	0,1963

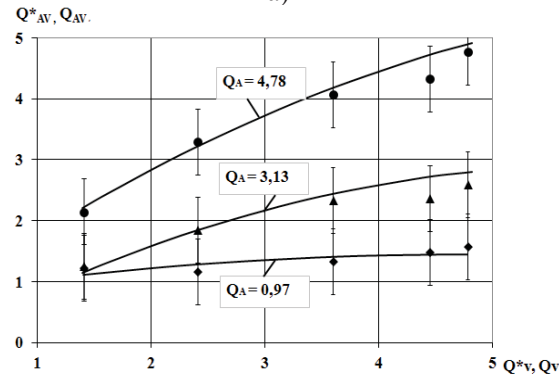
It can be noted that the multiplicative coefficients, reflecting the inter-modality interaction, have practically the same value for the three sequences under test. Their values are close to similar ones in the incomplete square model.

Figure 10 shows the relationships between monomodal video quality  $Q_V^*$  and sound quality  $Q_A^*$  and multimodal quality  $Q_{AV}^*$  for nonzero coefficients  $c_0, c_{A1}, c_{V1}, c_{AV}, c_{A2}, c_{V2}$  in polynomial approximation and experimental data for three sequences: news program, music video and sports report. The root mean square error is shown in the figure by the vertical lines. The corresponding sound quality values were chosen as the parameter of the curves. The lines are not straight. In this case, the model has the properties of mutual compensation of modalities described earlier, when one modality can complement and modify the perceptual experience created by the other.

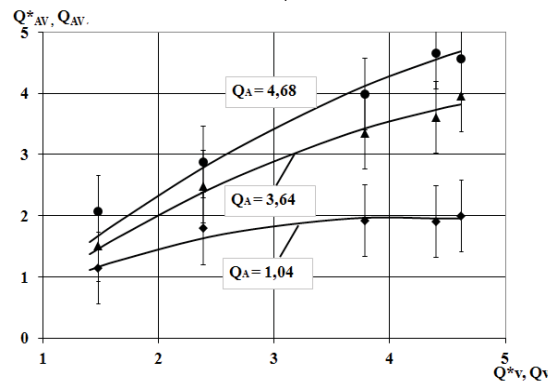
A polynomic of the third degree can be used to approximate the experimental data. In this case, the coefficients  $c_0, c_{A1}, c_{V1}, c_{AV}, c_{A2}, c_{V2}, c_{A2V}, c_{V2A}, c_{A3}, c_{V3}$  are non-zero, which are found in a similar way using the least squares method. The number of non-zero coefficients is almost doubled compared to the second-order polynomial model. The parameter set  $\{c_0, c_{A1}, c_{V1}, c_{AV}, c_{A2}, c_{V2}, c_{A2V}, c_{V2A}, c_{A3}, c_{V3}\}$  of the approximation of the experimental data  $Q_{AV} = f(Q_A, Q_V)$  by the polynomial of the third degree  $Q_{AV}^* = f^*(Q_A^*, Q_V^*)$  represent the multimodal quality perception. They are given in Table 9.



a)



b)



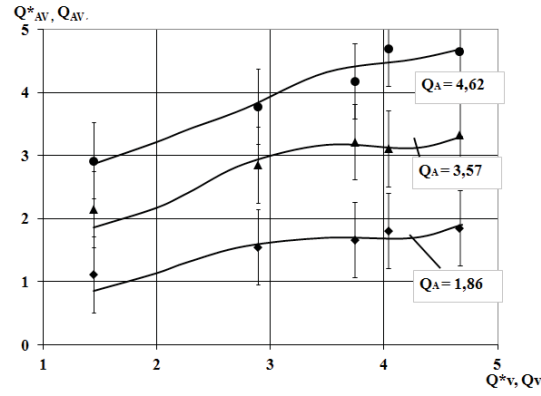
c)

**Figure 10.** Relationship among  $Q^*_v$ ,  $Q^*_{A'}$  and  $Q^*_{AV}$  for non-zero coefficients  $c_0$ ,  $c_{A1}$ ,  $c_{V1}$ ,  $c_{AV}$ ,  $c_{A2}$ ,  $c_{V2}$ :  
a – news program; b – music video; c – sport

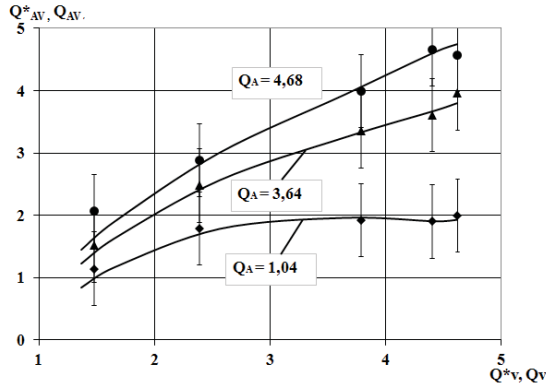
**Table 9.** Parameters of the empirical model of multimodal quality perception

	News	Music	Sport
$c_0$	-6,5779	1,1763	-2,4802
$c_{A1}$	-3,3591	-0,9556	0,7665
$c_{V1}$	8,8908	0,5734	2,9037
$c_{AV}$	0,2358	0,4339	-0,0215
$c_{A2}$	1,0676	0,0989	-0,2569
$c_{V2}$	-2,4707	-0,3008	-0,7334
$c_{V2A}$	-0,0897	-0,0671	0,0239
$c_{A2V}$	0,0845	0,0339	0,0114
$c_{A3}$	-0,1273	-0,0064	0,0276
$c_{V3}$	0,23988	0,0437	0,0570

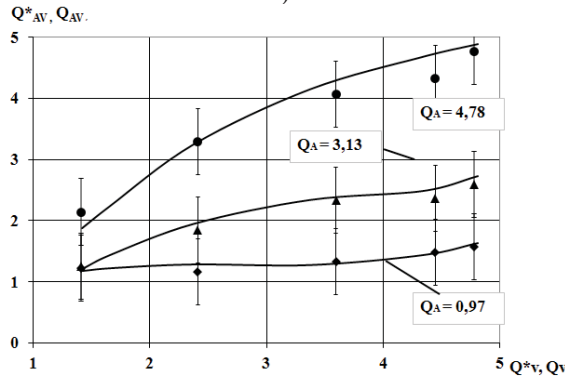
Figure 11 shows the relationships between monomodal video quality  $Q_v^*$  and sound quality  $Q_A^*$  and multimodal quality  $Q_{AV}^*$  for nonzero coefficients  $c_0, c_{A1}, c_{V1}, c_{AV}, c_{A2}, c_{V2}, c_{A2V}, c_{V2A}, c_{A3}, c_{V3}$  in polynomial approximation and experimental data for three sequences: news program, music video and sports report. The root mean square error is shown in the figure by the vertical lines. The corresponding sound quality values were chosen as the parameter of the curves. All the lines are not straight. The model has the properties of mutual compensation of modalities described earlier and can be used for approximation of experimental data.



a)



b)



c)

**Figure 11.** Relationship among  $Q_v^*$ ,  $Q_A^*$ , and  $Q_{AV}^*$  for non-zero coefficients  $c_0, c_{A1}, c_{V1}, c_{AV}, c_{A2}, c_{V2}, c_{A2V}, c_{V2A}, c_{A3}, c_{V3}$ : a – news program; b – music video; c – sport

The more non-zero coefficients in the polynomic, the closer the lines will be to the experimental points, and the accuracy of the description will be higher. However, monotonicity can be broken, which is undesirable for models describing user perception.

### 3.4. Assessment of model sufficiency using VQEG group criteria

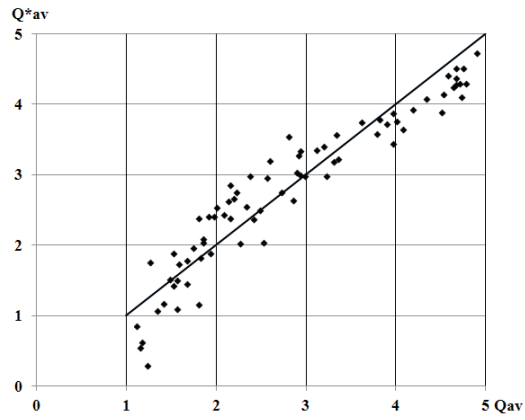
To analyze the possibility of using the obtained model to predict the assessment of the integral audiovisual quality of different multimedia content, encoded with different bit rates, it is necessary to check the relevance of the model using special criteria that take into account the requirements for subjective perception models. The VQEG (Video Quality Experts Group) group formulated special criteria: accuracy, monotonicity, and consistency and corresponding metrics characterizing the ability of the model to predict the perception of audiovisual materials (ITU-T. Technical paper. Tutorial on Objective perceptual assessment of Video Quality: Full Reference Television, 2004).

Accuracy is a parameter that characterizes the possibility of the selected model to predict the multimodal quality assessments obtained experimentally, according to the minimum average error value criterion. The model makes it possible to obtain the smallest difference between predicted and experimental values “in general” has the highest prediction accuracy. Accuracy is characterized by the mean square error and Pearson’s linear correlation coefficient, which makes it possible to measure the degree of linearity of the relationship between predicted and experimental data.

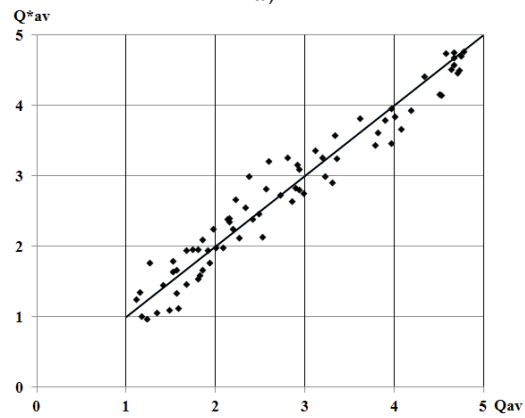
Ideally, the function describing the relationship of the predicted and experimental values should be monotonic. The increase in the predicted values should correspond to the increase in the experimental values of the audiovisual quality assessment. Spearman’s rank correlation coefficient of predicted and experimental values is a measure that describes well the monotonicity of the prediction. Spearman’s rank correlation coefficient refers to non-parametric indicators of the relationship between variables, which in this case represent the ranks of the compared values. The nature of random variables distribution does not matter when calculating Spearman’s rank correlation coefficient.

Consistency is a parameter that characterizes the ability of a model to provide consistently accurate predictions for all points. For example, one model provides a small value of the prediction error at almost all points but has several points with a large value of errors, and outliers. Another model provides a balanced set of error values, maybe numerically larger than the first one, but at all predicted points (no outliers). Thus, the model consistency can be characterized by the number of outliers relative to the total number of prediction points. An outlier is a point where an error value equal to the difference between the predicted and experimental values is two times greater than the root mean square error at this point. The value of the doubled root mean square error is used as a threshold value for determining outlier points. Accordingly, a smaller value of the outlier factor means that the predicted values are more consistent.

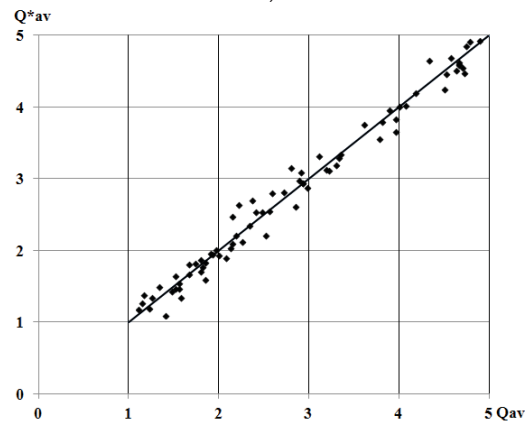
The relationship between the predicted  $Q_{AV}^*$  level of multimodal quality and experimental  $Q_{AV}$  data is shown in Figure 12 for different number of term of the polynomial approximation. The data in Figure 12 can be used to specify the accuracy, monotonicity, and consistency of the model as a tool for predicting the MOS value for sequences of different content coded with different bit rate values. The root-mean-square error specifying the accuracy of prediction, The Spearman rank-order correlation coefficient relating to the prediction monotonicity, the outlier ratio specifying the consistency are given in Table 10.



a)



b)



c)

**Figure 12.** Relationship between the predicted  $Q^*_{AV}$  level of multimodal quality and experimental  $Q_{AV}$  data: a – linear model; b – incomplete square model; c – polynomial of the second degree

**Table 10.** Calculated metrics characterizing the ability of the different models to predict the perception of audiovisual quality

	Linear model	Incomplete square model	Polynomic of the second degree	Polynomic of the third degree
The root-mean-square error	0,3550	0,2527	0,1568	0,1202
The Pearson correlation coefficient	0,9466	0,9758	0,9907	0,9955
The Spearman rank-order correlation coefficient	0,9568	0,9753	0,9883	0,9630

The first criterion introduced by the VQEG group is accuracy, which is characterized by the value of the mean square error and the Pearson linear correlation coefficient. As the number of non-zero coefficients in the polynomial increases, the model line will be more closely to experimental points, and the accuracy of the model will increase. Therefore, the minimization of the mean square error can be the first requirement for the model of subjective perception. It can be seen from the table that the value of the root-mean-square error noticeably decreases when the multiplicative term reflecting the effect of the modality's interaction (incomplete quadratic model) is non-zero. When factors  $c_{A2}$  and  $c_{V2}$  are non-zero (quadratic model) the value of the root-mean-square error noticeably decreases too. Further increase in the degree of the polynomial does not lead to a significant decrease of the mean square error.

The second criterion is monotonicity. As the number of non-zero coefficients in the polynomial increases, the model line will be more closely to experimental points, but at the same time, it will become a kinked curve. It breaks monotonicity, which is contrary to the principle of perception. Perception cannot be oscillatory. The monotonicity violation is noticeable in Figure 11 when the experimental data are approximated by a polynomial of the third degree. Also, it is confirmed by the decrease of the Spearman rank correlation coefficient value for the polynomial model of the third degree. The outlier ratio for all four cases of the polynomial model is 0.

The first requirement, an increase in accuracy, is met with an increase in the number of non-zero coefficients in the polynomial. The second requirement is monotonicity, which is violated as the degree of the polynomial increases, as the lines become oscillatory. An acceptable compromise between accuracy, monotonicity, and model complexity finds by using a polynomial model of the second degree. (Белозерцев, Гриненко, 2009; Перегудов, Белозерцев, Гласман, Гриненко, 2009; Peregudov, Glasman, Belozertsev, Grinenko, 2010).

### 3.5. Conclusions to section 3

1. A level decomposition is used in the work to develop a mathematical model of multimodal quality perception. At the first step empirical models of video and audio monomodal quality perception are presented describing the influence of video and audio bit rates on the perceived quality of image and sound. At the second step the model of subjective quality perception should describe the relationship between the integral multimodal quality score  $Q_{AV}^*$  as a function of video and audio quality scores  $Q_A^*$  and  $Q_V^*$  obtained during the video and audio estimation experiments made at the first level. The integral quality predicted by the model on the second level can be called multimodal.
2. On the first step all the experimental data can be approximated by exponential curves:

$$Q_V^* = k_{V1} + k_{V2} \cdot \exp(k_{V3} \cdot VBR)$$

$$Q_A^* = k_{A1} + k_{A2} \cdot \exp(k_{A3} \cdot ABR)$$

These models describe the influence of the video and audio bit rate on the picture and sound quality of video sequences respectively.

3. On the second step a trade-off between complexity of a model and correlation with the experimental results polynomial approximation of multimodal audiovisual quality is used:

$$Q_{AV}^* = c_0 + c_{A1} \cdot Q_A^* + c_{V1} \cdot Q_V^* + c_{AV} \cdot Q_A^* \cdot Q_V^* + c_{A2} \cdot Q_A^{*2} + c_{V2} \cdot Q_V^{*2} + c_{AV2} \cdot Q_A^* \cdot Q_V^{*2} + c_{A2V} \cdot Q_A^{*2} \cdot Q_V^* + c_{A3} \cdot Q_A^{*3} + c_{V3} \cdot Q_V^{*3}$$

The model describes the relationship between the multimodal quality level  $Q^*_{AV}$  and the monomodal quality levels ( $Q^*_A$  and  $Q^*_V$ ) that were calculated on the first step.

The non-zero multiplicative term allows to predict the properties of mutual compensation of modalities when one modality can complement and modify the perceptual experience created by the other ones.

4. To find the minimal number of terms in polynomial it was reasonable to consider the relationship between monomodal quality levels ( $Q^*_A$  and  $Q^*_V$ ) and the multimodal quality level ( $Q^*_{AV}$ ) for different numbers of terms and compare these relationships to the experimental data. The simplest model of multimodal quality perception is when only  $c_0$ ,  $c_{A1}$  and  $c_{V1}$  are non-zero coefficients. It is a linear model when the monomodal quality levels  $Q^*_A$  and  $Q^*_V$  are summed with weighting factors  $c_{A1}$  and  $c_{V1}$  to get the multimodal quality  $Q^*_{AV}$  in general describes the experimental data, but does not take into account cross-modalities interaction. The incomplete square model when factor  $c_{AV}$  is non-zero can be used as a simplest form to describe the effect of the interaction of two modalities. The polynomial model of the second degree is non-linear, and sufficiently reflects all the properties of the mutual compensation of modalities when one modality can complement and modify the perceptual experience created by other one.
5. A numerical check of the relevance of the model was performed to find the optimal number of non-zero coefficients in the polynomial model. Special criteria that take into account the requirements for visual perception models, are formulated by the VQEG (Video Quality Experts Group): accuracy, monotonicity, and consistency. The first requirement, an increase in accuracy, is met with an increase in the number of non-zero coefficients in the polynomial. The second requirement is monotonicity, which is violated as the degree of the polynomial increases, as the lines become oscillatory. An acceptable compromise between accuracy, monotonicity, and model complexity finds by using a polynomial model of the second degree.
6. The root-mean-square error specifying the accuracy of prediction is 0.1568 units. The Spearman rank-order correlation coefficient relating to the prediction monotonicity of the model is 0.9883. The outlier ratio specifying the consistency is 0. It is clear that the accuracy, monotonicity and consistency of the model are rather high, the polynomial approximation of the second degree can be used as a multimodal quality model. In this case, the model predicts the properties of mutual compensation of modalities when one modality can complement and modify the perceptual experience created by the other.

## 4. Optimization of compression codec parameters in digital television systems

### 4.1. Statement of the problem

Section 1 of this work showed the existence of cross-modal interaction in the process of forming the viewer's integral impression of the viewed audiovisual programs. It has been shown that cross-modal interaction exists when viewing programs of different content and in a wide range of bit rates of compressed image and sound data. Both modalities (visual and auditory) influence the perception of quality, but one of the modalities may take precedence depending on the content of the program. In section 3, a mathematical model of subjective perception was developed. With its use, it becomes possible to perform a quantitative description of the processes of cross-modal interaction between the components of an audiovisual program during their joint perception, when one of the components complements and changes the sub-

jective sensation formed by another component.

In television systems with data compression, including mobile television systems that use communication channels in which the target data rate of compressed data drops to one megabit per second and even lower values, special requirements are placed on the production and delivery of multimedia products. The main task is to ensure an acceptable level of quality in the face of severe technical restrictions on the delivery of audiovisual materials. To select the optimal values for the image and sound encoding parameters, it is important to analyze the quality requirements based on subjective perception. Considering the audiovisual program as a whole and taking into account the perception of its integral quality will solve the problem.

The paper proposes to solve two optimization problems important for practice.

1. The first optimization problem can be formulated as follows: to obtain the maximum integral quality of experience with a fixed bandwidth of the communication channel. The mathematical model of subjective perception can be applied for dynamic and adaptive to the content of broadcast materials allocation of the channel bandwidth between video and audio streams, which allows maximizing the level of integral quality of the experience of playing multimedia programs (Peregudov, Glasman, Belozertsev, Grinenko, 2009; 2010) under technical restrictions on the delivery of television materials (limited and fixed bandwidth of the communication channel).
2. The developed mathematical model can be used to solve the problem of minimizing the required bandwidth of the communication channel, provided that the integral quality of experience is not lower than the specified level. Thus, the second optimization problem can be formulated, which is solved within the framework of this work: to find the minimum throughput of the communication channel for a given value of the level of the integral quality of experience.

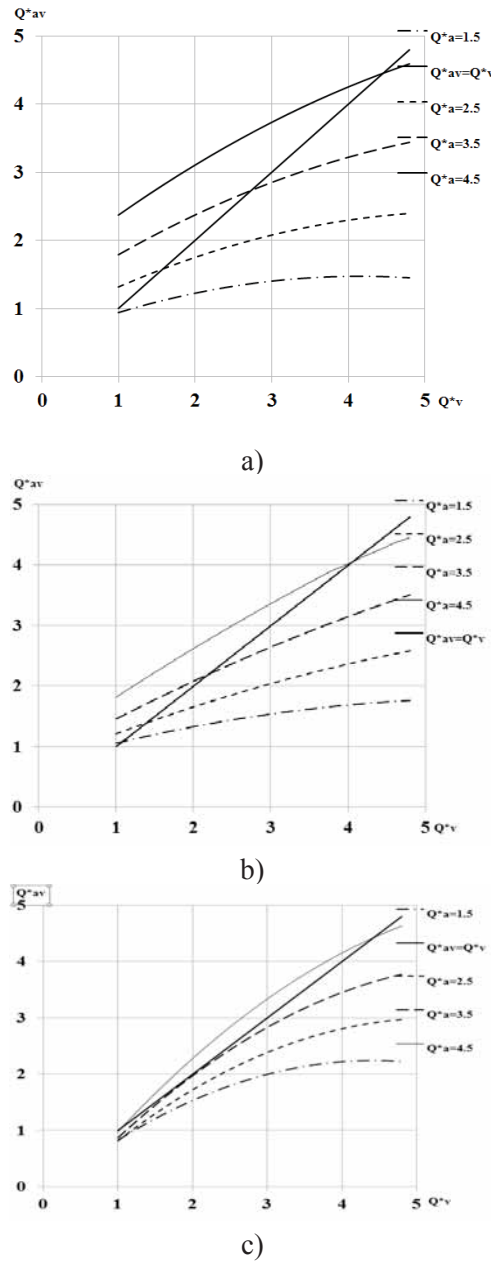
## **4.2. Cross-modal interaction of audiovisual program components**

### **4.2.1. Trade-offs between picture and sound quality**

In (Storms, 1998), an important question for practice is formulated about the interaction and mutual influence of video and audio components on each other in the process of forming an experience of integral quality: is a high/low quality of the sound accompaniment the cause of an increase/decrease in the perceived level of image quality (in relation to the level monomodal image quality played back without sound)? Conversely, does an image played back with sound cause a change in the perception of monomodal sound quality? The author of the work carried out experimental studies, the results of which give a qualitative answer to the question of the interaction of two modalities. Using the developed mathematical model of subjective perception, it becomes possible to give a quantitative assessment (Peregudov, Glasman, Belozertsev, Grinenko, 2010).

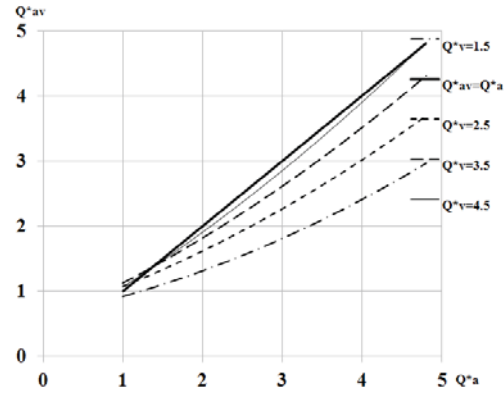
The graphs in Figure 13 illustrate the dependence of the level of integral quality predicted by the model and monomodal values of image and sound quality for three experimental scenes. Sound quality is selected as an option. An analysis of the presented characteristics allows us to conclude that there are so-called neutrality points. At these points, there is neither a decrease nor an increase in the level of subjectively perceived integral quality in comparison with monomodal estimates of image quality values. Consider Figure 18b, which shows the dependence of the integral quality level predicted by the model and the monomodal image and sound quality indicators for a music video. For example, if the values of the image and

sound quality indicators are equal to  $Q^*_v = 4$  and  $Q^*_a = 4.5$ , then the level of integral quality will correspond to the score “4”, i.e. equal image quality. All points of neutrality lie on the straight line  $Q^*_{AV} = Q^*_v$  (Гриненко, 2011; Перегудов, Белозерцев, Гласман, Гриненко, 2010; Peregudov, Glasman, Belozertsev, Grinenko, 2009). For low image quality (i.e., low video bit rates), the integrated quality perceptual level will always be higher than the perceptual image quality level for all three scenes. However, for medium and high values of image quality, the situation is not so clear and depends on the level of sound quality.

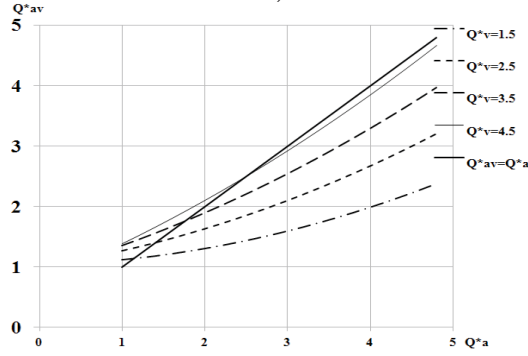


**Figure 13.** Relationship between monomodal image quality  $Q^*_v$  and integral quality  $Q^*_{AV}$  predicted by the model (the value of sound quality  $Q^*_a$  is a parameter) for a) news; b) music video; c) sports

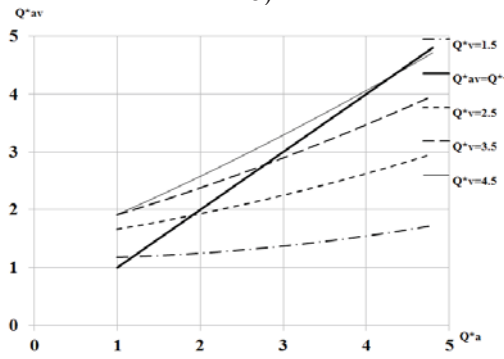
Figure 14 shows similar dependences of the integral quality level predicted by the model and monomodal image and sound quality indicators for three scenes. As a parameter for the curves, the corresponding image quality values are selected. There are also points of neutrality in the figure. At these points, there is neither a decrease nor an increase in the level of subjectively perceived integral quality in comparison with monomodal estimates of sound quality indicators.



a)



b)

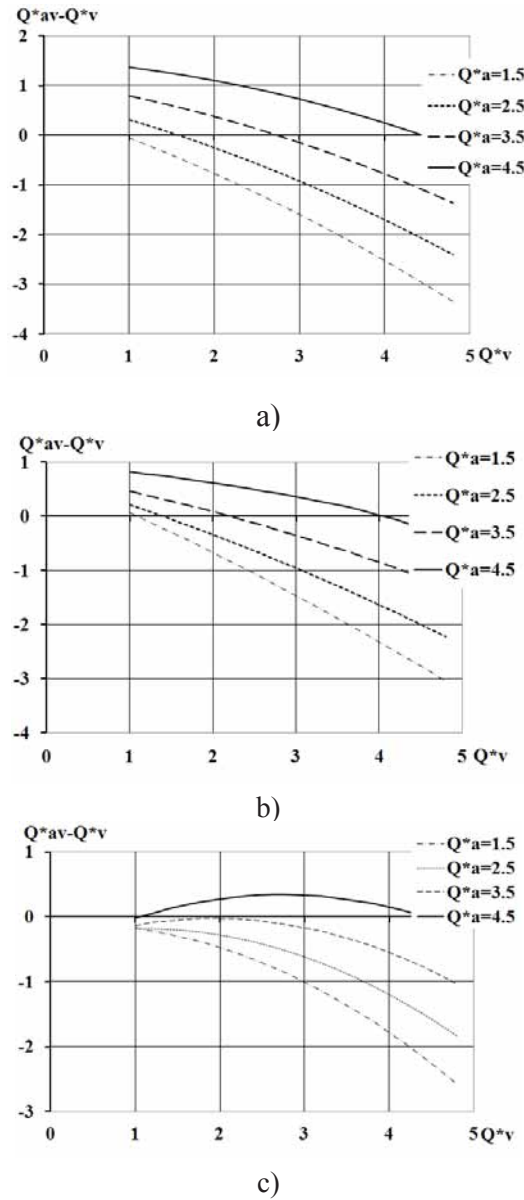


c)

**Figure 14.** Relationship between monomodal sound quality  $Q^*_A$  and integral quality  $Q^*_{AV}$  predicted by the model (the value of image quality  $Q^*_V$  is a parameter) for a) news; b) music video; c) sports

Figure 15 shows the dependence of the difference between the values of the integral quality estimate predicted by the model and the value of monomodal image quality as a function of the level of monomodal image quality. As a parameter for the curves, the corresponding values of the sound quality value are selected. A positive difference value means that the integral quality score will be higher than the monomodal image quality score. With the same degree of video compression (video data bitrate), the perception of video materials obtained by measuring the quality of only the image (monomodal quality) turns out to be lower than in the case when the expert evaluates the quality of the image and sound together (integral quality). The effect of increasing the level of perception is most noticeable in the region of low values of the image quality parameter. For a news program with values of image and sound quality indicators  $Q^*_V=1$  and  $Q^*_A=4.5$ , the level of perception of the integral quality will be higher by 1.4 points compared to the level of perception of the quality of only an image encoded with the same bit rate. With the same values of the image and sound quality indicators, the level of

perception of the integral quality for a music video will be higher by 0.9 points (Peregudov, Glasman, Belozertsev, Grinenko, 2010).

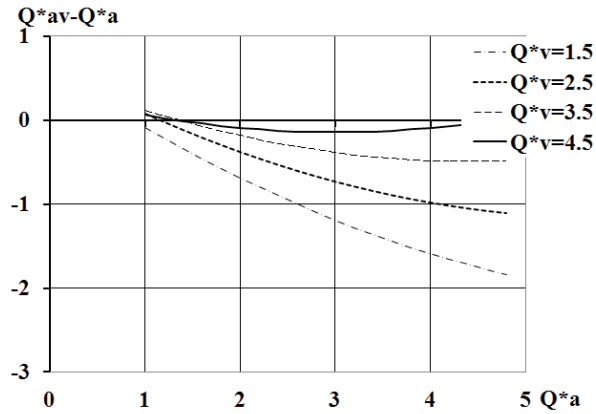


**Figure 15.** Change in the perception of integral quality in relation to monomodal image quality ( $Q^*_{av} - Q^*_v$ ) depending on the level of image quality ( $Q^*_v$ ) with different values of sound quality ( $Q^*_a$ ) for a) news program; b) music video; c) sports report

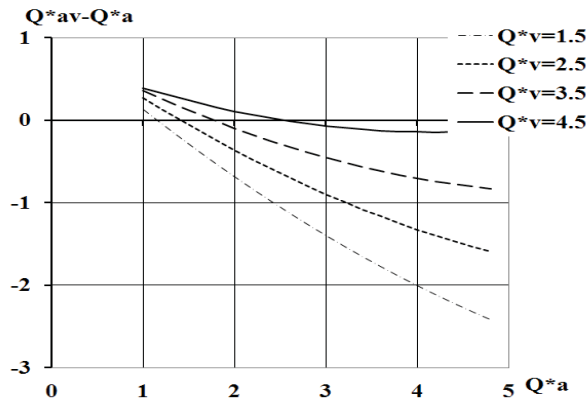
A negative difference value means that the integral quality score will be lower than the monomodal image quality score. With the same degree of compression (video data bitrate), the video perception estimates obtained by measuring only the image quality (monomodal quality) turn out to be higher than in the case when the expert evaluates both the image and sound quality (integral quality). The effect is most noticeable at high image quality (4.5 points) and low sound quality (1.5 points). At the same video data bitrate, the level of integral quality perception is 3 or more points lower than the level of monomodal image quality for a news program and a music video. Under the same conditions, the level of integral quality for a sports report is lower than the level of monomodal image quality by 2.5 points.

Figure 16 shows the dependence of the difference between the values of the integral quality estimate predicted by the model and the monomodal sound quality value as a function of the

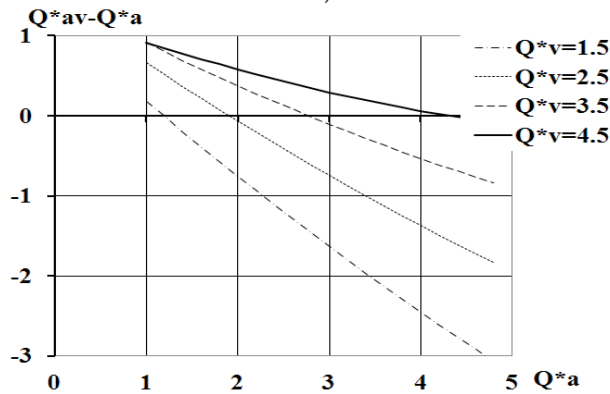
monomodal sound quality value. The level of integral quality of audiovisual materials (with integral assessment of image and sound quality) may turn out to be higher than the level of only sound quality (monomodal quality) for the same parameter values. High-quality sound reproduction enhances the overall integral level of the experience. This effect is most noticeable in the region of low data bitrates. The characteristics also indicate the existence of the opposite effect: an image played back with sound can reduce the integral experience of the quality of the video. At low values of the image quality and high sound quality, the integral quality score is from 1.5 to 3 points lower than the monomodal sound quality score encoded at the same audio data bitrate.



a)



b)



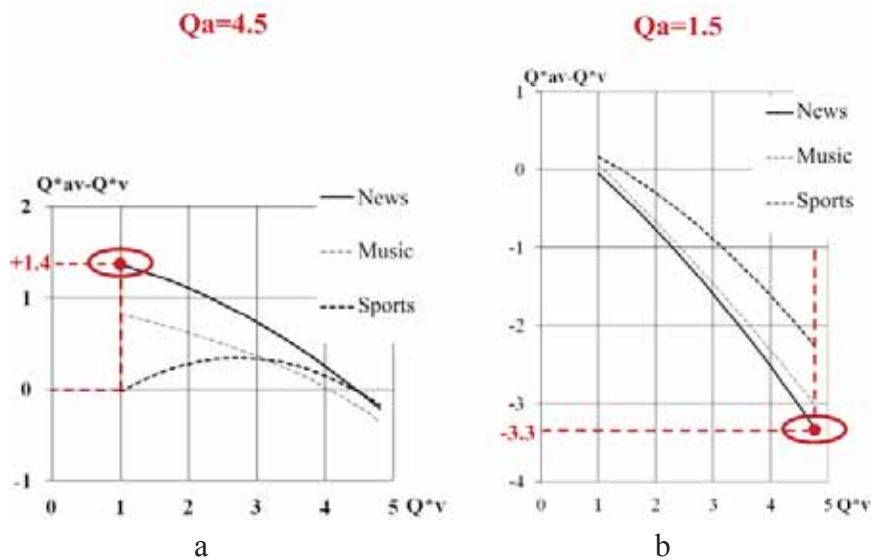
c)

**Figure 16.** Change in the perception of integral quality in relation to monomodal sound quality ( $Q^*_{AV} - Q^*_A$ ) depending on the level of sound quality ( $Q^*_A$ ) with different values of image quality ( $Q^*_v$ ) for a) news program; b) music video; c) sports report

#### 4.2.2. Influence of the genre and content of audiovisual materials on perceived quality

As noted, both components (image and sound) contribute to the integral perception of an audiovisual program, but one of them may be dominant. The influence of partial parameters of the video and audio series on the perception of quality as a whole depends on the genre and content of television materials. To illustrate cross-modal audiovisual interaction, we will use the experimental data obtained for three programs: news, music, and sports (Peregudov, Glasman, Belozertsev, Grinenko, 2010).

Figure 17(a) shows the relationship between the difference  $Q^*_{AV} - Q^*_v$  of the values of the integral experience quality assessment predicted by the model and the values of the monomodal image quality as a function of the monomodal image quality index  $Q^*_v$  for news, music and sports videos. The sound quality is quite high and corresponds to a score of 4.5 points. The characteristics obtained with a low sound quality, corresponding to a score of 1.5 points, are shown in Figure 17 (b).



**Figure 17.** Relationship between the difference  $Q^*_{AV} - Q^*_v$  of the values of the quality of experience predicted by the model and the values of the monomodal image quality as a function of the monomodal image quality value  $Q^*_v$  for news, music and sports videos: a) - monomodal sound quality  $Q^*_A=4.5$  points; b) - monomodal sound quality  $Q^*_A=1.5$  point

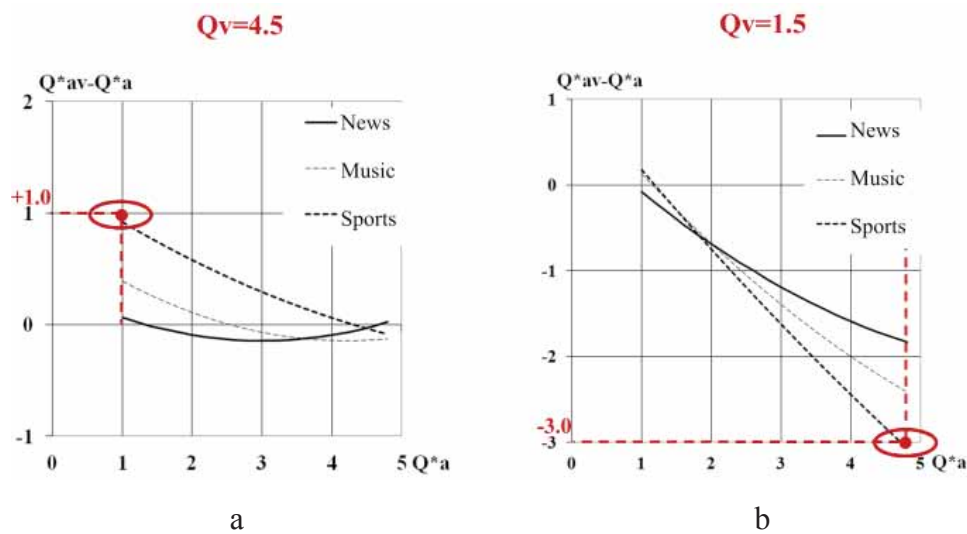
Similar curves illustrating the dependence of the difference  $Q^*_{AV} - Q^*_A$  of the predicted integral quality of experience score and the monomodal sound quality score as a function of the monomodal sound quality indicator  $Q^*_A$  are shown in Figures 18 (a, b). Graphs (Figure 18 a) correspond to high image quality on level 4.5 points. The characteristics presented in Figure 18 b are obtained with a low image quality value of 1.5 points.

It is easy to see that with a high quality of sound (Figure 17 a), the level of the integral quality of a news story increases (in relation to the monomodal image quality) to a greater extent than the level of the integral quality of a sports report or a music video. The excess of the integral assessment over the monomodal assessment of image quality reaches 1.4 points at low video data bitrates and, accordingly, low image quality. On the contrary, the integral quality of a news story is reduced by 3.3 points with low sound quality and high image quality, as shown in Figure 17 b.

This is quite understandable, since for a news story of the “talking head” type, the sound com-

ponent has a dominant value (Гласман, Белозерцев, Гриненко, 2011; Peregudov, Glasman, Belozertsev, Grinenko, 2010).

From the graphs in Figure 18 a, it can be seen that in the case of high video data bitrates, the integral quality score (compared to monomodal audio quality) increases to a greater extent for a sports report than for a news fragment or a music video. With low sound quality (Figure 17 b), the value of the integral quality of a sports report falls (in relation to the level of monomodal image quality) not as quickly as the quality of a news fragment and a music video. However, with high sound quality and low image quality, the level of integral quality falls (in relation to the level of image quality) faster for sports reports than for news and music videos (Figure 18 b). This trend can be explained by the high level of small details and high intensity of movement in sport reports, as well as the high semantic significance of the video in the perception of sports stories.



**Figure 18.** Relationship between the difference  $Q^*_{AV} - Q^*_A$  of the values of the quality of experience predicted by the model and the values of the monomodal sounds quality as a function of the monomodal sound quality value  $Q^*_A$  for news, music and sports videos: a) - monomodal image quality  $Q^*_v = 4.5$  points; b) - monomodal image quality  $Q^*_v = 1.5$  points.

Video clips with musical sound accompaniment occupy an intermediate position between a news story and a sports report, as it is characterized by an average levels of small image details and intensity of movement in the picture.

### 4.3. Adaptive bit-rate allocation between video and audio streams according to the genre of the audiovisual program

The first optimization problem, which is solved within the framework of the work, is formulated as follows: to obtain the maximum integral quality with a fixed bandwidth of the communication channel. The output signal of the developed integral quality perception model is used as an objective function in solving the optimization problem (Grinenko, Glasman, Belozertsev, 2014).

Qualitative reasoning shows that this problem has a solution. Both components, image and sound, contribute to the integral perception of an audiovisual program. If the sound or image quality is low, the integral quality level will decrease for all three experimental video types. Under conditions of limited bandwidth of the communication channel at low video data bitrates, the integral quality of audiovisual programs decreases due to poor image quality, al-

though the sound quality is quite high. As the video data bit-rate increases, the image quality level increases, but the audio data bit-rate decreases, therefore, the sound quality level drops. The level of integral quality of an audiovisual program is reduced due to poor sound quality. Therefore, there is an optimal allocation of the total data bitrate between the video and audio data.

In essence, the task is reduced to calculating the bitrates of audio and video data streams, at which the maximum value of the integral quality of experience  $Q_{AV}$  is achieved for a given total bit rate. The total bit rate of the data stream carrying audio video data, TBR (Total Bit Rate), is the sum of the audio data bitrate ABR (Audio Bit Rate) and the video data bitrate VBR (Video Bit Rate).

The mathematical formulation of the problem of finding the optimal solution can be written as follows (4.1):

$$\begin{cases} F = Q_{av}(ABR, VBR) \rightarrow \max \\ ABR + VBR = TBR \\ ABR > 0, VBR > 0 \end{cases} \quad (4.1)$$

In the formulation of the optimization problem, three components can be distinguished: the objective function, constraints, and boundary conditions. In this case, the optimization problem should be solved with constraints, that is, under the condition of a fixed total bit rate. The condition according to which the bitrates of the digital audio and video data streams can only be positive values is obvious, but necessary for the correct solution of the problem.

The objective function is described by an equation that is an integral perception model developed in Section 3:

$$Q_{AV}^* = c_0 + c_{A1} \cdot Q_A^* + c_{V1} \cdot Q_V^* + c_{AV} \cdot Q_A^* \cdot Q_V^* + c_{A2} \cdot Q_A^{*2} + c_{V2} \cdot Q_V^{*2},$$

where

$$Q_V^* = k_{V1} + k_{V2} \cdot \exp(k_{V3} \cdot VBR),$$

$$Q_A^* = k_{A1} + k_{A2} \cdot \exp(k_{A3} \cdot ABR),$$

$$ABR = TBR - VBR.$$

Then

$$Q_{AV}^* = c_0 + c_{A1} \cdot (k_{A1} + k_{A2} \cdot \exp(k_{A3} \cdot TBR) + k_{A2} \cdot \exp(-k_{A3} \cdot VBR)) + c_{V1} \cdot (k_{V1} + k_{V2} \cdot \exp(k_{V3} \cdot VBR)) + c_{AV} \cdot (k_{A1} \cdot k_{V1} + k_{V1} \cdot k_{A2} \cdot \exp(k_{A3} \cdot TBR) \cdot \exp(-k_{A3} \cdot VBR) + k_{V2} \cdot k_{A1} \cdot \exp(k_{V3} \cdot VBR) + k_{V2} \cdot k_{A2} \cdot \exp(k_{A3} \cdot TBR) \cdot \exp(k_{V3} - k_{A3} \cdot VBR)) + c_{A2} \cdot (k_{A1} + k_{A2} \cdot \exp(k_{A3} \cdot TBR) + k_{A2} \cdot \exp(-k_{A3} \cdot VBR))^2 + c_{V2} \cdot (k_{V1} + k_{V2} \cdot \exp(k_{V3} \cdot VBR))^2$$

The optimization problem is reduced to finding the extremum of the objective function. The extremum point is the point where the derivative of the objective function is equal to zero. Therefore, in order to find an analytical solution to the stated optimization problem, it is necessary to calculate the derivative of the objective function with respect to the video data bitrate at a fixed total data bitrate and equate it to zero.

The resulting equation does not have an analytical solution; it is possible to find the maximum of the objective function only by numerical methods. A large number of optimization methods have been developed, which can be divided into two groups: methods for narrowing the uncertainty interval and methods using derivatives. For the considered objective function,

the derivative exists, therefore, to solve the optimization problem, it makes sense to apply numerical methods using derivatives. One of the methods for solving the problem of such an optimization is the method of Lagrange multipliers, which was used in (85).

To solve the optimization problem, the data analysis software tools Microsoft Excel 2011.6, the Solver add-on (Бекр и др, 2005; Минько, 2004) were used in the work. The parameters and results of the solution search procedure in the Solver program are displayed using tables on three report sheets. Below are the parameters and results of the procedure for searching for the optimal bit rates of the digital audio and video data streams, at which the maximum value of the integral quality is achieved at a given total bit rate of 500 Kbps for the music video program. The Answer Report sheet (Figure 19) displays the start and end values of the target and variable cells, constraint cells, and constraint information. Sheet Sensitivity (Figure 20) contains information about how sensitive the solution is to small changes in the formula located in the target cell field in the Solver add-in dialog box or restrictions. This report is not generated for models that have integer constraints. For non-linear objective functions, the report contains the values of the reduced gradient and Lagrange multipliers. On the Limits sheet (Figure 21)–constraints, upper and lower limits and the target value of the cells of variables and the target cell, the optimal value of the objective function for the problem being solved are displayed.

Microsoft Excel 11.6 Answer Report

Worksheet: [Model calc\_v5\_opt.xls]Music-Opt

Report Created: 4/18/2014 8:56:06 PM

#### Target Cell (Max)

Cell	Name	Original Value	Final Value
\$O\$29	Qav max	4,525	4,623

#### Adjustable Cells

Cell	Name	Original Value	Final Value
\$B\$29	VBR	250,0	313,6
\$C\$29	ABR	250,0	186,4

#### Constraints

Cell	Name	Cell Value	Formula	Status	Slack
\$C\$29	ABR	186,4	\$C\$29=\$Q\$25-\$B\$29	Binding	0
\$B\$29	VBR	313,6	\$B\$29>=0	Not Binding	313,6
\$C\$29	ABR	186,4	\$C\$29>=0	Not Binding	186,4

**Figure 19.** Parameters and results of the procedure for finding a solution in the Solver program.

### Answer Report Sheet

Microsoft Excel 11.6 Sensitivity Report				
Worksheet: [Model calc_v5_opt.xls]Music-Opt				
Report Created: 4/18/2014 8:56:11 PM				
Adjustable Cells				
		Final	Reduced	
Cell	Name	Value	Gradient	
\$B\$29	VBR	313,6	0,0	
\$C\$29	ABR	186,4	0,0	
Constraints				
		Final	Lagrange	
Cell	Name	Value	Multiplier	
\$C\$29	ABR	186,4	0,0	

**Figure 20.** Parameters and results of the procedure for finding a solution in the Solver program. Sensitivity Report Sheet

Microsoft Excel 11.6 Limits Report							
Worksheet: [Model calc_v5_opt.xls]Music-Opt							
Report Created: 4/18/2014 8:56:15 PM							
		Target					
Cell	Name	Value					
\$O\$29	Qav max	4,623					
	Adjustable		Lower	Target		Upper	Target
Cell	Name	Value	Limit	Result		Limit	Result
\$B\$29	VBR	313,6	313,6	4,6		313,6	4,6
\$C\$29	ABR	186,4	186,4	4,6		186,4	4,6

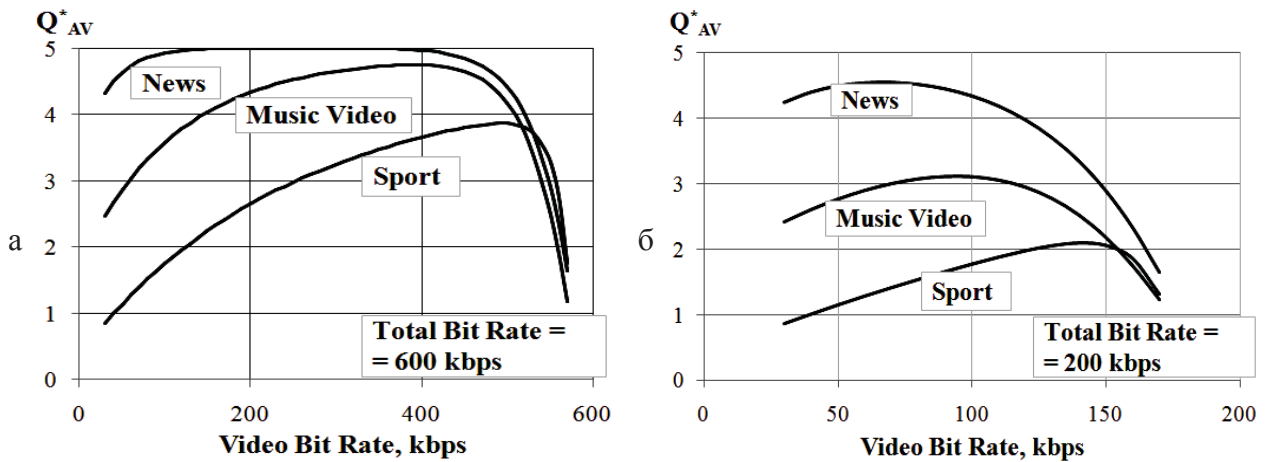
**Figure 21.** Parameters and results of the procedure for finding a solution in the Solver program. Limits Report Sheet

As the initial values (zero approximation) of the bitrates of the digital audio and video data streams, values equal to half of the total bandwidth of the communication channel are used. The available data transfer bitrate in the communication channel is divided in half between the bitrates of audio and video data, that is, equal shares of the total bandwidth of the communication channel are allocated for the transmission of audio and video information. For a music video with a ratio of video and audio bitrates of 250:250, the value of the integral quality that can be obtained is 4.525 points. The maximum integral quality for a music video, corresponding to  $Q^*_{AV}=4.623$  points, is achieved at the video data bitrate of 313.6 kbps and the audio data bitrate of 184.4 kbps. This distribution (313.6:186.4) can be considered optimal, with a limited total bit rate of 500 kBit/s.

The results of solving the optimization problem are shown in Figure 24 for a news story, a music video, and a sports report (at the total bitrates of 600 Kbps and 200 Kbps). The figure

clearly illustrates the change in the level of integral quality  $Q^*_{AV}$  with a change in the share of the bandwidth of the communication channel occupied for the transmission of information about the image and sound, at a fixed total bit rate. The figures clearly show the maxima of the integral quality, which can be obtained under conditions of a fixed bandwidth of the communication channel.

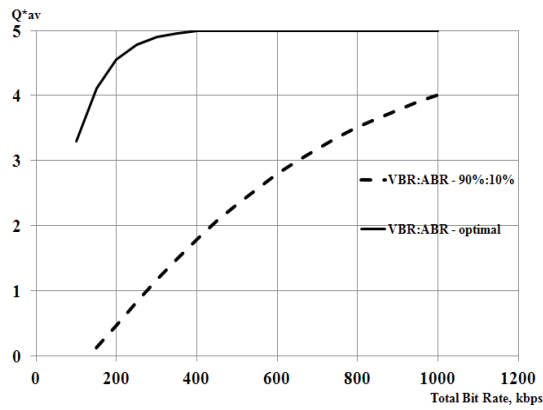
With the total bitrate of  $TBR=200$  kbps (Figure 22b), the maximum integral quality of a sports report, corresponding to  $Q^*_{AV} = 2.1$  points, is achieved at the video bitrate of 140 kbps and the audio bitrate of 60 kbps. Such a distribution (140:60) can be considered optimal. For a music video, the optimal ratio of video and audio bitrates is 90:110, which corresponds to the maximum integral quality value of 3.1 points. For a news program, the highest quality level of 4.5 can be achieved with the video/audio bit rate ratio of 70:130 (VBR = 70 kbps and ABR = 130 kbps).



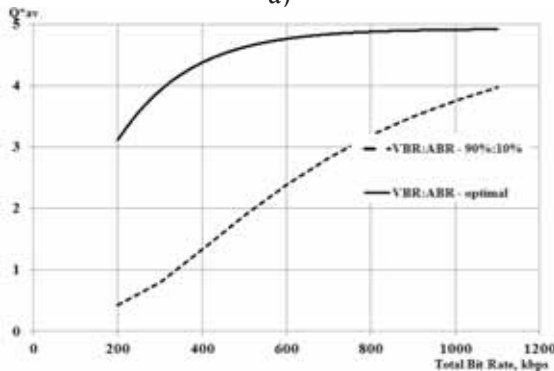
**Figure 22.** Relationships between the video data bit rate VBR and the predicted value of the integral quality  $Q^*_{AV}$  for a news story, a music video and a sports report: a – total bitrate of 600 Kbps; b – total bitrate of 200 Kbps.

With the total bitrate  $TBR = 600$  kbps (Figure 24a), the maximum value of the quality of a sports report  $Q^*_{AV} = 3.9$  points is achieved at the video bitrate  $VBR = 490$  kbps and the audio bitrate  $ABR = 110$  Kbps. The ratio of 490:110 can be considered as optimal for the transmission of highly dynamic sports reports at the total bit rate of 600 kbps. For a music video, the optimal ratio of 390:210 results in the quality score of 4.8. For a news report, the highest quality (5 points) will be provided with a video bitrate selected from the range from 150 kbps to 400 kbps (4, 81, 219, 220).

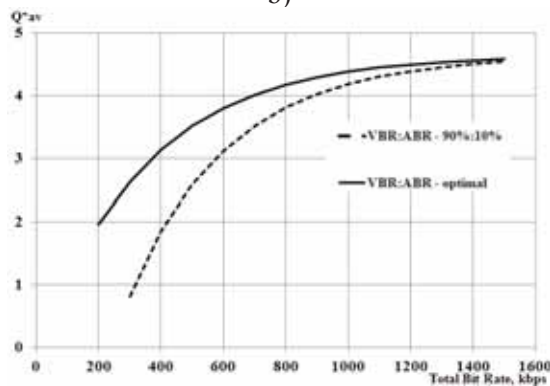
Traditionally, in digital television systems, most of the total bandwidth of the communication channel is allocated for the transmission of image information (for example, 90 %, then only 10 % remains for the transmission of sound information). The graphs in Figure 23 illustrate the relationship between the total bit rate of the data stream carrying audiovisual data and the integral quality value for a news story, a music video, and a sports report. The dotted line corresponds to the relationship showing the maximum value of the integral quality that can be obtained while maintaining a fixed ratio between the audio and video data bitrates at different values of the total bitrate. The solid line on the graphs corresponds to the relationship illustrating the maximum value of the integral quality that can be obtained with the optimal ratio of video and audio bitrates.



a)



b)



c)

**Figure 23.** The relationships between TBR available for the service and predicted audiovisual quality of experience in the cases of fixed (90% for video and 10% for audio) and optimum bit rate allocation for a) news, b) music video, c) sport program

The use of the optimal ratio between the bitrates of the digital stream of audio and video data when transmitting audiovisual programs can significantly increase the level of integral quality, especially when a narrow bandwidth of the communication channel is used. For a news story, provided that the total bit rate is 400 kbps, the integral quality level can be increased from 1.9 points to 5 points, provided that the bitrates are optimally allocated. The optimal ratio in this case would be 161.9:238.1 (video bit rate 161.9 kbps, audio bit rate 238.1 kbps). For a music video at the same total bit rate, the integral quality level can be increased from 1.3 points to 4.4 points at the video bitrate of 238.4 Kbps and the audio bitrate of 161.6 kbps. The ratio of 238.4:161.6 can be considered as optimal for music videos. For a sports report, the level of integral quality at the total stream rate of 400 kbps, with the fixed allocation, will be equal to 1.9 points. With the optimal ratio, it will be equal to 3.1 points (video bitrate 316.1 kbps,

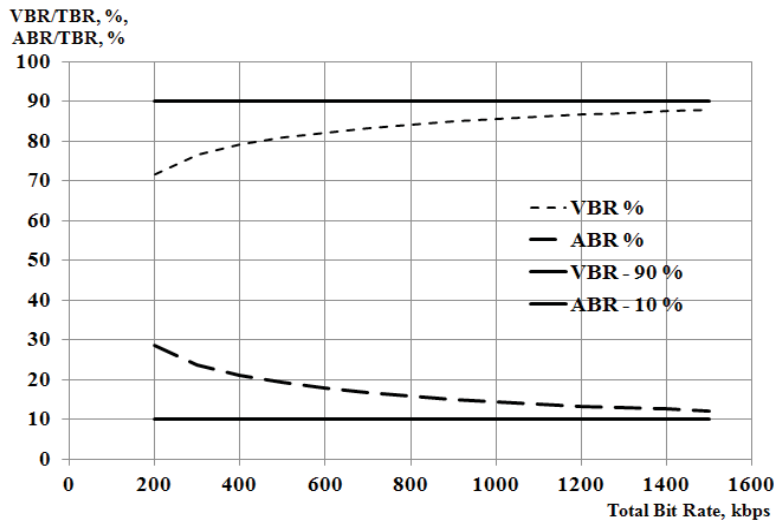
audio bitrate 83.9 kbps). If the total bit rate is 400 kbps, the use of the optimal ratio between the bitrates of the digital stream of audio and video data when transmitting news programs of the “talking head” type allows you to increase the level of integral quality by 3.1 points, when transmitting music programs—by 3.1 points, and when transmitting sports reports—by 2.1 points, compared with the transmission of the same programs with a fixed ratio between the bitrates of the digital stream of audio and video data, when 90 % of the bandwidth of the communication channel is allocated for the transmission of image information.

With the total bit rate of 1 Mbps, the quality level for a news story can be increased from 4 to 5 points. The optimal ratio in this case is 454.0:546.0 (video bitrate 454.0 kbps, audio bitrate 546.0 kbps). For a music video, the integral quality level can be increased from 3.8 points to 4.9 points at the video bitrate of 696.6 kbps and the audio bitrate of 303.4 kbps. The ratio of 696.6:303.4 can be considered optimal for transmitting music video programs, provided the total data transfer rate is 1 Mbps. For sports reports, the integral quality level can be increased from 4.2 to 4.4 points when using the optimal ratio of audio and video bitrates. So, provided that the total bit rate is 1 Mbps, the use of the optimal ratio between bitrates allows to increase the level of integral quality, when transmitting news program of the “talking head” type by 1 point, when transmitting a music program—by 1.1 points and when transmitting sports reporting—by 0.2 points (Grinenko, Glasman, Belozertsev, 2014).

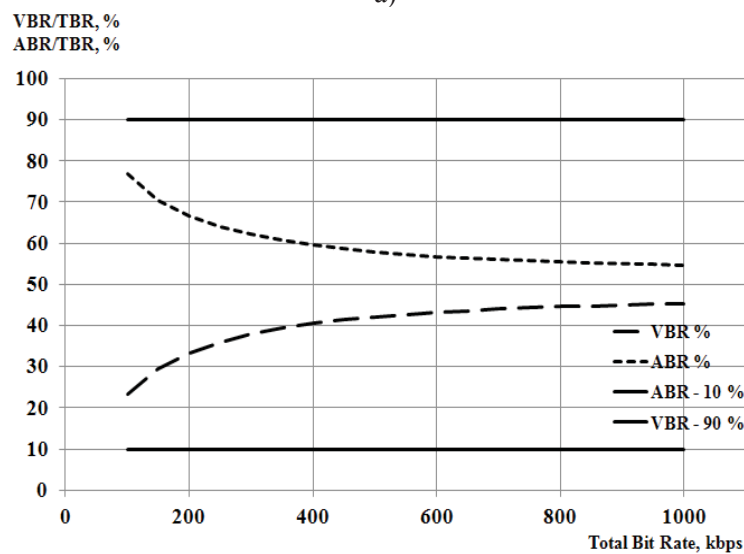
The increase in the level of integral quality for sports reports is the least significant. This is explained by the fact that the optimal ratio of audio and video data bitrates for highly dynamic scenes is closest to that often used in practice, where a large part of the total bandwidth of the communication channel is allocated for the transmission of image information. For news programs of the “talking head” type, with a relatively low overall bit rate, the allocation of most of the channel bandwidth for transmitting information about sound, as a semantically more important component, can significantly increase the integral quality.

The optimal ratio of audio and video bitrates for materials of different content, at which the maximum integral quality is achieved, is different for the same total bitrate of the digital data stream. Therefore, the use of channel bandwidth allocation between video and audio streams that is adaptive to the content of broadcast materials can be used to maximize the integral quality when playing multimedia programs under tight technical restrictions on the delivery of television materials, when there is a limited fixed bandwidth of the communication channel. Adaptive allocation of the bandwidth of the communication channel between video and audio data streams in accordance with the genre and content of the audiovisual program allows you to increase the level of integral quality of audiovisual programs at a given total bitrate of the transmitted data.

The optimal ratio of audio and video bitrates depends on the total bit rate available for the transmission of the multimedia program. Figure 24 shows the optimal shares of the channel bandwidth occupied for the transmission of information about the image and sound, as a functions of the total bandwidth of the communication channel.



a)



b)

**Figure 24.** Optimal channel bandwidth allocations for image and sound data transmission as functions of total channel bandwidth for a) sports; b) news

For sports reports, as the total bit rate of the audiovisual data stream increases, the share of the bandwidth of the communication channel for the transmission of image information will increase, while the share of the bandwidth of the communication channel allocated for the transmission of audio information will proportionally decrease. The optimal ratio of audio and video bitrates for sports reports with an increase in the total bitrate will tend to a value that is often used in practice. The optimal ratio of audio and video streams at the low total bitrate of 200 kbps is 70:30, while at the total bitrate of 1.5 Mbps, the optimal ratio is 90:10.

For a “talking head” news program with a low bandwidth of the communication channel (about 150 kbps), the ratio of 20:80 can be considered optimal. Most of the total bandwidth of the communication channel—80 % should be allocated for the transmission of information about the sound and only 20 % for the transmission of information about the image. This allocation of audio and video bitrates is not typical for television systems. The results obtained confirm the provisions formulated within the framework of the informational approach, which takes into account the presence in films and television programs of both semantic and aesthetic information. A larger share of the bandwidth of the communication channel should be allocated to a semantically more significant component of the audiovisual program. As it was shown,

for a news program of the “talking head” type, the quality of the sound accompaniment is of the greatest importance, therefore, a large share of the bandwidth of the communication channel should be allocated for the transmission of information about the sound. In rich and highly dynamic sports scenes, picture quality has relatively more weight than sound.

#### 4.4. Reducing the required bandwidth of the communication channel in accordance with the content of the multimedia program

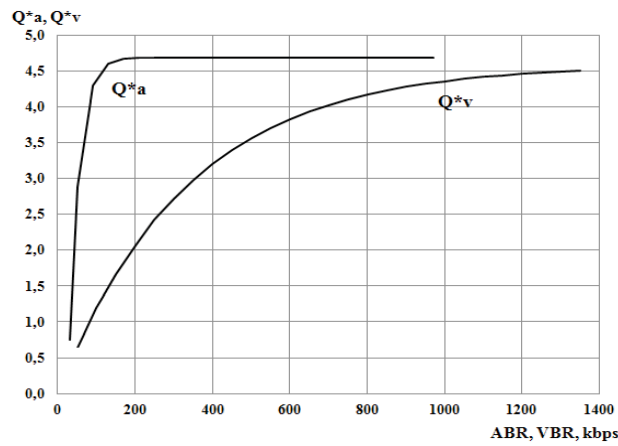
In digital television systems, methods and devices for compressing image and sound data are used in order to match the bitrate of the digital stream with the bandwidth of the communication channel used. As shown above, the total bitrate allocated for the transmission of an audiovisual program includes separate partial streams for the transmission of the image and sound signal. The analysis of the experimental data presented above suggests that considering the audiovisual program as a whole and taking into account polysensory perception of audiovisual information by a person will make it possible to achieve savings in the required communication channel bandwidth. The model can be applied to solve the problem of minimizing the required bandwidth of a communication channel, provided that the integral quality is at the level of a given value. Thus, the second optimization problem can be formulated, which is solved within the framework of this work: to obtain the minimum required bandwidth of the communication channel TBR (Total Bit Rate) for a fixed value of the integral quality level  $Q_{AV}$ . In fact, the problem is reduced to calculating the bitrates of audio ABR (Audio Bit Rate) and video VBR (Video Bit Rate) data streams, at which the minimum value of the total bitrate TBR is achieved at a given level of integral quality  $Q_{AV}$ .

The mathematical model of the problem of finding the optimal solution can be written as follows (4.2)

$$\begin{cases} F = TBR = (ABR + VBR) \rightarrow \min \\ Q_{av}(ABR, VBR) = Q_{av} \\ ABR > 0, VBR > 0 \end{cases} \quad (4.2)$$

It is a constrained optimization problem. The optimization problem was also solved using the Lagrange multiplier method using Microsoft Excel 2011.6 data analysis software tools Solver add-on (Бекр и др, 2005; Минько, 2004).

In known systems with compression, the proportions of partial data stream bitrates are fixed, which sets the output of the system to a fixed quality separately for the image and separately for the sound. The graph (Figure 25) shows the dependence of the monomodal image and sound quality indicators ( $Q_v^*$  and  $Q_a^*$ ) predicted by the model given in Section 3.2 on the bitrates of the video and audio stream when separately assessing the quality of image and sound for sports reports. As the initial conditions, that is, the values of the video and audio data stream bitrates of the zero approximation, the values of the video and audio bitrates are used, which provide a level of monomodal image and sound quality of 4.5 points. Monomodal image quality will reach the target level of 4.5 points at the digital video stream rate of 1310.3 kbps. Monomodal sound quality will be equal to 4.5 points at the digital audio stream rate of 108.3 kbps. Therefore, a total bit rate of 1418.6 kbps is required to transmit image and sound with a quality of 4.5 points.



**Figure 25.** Relationship between audio ABR and video VBR bitrates and predicted audio and video quality  $Q^*_A$  and  $Q^*_V$  for sports

If we consider the audiovisual program as a whole, then the integral quality for a sports report will reach 4.5 points at the video stream bitrate of 1058.4 Kbps and the audio stream bitrate of 161.7 Kbps. In this case, the minimum required total bitrate of the data stream carrying audiovisual data, provided that the integral quality is 4.5 points, will be equal to 1220.1 kbps. This value is almost 200 kbps less than the minimum total bitrate required to transmit the same sports report's image and audio at the monomodal image and audio quality level of 4.5. Consideration of an audiovisual program as a whole and taking into account the multimodal polysensory perception of audiovisual information will allow to achieve savings of the required bandwidth up to 14 % for the most critical to the bandwidth sports.

A significant reduction in the minimum required channel bandwidth can be achieved if one takes into account the fact that television programs of different genres and content can be played back with the same level of experience quality at different total program compressed data rates. Table 11 presents the values of the minimum required bandwidth of the communication channel for the transmission of programs of various content with a fixed quality level, obtained as a result of solving the optimization problem.

**Table 11.** Values of the minimum required bandwidth of the communication channel for the transmission of programs of various content

$Q_{AV}$	News Story	Musical Video	Sports Report
3.0	87,4 kbps	189,8 kbps	371,8 kbps
3.5	110,2 kbps	241,4 kbps	495,9 kbps
4.0	141,1 kbps	314,1 kbps	693,4 kbps
4.5	192,0 kbps	443,3 kbps	1220,1 kbps

For example, the minimum required total bitrate for a digital stream carrying audiovisual data, provided that the integral quality is 4.5 points, would be 1220.1 Kbps for a sports report (with a video stream rate of 1058.4 Kbps and an audio stream rate of 161.7 Kbps). At the same time, the minimum total bitrate of the digital data stream required to provide an integral quality level of 4.5 points for a music video and a talking head news story is 443.3 Kbps and 192.0 Kbps, respectively.

As can be seen, the minimum total data rate required to transmit three programs with the same level of integral quality is different, and the difference is quite significant. Therefore, the total bitrate of 1220.1 Kbps is redundant for the transmission of a news program and a music video

while maintaining a given level of integral quality. Adaptive bitrate control of video and audio data streams in accordance with the genre and content of audiovisual programs will achieve a reduction in the amount of transmitted data and the average bandwidth of the communication channel at a given level of integral quality of playback of audiovisual programs.

The proposed method of adaptive control of stream bitrates in accordance with the genre and content of audiovisual programs outwardly resembles the technology of statistical multiplexing, which is currently used in digital television systems. The applied statistical multiplexing technologies allow transmitting a certain number of video services within a fixed bandwidth of a communication channel, allocating a large share of the stream bitrate to channels whose images are highly detailed and dynamic. If at a certain point in time, programs with low dynamics (news like “talking head” or talk shows) are transmitted on all channels included in the multiplex, then they do not require a high bitrate. However, in any case, the entire available bit rate will be allocated to the transmission of these programs, and the bitrate allocated to each of them may be excessive for the transmission of a given program, even at the highest level of quality. When highly dynamic programs are transmitted simultaneously on all channels, the opposite situation may occur. The bitrate allocated to each program may not be sufficient, the quality of audiovisual programs will decline, and the viewer will observe defects and artifacts of compression.

The proposed method is based on taking into account the integral audiovisual quality of experience and considering an audiovisual program as a whole. It's not just about allocating more bit rates to video programs with more complex and dynamic picture and sound. All programs are transmitted with a fixed level of integral quality, and the minimum total bitrate for each individual program, necessary to maintain a given level of quality. The minimum data rate required to provide a given quality level for each program is set taking into account the optimal bitrate allocation between video and audio streams. The proposed method aims to maintain a constant quality level for all transmitted programs while transmitting each program with the minimum bitrate necessary to maintain the established quality level. Therefore, it is aimed at minimizing the amount of transmitted data, which indicates the economic effect of the implementation of the method.

Currently, the leading Russian telecom operators have already launched the Mobile TV service. Depending on the selected service package, more than 200 TV channels are available to users of mobile and portable devices. Available channels can be divided into several groups depending on their specialization: general thematic; news and information; sports; musical channels; channels specializing in educational programs; channels specializing in children's audiences and programs for family viewing; channels specializing in broadcasting movies, series and entertainment channels. The topics of broadcast audiovisual programs depend on the specialization of the TV channel.

In the course of the work, an analysis was carried out for a day (weekday—Wednesday) of the programs of three leading Russian general-thematic channels (“Channel 1”, “Russia 1”, “Russia 2”) included in the standard package of services of the Mobile TV service. The channels are aimed at a large audience, so the TV program covers a wide range with different content. The programs broadcast on TV channels can be conditionally divided into three groups according to the degree of compliance with the audiovisual characteristics of the three test videos used in the work.

The program guide of Channel 1 mainly consists of news releases, all kinds of talk shows and serial films. Programs that are similar in audiovisual characteristics to the “talking head” test

story (short news releases, talk shows) occupy almost 48 % of the broadcast time during the day, which is about 11.5 hours. The broadcasting of programs with medium detail and image dynamics (entertainment shows, melodramas) accounts for about 45 % of the channel's total broadcast time during the day, which is about 11 hours. Thus, the broadcasting of programs similar in audiovisual characteristics to a highly dynamic test sports video is about 7 %, or 1.5 hours per day. It is possible to transmit programs of various content with a constant bit rate of 1220 kbps, which is necessary to provide an integral quality level of 4.5 points for the most critical to the bandwidth of the communication channel, like a highly dynamic sports video. But to transmit a program of the "talking head" type with an integral quality level of 4.5 points, a minimum total data rate of 192 Kbps is required. Consequently, within 11.5 hours a day, audiovisual programs of this type can be transmitted at a bit rate of 192 kbps without loss of quality of experience. Similarly, for broadcasting programs with medium dynamism and image detail in the image for 11 hours a day, a bit rate of 443 kbps can be used. With this control of the bitrate of digital data streams, depending on the content of the video, all audiovisual programs will be transmitted with an integral quality level of 4.5 points, and the average bitrate per day will be 380 kbps. The obtained value of the digital stream bitrate is 69 % (3.2 times) less than the stream bitrate required to transmit the most critical to the bandwidth of the communication channel, a highly dynamic video (1220 kbps). It can be concluded that by controlling the bitrate of digital data streams, depending on the content of audiovisual programs, it is possible to achieve significant savings in the required bandwidth of the communication channel.

Similar data were obtained for the TV channels "Russia 1" and "Russia 2". On the "Russia 1" TV channel, programs that are similar in audiovisual characteristics to the test "music video" occupy 54 % of the broadcast time during the day, which is about 13 hours. About 27 % and 19 %, or 6.5 hours and 4.5 hours per day, respectively, fall on the broadcasting of "talking head" programs and highly dynamic stories. The average digital stream bitrate per day is 525 kbps, which is 57 % less than the bitrate required to transmit the most critical to the bandwidth of the communication channel, a highly dynamic video with an integral quality level of 4.5 points.

The TV channel "Russia 2" was previously a sports channel "Sport", but now it has expanded the subject, including a large number of educational programs, documentaries and series in the program. Broadcasting of highly dynamic programs and programs with medium detail and image dynamics is about 40 % and 50 %, or 10.5 hours and 13.5 hours per day, respectively. The average bit rate per day is 780 kbps, which is 36 % less than the speed required to transmit the most bandwidth-critical highly dynamic video.

If we assume that some subscriber has subscribed to view the analyzed three channels, then the daily average digital stream bitrate for the three channels will be 561 kbps when using adaptive control of video and audio data stream bitrates at the integral quality level of 4.5 points. This value is 54 % less than the speed required to transmit the most bandwidth-critical sports video (1220 Kbps).

There are TV channels where news or talk shows form the basis of broadcasting, such as the "RBC business TV" channel. In this case, it can be assumed that the average data bitrate per day will be comparable to the bitrate required for "talking head" type programs with the integral quality level of 4.5 points. Similarly, there are sports TV channels, such as "Sport 1", where broadcasts of sports competitions form the basis of broadcasting. In this case, it is advisable to use the digital stream bitrate of 1220 Kbps for the transmission of audiovisual data,

which makes it possible to provide the integral quality level of 4.5 points for the most critical to the bandwidth of the communication channel, a highly dynamic sports report.

Similar calculations were performed for the full package of programs provided by the mobile television operator SPB TV, one of the most popular TV services in Russia. With its help, more than 200 popular Russian and foreign TV channels can be viewed both on mobile devices with any software platform and on a computer. The average transmission bitrate per day in this case is 720 kbit/s, which is 41 % less than the bitrate required to transmit the most critical to the bandwidth sports reports with an integral quality level of 4.5 points. This result means a 41 % reduction in the amount of data transmitted per day, which indicates the economic effect of applying overall data bitrate control in accordance with the genre of the audiovisual program and adaptive allocation of the bandwidth of the communication channel between video and audio data streams.

#### **4.5. Conclusions to section 4:**

1. Using the developed mathematical model of subjective perception, quantitative answers are given to an important for practice question about the interaction and mutual influence of video and audio components on each other in the process of forming a feeling of integral quality, when one of the components complements and changes the subjective feeling formed by another component.
2. An analysis of the results obtained allows us to conclude that there are so-called neutrality points. At these points, there is neither a decrease nor an increase in the level of subjectively perceived integral quality in comparison with monomodal estimates of image or sound quality indicators.
3. With the same degree of video compression (video stream bitrate), the video perception estimates obtained by measuring the quality of only the image (monomodal quality) turn out to be lower/higher than in the case when the expert evaluates the quality of the image and sound together (integral quality). The effect of increasing the level of integral quality (that is, the estimates of the perception of video materials obtained by assessing the quality of only the image are lower) is most noticeable in the region of low values of the image quality parameter. The effect of reducing the level of perception of the integral quality is most noticeable at high image quality (4.5 points) and low sound quality (1.5 points).
4. The influence of partial parameters of video and audio series on the perception of quality as a whole depends on the genre and content of television materials. The level of the integral quality of a news story (in relation to the monomodal image quality) increases with a high quality of the sound to a greater extent than the level of the integral quality of a sports report or a music video. In the case of high image quality, the integral quality score increases to a greater extent for a sports report than for a news fragment or a music video.
5. The developed mathematical model of subjective perception can be applied for dynamic, adaptive to the content of broadcast materials, allocation of the channel bandwidth between video and audio streams, which allows maximizing the level of integral quality of playback of multimedia programs under severe technical restrictions on the delivery of television materials, when there is a limited fixed bandwidth of the communication channel.

6. Application of the optimal ratio between the bitrates of the digital stream of audio and video data when transmitting audiovisual programs can significantly increase the level of integral quality. Adaptive allocation of the bandwidth of the communication channel between video and audio data streams in accordance with the genre and content of the audiovisual program allows you to increase the level of integral quality of playback of audiovisual programs (by 2–3 points on the five-point scale, depending on the genre of the program and the stream bitrate) at a given total stream rate transmitted data.
7. The optimal ratio of audio and video bitrates depends on the total bitrate available for the transmission of a multimedia program, and is different for audiovisual programs of different content. The optimal ratio of audio and video bitrates at the low total bit rate of 200 kbps for a sports report is 70:30, while at the total bit rate of 1.5 Mbps, the optimal ratio is 90:10. For a talking head news program with a small bandwidth of 200 kbps, the ratio of 20:80 can be considered optimal.
8. A large share of the bandwidth of the communication channel should be allocated to the semantically more significant component of an audiovisual program. For a “talking head” news program, the quality of the sound accompaniment is of the greatest importance, as a semantically more significant component, therefore, a large share of the communication channel bandwidth should be allocated for transmitting information about sound. In richly detailed and highly dynamic sports reports, a large share of the bandwidth is dedicated to transmitting image information.
9. The developed mathematical model of subjective perception can be used to find the minimum required bandwidth of the communication channel to ensure the integral quality at the level of a given value.
10. The minimum total data bitrate required to transmit three videos of different genres with the same level of integral quality varies significantly. When broadcasting a news program with a quality level of 4.5 points, the total bit rate of 192 kbps will be required, for a music video—443 kbps, for a sports report—1220 kbps.
11. The use of adaptive bitrate control for video and audio data streams in accordance with the genre and content of audiovisual programs makes it possible to achieve a reduction in the amount of transmitted data and the required average bandwidth of the communication channel at a given level of integral quality of playback of audiovisual programs.

## **5. Limits of Energy and Information processing in human brain**

The ideal information technology development increases amount and / or speed of information processing with minimizing of the energy cost. Technologies for compressing video and audio television signals are due to resource limitations, both energy, material and time. Existing technologies for compressing television signals are consistent with the fundamental limitations of human sensory perception—to do more and spend less. These limitations, are the result of evolution evoked by the limited resources to ensure the reception and processing of information. The biological sensory systems and information technology have to compress the amount of information transmitted from the sense organs to the brain for processing.

At the first stage of any of the biological sensor systems the restrictions begin at its input. There are limitations on the perceived frequency band of both video and audio signals. For the human visual system, the visible frequency range is limited in the short-wave region of the electromagnetic spectrum, as ultraviolet light is harmful to all living cells, and the long-wave

limitation region introduces infrared light to suppress thermal radiation in “warm-blooded” animals,

In further processing of sensory information, the mechanism of selective attention introduces limitations from one hand, but from the other it permit to distribute resources (Shapiro editor 2002, Tsotsos et al. 2021; Verghese, Pelli, 1992). We must comment that there are two types of attention, namely general attention, ensuring the body is ready for anxiety, and it mobilize all human resources and very economical the selective attention carrying out by our goal selection. Since it is selective attention that limits the cognitive capabilities of a person. The presented in this chapter television signal compression technologies, used consumer’s selective attention mechanisms.

The mechanism of selective attention ensures the compression of the processed information saves resources, but its side effect is localized, selective inattention blindness. This “blindness” is transient and manifests itself in the fact that we do not notice (see) everything in the visual field and do not hear everything that sounds around us at a given moment of the time. But at some other moments and in different environments, we will detect and recognize this signal. This means that the object observed even with a good signal-to-noise ratio, but in a different context, can be both visible and invisible, conscious and unconscious. Switching and focusing attention depends on many factors, the goals facing the individual, his motivations, and needs of his body. The physiological mechanisms of switching attention when a person perceives audio and video signals is carried out by switching the supply of blood to different parts of the brain that perform different functions. while maintaining an almost constant blood supply to the brain. The temporal processes of information processing clearly demonstrate the limitations introduced by the operation of the selective attention mechanism (Shapiro editor, 2001).

Now we know that the human brain represents only 2–3 % of total body mass, which used 20 % of arterial blood. Investigations of cerebrovascular relationships was started long ago from the investigations Alexander Monro (1783), then he compare the brain and the hand mass and amount of blood supply. According to hydraulic model there was no active control of cerebral blood flow, as cerebral blood flow passively followed blood pressure. Only F.C. Donders (1868) proposed the active regulation of cerebral blood flow in response to cognitive activation an evoked even small local hyperemia as result of brain metabolism neurovascular coupling. Angelo Mosso experimentally demonstrate the regulation of cerebral blood flow in humans. Mosso recorded pulsations of the brain in a patient with skull defects, and introduced the human balanced table that tipped downwards in the head side if the weight of the head was increased during increasing of emotional or intellectual activity. Roy and Sherrington experiments (1890) develop the metabolic hypothesis.

In the end of XX century and first decades of XXI the new technology permit to measure the micro circulation of the blood in the brain. Using oxygen microelectrodes, Profiles of pO<sub>2</sub> between neighboring arterioles and venules were for the first time measured. The data clearly evidenced that O<sub>2</sub> diffusional shunting took place between cortical arterioles and venules, provided they were distanced from each other for not over 80–100 microns (Vovenko et al., 2002). The length of elementary arteriole-venula junction is about the same as neurons ansambel the cortical column diameter (Hubel, Wiesel, 1977), functional elementary module of information processing, coding and transmission in the visual cortex (Glezer, 1995).

The synchronization of this modules work during visual cognitive task is possible to see by using the positron emission tomography–PET, single-photon emission computed tomogra-

phy- SPECT, functional magnetic resonance imaging–MRI for measurements of blood circulation in the local areas of the brain (Sokoloff 1964; 1985). The large-scale brain networks work during cognition are investigated by fMRI and diffusion tractography methods (Bressler, Menon 2010).

General blood circulation and supply of the whole brain, remains almost constant for most of the life of the organism, as shown by both old and new methods of measuring cerebral blood flow ([Fantini, 2016](#); Claassen et al 2021).

An increase in heart rate and an increase in blood pressure also affect blood flow. However, these are extreme conditions for the activity of the organism. The delivery of energy to the brain is the delivery of glucose and oxygen through the blood supply. Intellectual or emotional stress causes a certain rush of blood to the brain, which can be seen if the subject is placed on a balance bar (Field, Inman, 2014) Weighing brain activity with the balance: a contemporary replication of Angelo Mosso's historical experiment only partly support the Roy–Sherrington hypothesis for all the brain (Sándor et al., 2002; Field, Inman, 2014), as it change activation not for all but only in different areas.

The brain, despite the principle of parallel processing of information, cannot solve all problems at once. In order to maintain an approximate constancy of the overall energy supply, the activity of cellular metabolism and blood supply to different brain structures is redistributed. “A neurophysiological nature of BOLD signal in fMRI clearly show that the local increase in the BOLD contrast directly and monotonically reflects local increase of the neural activity” (Logothetis et al., 2001) “. This rise can capture both significant areas of the cortex corresponding with the Brodmann zones, united in large-scale neural networks, and local and local corresponding with the modules of the visual cortex. (Frostig et al., 1990).

The attention switching mechanism ensures the redistribution of neural activity and blood flow in accordance with the need to achieve a particular goal. The brain's self-regulatory mechanisms ensure that less energy is “fed” to neurons that respond to information that is out of our focus when our task becomes more difficult. In the presence of a severe restriction of energy supply, the brain can cope with complex tasks by switching the supply of oxygen and nutrients (“energy”) from one structure to another, changing priorities and focus of attention. A change in the priority of either the auditory or the visual system is clearly manifested, for example, during informational and/or gravitational overloads. Inattentional blindness developed as a result of a long evolutionary development and provided effective problem solving. It does not always solve problems successfully, and then this decision ends tragically.

When attention is switched, changes in blood microcirculation occur in different parts of the brain. These small but significant differences in local blood flow in the brain using fMRI and fNIRS can be recorded indirectly based on changes in the concentration of different forms of hemoglobin (with or without oxygen) during human activity. But new technologies make it possible to study activity directly in neurons, and not indirectly through the bloodstream. Changes in neuronal metabolism were shown using a modified fNIRS technique. The modified method of broadband spectroscopy in the near infrared region made it possible to measure the level of oxidation of cytochrome C oxidase, as an intracellular marker of cellular metabolism (Bruckmaier et al., 2020). The level of cytochrome C-oxidase oxidation in the areas of the visual cortex that respond to stimuli increases with a high perceptual load and decreases with a decrease in the load. Broadband spectroscopy in the near infrared region makes it possible to record the level of oxidation of cytochrome C oxidase in the human

visual cortex. The level of cytochrome C-oxidase oxidation in stimulus-responsive areas of the visual cortex was increased with high perceptual load compared with low, with low skill, while the levels of cytochrome C-oxidase oxidation were reduced with skill acquisition. The redistribution of attention and activation of various structures decreases with the development of the skill of performing the task facing the subject. But the overall metabolic rate of the human brain remains constant. (Bruckmaier et al., 2020). The results of a study of the characteristics of visual search with skill development also showed a decrease in costs (Skuratova et al., 2022). Established an increased level of cellular metabolism associated with the processing of complex tasks, and a decrease in the level with the development of the skill (Derbyshire, 2018).

The processing of information in the brain is limited, and the increased demands on the body with a sharp complication of the task in a difficult environment leads to the need for quick and repeated switching of attention. This condition can lead to fatigue, an increase in the level of errors. During long-term monotonous work, fatigue may also be due to the fact that, due to the opponent relationship of various brain structures, a long-term continuous activation of some brain structures occurs, leading to the “robbing” of others. That is why people intuitively tend to switch activities. Although it is often difficult to organize, for example, during intensive work on the conveyor.

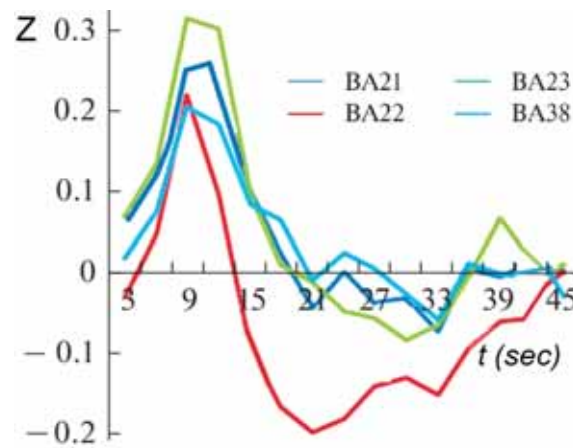
The constancy of nutrition of the whole brain is determined by the principle of least action, since it requires minimization of costs, which leads to a redistribution of functions between neural networks. Based on the interaction of different neural networks, it determines the principle of the opponent organization of physiological processes. Opponent (reciprocal) relationships of cellular activity between neurons of different networks, between neurons and glia, between different large-scale neural networks, between the processing of information from different sensory systems are the general principle of organizing the body's work. Basic, for brain function, the relationship between neurons and glial cells are opponents even with constant blood flow (Shelepin, 1970).

On the figure 26 examples of visual attention work is demonstrated.

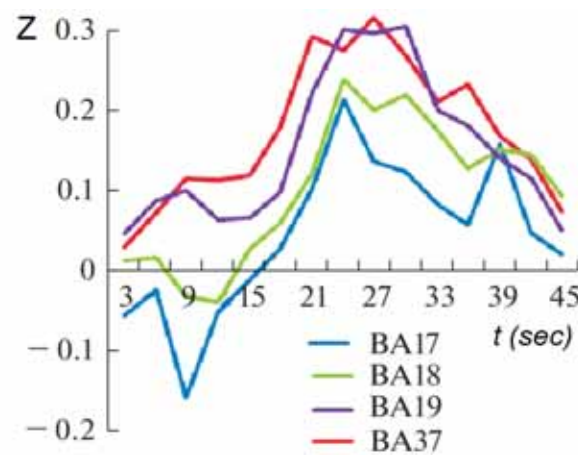
The guessing at 10 % of counter presented and the correct decision at 20 % counter length presented In experiments provided by K. Shelepin and Yu. Shelepin (2020) was used computerized version of Gollin test with dynamic presentation of incomplete images. On the figure 28 is shown only three levels of filling the contour of the image of the test object. The abscise axis (X) the time of presentation is numerically corresponds, to the counter length presented. This correspondent, was obtained by using proper speed of counter grout. On the ordinate (Y) the BOLD signal in relative to the rest amplitude (Z). On the bottom of figure 28 there are examples of three images with 10, 20 and 40 percent (%) of presented contours. If only less than 10 % of the contour is presented it is impossible to recognize the object. At this situation the observer has really free fill choice. For a future experiments it is necessary to mention that the relation the threshold percent of contour strongly depends on the thickness of the contour lines in display pixels to the width of the windows in same display pixels in which pieces of this contour are visible.

Switching the activity of different areas of the brain and the corresponding redistribution of blood supply to them do not solve all the tasks of brain life support. That is why, when tasks become more difficult, blood pressure rises, and the pulse quickens. But these extreme states are beyond the measured work of the operator, which we are discussing.

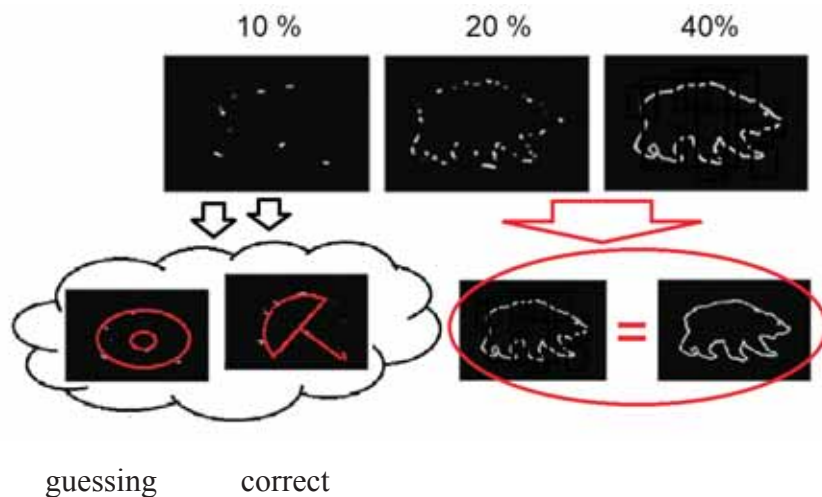
A “Auditory” brain cortex areas BOLD signal to Gollin test



B “Visual” brain cortex areas BOLD signal to Gollin test



C The different visualization of the counter give guessing or the correct decision



**Figure 28.** The Z estimation of the BOLD signal simultaneously changes in different “visual” and “auditory” cortex Brodman areas to the Gollin test presentation during first 45 sec after start. The abscise axis is a time of dynamic grows of percent length of test image counter presentation, and the same time numbers on abscissa are the length of the figure presented in percent. The ordinate axis (Z) the relative to the “rest” intensity of blood flow. (Shelepin & Shelepin, 2020).

There are still limitations in the processing of information. They occur at the level of synapses. The information flow is coordinated with energy costs. Therefore, in the visual system, the flow of information from the retina to the primary visual cortex must be counted not only in bits per second, but also as the number of bits of information transmitted by a certain amount of molecules necessary to ensure synaptic transmission. A model of energetically optimal transmission of information in “serotonin” synapses of thalamic relay neurons has been developed. The evaluation was based both on the entropy of transmission and on the evaluation of mutual information. The transfer entropy, as a non-parametric assessment of the amount of directed transfer of information between two random processes, often turns out to be better for assessing the work of the brain than the mutual information associated with it. But the authors made an assessment by both methods (Conrad & Jolivet, 2020). Mutual information and a Bayesian approach. The Bayesian approach can be used to estimate the mutual information of that distribution. The first work to do this, which also showed how to do Bayesian estimation of many other information-theoretic properties besides mutual information (Wolpert and Wolf, 1995; Hutter and Zaffalon, 2001). A mathematical model was created for the ability to process information in the brain area in terms of neuronal activity, input data storage capacity and afferent information arrival rate based on the Bayesian strategy for optimal decision making (Tongtong Li et al., 2022).

## 6. Conclusions

1. A new method was proposed to conduct subjective assessments of multimodal quality of audiovisual programs, based on the perception of semantic and aesthetic information in television broadcasting. The proposed method extends the boundaries of the usage of standardized methods, applying them to study the effect of cross-modal interaction between the components of an audiovisual program. It allows to carry out experiments on multimodal quality evaluation, when audio and video content interact to produce the perception of quality in digital television systems with data compression.
2. New scale for cross-modal interaction in assessing the quality of experience – integral assessment of the quality of the image and sound was proposed in the study. The scale contains an explicit category of evaluation, equivalent to a zero quality value, and it has the properties of a ratio scale.
3. It was shown that the multimodal quality is content dependent. Quantitative assessments were obtained to describe the influence of genre of audiovisual program on multimodal quality.
4. It is proposed to use level decomposition to develop a mathematical model of multimodal quality assessment of compressed television materials. At the first level the influence of the video and audio bit rate on the picture and sound quality of video sequences respectively was described. At the second level the model describes the influence of the video and audio bit rates on the multimodal audiovisual perceptual quality.
5. A polynomial function of the second degree of two variables – image and sound quality indicators, can be used to describe the integral quality assessment of audiovisual programs at the second level of the mathematical model of subjective perception. The effect of cross-modal interaction is displayed by including a multiplicative term – the product of image and sound quality indicators.
6. The effect of cross-modal interaction between the components of an audiovisual pro-

gram is studied, when one component changes the subjective perception created by the other component and when the information received within one sensory modality affects the perception of information received within another sensory modality, for programs of different genres and different content. Quantitative assessments of exchange relations between auditory and visual modalities have been obtained and the mechanisms of interdependence of perception of audio and video information, for which only a qualitative description was known before, have been found.

7. Practical recommendations are given for adaptive bandwidth allocation between video and audio data streams according to the genre and content of the audiovisual program, which allows increasing the level of integral quality of audiovisual program (by 2–3 points on a five-point scale, depending on the program genre and bit rate rate) at a given total bit rate of the transmission channel.
8. Practical recommendations are given for adaptive control of video and audio bit rates according to the genre and content of audiovisual programs, which allows reducing the amount of transmitted data and the required average bandwidth of the communication channel at a given level of integral quality of audiovisual programs playback (by 35–41 % depending on the quality level).
9. The energy limit for information processing in the brain is the typical problem for all information technologies. The different neuronal nets need more energy during different work. The way to supply different group of brain cells is not only to increase speed of the blood flow but to provide redistribution of the blood flow in the brain by selective attention.

## References:

1. Адлер, Е.П., Маркова, Е.В., Грановский, Ю.В. Планирование эксперимента при поиске оптимальных условий. 2-е изд., перераб. и доп. – М.: Наука, 1976. – 280 с.
2. Бекр, Кеннет, Кэйри, Патрик. Анализ данных с помощью Microsoft Excel.: пер. с англ. – М.: Издательский дом «Вильямс», 2005. – 560 с.
3. Белозерцев, А.В., Гриненко, Е.Н. Мультимодальная оценка качества мультимедийных материалов в системах мобильного телевидения // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. – 2009. – № 4(62). – С. 102–107.
4. Быструшкин, К., Степаненко, Л. Звук ультравысокой четкости. О новых форматах многоканального звука Dolby AC-4 и DTS-UHD. // Журнал «Media Vision». – № 06/46. – 2014 г. – с. 77–80
5. Ван дер Варден, Б.Л. Математическая статистика. Перевод с немецкого Л. Н. Большева; под ред. Н. В. Смирнова. – М.: Изд-во Иностранной литературы, 1960. – 436 с.
6. Вентцель, Е.С. Теория вероятностей. – 4-е изд., стереотипное. – М.: «Наука», 1969. – 576 с.
7. Вентцель, Е.С., Овчаров, Л.А. Теория вероятностей и ее инженерные приложения. – 2-е изд., стер. – М.: Высш. шк., 2000. – 480 с.
8. Воробьев, В.П., Степанов, В.А. Проблемное поле медиаэкологии: опыт демаркации научного направления [Электронный ресурс] // Веснік БДУ. – 2011. Сер. 4. – No 2. <http://www.elib.bsu.by/bitstream/123456789/13892/1/86–90.pdf>
9. Гласман, К. Конференция IBC2006 // Журнал «625». – 2006. – № 10.
10. Гласман, К. Конференция IBC'98: теория и практика цифрового вещания // «625». – 1998. – № 9. – с. 38–44
11. Гласман, К.Ф., Белозерцев, А.В., Гриненко, Е.Н. Распределение пропускной способно-

- сти канала между потоками видео- и аудиоданных в системах вещания на мобильные терминалы // Сборник материалов 8-й Международной конференции «Телевидение: передача и обработка изображений». – СПб.: Изд-во СПбГЭТУ «ЛЭТИ». – 2011
12. Гмурман, В. Е. Теория вероятностей и математическая статистика: учебное пособие для вузов. – 9-е изд., стер. – М.: Высшая Школа, 2003. – 479 с.
  13. Гриненко, Е. Мультимодальное качество мультимедийных программ: Методы оценки // LAP LAMBERT Academic Publishing GmbH & Co. – 2011. – 88 стр. ISBN: 978-3-8433-1809-9
  14. Гриненко, Е.Н., Поддубная, Ю.С., Гласман, К.Ф., Белозерцев, А.В. Оптимальное распределение пропускной способности канала связи между сигналами изображения и звука в системах стереоскопического мобильного телевидения // Материалы и доклады IV международной научно-технической конференции запись и воспроизведение объемных изображений в кинематографе и других отраслях. – М.: МКБК. – 2013. – С. 219–231
  15. Гришин, В.К. Статистические методы анализа и планирование экспериментов. – М.: Издательство Московского университета, 1975. – 146 с.
  16. Джонсон, Н., Лион, Ф. Статистика и планирование эксперимента в технике и науке. Методы обработки данных. Перевод с английского под редакцией канд. техн. Наук Э.К. Лецкого. – М.: Мир, 1980–620 с.
  17. Закс, Л. Статистическое оценивание / Пер. с нем. В.Н. Варыгина; научное редактирование предисловие Ю.П. Адлера и В.Г. Горского. – М.: Статистика, 1976. – 598 с.
  18. Кобзарь, А.И. Прикладная математическая статистика. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
  19. Кухта, М.С. Методология моделирования восприятия визуальной информации: автореферат дис. док.фил.наук: 09.00.08 [Электронный ресурс] – Томск, 2004. – Режим доступа: <http://www.dissercat.com/content/metodologiya-modelirovaniya-vospriyatiya-vizualnoi-informatsii>.
  20. Левченко, А.Е. Формы и методы воздействия СМИ на общественное сознание [Электронный ресурс] // Пресс-Служба. Электрон, дан. – 2005. – № 9. – Режим доступа: <http://www.pr-club.com/PRLib/PRRabots.shtml>.
  21. Минько, А.А. Статистический анализ в Microsoft Office Excel. – М.: Издательский дом «Диалектика», 2004. – 448с.
  22. Митина, О.В. Математические методы в психологии: Практикум. – М.: Аспект Пресс, 2008. – 238 с.
  23. Моль, А. Социодинамика культуры. Пер. с фр. / Вступ. статья, редакция и примечания Б.В. Бирюкова, Р.Х. Зарипова и С.Н. Плотникова. – М.: Прогресс, 1973
  24. Моль, А. Теория информации и эстетическое восприятие. Перевод с французского Б.А. Власюка, Ю.Ф. Кичатова и А.И. Теймана, под редакцией, с послесловием и примечаниями Р.Х. Зарипова, В.В. Иванова, вступительная статья Б.В. Бирюкова, С.Н. Плотникова. – М.: «МИР», 1966. – 352 с.
  25. Морозов, В.П. Искусство и наука общения: невербальная коммуникация. Под общей редакцией члена-корреспондента РАН В.И. Медведева. – М.: ИП РАН, Центр «Искусство и наука», 1998. – 164 с.
  26. Грицанов А.А. Новейший философский словарь. – Мн.: Изд. В.М. Скакун, 1998. – 896 с.
  27. Перегудов, А.Ф., Белозерцев, А.В., Гласман, К.Ф., Гриненко, Е.Н. Мультимодальная оценка качества мультимедийных материалов в системах мобильного телевидения // Сборник материалов Недели науки и творчества. – СПб.: Изд-во СПбГУКиТ. – 2009. – С. 83–85.
  28. Перегудов, А.Ф., Белозерцев, А.В., Гласман, К.Ф., Гриненко, Е.Н. Модель оценки мультимодального качества мультимедийных материалов // Сборник материалов Недели науки и творчества. – СПб.: Изд-во СПбГУКиТ. – 2010. – С. 88–89.
  29. Перегудов, А.Ф., Белозерцев, А.В., Гласман, К.Ф., Гриненко, Е.Н., Мобильное телевидение: мультимодальная оценка качества аудиовизуальных материалов // Сборник материалов 7-й Международной конференции «Телевидение: передача и обработка изобра-

- жений». – СПб.: Изд-во СПбГЭТУ «ЛЭТИ». – 2009. – С. 115–122.
30. Перегудов, А.Ф., Гласман, К.Ф., Белозерцев, А.В., Гриненко, Е.Н. Мульти模альная оценка качества аудиовизуальных материалов в системах мобильного телевидения // «Вопросы радиоэлектроники». Серия «Техника телевидения». – 2008. – Вып. 2. – С. 96–105.
  31. Перегудов, А.Ф., Гласман, К.Ф., Белозерцев, А.В., Гриненко, Е.Н. Оценка качества телевизионных материалов в системах вещания на мобильные терминалы // Научно-технический вестник СПбГУ ИТМО. – 2009. № 04 (62). – С. 102–108.
  32. Перегудов, А.Ф., Гласман, К.Ф., Белозерцев, А.В., Гриненко, Е.Н. Модель оценки мульти模ального качества аудиовизуальных программ, транслируемых на мобильные терминалы // «Вопросы радиоэлектроники». Серия «Техника телевидения». – 2010. – Вып. 1. – С. 21–33.
  33. Поцелуев, С.П. Символическая политика: констелляция понятий для подхода к проблеме // Политические исследования. – 1999. – № 5. – с. 62–71
  34. Шелепин Ю.Е.. К математической интерпретации нейроно-глиальных взаимоотношений. Доклады АН СССР, 1970, том 192, № 3, стр. 698–701.
  35. Шмаков, П.В. Основы цветного и объемного телевидения. – М.: Советское радио, 1954. – 304 с.
  36. Ядов, В.А. Социологические исследование: методология, программа, методы. – 2-е изд., перераб. и доп. – М.: Наука, 1987. – 248 с.
  37. Archer, E.; Park, I.M.; Pillow, J. (2013). “Bayesian and Quasi-Bayesian Estimators for Mutual Information from Discrete Data”. *Entropy*. 15 (12): 1738–1755. Bibcode:2013Entropy..15.1738A. CiteSeerX 10.1.1.294.4690. doi:10.3390/e15051738.
  38. Apteker, R.T., Fisher, J.A., Kisimov, V.S., and Neishlos H. Video Acceptability and Frame Rate // *IEEE Multimedia*. – 1995. – Vol. 2, No. 3. – Pp. 32–40
  39. Bajarin, Ben. Why Only Apple Can Promise a Better Experience to Customers Every Year? [Электронный ресурс] // *Time Tech*. – June 12, 2012. – Режим доступа: <http://techland.time.com/2012/06/12/why-only-apple-can-promise-a-better-experience-to-customers-every-year/>
  40. Barber, P.J. and Laws, J.V. Image quality and video communication // *Proceedings of the IEEE International Symposium*; edited by R.I. Damper, W Hall and J. W. Richards. – London: Pentech Press, 1994. – Pp 163–178.
  41. Bau, N.D. Experience delivery networking: the network ahead [Электронный ресурс] // *Global Telecoms Business*. – 26 September 2011. <http://www.globaltelecomsbusiness.com/article/2907010/Experience-delivery-networking-the-network-ahead.html?ArticleID=2907010>
  42. Belk, R. W. Situational variables and consumer behavior // *Journal of Consumer Research Inc.* – 1975. – Volume 3, Issue 3. – Pp. 157–164
  43. Bench, A., Sasse, M. A., and H. de Meer. Of packets and people: A User-centered Approach to Quality of Service // *Proceeding of 8th International Workshop on Quality of Service*. – 2000. – Pp. 189–197.
  44. Bermant, Robert I., and Welch, Robert B. Effect of Degree of Separation of Visual-Auditory and Eye Position upon Spatial Interaction of Vision and Audition // *Perceptual and Motor Skills*. – 1976. – Vol. 43. – Pp. 487–493.
  45. Bevan, N. What is the difference between the purpose of usability and user experience evaluation methods? [Электронный ресурс] // UXEM’09 Workshop, INTERACT 2009, Uppsala, Sweden. – Режим доступа: <http://www.nigelbevan.com/cart.htm>
  46. Bonneel, N., SinVri, C, Vijuid-Deltiuni. I., DrettakLs, CI. Bimodal perception of audio-visual material properties for virtual environments // *ACM Transactions on Applied Perception (TAP)*. – 2010. – Volume 7 Issue 1, January 2010 Article No.1
  47. Bouch, A, & Sasse, M.A. Network Quality of Service—An Integrated Perspective. [Электронный ресурс] // *Proceedings of the Tenth International Conference on Rewriting Techniques and Applications, Vancouver*. – 1999. – Режим доступа: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.43.2094&rep=rep1&type=pdf>

48. Bouch, A. and Sasse, M.A. The case for predictable media quality in networked multimedia applications. [Электронный ресурс] // Proceedings of the ACM/SPIE Multimedia Computing and Networking. –USA, 2000. –Режим доступа: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.730>
49. Bruckmaier M., Tachtsidis I., Phan Ph., and Lavie N.. Attention and Capacity Limits in Perception: A Cellular Metabolism Account // Journal of Neuroscience 26 August 2020, 40 (35) 6801–6811; DOI: <https://doi.org/10.1523/JNEUROSCI.2368-19.2020>
50. Burrows, David, and Solomon, Barry A. Parallel scanning of auditory and visual information // Memory & Cognition. – 1975. – Vol. 3, No. 4. – Pp. 416–420.
51. Claassen J., Thijssen D., Panerai R., Faraci F. // Regulation of cerebral blood flow in humans: physiology and clinical implications of autoregulation// Review Physiol Rev. 2021 Oct 1;101(4):1487–1559. doi: 10.1152/physrev.00022.2020. Epub 2021 Mar 26. PMID: 33769101 PMCID: PMC8576366 DOI: 10.1152/physrev.00022.2020
52. Claypool, M., Tanner, J. The effects of jitter on the perceptual quality of video // Proceedings of the seventh ACM international conference on Multimedia (Part 2). – 1999. – Pp 115–118.
53. Colavita, Francis B. Human sensory dominance // Perception & Psychophysics. – 1974. – Vol. 16. – Pp. 409–412.
54. Conrad M. & Jolivet R.B. // Comparative performance of mutual information and transfer entropy for analyzing the balance of information flow and energy consumption at synapses (Short title: A comparison of information theory measures for spike trains). bioRxiv preprint 2020. DOI: 10.1101/2020.06.01.127399 //LicenseCC BY4.0
55. De Koning, T.C.M, Veldhoven, P., Knoche, H. Kooij, R.E. Of MOS and men: bridging the gap between objective and subjective quality measurements in mobile TV. // Proceedings of Multimedia on Mobile Devices 2007, IS&T/SPIE Symposium on Electronic Imaging, Vol 6507. – 2007.
56. EBU Document Tech. 3286–1997. Assessment methods for the subjective evaluation of the quality of sound programme material–Music [Электронный ресурс]. – 1997. –: <https://tech.ebu.ch/docs/tech/tech3286.pdf>
57. EBU Technical Recommendation R90–2000. The subjective evaluation of the quality of sound programme material, 2000.
58. ETSI EG 202 534. Human Factors (HF): Guidelines for real-time person-to-person communication services. ETSI Guide. –European Telecommunications Standards Institute, 2007.
59. ETSI TR102 535. Human Factors (HF): Guidelines for real-time person-to-person communication services; future requirements. ETSI Technical Report. –European Telecommunications Standards Institute, 2007.
60. ETSI TS133 220 (3GPP SA 3). Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Generic Authentication Architecture (GAA); Generic bootstrapping architecture (3GPP TS33.220). Technical Specification. – European Telecommunications Standards Institute, 2009
61. Bressler S.L., Menon V. Large-scale brain networks in cognition: emerging methods and principles //Trends in cognitive sciences. – 2010. – T. 14. – № . 6. – C. 277–290.
62. Fantini, S.; Sassaroli, A.; Tgavalekos, K.; Kornbluthb J.; // Cerebral blood flow and autoregulation: current measurement techniques and prospects for noninvasive optical methods// Neurophotonics. 2016 Jul; 3(3): 031411. PMCID: PMC4914489// PMID: 27403447// <https://doi.org/10.1101/2020.06.01.127399> // Published online 2016 Jun 21. doi: 10.1117/1.NPh.3.3.031411
63. Field, D. T.; Inman, L.A. (February 2014). “Weighing brain activity with the balance: a contemporary replication of Angelo Mosso’s historical experiment”. Brain. 137 (2): 634–639. doi:10.1093/brain/awt352. PMID24408614.
64. Fong, A.C.M. and Hui, S. C. Low-bandwidth Internet streaming of multimedia lectures // Engineering Science and Education Journal. –2001. –Volume 10, Issue 6. –Pp. 212–218
65. Forlizzi, J., Battarbee, K. Understanding experience in interactive systems // Proc. of the

- 2004 conference on Designing interactive systems: processes, practices, methods, and techniques.–2004.–Pp. 261–268
66. Frostig R., Lieke E., Ts'o D., Grinvald A.. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high-resolution optical imaging of intrinsic signals //Proc. Natl. Acad. Sci. USA, (Neurobiology), Vol.87, pp.6082–6086, August 1990
  67. Derbyshire K. How The Brain Saves Energy By Doing Less // Semiconductor Engineering. MARCH 15TH, 2018 <https://semiengineering.com/how-the-brain-saves-energy-by-doing-less/>
  68. Ganglbauer, E., Schrammel, J., Deutsch, S., & Tscheligi, M. Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. [Электронный ресурс] // Workshop on Measuring Mobile Emotions: Measuring the Impossible? at Mobile-HCI 2009.–Режим доступа: [http://jschrammel.files.wordpress.com/2010/10/interact\\_workshop\\_uxevaluation-eg.pdf](http://jschrammel.files.wordpress.com/2010/10/interact_workshop_uxevaluation-eg.pdf)
  69. Ghinea, G. and Thomas, J. P. Impact of Protocol Stacks on Quality of Perception // Proc. IEEE International Conference on Multimedia and Expo.– 2000.–Vol. 2.–Pp. 847–850.
  70. Ghinea, G., and Thomas, J.P. QoS Impact on User Perception and Understanding of multimedia Video Clips // Proceedings of the 6th ACM International Conference on Multimedia '98.– 1998.–Pp. 49–54
  71. Glezer V.D.. Vision and Mind. Modeling Mental Functions. ISBN9781138986787, 1995
  72. Gregg, Lee W., and Brogden, W.J. The Effect of Simultaneous Visual Stimulation on Absolute Auditory Sensitivity // Journal of Experimental Psychology.– 1952.–Vol. 43.–Pp. 179–186.
  73. Grinenko, E., Poddubnaya, J., Glasman, K., Belozertsev, A. Optimal Bitrate Allocation between Audio and Video in Mobile 3DTV Application // Conference Publication International Broadcasting Convention 2011.
  74. Grinenko, Evgenia, Glasman, Konstantin, Belozertsev, Alexander. Content-adaptive Bitrate Reduction in Mobile Multimedia Applications [Электронный ресурс] // Proceedings of 4th IEEE International Conference on Consumer Electronics–Berlin.–2014.– 1 электрон. опт. диск (CD-ROM)
  75. Gulliver S.R. and Ghinea G. How Level and Type of Deafness Affects User Perception of Multimedia Video Clips // Universal Access in the Information Society.– 2003.– Vol. 2, (4).– Pp. 374–386
  76. Gulliver, S. R., Ghinea, G. Defining the users perception of distributed multimedia quality // ACM Trans. on Multimedia Computing, Communications and Applications (TOMM).– 2006.– Volume 2 Issue 4.–Pp 241–257
  77. Hands, D. S. A Basic Multimedia Quality Model // IEEE Transactions on Multimedia. 2004.– Vol.6, No.6.–Pp. 806–816.
  78. Harman, G. The Intrinsic Quality of Experience // Philosophical Perspectives.– 1990.– Vol. 4 Action Theory and Philosophy of Mind.–Pp 31–52.
  79. Hassenzahl, M., Tractinsky, N. User Experience—a Research Agenda [Электронный ресурс] // Behaviour and Information Technology.–2006.–Vol. 25, No. 2.–Pp. 91–97.–Режим доступа: <https://ccrma.stanford.edu/~sleitman/UserExperienceAResearchAgenda.pdf>
  80. Hiltunen, M., Laukka, M., Luomala, J. Mobile User Experience.–Finland: Edita Publishing Inc. IT Press, 2002.– 214 pp.
  81. Howard, I. P., and Templeton, W. B. Human Spatial Orientation.–New York: Wiley, 1966.–541 pp.
  82. Hubel, D.H., Wiesel, T.N.: Functional architecture of macaque monkey visual cortex. Proc. R. Soc. London, Ser. B198, 1–59 (1977).
  83. Hutter, M. Zaffalon M. (2005). Distribution of Mutual Information from Complete and Incomplete Data// Computational Statistics & Data Analysis 48(3):633–657 DOI:10.1016/j.csda.2004.03.010
  84. Hulusic, V., Aranha, M. and Chalmers, Alan. The influence of cross-modal interaction on

- perceived rendering quality thresholds // Journal of WSCG2008.–2008.–Vol.16.–Pp. 41–48.
85. ITU-R REPORT BT.1082–1: Studies toward the unification of picture assessment methodology, 1990.
  86. ITU-R WP6C–Programme Production and Quality Assessment. Contribution: 97. New quality measures for the next generation of broadcasting and audio content delivery [Электронный ресурс], 2012.–Режим доступа: <http://www.itu.int/md/R12-WP6C-C-0097/en>
  87. ITU-T. Technical paper. Tutorial on Objective perceptual assessment of Video Quality: Full Reference Television, 2004.
  88. Jones, Bill, and Kabanoff, Boris. Eye movements in auditory space perception // Perception & Psychophysics.– 1975.–Vol. 17, No. 3.–Pp. 241–245.
  89. Jumisko-Pyykkö, S., Häkkinen J. I would like see the face and at least hear the voice”: Effects of Screen Size and Audio-video Bitrate Ratio on Perception of Quality in Mobile Television. [Электронный ресурс] // The 4th European Interactive TV Conference, Athens, Greece, 2006.–Режим доступа: <http://s3.amazonaws.com/publicationslist.org/data/jukka.hakkinen/ref-44/Jumisko%202006.pdf>
  90. Jumisko-Pyykkö, S., Häkkinen, J. Evaluation of subjective video quality of mobile devices // Proceeding MULTIMEDIA ‘05 Proceedings of the 13th annual ACM international conference on Multimedia.– 2005.–Pp 535–538
  91. Jumisko-Pyykkö, S., Hakkinen, J., Nyman, G. Experienced Quality Factors–Qualitative Evaluation Approach to Audiovisual Quality// Proceedings of the IS&T/SPIE’s International Symposium on Electronic Imaging: Multimedia on Mobile Devices 2007.
  92. Jumisko-Pyykkö, S., Hannuksela, M. M. Does Context Matter in Quality Evaluation of Mobile Television? [Электронный ресурс] // MobileHCI 2008. Proceedings of the 10th international conference on Human computer interaction with mobile devices and services.– 2008.–Pp 63–72.–Режим доступа: <http://www.students.tut.fi/~jumisko/publications/p63-jumisko-pyykko.pdf>.
  93. Jumisko-Pyykkö, S., Reiter, U., Weigel, C. Produced Quality is Not Perceived Quality–A Qualitative Approach to Overall Audiovisual Quality. // Proceedings of the 1st International Conference on 3DTV 3DTV–CON ‘07.– 2007.–Pp. 1–4
  94. Jumisko-Pyykkö, S., Strohmeier, D. Report on research methodologies for the experiments [Электронный ресурс] // Technical Report Project Mobile 3DTV.– 2008.–Режим доступа: [http://sp.cs.tut.fi/mobile3dtv/results/tech/D4.2\\_Mobile3dtv\\_v1.0.pdf](http://sp.cs.tut.fi/mobile3dtv/results/tech/D4.2_Mobile3dtv_v1.0.pdf)
  95. Jumisko-Pyykkö, S., Utriainen, T. User-centered quality of experience: Is mobile 3D video good enough in the actual context of use? [Электронный ресурс] // Proceedings of Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics: VPQM 2010.–Режим доступа: [http://satujumiskopyykkko.net/wb/media/conferences/publications/JumiskoPyykkoEtUtriainen\\_User-Centered\\_IsMobile\\_2010.pdf](http://satujumiskopyykkko.net/wb/media/conferences/publications/JumiskoPyykkoEtUtriainen_User-Centered_IsMobile_2010.pdf)
  96. Jumisko-Pyykkö, S., Weitzel, M., Strohmeier, D. Designing for user experience: what to expect from mobile 3d tv and video? // Proceedings of the 1st international conference on Designing interactive user experiences for TV and video.– 2008.–Pp. 183–192
  97. Kawalek, J. A User Perspective for QoS Management [Электронный ресурс] // Proceedings of the QoS Workshop aligned with the 3rd International Conference on Intelligence in Broadband Services and Network.– 1995.–Режим доступа: [http://www.researchgate.net/publication/243526642\\_A\\_user\\_perspective\\_for\\_qos\\_management](http://www.researchgate.net/publication/243526642_A_user_perspective_for_qos_management)
  98. Kies J.K., Williges, R.C. and Rosson, M.B. Evaluating desktop video conferencing for distance learning // Computers and Education.– 1997.–Volume 28, Number 2.–Pp. 79–91
  99. Kilkki, K. Quality of Experience in Communications Ecosystem // Journal of Universal Computer Science.– 2008.–Vol. 14, no. 5.–Pp 615–624.
  100. Knoche, H., McCarthy, J. D., Sasse, M.A. Reading the fine print: the effect of text legibility on perceived video quality in mobile TV // Proceedings of the 14th ACM International Conference on Multimedia.– 2006.–Pp. 727–730.
  101. Knoche, Hendrik and Sasse, M. Angela. Getting the Big Picture on Small Screens: Quality of

- Experience in Mobile TV [Электронный ресурс] // In Ahmad, A. M.A. & Ibrahim, I.K. (eds.) Multimedia Transcoding in Mobile and Wireless Networks.– 2008.–Chapter 3.–Pp.31–46. Режим доступа: <http://www.irma-international.org/viewtitle/27194/>
102. Knudsen, E. I., and Brainard, M. S. Creating a Unified Representation of Visual and Auditory Space in the Brain // *Annual Review of Neuroscience*.– 1995.– Vol. 18.–Pp.19–43.
  103. Kontola, K., Makinen, J. M., Taleb, A., Bruhn, S., Besette, B., Salami, R. AMR-WB+: Low Bit Rate Audio Coding for Mobile Multimedia. // *IEEE International Symposium on Broad-band Multimedia Systems and Broadcasting Preliminary Program*, March 2006.
  104. Kravkov, S. V. The Influence of Sound upon the Light and Color Sensibility of the Eye // *Acta Ophthalmologica Scandinavica*.– 1936.– Vol. 14.– Pp. 348–360.
  105. Larsson, P., Västfjäll, D., Kleiner, M. Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments. // *Proceedings of the 2001 International Conference on Auditory Display*, 2001
  106. Lee, Jong-Seok; De Simone, Francesca; Ebrahimi, Touradj. Influence of audio-visual attention on perceived quality of standard definition multimedia content // *Proceedings of the International Workshop on Quality of Multimedia Experience*.– 2009.– Pp. 13–18.
  107. Li-yuan, L., Wen-an, Z., Jun-de, S. The Research of Quality of Experience Evaluation Method in Pervasive Computing Environment [Электронный ресурс] // 1st International Symposium on Pervasive Computing and Applications.– 2006.–Режим доступа: [http://ieeexplore.ieee.org/xpl/tocresult.jsp?sortType%3Dasc\\_p\\_Sequence%26filter%3DAND%28p\\_IS\\_Number%3A4079022%29%26rowsPerPage%3D100&pageNumber=2](http://ieeexplore.ieee.org/xpl/tocresult.jsp?sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A4079022%29%26rowsPerPage%3D100&pageNumber=2)
  108. Logothetis, N. K.; Pauls, Jon; Auguth, M.; Trinath, T.; Oeltermann, A. (July 2001). “A neurophysiological investigation of the basis of the BOLD signal in fMRI”. *Nature*. 412 (6843): 150–157. doi:10.1038/35084005. PMID11449264. Our results show unequivocally that a spatially localized increase in the BOLD contrast directly and monotonically reflects an increase in neural activity. Lopez, D., Gonzalez, F., Bellido, L., Alonso, A. Adaptive multimedia streaming over IP based on customer oriented metrics [Электронный ресурс] // *International Symposium on Computer Networks*.– 2006.–Режим доступа: <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=11037>
  109. Manin Y., Marcolli M.. Homotopy. Theoretic and Categorical Models of Neural Information Networks Computer Science. Computer Science > Logic in Computer Science// 24 Aug 2022, v2, arXiv:2006.15136// arXiv:2006.15136v2
  110. Marks, Lawrence E. On Cross-Modal Similarity: The Perceptual Structure of Pitch, Loudness, and Brightness // *Journal of Experimental Psychology: Human Perception and Performance*.–1989.– Vol. 15, No. 3.–Pp. 586–602.
  111. Martinussen, Inda. A cross-modal enhanced multi-dimensional visualization: master thesis / Inda Martinussen.–Aalborg University Copenhagen, 2007.– 68 pp.
  112. Masry, M., Hemami, S. S., Rohaly, A. M., and Osberger, W. Subjective quality evaluation of low bit rate video // *Human vision and electronic imaging VI–Proceedings of the SPIE*. Edited by Rogowitz B.E. and Pappas T.N. 2001.–Pp. 195–195
  113. MediaCoder–a free universal media transcoder // [www.mediacoder.net](http://www.mediacoder.net).
  114. Mitagvaria Nodar P., Bicher Haim I. // *Cerebral Blood Flow Regulation (Nova Biomedical)*. 10 April 2012// [file:///C:/Users/123/Downloads/Cerebral blood flow regulation.pdf](file:///C:/Users/123/Downloads/Cerebral%20blood%20flow%20regulation.pdf)
  115. Moore, M. S., Foley, J. M. and Mitra, S.K. Comparison of the detectability and annoyance value of embedded MPEG-2 artifacts of different type, size, and duration // *Proc. SPIE Conference on Human Vision and Electronic Imaging*.–2001.–VI.–Pp.90–101
  116. Moorsel, V.A. Metrics for the Internet Age: Quality of Experience and Quality of Business [Электронный ресурс] // *HP Labs Technical Report HPL-2001–179*.–July 2001.–Режим доступа: <http://www.hpl.hp.com/techreports/2001/HPL-2001-179.html>.
  117. Muhammad, N., Chiavelli, D., Soldani, D., and Li, M. QoS and QoE Management in UMTS Cellular System [Электронный ресурс] // *John Wiley & Sons*.– 2006.–Режим доступа: <http://onlinelibrary.wiley.com/doi/10.1002/9780470034057.ch1/summary>

118. Neisser, U. *Cognition and Reality, Principles and Implications of Cognitive Psychology*. – San Francisco: W.H. Freeman & Company, 1976. – 230 pp.
119. Neuman, W. Russell. *Beyond HDTV: Exploring Subjective Responses to very High Definition Television* // MIT Media Library, Massachusetts Institute of Technology, Cambridge, Massachusetts. – 1990.
120. Neumann, W. R., Crigler, A. N., Bove, V. M. *Television Sound and Viewer Perceptions* // Proc. of Audio Eng. Soc. Ninth International Conference. – 1991. – Pp. 101–104.
121. Oksman, V., Noppari E., Tammela, A., Miikinen, M. and Ollikainen, V. *News in Mobiles. Comparing text, audio and video* [Электронный ресурс] // VTT Tiedotteite-Research Notes 2375, 2007. – Режим доступа: <http://www.vtt.fi/inf/pdf/tiedotteet/2007/T2375.pdf>
122. *Optimal Experience: Psychological Studies of Flow in Consciousness*. In Csikszentmihalyi, M. and Csikszentmihalyi, I. (eds.). – Cambridge University Press, 1992. – 416 pp.
123. Peregudov Alexander, Grinenko Eugenie, Glasman Konstantin, Belozertsev Alexander. *An Audiovisual Quality Model of Compressed Television Materials for Portable and Mobile Multimedia Applications* // Proceedings of IEEE ISCE2010.
124. Peregudov, A., Glasman, K., Belozertsev, A. and Grinenko, E. *A multimodal quality model for content adaptive bitrate allocation in mobile multimedia applications* [Электронный ресурс] // Conference Publication International Broadcasting Convention 2009. – 2009. – 1 электрон. опт. диск (CD-ROM).
125. Peregudov, A., Glasman, K., Belozertsev, A. and Grinenko, E. *Multimodal Quality Assessments of Compressed Television Materials for Portable and Mobile Devices*. // SMPTE Motion Imaging Journal. January/February 2010. – Pp.42–51.
126. Peregudov, A., Glasman, K., Belozertsev, A., and Grinenko, E. *Multimodal Quality Assessments of Compressed Television Materials for Portable and Mobile Devices* // Conference Publication International Broadcasting Convention 2008. – Pp. 396–404.
127. Peregudov, Alexander, Grinenko, Eugenie, Glasman, Konstantin, Belozertsev, Alexander. *An Audiovisual Quality Model of Compressed Television Materials for Portable and Mobile Multimedia Applications* [Электронный ресурс] // Proceedings of 14th IEEE International Symposium on Consumer Electronics. – 2010. – 1 электрон. опт. диск (CD-ROM).
128. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T. *Human-Computer Interaction*. Addison-Wesley, Wokingham (1994). Chapter 29–34. 1994.
129. Procter, R., Hartwood, M., McKinlay, A. and Gallacher, S. *An investigation of the influence of network quality of service on the effectiveness of multimedia communication* // Proc. of the international ACM SIGGROUP conference on supporting group work. – 1999. – Pp. 160–168
130. Ragot, Richard, Cave, Christian, and Fano, Michel. *Reciprocal effects of visual and auditory stimuli in a spatial compatibility situation* // Bulletin of the Psychonomic Society. – 1988. – Vol. 26, No. 4. – Pp. 350–352.
131. Rec ITU-R BT.1129–2. *Subjective Assessment of Standard Definition Digital Television (SDTV) Systems*, 1998.
132. Rec ITU-R BT.1788–0. *Methodology for the subjective assessment of video quality in multimedia applications*, 2007
133. Rec ITU-R BT.2021. *Subjective methods for the assessment of stereoscopic 3DTV systems*, 2012.
134. Rec. ITU-R BS.1116–1. *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*, 1997.
135. Rec. ITU-R BS.1284–1. *General methods for the subjective assessment of sound quality*, 1997–2003.
136. Rec. ITU-R BT.1128–2. *Subjective Assessment of Conventional Television Systems*, 1997.
137. Rec. ITU-R BT.1438. *Subjective assessment of stereoscopic television pictures*, 2000.
138. Rec. ITU-R BT.500–13. *Methodology for the subjective assessment of the quality of television pictures*, 2012.
139. Rec. ITU-R BT.710–4. *Subjective Assessment Methods for Image Quality in High-Definition*

- Television, 1998.
140. Rec. ITU-T P. 800. Methods for subjective determination of transmission quality, 1996.
  141. Rec. ITU-T P. 800.1. Mean Opinion Score (MOS) terminology, 2006.
  142. Rec. ITU-T P. 800.2. Mean Opinion Score (MOS) interpretation and reporting, 2013.
  143. Rec. ITU-T P. 805. Subjective evaluation of conversational quality, 2007
  144. Rec. ITU-T P. 910. Subjective video quality assessment methods for multimedia applications, 1999.
  145. Rec. ITU-T P. 911. Subjective audiovisual quality assessment methods for multimedia applications, 1998.
  146. Reichl, P. From 'Quality-of-Service' and 'Quality-of-Design' to 'Quality-of-Experience': A Holistic View on Future Interactive Telecommunication Services [Электронный ресурс] // Conference publications 15th International Conference on Software, Telecommunications and Computer Networks. –2007. –Режим доступа: <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4446051>
  147. Reiter, U., Köhler, T. Criteria for the Subjective Assessment of Bimodal Perception in Interactive AV Application Systems // Proceedings of the Ninth International Symposium on Consumer Electronics. – 2005. – Pp. 186–192
  148. Ries, M., Puglia, R., Tebaldi, T., Nemethova, O., Rup, M. Audiovisual Quality Estimation for Mobile Streaming Services // Proc. the 2nd International Symposium on Wireless Communication Systems. Siena, 2005.
  149. Rimell, A. Hollier, M. The significance of cross-modal interaction in audio-visual quality perception [Электронный ресурс] // 1999 IEEE Third Workshop on Multimedia Signal Processing. – 1999. – Pp. 509–514. –Режим доступа: [http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=794134&punumber%3D6434%26sortType%3Dasc\\_p\\_Sequence%26filter%3DAND%28p\\_IS\\_Number%3A17174%29%26pageNumber%3D3](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=794134&punumber%3D6434%26sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A17174%29%26pageNumber%3D3)
  150. Roto, V. Obrist, M., Väänänen-Vainio-Mattila, K. User Experience Evaluation Methods in Academic and Industrial Contexts // Workshop on User Experience Evaluation Methods, in conjunction with Interact'09 conference. –2009. –Режим доступа: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.1764&rep=rep1&type=pdf>
  151. Roto, V. Web Browsing on Mobile Phones – Characteristics of User Experience. Doctoral dissertation, TKK Dissertations 49. (<http://lib.tkk.fi/Diss/2006/isbn9512284707/isbn9512284707.pdf>). // Helsinki University of Technology, Finland, 2006.
  152. Roy C. S., Sherrington C. S. // On the Regulation of the Blood-supply of the Brain *J Physiol.* 1890 Jan; 11(1–2): [85]–108, 158–7–158–17. PMCID: PMC1514242, PMID: [16991945](https://pubmed.ncbi.nlm.nih.gov/16991945/), doi: [10.1113/jphysiol.1890.sp000321](https://doi.org/10.1113/jphysiol.1890.sp000321)
  153. Sándor P., Benyó Z., Erdős B., Lacza Z., Komjáti K. // The Roy–Sherrington hypothesis: facts and surmises // International Congress Series. Volume 1235, July 2002, Pages 325–335. [https://doi.org/10.1016/S0531-5131\(02\)00201-7](https://doi.org/10.1016/S0531-5131(02)00201-7)
  154. Shapiro K., (Editor) The limits of attention. Temporal constraints in human information processing. 2001, Oxford university Press. ISBN0 19 850516 7 Hbk; ISBN0 19 850515 9 Pbk
  155. Shelepin Yu. E., Mathematical interpretation of neuron glial interaction. // Proceedings of the USSR Academy of Sciences, 1970, vol. 192, N3, pp. 698–701. // (Шелепин Ю. Е.. К математической интерпретации нейроно-глиальных взаимоотношений. Доклады АН СССР, 1970, том 192, № 3, стр. 698–701).
  156. Shelepin Yu. E., Fokin V. A., Harauzov A. K. Pronin, S. V., and Chikhman V. N. // Location of the Decision Making Centre during Image Shape Perception // ISSN00124966, Proceedings of the Russian Academy of Sciences (Doklady, Russian Academy of Science) Biological Sciences (Series), 2009, Vol. 429, pp. 1–3. © Pleiades Publishing, Ltd., 2009.
  157. Shelepin, Yu. E., Kharauzov, A. K., Vakhrameeva, O. A., Zhukova, O. V., Pronin, S. V., Tsvetkov, O. V., Skuratova, K. A., Shelepin, E. Yu. (2021) Unconscious visual signals and involuntary human reactions. Integrative Physiology, vol. 2, no. 4, pp. 352–377. <https://www.doi.org/10.33910/2687-1270-2021-2-4-352-377>

158. Shelepin, K.Y., Shelepin, Y.E. Rearrangement of the Activity of Neural Networks in the Human Brain on Reaching the Recognition Threshold for Fragmented Images. *Neurosci Behav Physi* (2021).
159. Skuratova K.A., Shelepin E. Yu., and Yarovaya N.P. // Optical search and visual expertise// *J. Opt. Technol.* 88(12), 700–705 (2021)
160. Sokoloff L: Cerebral blood flow measured with radioisotopes. In *Dynamic Clinical Studies with Radioisotopes*, US Atomic Energy Comission report June 1964
161. Sokoloff L: Application of quantitative autoradiography to the measurement of biochemical processes in vivo, in positron emission tomography. In: Reivich M, Alavi A, (eds). Alan R. Liss, Inc, New York, NY, 1985.
162. Stein, Barry E., and Meredith, M. Alex. *The Merging of the Senses* // The MIT Press, Cambridge, Massachusetts.– 1993.
163. Steinmetz R. Human perception of jitter and media synchronization. // *IEEE Journal on Selected Areas in Communications* Volume:14, Issue: 1, Jan 1996, Pp. 61–72.
164. Storms, R. L. and Zyda, M.J. Interactions in perceived quality of auditory-visual displays // *Presence*.–2000.–vol. 9.–Pp. 557–580.
165. Storms, R. L. *Auditory-Visual Cross-Modal Perception Phenomena*: Ph.D. thesis in Computer Science.–Naval Postgraduate School, Sept. 1998.
166. Strohmeier, D. Jumisko-Pyykkö, S. and Kunze, K. Open Profiling of Quality: A Mixed Method Approach to Understanding Multimodal Quality Perception [Электронный ресурс] // *Advances in Multimedia*.–2010.–Режим доступа: [downloads.hindawi.com/journals/am/2010/658980.pdf](http://downloads.hindawi.com/journals/am/2010/658980.pdf)
167. Tamminen, S., Oulasvirta, A., Toiskallio, K., Kankainen, A. Understanding mobile contexts [Электронный ресурс] // *Personal and Ubiquitous Computing*.– 2004.– Volume 8 Issue 2.– Pp. 135–143.–Режим доступа: [http://www.hiit.fi/u/oulasvir/scipubs/UMC\\_Puc2004\\_Oulasvirta.pdf](http://www.hiit.fi/u/oulasvir/scipubs/UMC_Puc2004_Oulasvirta.pdf)
168. Thurstone, L. L., Chave, E.F. *The Measurement of Attitudes. A Psychophysical Method and Some Experiments with a Scale for Measuring Attitude toward Church*. 7th ed.–Chicago: University of Chicago Press.–1929.
169. Thurstone, Louis L. A Method of Scaling Psychological and Educational Tests [Электронный ресурс] // *Journal of Educational Psychology*.– 1925.–16.–Pp.433–451.–Режим доступа: [https://www.brocku.ca/MeadProject/Thurstone/Thurstone\\_1925b.html](https://www.brocku.ca/MeadProject/Thurstone/Thurstone_1925b.html)
170. Thurstone, Louis L. Rank Order as a Psychophysical Method [Электронный ресурс] // *Journal of Educational Psychology*.– 1931.–14.–Pp. 187–201.–Режим доступа: [https://www.brocku.ca/MeadProject/Thurstone/Thurstone\\_1931a.html](https://www.brocku.ca/MeadProject/Thurstone/Thurstone_1931a.html)
171. Tongtong Li, Yu Zheng, Zhe Wang, David C. Zhu, Jian Ren, Taosheng Liu & Karl Friston// Brain information processing capacity modeling <https://www.nature.com> // February 2022, *Scientific Reports* 12(1) // DOI:10.1038/s41598-022-05870-z,
172. Tsotsos, J. K., Abid, O., Kotseruba, I., Solbach, M. D. (2021) On the control of attentional processes in vision. *Cortex*, vol. 137, no. 5, pp. 305–329. <https://doi.org/10.1016/j.cortex.2021.01.001>
173. Turner, P., Turner, S., McGregor, I. Listening, Corporeality and Presence // *Proceedings of the 10th Annual International Workshop on Presence*.– 2007.–pp. 43–49.
174. Utriainen, T., Jumisko-Pyykko, S. Experienced audiovisual quality for mobile 3D television. [Электронный ресурс] // *3DTV-Conference: The True Vision–Capture, Transmission and Display of 3D Video (3DTV-CON)*.– 2010.–Pp. 1–4.–[http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5503122&punumber%3D5503122%26sortType%-3Dasc\\_p\\_Sequence%26filter%3DAND%28p\\_IS\\_Number%3A5506201%29%26pageNumber%3D2&pageNumber=3](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5503122&punumber%3D5503122%26sortType%-3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A5506201%29%26pageNumber%3D2&pageNumber=3).
175. Väänänen-Vainio-Mattila, K., Segerståhl, K. A Tool for Evaluating Service User eXperience (ServUX): Development of a Modular Questionnaire. [Электронный ресурс] // *User Experience Evaluation Methods in Product Development (UXEM'09)*. Workshop in Interact'09

- conference.– 2009.– Phttp://www.cs.tut.fi/~kaisavvm/UXEM\_ServUX\_questionnaire\_V1.0\_KVVM\_KS\_update\_300609.pdf
176. Verghese, P., Pelli, D. G. (1992) The information capacity of visual attention. *Vision Research*, V. 32, 5, 983–995. [https://doi.org/10.1016/0042-6989\(92\)90040-p](https://doi.org/10.1016/0042-6989(92)90040-p)
  177. Verscheure, O., Frossard, P. and Hamdi, M. Joint impact of MPEG-2 encoding rate and ATM cell losses on video quality [Электронный ресурс] // Proceedings of Global Telecommunications Conference. The Bridge to Global Integration. IEEE.–1998.–Volume 1.–Pp. 71–76. <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6295>
  178. Vleeschouwer, C. De, Delmot, T., Marichal, X., Macq, B. A Fuzzy Logic System for Content-Based Bitrate Allocation [Электронный ресурс] // SP: IP (10).– 1997.–No 1–3.–Pp 115–141. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.17.3650&rep=rep1&type=pdf>
  179. Vovenko E.P., Sokolova LB., Loshchagin O.V.. Oxygen diffusion through a wall of the rat brain cortex venule in breathing with pure oxygen. Oxygen diffusion through the venule walls in the rat cerebral cortex during breathing with pure oxygen// *Russian Physiological Journal*. V. 88. N3. P. 372–380. 2002// <https://pubmed.ncbi.nlm.nih.gov/12013732/>
  180. Waltl, M., Timmerer, C., Hellwagner, H. A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects. [Электронный ресурс] // Proc. 1st International Workshop on Quality of Multimedia Experience.– 2009.–Режим доступа: [http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5235203&punumber%3D5235203%26sortType%3Dasc\\_p\\_Sequence%26filter%3DAND%28p\\_IS\\_Number%3A5246936%29%26pageNumber%3D2&pageNumber=1](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5235203&punumber%3D5235203%26sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A5246936%29%26pageNumber%3D2&pageNumber=1)
  181. Wang, Y., Claypool, M., and Zuo, Z. An empirical study of RealVideo performance across the internet // Proc. of the First ACM SIGCOMM Workshop on Internet Measurement: ACM Press.– 2001.–Pp. 295–309
  182. Wang, Y., Lv, K., Huang, R. et al. (2020) Glance and focus: A dynamic approach to reducing spatial redundancy in image classification. In: 34th Conference on Neural Information Processing Systems (NeurIPS2020).[Online]. Available at: <https://arxiv.org/pdf/2010.05300.pdf> (accessed 01.10.2021).
  183. Watson, A., Sasse, M.A. Measuring perceived quality of speech and video in multimedia conferencing applications // Proceedings of the sixth ACM international conference on Multimedia.– 1998.–Pp. 55–60.
  184. Wikstrand, G. Improving user comprehension and entertainment in wireless streaming media [Электронный ресурс] // Licentiate thesis.–Umea: Umea University, 2003.– 47 Pp.– Режим доступа: <http://umu.diva-portal.org/smash/get/diva2:145035/FULLTEXT01>
  185. Wikstrand, G., and Eriksson, S. Football Animation for Mobile Phones // Proc. of NordiCHI.– 2002.–Pp. 255–258
  186. Williams, A.A. and Arnold, G.M. The use of free-choice profiling for the evaluation of commercial ports // *J. Sci. Food Agric.*, 35, 558–568, 1984.
  187. Wilson G.M. and Sasse M.A. Do users always know what’s good for them? Utilising physiological responses to assess media quality // Proceedings of HCI2000. Edited by McDonald S., Waern Y. and Cockton.–London: G. Springer-Verlag.–, 2000.–Pp 327–341.
  188. Winkler, S.; Symmetricom, San Jose, CA; Mohandas, P. The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics // *Broadcasting, IEEE Transactions*.– 2008.–Volume 54, Issue 3.–Pp. 660–668.–Режим доступа: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4550731>
  189. Winkler, S. Visual fidelity and perceived quality: Toward comprehensive metrics // B. E Rogowitz, T.N. Pappas (Eds.). *Human Vision and Electronic Imaging VI*–Proceedings of SPIE.– 2001.–Pp. 114–125
  190. Winkler, S., Faller, Ch. Maximizing audiovisual quality at low bitrates // Proc. Workshop on Video Processing and Quality Metrics for Consumer Electronics. Scottsdale, AZ, 2005.
  191. Winklerand, S., Faller, C. Audiovisual quality evaluation of low-bitrate video // Proc. SPIE

- Human Vision and Electronic Imaging. – 2005. – vol.5666. – Pp. 139–148.
192. Wolpert D. H. and Wolf D. R. Estimating functions of probability distributions from a finite set of samples// *Phys. Rev. E*52, 6841; 1-December 1995; Erratum *Phys. Rev. E*54, 6973 (1996)
  193. Yamagishi, Kazuhisa, Tominaga, Toshiko, Hayashi, Takanori, and Takahashi, Akira. Objective Quality Evaluation Model for Videophone Services [Электронный ресурс] // *NTT Technical Review*. – 2007. – Vol. 5, No. 6. – Режим доступа: <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200706le2.html>
  194. Zaman, B. Introducing a Pairwise Comparison Scale for UX Evaluations with Preschoolers // *IFIP Conference on Human-Computer Interaction – INTERACT*. – 2009. – Pp. 634–637.
  195. Zhukova, O. V., Malakhova, E. Yu., Shelepin, Yu. E. (2019) La Gioconda and the indeterminacy of smile recognition by a person and by an artificial neural network. *Journal of Optical Technology*, vol. 86, no. 11, pp. 706–715. <https://doi.org/10.1364/JOT.86.000706References>

## **Chapter 6. Genomics of cognitive functions: from molecular instability to mental disability**

**Iourov I. Y.**

*Yurov's Laboratory of Molecular Genetics and Cytogenomics of the Brain, Mental Health Research Center, Moscow, Russia*

*Vorsanova's Laboratory of Molecular Cytogenetics of Neuropsychiatric Diseases, Veltischev Research and Clinical Institute for Pediatrics of the Pirogov Russian National Research Medical University, Moscow, Russia*

*Russian Medical Academy of Continuous Postgraduate Education, Moscow, Russia*

### **Abstract**

Genomic studies of human cognition repeatedly demonstrate significant contribution of genome variability to the development of cognitive functions. Genetic defects are systematically associated with cognitive deficits in a wide spectrum of brain diseases. However, there is a type of abnormal genome behavior—genome/chromosome instability,—which requires a special focus in the cognitive genomics context. Despite data on causative role of chromosome and genome instability in non-cancerous brain dysfunction, cognitive genomics rarely addresses molecular instability causing chaotic genome behavior in the diseased brain. Here, an overview of the impact of genome and chromosome instability in the brain (neurogenomic instability) on cognitive functions is provided.

### **Introduction**

The development of the human brain is highly dependent on genomic constitution of neural cells. Genomic variability affecting neurodevelopmental pathways is the essential source for brain diseases (Kelley, Paşca, 2022). Thus, it is not surprising that cognitive functions are affected when genomic pathology alters brain development and functions. Currently, almost all types of pathological genomic changes (from single-gene mutations to chromosomal abnormalities) have been associated with neurological and psychiatric diseases exhibiting cognitive deficits (Nokelainen, Flint, 2002; Corvin et al 2012; de la Fuente et al 2021; Iourov et al 2022). These findings have allowed to propose an emerging field of biomedicine and neuroscience referred to as cognitive genetics or cognitive genomics (Corvin et al 2012; Berto et al 2020; Fitzgerald et al 2020).

Cognitive genomics systematically brings insights into understanding of relationships between molecular processes and brain functioning. Discoveries in this field has provided us with knowledge about ontogenetic (ontogenomic) changes affecting brain functioning (Yurov et al 2010; Iourov et al 2019b, 2021b; Dai, Guo, 2021), genomic pathways to cognitive deficits (Iourov et al 2019a; Exposito-Alonso, Rico, 2022) and pharmacogenomic aspects of treating the altered cognition (Cacabelos et al 2012). In total, the idea that genomic analyses are required for uncovering molecular processes shaping the cognition has become dominant (Konopka, Geschwind, 2010; Fitzgerald et al 2020). Accordingly, cognitive deficits have been selected as an important target for studies in medical genomics.

Taking into account numerous reviews describing molecular genetics of diseases associated with cognitive deficits (Nokelainen, Flint, 2002; Konopka, Geschwind, 2010; Corvin et al 2012; Doherty, Owen, 2014; Berto et al 2020; Fitzgerald et al 2020), I have selected genome

and chromosome instability affecting the brain as a prime focus of the review. The underlying idea of this communication is based on a series of studies reporting on recent analyses demonstrating this type of genome behavior as a mechanism for brain dysfunction. Surprisingly, genome and chromosome instability is rarely addressed in the cognitive genomics context. Thus, the communication's focus on genome and chromosome instability seems to be appropriate.

### **Genome and chromosome instability in the human brain**

Genome and chromosome instability is able to be a molecular/cellular mechanism for brain diseases. This becomes even more significant when brain cells are affected. Tissue-specific mosaicism is an important source for human morbidity including brain disorders (Iourov et al 2019b, 2022). Genomic uniqueness of single neuronal cells determines the functional activity, homeostasis and cellular fate (Gupta et al 2020). Actually, the overwhelming majority of genomic studies are performed using blood samples, because these are easily available comparing with samples acquired from other tissues. Obviously, the human brain represents a difficult target for direct genomic analysis. However, an appreciable amount of neurogenomic studies (analysis of neural cell genomes) using post-mortem brain samples have shown that somatic mosaicism for chromosomal abnormalities, single-gene mutations, copy number variations (CNVs), genome/chromosome instability and other types of genomic and epigenomic changes are associated with a wide spectrum of non-cancerous brain disorders. Data on somatic mosaicism and genome instability in the diseased brain have been recently reviewed in details (Iourov et al 2022). These data have underlain new fields of medical genetics/genomics and neuroscience, i.e. neurocytogenetics or neurocytogenomics and somatic brain genetics. The studies in these fields have demonstrated that a wide spectrum of brain diseases accompanied by cognitive deficits is associated with brain-specific somatic genomic variations.

Psychiatric disorders associated with cognitive deficits (autism and schizophrenia) have been systematically reported to demonstrate brain-specific genomic variations manifested as single-gene mutations, CNV and chromosomal abnormalities. The commonest type of genomic changes affecting the brain is aneuploidy (loss or gain of chromosome(s) in a cell) affecting small but significant cellular populations. Actually, somatic aneuploidy is considered as a subtype of chromosome instability, which profoundly affects cellular homeostasis and leads to cell death under certain circumstances (Yurov et al 2009; Tiganov et al 2012; D’Gama, Walsh 2018; Potter et al 2019; Iourov et al 2019b, 2021a, 2022). The latter is suggested to be molecular and cellular pathway to neurodegeneration. Indeed, early- and late-onset neurodegenerative diseases demonstrate high rates of chromosomal and genomic instability in the affected brain. Aneuploidy and other types of genomic and chromosomal instability are detectable in the brain of individuals suffering from Alzheimer’s disease, Parkinson disease, Niemann-Pick disease (type C1), ataxia telangiectasia, and frontotemporal lobar degeneration (Yurov et al 2009, 2019; Iourov et al 2013, 2019b, 2022; Leija-Salazar et al 2018; Zhang, Vijg, 2018; Potter et al 2019; Vorsanova et al 2020). The impact of chromosome instability on cellular homeostasis is highly appreciable and has the potential to cause cell death. The model of chromosome instability involvement in neurodegeneration suggests that cells with abnormal genomes are susceptible to programmed cell death, which is triggered by environmental or endogenous factors producing, thereby, a kind of a self-protection effect under natural selection. Unfortunately, the amount of neuronal cells affected by the instability, which are cleared by this “protection”, is generally high enough to result in a neurodegenerative process threatening patient’s life (Iourov et al 2013; Yurov et al 2019; Vorsanova et al 2020). The last but not the least disease group in the list of non-cancerous pathology associated with genomic

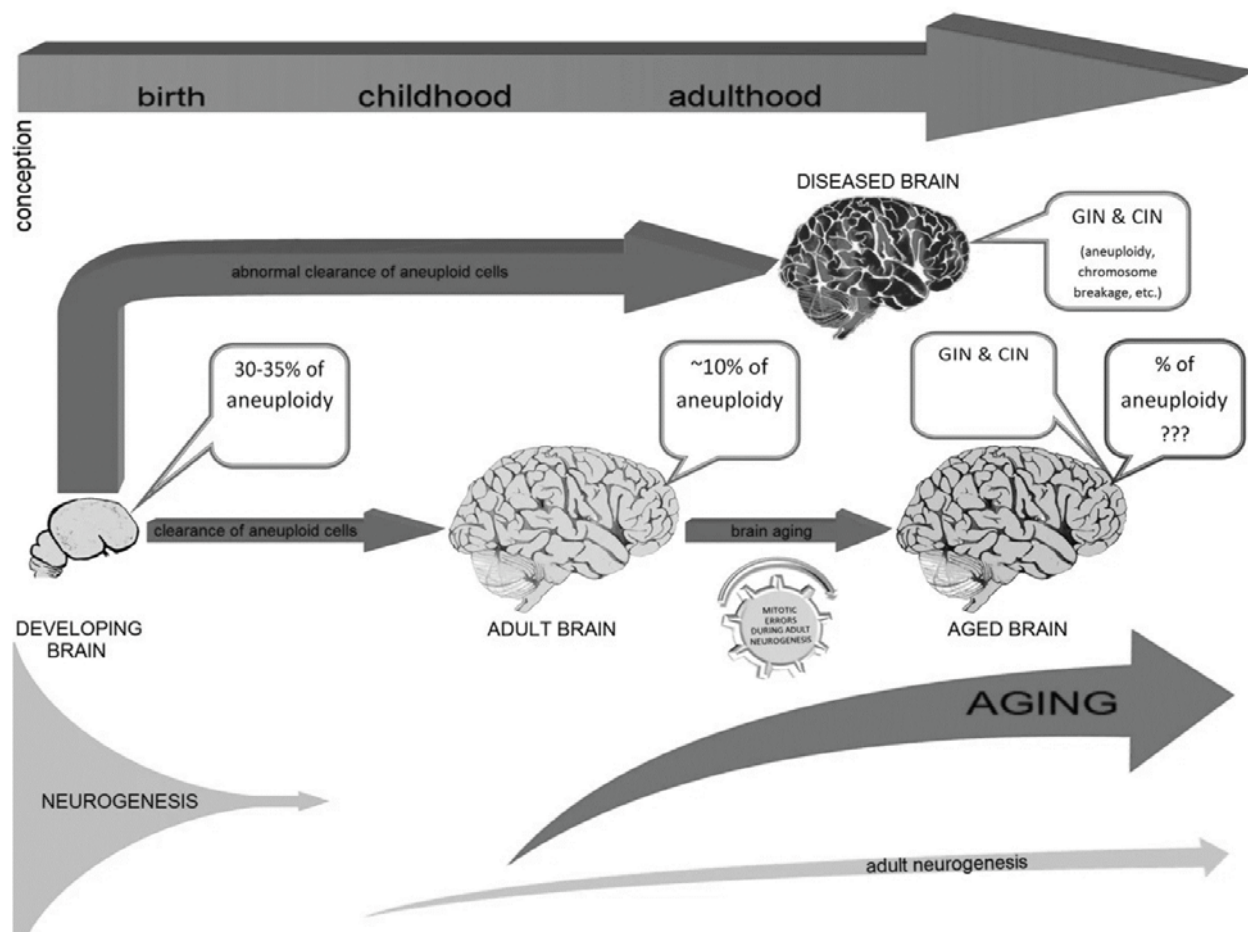
changes in the brain is epilepsy (Ye et al 2019). Unfortunately, specific analysis of chromosomal and genomic instability in the epilepsy brain has not been systematically performed. Hopefully, future studies would fill this gap in our neurogenomic knowledge.

Chromosome and genome instability is a mechanism for abnormal processes in the diseased brain. Alterations to cell cycle regulation, chromosomal segregation, DNA repair and replication, programmed cell death and other pathways involved in safeguarding genome stability are observed in the diseased brain. The mechanism for somatic mosaicism and chromosome/genome instability implicating the aforementioned processes is the most probable (Iourov et al 2013, 2020, 2021b, 2022; Alt et al 2018; Yurov et al 2009, 2019). Pathway-based analysis of abnormal molecular and cellular processes has given insights into the origins of molecular instability producing chaotic genome behavior in the diseased brain (Iourov et al 2020, 2021a; Nelson et al 2022). It is highly likely that high rates of chromosome and genome instability are generated during prenatal development of the central nervous system. Moreover, developmental chromosome instability is abundant in the developing human brain (for review, see Yurov et al 2010; Iourov et al 2021a, 2021b, 2022). Alterations to pathways, which are required for clearance of abnormal cells (programmed cell death, cell cycle regulation) and/or for DNA repair (DNA replication), should lead to the persistence of cells with chromosomal and genomic aberrations in the brain after birth. Postnatal brain-specific instability, in its turn, affects multiple molecular and cellular processes, which shape human cognition.

It is noteworthy that rates of somatic mosaicism and chromosomal/genomic instability increase during ontogeny. Brain aging is associated with slight, but significant increasing of neural cells affected by chromosomal and/or genomic aberrations (Yurov et al 2009, 2010; Zhang, Vijg, 2018; Potter et al 2019; Iourov et al 2019b, 2021b). Since human aging is associated with brain deterioration associated with cognitive deficits (Elliott et al 2021), the contribution of chromosome and genome instability to the altered cognition has been proposed (Iourov et al 2021b). These observations clearly indicate that human cognition is dependent on rates of genome and chromosome instability even in unaffected population through ontogeny. Ontogenetic aspects of chromosome and genome instability in the human brain are schematically depicted by figure 1.

Environmental effects are not usually considered when genomic causes of cognitive dysfunctions are described. However, somatic cell genomics of brain dysfunction is unable to leave aside genetic-environmental interactions. These may affect neuronal homeostasis either by changing the genomic landscape of single neuronal cells (i.e. causing genomic (chromosomal) aberrations) or by changing the functional landscape of single neuronal cells (alterations to genome organization at chromosomal or chromatin level affecting DNA repair and transcriptional activity of cellular genomes) (Iourov et al 2013; Rajarajan et al 2016; Fernando et al 2020). The latter type of instability has not been profoundly addressed. Taking into account the availability of technologies for studying the functional landscape of single neuronal cells, future studies will certainly reveal the way how such kind of molecular instability affects human cognitive functions.

According to neurogenomic studies of the human brain in health and disease, one may conclude that genome and chromosome instability does affect cognitive functions. Aging-related changes in cognitive functions seem to be mediated by chromosome/genome instability in cells of the central nervous system. Once origins and consequences of the molecular instability are unveiled, successful therapeutic interventions in devastating neurological and psychiatric disorders are available.



**Figure 1.** Schematic representation of ontogenetic aspects of chromosome and genome instability in the human brain. The developing human brain (12–15 weeks' gestation) exhibit 30–35 % of cells affected by the instability, which are formed during neurogenesis (prenatal brain development). This process becomes exhausted soon after birth. At later developmental stage, adult neurogenesis starts, being, however, significantly less productive in terms of the amount of cells formed. Abnormal clearance of aneuploid cells leads to postnatal brain diseases, which are featured by GIN and CIN confined to the brain. Some of these diseases are associated with accelerated aging (i. e. Alzheimer's disease and ataxia-telangiectasia). Normal brain development leads to decrease of aneuploidy rates. The presence of abnormal cells in the brain from the early prenatal development to the late ontogeny is hypothesized to give rise to genome and chromosome instability in the brain of elderly individuals. Mitotic errors during adult neurogenesis can also produce aneuploid cells throughout aging. (Copyright © Yurov et al 2009; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License, <http://creativecommons.org/licenses/by/2.0>)

### Molecular instability in the light of mental disability: opening the black box

Chromosome and genome instability is suggested to have a broad spectrum of origins. Pathway-based analyses (Iourov et al 2019a, 2021a; Nelson et al 2022; Savojardo et al 2022) have shown that processes regulating cell cycle, chromosomal segregation, DNA repair and replication, programmed cell death and general pathways safeguarding genome stability are repeatedly altered in neurological and psychiatric diseases. More specifically, alterations are

observed in following pathways: mTOR, PI3K-Akt, p53, PTEN, and MAPK (Iourov et al 2021a). These pathways are required to grant genome stability. Therefore, it is highly likely that susceptibility to neuronal genome instability exists in individuals suffering from devastating brain diseases exhibiting cognitive deficits. It is more likely that such susceptibility results from gene mutations, CNVs and chromosome abnormalities, which affects genes implicated in the aforementioned pathways.

A recent study has demonstrated that the susceptibility to molecular instability may be produced by mutational burden causing alterations to safeguarding genome stability via saturation in genomic changes encompassing genes involved in corresponding pathways. The instability occurred due to these pathway alterations and CNVs or chromosomal imbalances has progressive and dynamic nature and is associated with a wide spectrum of neurodevelopmental diseases and cognitive deficits. In that study, we termed such kind of chromosomal instability as chromohelkosis (from the Greek words chromosome ulceration) and the model of the genome instability as the “cytogenomic theory of everything”. Since the model is able to explain numerous phenomena medical genomics or genomics of cognition, it mimics similar theory in physics (Iourov et al 2020). Thus, the model which suggests the existence of susceptibility to genome instability resulted from inherited or sporadic mutations affecting specific molecular and cellular pathways may be a valuable theoretical tool for addressing cellular processes and phenotypical outcomes in morbid conditions associated with genome and chromosome instability.

The analysis of causes and consequences of genome variability producing cognitive deficits in brain disorders may be described using the well-known model of black box. The underlying idea of the model is referred to the existence of a black box, in which enigmatic processes occur. However, we are able to observe data entering the box and processes/phenomena exiting the box. In the medical genomics context, genomic data enters the black box and phenotype or clinical outcome are at the exit of the box. Processes, which are required for a genomic change to produce the specific phenotype, are enigmatic, i.e. hidden in the black box. Nonetheless, to uncover the mechanism of brain disorder and cognitive deficits in an individual, we need to know the whole sequence of pathogenetic events. Once the knowledge is acquired, new opportunities for treatment and rehabilitation of individuals suffering from brain diseases are available. The introduction of genome and chromosome instability in the pathogenetic cascade of brain disorders causing cognitive deficits seems to help in uncovering processes occurring in the black box. In fact, generation and propagation of molecular instability is one of the most important processes inside the black box. Originating from mutational burden causing alterations to corresponding pathways and genetic-environmental interactions, the instability, which naturally occurs during prenatal neurodevelopment, persists after birth. The increase of instability rates adversely affects brain functioning by genomic changes involving genes required for cognition. During ontogeny, the rates of instability may increase leading to worsening of the condition and cognitive performance. In fact, the black box hides chaotic genome behavior and chaotic processes occurring in neuronal cells, which produce, in its turn, chaotic mental behavior or mental disability.

According to the previous observations (Yurov et al 2010, 2019; Iourov et al 2013, 2021a, 2021b, 2022; Potter et al 2019; Vorsanova et al 2020; Ye et al 2020; Dai, Guo, 2021), the sequence of events leading to genome and chromosome instability in the brain and producing cognitive deficits is as follows:

- 1) Susceptibility of neural genomes to instability and defective changes in programmed

cell death pathway are either inherited (parents have genomic changes, which are unable to cause this susceptibility, but the combination of parental genomic changes inherited by their infant are able to cause it) or sporadically generated in parental germ cells.

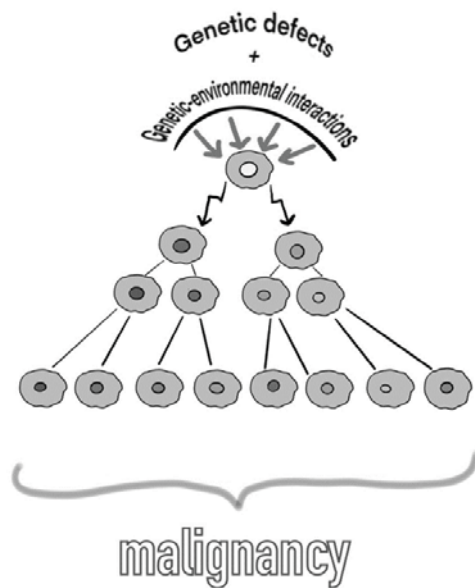
- 2) Chromosomal and genome instability is naturally generated during early brain development affecting up to 30–35 % cells.
- 3) Cells affected by the instability are not cleared due to defects in programmed cell death pathway; instability rates increase because of neural genome susceptibility to the instability.
- 4) Chromosomal and genomic chaos is observed.
- 5) Chromosomal and genomic instability in the diseased brain alters genomic loci containing genes implicated in molecular and cellular pathways required for performing cognitive functions.
- 6) The persistence of instability and alterations to molecular and cellular pathways required for proper brain functioning increases the impact of chromosomal and genomic chaos on the central nervous system.
- 7) Mental disability (mental chaos) is observed.

Chaotic genome behavior has various adverse effects on cellular populations. It may produce malignancy or cause progressive cell death, e.g. neurodegeneration, when cells of the central nervous system are affected (Yurov et al 2019; Ye et al 2020). There are underlying differences between cancerous and non-cancerous (brain-specific) chromosome (genome) instability. Figure 2 describes the model for chromosomal instability mediating cancer and neurodegeneration.

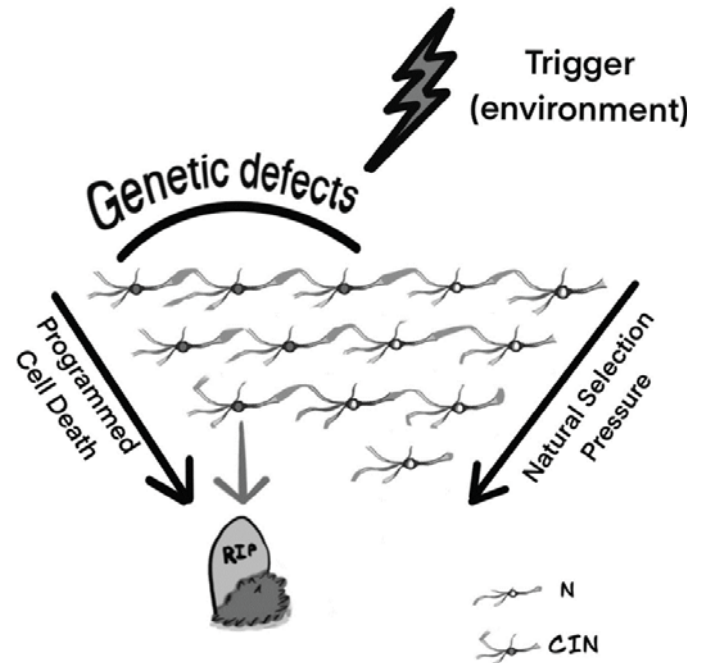
As one may see, chaotic genome behavior manifesting as chromosomal instability possesses various effects. Chromosomal instability in the brain is likely to result in neurodegeneration, when programmed cell death pathways are altered. It is to note that neurodegenerative diseases demonstrate striking cognitive deficits (Fitzgerald et al 2020). On the other hand, the instability may affect pathways of neural cell functioning producing cognitive deficits associated with intellectual disability, autism, epilepsy, schizophrenia. In other words, molecular instability producing chaotic genome behavior results in mental disability manifesting as cognitive deficits (i.e. mental chaotic behavior). It is important to keep in mind that genome and chromosome instability is an element of pathogenetic cascade resulting in brain diseases or altered cognition. Factually, genome and chromosome instability is a missing link between mutations affecting all the cells of an individual (gene and chromosomal mutations, CNVs), genetic-environmental interactions and phenotypic manifestations of brain diseases (Iourov et al 2013, 2021a, 2022). Therefore, the black box mentioned previously becomes less enigmatic due to the application of genome/chromosome instability model to brain disorders associated with cognitive deficits.

Finally, the overlap between psychiatric disorders, a major problem for medical genomics and psychiatric genetics (Doherty, Owen, 2014; Romero et al 2022), may be explained by the chromosome/genome instability in the affected brain. Chromosomal and genomic instability may have similar (or even identical) causes, but the effect is generally different in each affected individual inasmuch as the instability is able to alter any molecular or cellular process in a spontaneous fashion. Thus, molecular instability may alter “autistic pathways” in one case, in contrast to another case, which demonstrate molecular instability altering “neurodegenerative pathways”. Certainly, this suggestion requires further comparative studies.

## A Cancer



## B Neurodegeneration



**Figure 2.** Theoretical model for chromosomal instability mediating (A) cancer and (B) neurodegeneration. (A) Genetic defects and genetic-environmental interactions may cause chromosomal/genomic changes, which produce chromosomal instability; alternatively, cell populations may adapt to aneuploidy and chromosomal instability evolving to a cell population with a fitness advantage. Cells affected by chromosomal instability and tolerating deteriorating effects of chromosomal instability on cellular homeostasis are able to evolve clonally to produce malignancy. (B) Chromosomal instability and/or somatic mosaicism affecting a significant proportion of cells interacting with environmental triggers may result into progressive neuronal cell loss (neurodegeneration) under natural selection pressure and through the programmed cell death (N, normal neurons; CIN, neuronal cell affected by chromosomal instability). The model is based on the observations of chromosomal instability in the neurodegenerating brain and cancers. (Copyright © Yurov et al 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution License, CC BY)

### Concluding remarks

A variety of views on genomic mechanisms for human cognition are united in acknowledging the fact that genomic variations shape cognitive functions in health and disease. However, decades of studying complex brain diseases associated with cognitive deficits have not brought complemented insights into the mechanisms for genomic origins of brain functioning including the cognitive functions. Moreover, molecular and cellular diversity which shapes the cognition at the single-neuron level is incompletely understood. Accordingly, one may suppose that there are elements (missing links) in molecular and cellular pathways to human cognitive functions, which we are, as yet, unable to highlight. Chromosome or genome instability is a promising candidate representing an important part of pathogenic cascades producing mental disability in the widest sense (Yurov et al., 2010; Tiganov et al 2012; Iourov et al 2021a).

Cognitive deficit is an integrated part of pathogenesis in neurodevelopmental, neuropsychiatric and neurodegenerative diseases caused by genetic defects (genomic variations). Therefore,

to get further insights into the development of normal and altered cognition, one has to focus on molecular and cellular processes in brain cells shaped by the genomic variations (Berto et al 2020; Fitzgerald et al 2020). Numerous studies mentioned here demonstrate significant contribution of somatic mosaicism, genome and chromosome instability to the etiology of brain disorders. Thus, there is a need to integrate these data into current understanding of molecular processes underlying mental disability. In fact, genome and chromosome instability in the brain or neurogenomic instability represent a key element of the pathogenic cascades of brain diseases exhibiting cognitive deficits. Originating from heritable or sporadic mutations in genes implicated in genome safeguarding pathways, neurogenomic instability affects a wide spectrum of processes in neuronal and glial cells. As a result, each affected cell demonstrates altered functioning. Affecting large populations of neural cells, this molecular instability generally leads to mental disability (inability). This sequence of molecular and cellular events represents a highly likely mechanism for cognitive deficits in brain disorders. Alternatively, as shown by genomic analyses of the developing human brain (Iourov et al 2019b; Costantino et al 2021), the molecular instability may be a mechanism for natural selection in neural cell populations, when affected cells are cleared by programmed cell death.

To this end, it is noteworthy that studies of neurogenomic instability evidence that chaotic genome behavior mediates chaotic processes in neuronal cells, which are likely to result in altered cognition or mental disability (Iourov et al 2020; Ye et al 2020). Consequently, one may propose that further theoretic studies dedicated to interplay between molecular instability and mental inability should apply “cytogenomic theory of everything” (for more details, see Iourov et al 2020) and/or chaos theory to provide a sophisticated view on genomic causes of cognitive deficits.

This communication is dedicated to professors Yuri B. Yurov and Svetlana G. Vorsanova, two outstanding researchers in the field of medical genomics, cytogenetics, cytogenomics and chromosome biology. Neurocytogenetics (neurocytogenomics) and somatic brain genetics have been emerged largely due to their continuous efforts. Their longstanding research in the field of brain genetics has been accompanied by numerous discoveries, among which brain-specific aneuploidy and chromosome instability causing cognitive deficits. Moreover, their original research provided us with technological basis for studying genome variations in brain tissues and deciphering the outcome of genomic changes. For more details about these brilliant and prodigious scientists, one may address the papers describing their walk of life and original ideas (Iourov, Vorsanova, 2018; Iourov, 2022).

*Author is supported by the Government Assignment of the Russian Ministry of Science and Higher Education, Assignment no. AAAA-A19-119040490101-6 and by the Government Assignment of the Russian Ministry of Health, Assignment no. 121031000238-1.*

## References

1. Alt FW, Schwer B. DNA double-strand breaks as drivers of neural genomic change, function, and disease. *DNA Repair (Amst)*. 2018. 71: 158–163. <https://doi.org/10.1016/j.dnarep.2018.08.019>
2. Berto S, Liu Y, Konopka G. Genomics at cellular resolution: insights into cognitive disorders and their evolution. *Hum Mol Genet*. 2020. 29(R1): R1-R9. <https://doi.org/10.1093/hmg/ddaa117>
3. Cacabelos R, Martinez-Bouza R, Carril JC, Fernandez-Novoa L, Lombardi V, Carrera I, Corzo L, McKay A. Genomics and pharmacogenomics of brain disorders. *Curr Pharm Biotechnol*. 2012. 13(5): 674–725. <https://doi.org/10.2174/138920112799857576>

4. Costantino I, Nicodemus J, Chun J. Genomic mosaicism formed by somatic variation in the aging and diseased brain. *Genes (Basel)*. 2021. 12(7): 1071. <https://doi.org/10.3390/genes12071071>
5. Corvin A, Donohoe G, Hargreaves A, Gallagher L, Gill M. The cognitive genetics of neuropsychiatric disorders. *Curr Top Behav Neurosci*. 2012. 12: 579–613. [https://doi.org/10.1007/7854\\_2011\\_188](https://doi.org/10.1007/7854_2011_188)
6. Dai X, Guo X. Decoding and rejuvenating human ageing genomes: lessons from mosaic chromosomal alterations. *Ageing Res Rev*. 2021. 68: 101342. <https://doi.org/10.1016/j.arr.2021.101342>
7. de la Fuente J, Davies G, Grotzinger AD, Tucker-Drob EM, Deary IJ. A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nat Hum Behav*. 2021. 5(1): 49–58. <https://doi.org/10.1038/s41562-020-00936-2>
8. D’Gama AM, Walsh CA. Somatic mosaicism and neurodevelopmental disease. *Nat Neurosci*. 2018. 21(11): 1504–1514. <https://doi.org/10.1038/s41593-018-0257-3>
9. Doherty JL, Owen MJ. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med*. 2014. 6(4): 29. <https://doi.org/10.1186/gm546>
10. Elliott ML, Belsky DW, Knodt AR, Ireland D, Melzer TR, Poulton R, Ramrakha S, Caspi A, Moffitt TE, Hariri AR. Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. *Mol Psychiatry*. 2021. 26(8): 3829–3838. <https://doi.org/10.1038/s41380-019-0626-7>
11. Exposito-Alonso D, Rico B. Mechanisms underlying circuit dysfunction in neurodevelopmental disorders. *Annu Rev Genet*. 2022. 56: 391–422. <https://doi.org/10.1146/annurev-genet-072820-023642>
12. Fernando MB, Ahfeldt T, Brennand KJ. Modeling the complex genetic architectures of brain disease. *Nat Genet*. 2020. 52(4): 363–369. <https://doi.org/10.1038/s41588-020-0596-3>
13. Fitzgerald J, Morris DW, Donohoe G. Cognitive genomics: recent advances and current challenges. *Curr Psychiatry Rep*. 2020. 22(1): 2. <https://doi.org/10.1007/s11920-019-1125-x>
14. Gupta P, Balasubramaniam N, Chang HY, Tseng FG, Santra TS. A single-neuron: current trends and future prospects. *Cells*. 2020. 9(6): 1528. <https://doi.org/10.3390/cells9061528>
15. Iourov IY, Svetlana G, Vorsanova (1945–2021). *Mol Cytogenet*. 2022. 15(1): 35. <https://doi.org/10.1186/s13039-022-00613-1>
16. Iourov IY, Vorsanova SG, Yuri B, Yurov (1951–2017). *Mol Cytogenet*. 2018. 11: 36. <https://doi.org/10.1186/s13039-018-0383-3>
17. Iourov IY, Vorsanova SG, Yurov YB. Somatic cell genomics of brain disorders: a new opportunity to clarify genetic-environmental interactions. *Cytogenet Genome Res*. 2013. 139(3): 181–188. <https://doi.org/10.1159/000347053>
18. Iourov IY, Vorsanova SG, Yurov YB. Pathway-based classification of genetic diseases. *Mol Cytogenet*. 2019a. 12: 4. <https://doi.org/10.1186/s13039-019-0418-4>
19. Iourov IY, Vorsanova SG, Yurov YB, Kutsev SI. Ontogenetic and pathogenetic views on somatic chromosomal mosaicism. *Genes (Basel)*. 2019b. 10(5): 379. <https://doi.org/10.3390/genes10050379>
20. Iourov IY, Vorsanova SG, Yurov YB, Zelenova MA, Kurinnaia OS, Vasin KS, Kutsev SI. The cytogenomic “theory of everything”: chromohelkosis may underlie chromosomal instability and mosaicism in disease and aging. *Int J Mol Sci*. 2020. 21(21): 8328. <https://doi.org/10.3390/ijms21218328>
21. Iourov IY, Vorsanova SG, Kurinnaia OS, Zelenova MA, Vasin KS, Yurov YB. Causes and consequences of genome instability in psychiatric and neurodegenerative diseases. *Mol Biol*. 2021a. 55(1): 37–46. <https://doi.org/10.1134/S0026893321010155>
22. Iourov IY, Yurov YB, Vorsanova SG, Kutsev SI. Chromosome instability, aging and brain diseases. *Cells*. 2021b. 10(5): 1256. <https://doi.org/10.3390/cells10051256>
23. Iourov IY, Vorsanova SG, Kurinnaia OS, Kutsev SI, Yurov YB. Somatic mosaicism in the

- diseased brain. *Mol Cytogenet.* 2022. 15(1): 45. <https://doi.org/10.1186/s13039-022-00624-y>
24. Kelley KW, Paşca SP. Human brain organogenesis: toward a cellular understanding of development and disease. *Cell.* 2022. 185(1): 42–61. <https://doi.org/10.1016/j.cell.2021.10.003>
  25. Konopka G, Geschwind DH. Human brain evolution: harnessing the genomics (r)evolution to link genes, cognition, and behavior. *Neuron.* 2010. 68(2): 231–244. <https://doi.org/10.1016/j.neuron.2010.10.012>
  26. Leija-Salazar M, Piette C, Proukakakis C. Review: somatic mutations in neurodegeneration. *Neuropathol Appl Neurobiol.* 2018. 44(3): 267–285. <https://doi.org/10.1111/nan.12465>
  27. Nelson RS, Dammer EB, Santiago JV, Seyfried NT, Rangaraju S. Brain cell type-specific nuclear proteomics is imperative to resolve neurodegenerative disease mechanisms. *Front Neurosci.* 2022. 16: 902146. <https://doi.org/10.3389/fnins.2022.902146>
  28. Nokelainen P, Flint J. Genetic effects on human cognition: lessons from the study of mental retardation syndromes. *J Neurol Neurosurg Psychiatry.* 2002. 72(3): 287–296. <https://doi.org/10.1136/jnnp.72.3.287>
  29. Potter H, Chial HJ, Caneus J, Elos M, Elder N, Borysov S, Granic A. Chromosome instability and mosaic aneuploidy in neurodegenerative and neurodevelopmental disorders. *Front Genet.* 2019. 10: 1092. <https://doi.org/10.3389/fgene.2019.01092>
  30. Rajarajan P, Gil SE, Brennand KJ, Akbarian S. Spatial genome organization and cognition. *Nat Rev Neurosci.* 2016. 17(11): 681–691. <https://doi.org/10.1038/nrn.2016.124>
  31. Romero C, Werme J, Jansen PR, Gelernter J, Stein MB, Levey D, Polimanti R, de Leeuw C, Posthuma D, Nagel M, van der Sluis S. Exploring the genetic overlap between twelve psychiatric disorders. *Nat Genet.* 2022. 54(12): 1795–1802. <https://doi.org/10.1038/s41588-022-01245-2>
  32. Savojardo C, Baldazzi D, Babbi G, Martelli PL, Casadio R. Mapping human disease-associated enzymes into Reactome allows characterization of disease groups and their interactions. *Sci Rep.* 2022. 12(1): 17963. <https://doi.org/10.1038/s41598-022-22818-5>
  33. Tiganov AS, Iurov IuB, Vorsanova SG, Iurov IYu. Genomic instability in the brain: etiology, pathogenesis and new biological markers of psychiatric disorders. *Vestn Ross Akad Med Nauk.* 2012. (9): 45–53.
  34. Vorsanova SG, Yurov YB, Iourov IY. Dynamic nature of somatic chromosomal mosaicism, genetic-environmental interactions and therapeutic opportunities in disease and aging. *Mol Cytogenet.* 2020. 13: 16. <https://doi.org/10.1186/s13039-020-00488-0>
  35. Ye CJ, Sharpe Z, Heng HH. Origins and consequences of chromosomal instability: from cellular adaptation to genome chaos-mediated system survival. *Genes (Basel).* 2020. 11(10): 1162. <https://doi.org/10.3390/genes11101162>
  36. Ye Z, McQuillan L, Poduri A, Green TE, Matsumoto N, Mefford HC, Scheffer IE, Berkovic SF, Hildebrand MS. Somatic mutation: the hidden genetics of brain malformations and focal epilepsies. *Epilepsy Res.* 2019. 155: 106161. <https://doi.org/10.1016/j.eplepsyres.2019.106161>
  37. Yurov YB, Vorsanova SG, Iourov IY. GIN'n'CIN hypothesis of brain aging: deciphering the role of somatic genetic instabilities and neural aneuploidy during ontogeny. *Mol Cytogenet.* 2009. 2: 23. <https://doi.org/10.1186/1755-8166-2-23>
  38. Yurov YB, Vorsanova SG, Iourov IY. Ontogenetic variation of the human genome. *Curr Genomics.* 2010. 11(6): 420–425. <https://doi.org/10.2174/138920210793175958>
  39. Yurov YB, Vorsanova SG, Iourov IY. Chromosome instability in the neurodegenerating brain. *Front Genet.* 2019. 10: 892. <https://doi.org/10.3389/fgene.2019.00892>
  40. Zhang L, Vijg J. Somatic mutagenesis in mammals and its implications for human disease and aging. *Annu Rev Genet.* 2018. 52: 397–419. <https://doi.org/10.1146/annurev-genet-120417-031501>

## Chapter 7. Monitoring of psychophysical reactions of a person in the course of professional activity

*Gorelik S., Khaskelberg E., Leshchinskiy V.*

*Soft Master Ltd (Netania, Israel), ITMO University (St.Petersburg, Russia)*

### Abstract

The development of a method for predicting the human condition for the tasks of preventive medicine and its approbation are the subject of this study. The method is based on a mathematical model of the “human-observer” system. The role of an observer is performed by an IT system that includes measuring instruments and intelligent processing algorithms with self-learning. The result of monitoring is an assessment of integral indicators, on the basis of which forecasts of the state are formed and potential risks to human health are determined. The criterion by which decisions are made is the dynamics of the information entropy of the “human observer” system. Testing was carried out in target audiences to assess the psychophysical and general functional state of the body (to identify signs of fatigue, stress, diseases). In all cases, the dynamics of the information entropy of the “human-observer” system is analyzed. In the case of statistically significant deviations of the measurement results from the predicted value, a recommendation is formed to adjust the modes of human behavior and / or the need for more detailed medical research and treatment. The correlation between driving style and psychophysical reactions of the driver has been experimentally confirmed. This makes it possible to automatically assess the driver’s condition based on the current assessment of the car’s movement.

### 1. General considerations

**«The doctor of the future will not give medicine, but will draw patient attention to care of the body, to a diet and to search of the reasons and ways of prevention of a disease» (Thomas Edison,1908) ».**

1.1. This idea of Thomas Edison at the heart of preventive medicine—a direction whose main goal is to preserve human health by preventing the development of various diseases and pathologies (Haken G. 1982.). Modern medicine, has mastered a large number of instrument methods for constant control of various health parameters (S. Levine, Malone E, Lekachvili A. Briss P.,2019). In order to effectively and timely identify threats and risks, a wide range of special sensors are usually used with which a person must be in regular contact. At the same time, the analysis of measured indicators is mainly reduced to a green-yellow-red assessment and, most often, is carried out independently for each of the areas of medicine based on existing experience reflecting practice. This approach is very limited, because it does not allow taking into account the complex nature of health problems and the correlation between various factors affecting general well-being and does not allow timely prediction of possible crisis states.

1.2. Control and information theories studying the stability and equilibrium of systems provide new opportunities for predicting the state of each person in real life by assessing the dynamics of their integral indicators, which can be relatively easily measured during production and/or household conditions. A person is a single and integral system capable of self-regulation and adaptation to multiple changing external and internal factors, while maintaining the effectiveness of the necessary life support mechanisms. obtaining the greatest efficiency with the

least energy consumption (Винер Н. 1983). All processes in the human body are built on the principle of ensuring maximum stability (Lawrie, G. and Cobbold, I. (2004)). This is a very important statement—the main criterion is not optimization, but stability (Каплан Р. Нортон Дэвид П., 2003). Since optimal systems are very sensitive in random influences and can quickly come to a crisis state because of this.

1.3. Adaptation mechanisms have a resource determined by the balance between the supply of energy and the energy costs of the body. The greatest efficiency is observed in the equilibrium of the system, then the resource for adaptation is not reduced, however, with the imbalance of the system, the depletion of stabilizing factors is observed and the adaptation resource is significantly reduced or disappears. Depletion of adaptive potential inevitably leads to certain disease states.

## **2. Biorhythms and functional state of the body**

2.1. Biorhythms determine the cyclicity of biological processes in the human body and are formed under the influence of wave processes in the surrounding space. The human brain operates on electrical signals called brain waves. There is a dominant brain wave frequency that determines the general state of the brain (an analogue of carrier frequencies in radio communication) and there are frequencies that occur in different circuits of the body control system. All signals providing the transmission of control commands feedback signals are created with a chain of biochemical reactions and electrical interactions.

2.2. The transfer of information inside the body from the brain to the actuator elements and from sensors to the control center is similar to the corresponding processes in communication networks. This process occurs at frequencies that are determined by

- Rate of propagation of chemical and electrical interactions
- Time spent processing primary data in sensor systems
- Command execution inertia
- The duration of processes that determine the vital activity of a person (for example, motor functions, metabolic processes).

2.3. Each control circuit in the body has its own oscillation frequency, and their set determines a set of frequencies that are synchronized with the body's biorhythms.

Synchronization of various subsystems takes place in the process

- Human evolution as a biological species,
- Adaptation of the individual to the influence factors changing in the personal mashtab.

For example, the gravitational and light fields generated by cosmic processes create appropriate rhythms into which the human body is drawn to obtain energy. Therefore, such subsystems of the body as the nervous system, brain, digestive, cardiovascular, respiratory, endocrine and others are synchronized with annual, seasonal, monthly, daily, hour cycles. They form “natural” frequencies in each control loop that connect a person to the outside world synchronizes with their rhythms.

The natural frequencies of the internal control circuits may differ upward compared to the systems listed above. This is due to influence factors, which are determined by biochemical and electrical circuits of information transmission within the body. It is at these frequencies

that the systems that provide psychophysical behavior and a number of others work (hearing, speech, vision, touch, smell and some others)

2.4. The most important conclusion that follows from modern ideas about the mechanisms of control of the body and its individual subsystems is the inefficiency of considering each system separately from others.

### **3. Analogy with information systems**

3.1. Subsystems of communication, information processing and storage, actuating elements form a complex system in the human body. Subsystems interact at the level of information and energy (this is the same, but in different forms of representation) exchange with external objects.

In other words, the human body is an open (in the mathematical sense) system in which there is an intensive information interaction between internal subsystems and external objects and a strong correlation between the observed parameters by which modern medicine is used to assessing a person's condition and his ability to perform any actions in the external environment (for example, professional duties, sports exercises, etc.).

3.2. Most of the data that describe interacting systems are time variables (dynamic) and cannot be precisely determined at the now. On the other hand, the model of the human body in the external environment is "incomplete" in the mathematical sense (a typical situation in the theory of "fuzzy sets") (Дуплик С. 2004). Therefore, it is not possible, based on one-off or rare tests or studies, to assess whether a person's condition is "conditionally healthy" or corresponds to a disease or, even, threatens their life (mathematically, this means that any life-critical subsystem or several subsystems lose resistance (Горелик С. Пат. РФ 2209582)

3.3. The stability of the entire system is determined by the corresponding criteria of control theory and equations of mathematical logic, and it can only be estimated by knowing previous state estimates (that is, it is necessary to have a "history" of the state that preceded the current moment). That is why in the practice of modern medicine in such situations, a consultation of different specialists is created for decision-making and/or a doctor with extensive experience in functional diagnostics is involved.

3.4. Following the principles of analogy between control systems in the animal kingdom and machines (Винер Н.1983) and methods of constructing a phenomenological model, the following statements can be programmed:

- The human body is an "open" information system with fuzzy source data and a mathematically incomplete description (fuzzy logic). The phenomenological model of the organism describes the empirical relationship of different subsystems of the organism with each other in a way that is consistent with the fundamental theory of control and physiology of the organism, but is not directly derived from the known data, due to their incompleteness. We will not explain in detail why variables that mathematically describe the components of a system interact exactly as they do. We will simply describe the information interactions between the organism and the external environment on the assumption that the dynamics of the measured parameters of the external environment can be used to estimate the dynamics of variables that are arguments in some functions that are not exactly known to us.
- In other words, we linearize the prediction by choosing a relatively short time interval

to predict. The linear prediction is described by the superposition of the value of derivatives over time of the values of the measured indicator. Weight coefficients determine degree of influence of one or another parameter of organism on indicator. In some cases, we add nonlinear components to the model, for example, logical operators, which take into account the nature of the correlation between influence factors. Regression analysis is used to create a statistical model that takes into account the stochastic nature of all processes.

- Control of the human body is carried out simultaneously using several circuits, each of which has its own frequencies, which determine the dynamics of psychophysical processes (the adoption of intelligent (requiring the connection of a central control system (brain) and spontaneous decisions (fine and large motor skills, for example), which are taken at the level of peripheral (sensory) systems.
- All control circuits are connected in an information sense to each other, and the parameters describing the state of the body are always correlated
- A reasonable scientific criterion for predicting the state of the body at this level of technology is the prediction of the state of the entire system based on the dynamics of the information entropy of the system of the body of a particular person using the principle of linearization of the model. Sometimes additional conditions are introduced that take into account the non-linearity of processes.
- Measuring instruments, with the help of which a forecast of the current state of the body for the next monitoring cycle is formed, should provide a sufficiently high accuracy, and the parameters should be “measurable” (that is, under the same external conditions, the measurement should give an error within the specified statistical limits). For this purpose, in the process of measurements, robust estimation algorithms with a sliding sample should be used, the number of samples in which (approximately 5–7 cycles of the natural frequency of the corresponding control loops) makes it possible to exclude large random emissions and, at the same time, not lead to critical delays in the feedback circuit.
- The monitoring period is related to the proper frequency of the corresponding control loop and the dynamics of external influence processes. The higher the frequency, the shorter the interval between samples. If the measurement causes too much delay, then the system cannot be controlled by means of feedbacks.
- To implement this approach, it is necessary to regularly monitor several complex indicators that depend on the state of the body at the frequency of the corresponding natural frequency of control circuits that ensure the stability and stability of the human body. In this case, it is understood that the measuring system (observer) has a high enough accuracy so as not to affect the entropy of the observer-object (human) system. Then the state forecast for the next measurement cycle will be within the statistical error (determined by the level of fuzziness of the data and incomplete model).
- This would mean that the entropy of the observer-human system does not increase (as an observer, a measuring instrument). But, if errors are repeated over several periods, then this will be a sign that stability is lost (that is, entropy increases). This may be due to the violation of internal psychophysiological processes in the body. Such a situation is assessed as dangerous and is usually associated with disorders of interaction or damage to individual organs or subsystems of the body.

- One of the most common psychophysical problems is stress (Chiba T, Kanazawa T, Koizumi A, 2019). Stress affects all vital aspects of human activity. In terms of this work, stress is a violation of the synchronization of biorhythms of different control loops. Such a situation can be a consequence of functional disorders, changes in behavior, work, sleep, nutrition, physical activity (for example, when changing the time zone, harphic of work), etc. Frequent repetitions of the systematization of the change of one or more factors associated with biorhythms, in which a sufficiently complete adaptation of the body does not occur, are especially dangerous. The repetition of such situations leads to the “wear and tear” of body control systems. Each transition process will occur longer and with large losses leading to a decrease in the performance and deterioration of the patient’s psychophysical state.
- Internal electromagnetic waves are synchronized during evolution with external fields of different origin (gravitational, electromagnetic, acoustic). Violations, defects of one and/or several components of the life cycle are usually persistent. Disorders can potentiate each other, and the failure to communicate the circle of disorders over time leads to the formation of a pathological conglomerate and gross disease states.

#### **4. Preventive medicine based on regular monitoring on the dynamics of human information entropy as an information object**

4.1. As noted above, it is impossible to create a model of the human body that is sufficiently complete and accurate for forecasting. It is also impossible to increase the frequency and number of complex chemical and physical analyzes of individual parameters of each person’s body. The solution to this problem is to use the procedure of regular monitoring of integral indicators related to the performance of production or household tasks with repeating elements.

4.2. The basic hypothesis is that if the external conditions are constant, then the reactions and behavior of a person will be predictable quite accurately. If the forecast is not confirmed, then the person, based on the assessments of experts and/or neural network algorithms trained on assessments of similar situations earlier, the patient is offered to make small short-term modifications of behavior or undergo special tests to clarify the factors that caused deviations of the measured integral indicators from the forecast values. In critical cases, when it is not possible to stabilize forecasts, the observation materials (history) are transmitted by medical specialists and they make decisions about treatment. Mathematically, this means that the entropy of the body control system continues to grow uncontrollably and outside exposure (treatment) is required (Мартин Н., Ингленд Дж. *Математическая теория энтропии*. М.: Мир, 1988.)

4.3. Obviously, the proposed approach can only be used for disorders having a sufficiently long latent period. Acute diseases are diagnosed and treated by other methods. Therefore, we position the considered approach only as one of the elements of preventive medicine and self-education of a person in a healthy lifestyle, which includes an analysis of his own condition and the factors that determine it.

4.4. Prevention and rehabilitation are the main areas of use of the offered IT and mathematical solutions. They are aimed at maintaining and restoring human adaptation potential, in order to preserve, improve, restore the quality of life (Figure 1).

4.5. Today, life in the metropolis has many risk factors that contribute to the development of psychophysical diseases, in the absence of effective approaches to psychohygiene. High

speed and large flow of information, high stress load, sleep disorders, nutritional defects, lack of time for physical activity – all these factors form the conditions of insufficient recovery and repeat from day to day asthenize the body, disrupting the adaptation resource.

4.6. Based on the above theoretical prerequisites, IT solutions for predicting and preventing various deviations from standard processes occurring in the subject's body were developed and tested in practice. The basis of the solution is remote regular monitoring of the dynamics of human information entropy, with a frequency determined by the corresponding biorhythms. Based on the results of each monitoring cycle, a probabilistic forecast of the results of the next measurement cycle is formed. If a trend towards an increase in entropy is detected, draft corrective decisions are developed, which are recommended to the tested person directly or through a medical specialist. The adjustment results are controlled by a similar technique and allow control of the process of recovery of psychophysical behavior after treatment (Figure 1). Some results and conclusions of the experimental study are given below.

### Problems and solutions

Problem	Consequence	«InvitaMotus» solution	Result
Body aging of the body may lead to certain motor dysfunctions (early signs of dementia, Alzheimer's disease, vibration disease, etc.)	Late prevention and treatment Quality of life decrease	Regular monitoring on special app-based tests Availability of monitoring tools in home/work environment	Early symptom detection Reduction of costs for patient and insurance company Extension of high quality of life period
The risk of "missing" the transition from a "healthy" equilibrium state of the body to a "diseased" one.	Decrease of success treatment rates High treatment costs	Mathematical analysis based on regular monitoring data	Timely alert in case of statistically significant deviations Early referral to a specialist doctor
High probability of diagnostic errors as a result of anamnesis subjectivity	Harm from inaccurate treatment Extra costs	IM-passport – is a base of objective dynamic knowledge of patient's health condition over extended monitoring periods	Timely diagnosis Effective treatment Cut of treatment costs

im\*

3. System for motor function monitoring and training

**Figure 1.** Directions of application of technology for prediction and control of psychophysical disorders

4.7. Other areas of use of the results of the present work are correction of eating behavior, balancing of physical activity, harmonization of metabolism and biorhythms of eating behavior.

4.8. The use of the obtained results is proposed for preclinical prevention of the development of neuro-degenerative diseases, increasing the level of adaptation and stress resistance, maintaining a high level of efficiency and fulfilling professional duties related to long-term psychophysical loads (for example, drivers, dispatcher, etc.).

4.9. The ideology of the project allows the formation of platforms used to control physiological functions in working groups, for example, monitoring the level of fatigue in drivers of regular buses, monitoring the dynamics of stress levels in employees of large companies, and much more (Figure 2)

## Medical "post-trip" control

During the "Post-trip" inspection, the data obtained in the monitoring process is analyzed.

The ascent of the route today

Part of the route	Blood pressure	
	Systolic	Diastolic
36	Rise	Rise
37	Norm.	Rise
38	Norm.	Norm.

The ascent of the route yesterday

Part of the route	Blood pressure	
	Systolic	Diastolic
36	Norm.	Norm.
37	Norm.	Norm.
38	Norm.	Norm.



Further data of medical monitoring are correlated with data of the driving quality that is collected by the system of safety driving.

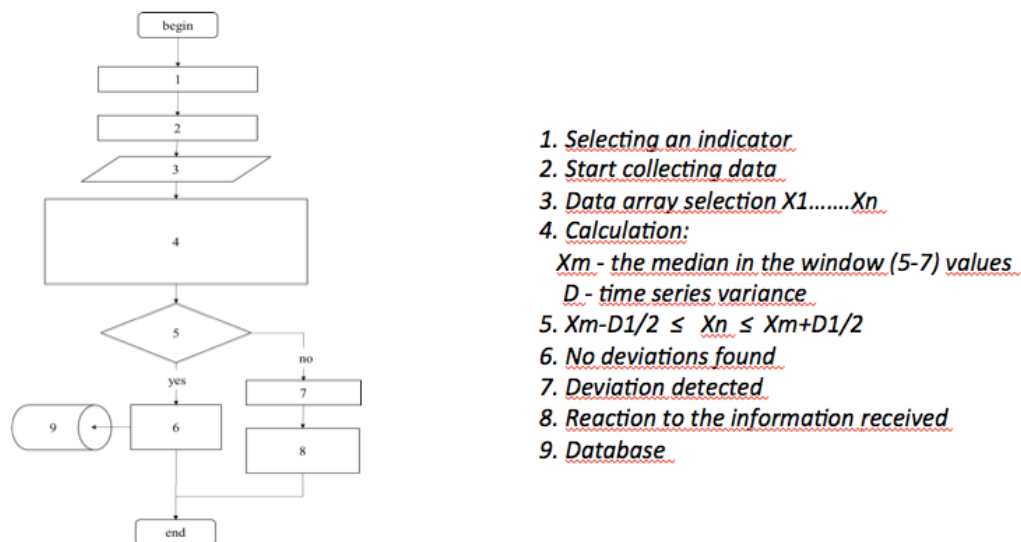
The hypotheses for the knowledge base and further verification are formed.

**Figure 2.** Monitoring of psychophysical state of drivers of regular buses

4.10. An important feature of the proposed solution is that a mathematical model has been created that allows you to assess a person's condition not by general signs, but by individual ones. These signs are formed on the basis of observation of integral indicators, which make it possible to notice in the early stages how the body of the observed client goes out of balance and does not recover in equilibrium over a certain period of time (this period has an individual value for each organism).

4.11. The transition from one equilibrium state to another can be associated with the influence of changing external and internal factors. If the body does not come to an equilibrium state during the above period, then it needs help from outside. Those who studied thermodynamics and did not forget it will immediately understand that we are talking about Prigozhin's theory of equilibrium states and entropy (Пригожин И., Стенгерс И. 2003, Коротаев С. 2016)). We use the analogy between informational and thermodynamic entropy to predict possible problems that prevent equilibrium recovery and loss of stability based on its dynamics for the observed

4.12. The observed indicators are very simple and available for measurement in domestic or industrial conditions. Numerous sensors and mobile applications are already emerging that provide periodic information about the human body, which can be processed using a mathematical model that allows you to make forecasts of the state of the body. The functional diagram of the algorithm implemented in the considered solution is shown in Figure 3.



**Figure 3.** The functional scheme of the algorithm

4.13. The purpose of the algorithm is to identify a statistically significant trend towards increasing the entropy of the system (S. Gorelik, V. Grudin, E. Khaskelberg, 2020). It is carried out by comparing the forecast and measured values of integral indicators. To filter random outliers and ensure the reliability of the estimate on a small sample, robust statistics methods are used. The key is to identify a situation where the discrepancies between the forecast and current estimates become statistically significant (the magnitude exceeds the random error of the estimate) and repeatable. Such a situation means the need for corrective actions in order to return the system to its previous equilibrium state or to a new one (for example, corresponding to changed external factors). The inability to move to one of the equilibrium states is a sign of “critical wandering of the body’s indicators,” a period of the possibility of undesirable changes in the functioning of systems.

## 5. Experimental results

5.1. The tests were carried out in several directions.

- To study the psychophysical behavior of professional passenger bus drivers and to identify the correlation of psychophysical behavior with the results of professional activities
- To control metabolism and identify the correlation between weight characteristics and the psychological state of the patient and his eating behavior.

5.2. In both cases, the following key operations were provided:

- Ensure regular collection of indicators without attendance at specialized agencies.
- Automatic analysis of the dynamics of key integral indicators and identification of statistically significant excesses.
- Provision of medical support for the observed (remote and face-to-face).
- Preparation of a documented history for seeking medical attention.

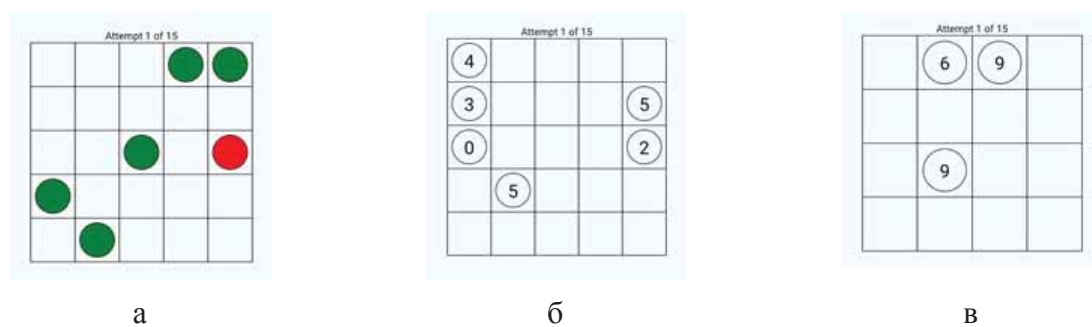
5.3. We used the following key resources:

- algorithm and software and hardware for data monitoring (online, off-line), based on available hardware and mobile solutions (Figure 3).

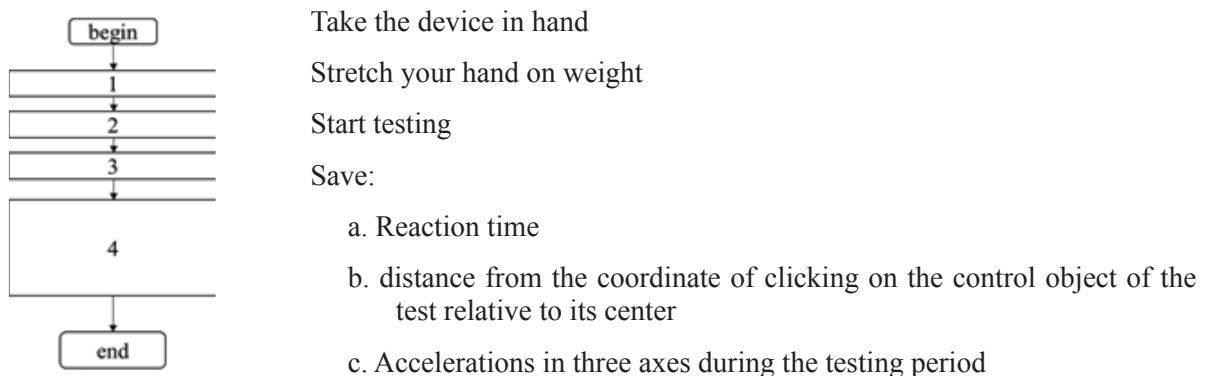
- methodology for generating conclusions based on an automated expert system and an accumulated knowledge base
- healthcare professionals who monitored and consulted patients during the experiments

5.4. The general task of the experimental study was to identify the correlation between the dynamics of the information entropy of the human-observer system (an observer is a set of measuring tools that implement the monitoring functionality) and medical indicators that determine potential threats to health. If the predicted values were confirmed at the next measurement cycle taking into account the error, then the state of the body was assessed as normal within the framework of the specific task. For example, aimed at preserving psychophysical reactions, detecting a stress state and its probable cause, disturbance of rhythms associated with metabolism, etc.

5.5. Experimental verification of the developed algorithm was carried out on the task of monitoring the psychophysical state of drivers of passenger transport. The purpose of the control is to identify signs in the driver's behavior that affect the quality and safety of driving a vehicle. Before the start of the next working cycle, medical control of the general condition of the driver was carried out and on the basis of the conclusion of a medical specialist, the possibility of his admission to driving was determined. At the same time, using special tests (examples of performed tests displayed on the smartphone screen are presented in Figure 4, individual tests within the testing session were combined into groups). parameters were measured that determine the driver's psychophysical response During the work cycle (during regular transportation along the route), repeated testing of indicators and assessment of current values of the same parameters were carried out every 4 hours. The results obtained were synchronized and matched with driving quality scores. Numerical parameters of this model were evaluated during the monitoring process and used to identify correlations between test results and driving quality measures.



**Figure 4**



**Figure 5**

5.6. The testing process is performed according to the algorithm shown in Figure 5. According to him, the test on the outstretched hand holds the smartphone and with the finger of the other hand presses the control object: a red circle or a circle with a number that periodically appears in various cells among the circles of “interference.”

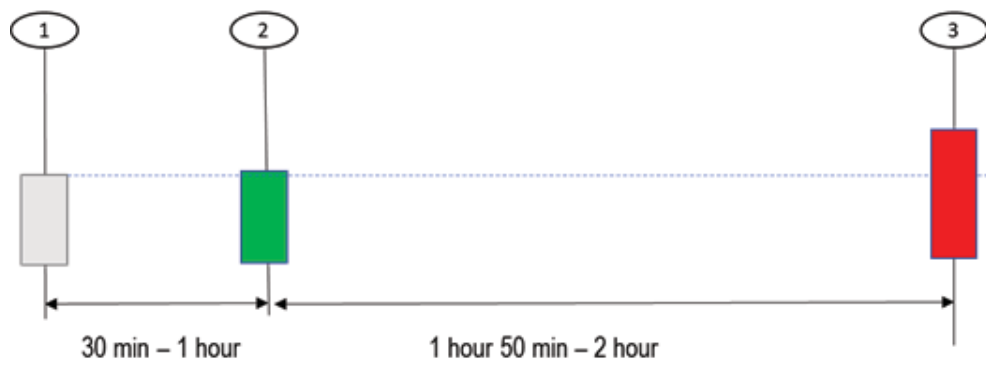
During testing, the following are saved in the test log:

- Data on the time (in milliseconds) spent on the “correct” press for each configuration after changing it.
- Accuracy of finger contact on the test circle
- Support arm accelerations.
- The test group for each test session is 3 to 5 tests. The duration of each test statistically takes no more than 20 seconds. The total duration of the test session takes no more than 100 seconds. Practical test results showed an average result of no more than 1 minute per session. The testing program provided for the rotation of tests during the testing period to eliminate the addictive factor.
- During testing, the following were recorded:
  - Data on the time (in milliseconds) spent on the “correct” press for each configuration after changing it.
  - Accuracy of finger contact on the test circle
  - Support arm accelerations.

Five different tests were used for each testing session. The duration of each test is on the order of 20 seconds. The total duration of the test session takes no more than 100 seconds (in practice, on average 1 minute per session), which fits conveniently into the technology of the production process. To eliminate the factor of driver’s “habituation” to tests, the testing program provided for the rotation of tests during the testing period. The number of errors in the execution of test tasks depends on the current state of the cognitive subsystems. In this experiment, the fine motor system, the subcortex of the brain system, the visual system, which affect the number of decision-making errors, were monitored.

5.7. During the experiments, hypotheses important for decision-making were tested:

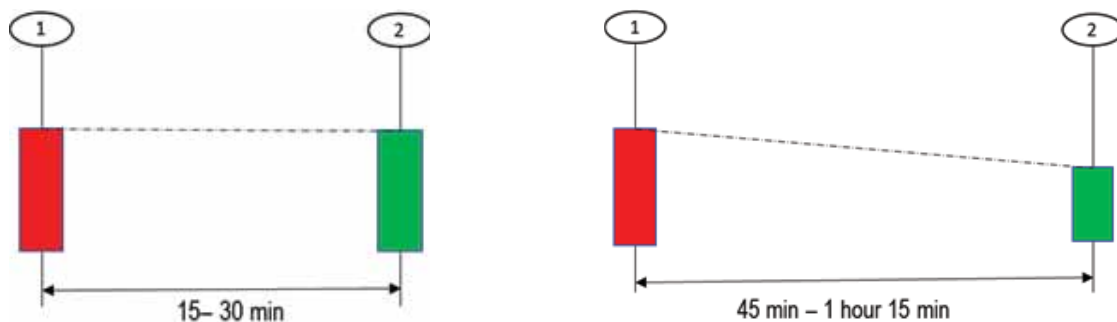
- Estimates of the level of errors in the performance of the test task and the assessment of the quality indicators of the vehicle control obtained when processing accelerometer data installed on the vehicle (between tests) are statistically significantly correlated. This assumes that the test results can determine during the working day the level of the value of the “psychophysical fatigue” factor that prevents driving safety and make appropriate decisions.
- The level of error in performing test tasks is statistically significantly dependent on the duration of rest, which makes it possible to determine and plan the optimal mode of operation and rest of the driver. The assessments of the psychophysical state of the driver obtained during testing have a clear individual character, which can be used in determining readiness for driving a vehicle. Typical situation of change of control indicators before and after the flight is shown in Figures 6–7. The green rectangle shows the values of the complex indicator of the driver’s psychophysical state before the start of the driving, red – after in the end.



1. Initial (calibration) measurement
2. Measurement before trip
3. Measurement at the end of the trip

**Figure 6**

- Regular testing at intervals of 4 hours is sufficient to obtain statistically significant metric values showing how the performance of passing tests changes as a result of driving a vehicle by changing the driver's psychophysical state. This fact allows you to compare with the possible risks of deterioration of driving quality to a critical level.



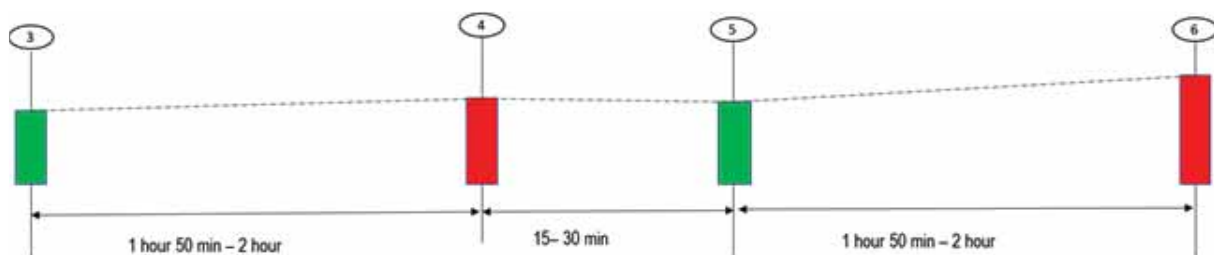
a

b

1. Measurement at the end of the trip

2. Measurement before the trip

**Figure 7–1**



c

3. Measurement before trip 1

4. Measurement at the end of trip 1

5. Measurement before trip 2

6. Measurement at the end of trip 2

**Figure 7–2**

5.8. Based on the results of the testing, important conclusions were drawn for the management of transportation (Gorelik S, Grudin V, Khaskelberg E., 2020). In particular, recommendations have been formulated to select the interval of periodic rest, the duration of inter-trip

intervals necessary for restoration. Other conclusions from the test results allow you to clarify some hypotheses about the state of health of the driver. For example, some of the subjects were found to have a high level of hand tremor and a significant dependence of tremor on the duration of the load that occurs during the driving process. It is also possible to detect visual impairment during the day, individual susceptibility of participants to the same type of stress and the manifestation of signs of stress from long-term driving.

Robust statistics methods were used to filter random outliers and ensure the reliability of the estimate on a small sample. The key is to identify a situation where the discrepancies between the forecast and current estimates become statistically significant (the magnitude exceeds the random error of the estimate) and repeatable. Such a situation means the need for corrective actions in order to return the system to its previous equilibrium state or to a new one (for example, corresponding to changed external factors).

The inability to move to one of the equilibrium states is a sign of “critical wandering of the body’s indicators,” a period of the possibility of undesirable changes in the functioning of systems.

5.9. Experiments conducted with drivers of passenger transport confirmed important hypotheses that were necessary for making decisions on the practical use of the method and algorithms in the production process:

Estimates of the level of errors in the performance of the test task and the assessment of the quality indicators of the vehicle control obtained during the processing of accelerometer data installed on the vehicle (in the period between tests) are statistically significantly correlated. This assumes that the test results can determine during the working day the level of the value of the “psychophysical fatigue” factor that prevents driving safety and make appropriate decisions.

The level of error in performing test tasks statistically significantly depends on the duration of rest, which makes it possible to determine and plan the optimal mode of operation and rest of the driver. The assessments of the psychophysical state of the driver obtained during testing have a clear individual character, which can be used in determining readiness for driving a vehicle. A typical situation of change in control indicators before and after the flight showed that regular testing at intervals of 4 hours gave statistically significant values of metrics showing how the performance of passing tests and the results of measuring the performance of the vehicle control by changing the psychophysical state of the driver (fatigue) and compare the data with possible risks of deterioration of driving quality to a critical level.

From the tests carried out, recommendations important for practice were also formulated on the selection of the interval of periodic rest of drivers and the duration of inter-trip intervals necessary for restoration. Other test findings confirm the ability to formulate and refine some predictions about the driver’s health. For example, some of the subjects had a high level of hand tremor and a significant dependence of the tremor on the duration of the load that occurs during the driving process, which served as the basis for the decision to change its operating mode. Also, it is possible to detect visual impairment during the day, individual susceptibility of participants to the same type of stress and the manifestation of signs of stress from driving.

5.10. Based on the results obtained, the use of the developed method and algorithm in other similar tasks can also be recommended, where regular monitoring of the psychophysical state allows timely identification of potential threats to health and/or the technological process in which the monitoring object is involved. For example, to control and train fine motor skills

and children, manage the recovery process of stroke patients, identify signs of diseases associated with brain activity in adulthood and old age.

Experiments were also conducted to investigate a group of patients with clinically confirmed signs of stress by monitoring the spectral characteristics obtained from EEG data. The hypothesis of correlating measures based on signal spectrum data in the alpha rhythm range (usually 10–14 Hz) with clinical state assessments was tested. The dynamics of the improvement of the condition during the treatment process by exposure to acoustic waves (a process known in science as feedback through the effect on the neural network (neurofeedback)) were also noted (10–12). The experiments analyzed the signals obtained by EEG.

Characteristic signs were identified and biorhythms accompanying stress were assessed. The results were used to obtain initial data on individual human biorhythms in order to adjust the appropriate instruments used to generate sound waves that suppress signs of stress based on methods of non-chemical correction of various medical factors (13–14).

The model of the object tested in the process of monitoring and the algorithm for monitoring the integral parameters of the human body by the dynamics of the information entropy of the human-observer system presented in this article make it possible to identify and predict critical situations in the state of various functional subsystems of the body in everyday life and/or conditions of professional activity and are aimed at use in preventive and telemedicine.

The developed solutions can be used for the general assessment of a person's psychophysical state, as well as for the detection of stresses and other critical conditions, as well as for the development of individual recommendations for treatment and prevention. Thus, the possibilities of implementing the basic requirement of modern medicine are being realized: the fight against consequences should be combined as much as possible with the prediction of potential threats (14).

## 6. Conclusion

6.1. The current state of mathematical science and information technology makes it possible to move on to regular health control, prediction of possible crises, development of draft solutions for medical specialists.

6.2. A person is a single and integral system capable of self-regulation and adaptation to multiple changing external and internal factors, while maintaining the effectiveness of the necessary life support mechanisms. All processes in the human body are built on the principle of ensuring maximum stability (*Хакем Г.*, 2001., Adrian, E., Matthews, B, 1934. Chiba T, Kanazawa T, Koizumi A, 2019)

6.3. The adaptation mechanisms that exist in the body have a resource determined by the balance between the supply of energy and the energy costs of the body. The greatest efficiency is observed in the equilibrium of the system, then the resource for adaptation is not reduced, however, with the imbalance of the system, the depletion of stabilizing factors is observed and the adaptation resource is significantly reduced or disappears.

6.4. Depletion of adaptive potential inevitably leads to loss of resistance (i.e., certain disease states)

6.5. Internal biorhythms are of decisive importance in the formation of a stable system and are formed under the influence of external fields of different types. The spectrum of external fields is formed under the influence of low-frequency factors (annual, monthly, daily cycles),

medium-frequency—with wavelengths commensurate with the dimensions of elements of sensory systems and internal organs, high-frequency—under the influence of processes occurring at molecular and submolecular levels.

6.6. Violations, defects of one and/or several components of the life cycle are usually persistent. Disorders can potentiate each other and non-communication of the circle of disorders over time leads to the formation of a pathological conglomerate and gross disease states.

6.7. The results of theoretical and experimental studies published in the present work made it possible to formulate a modern approach to preventive medicine, in which it is possible to predict crisis states and their early prevention and maintain adaptation potential.

6.8. The model of the object tested in the process of monitoring and the algorithm for monitoring the integral parameters of the human body by the dynamics of the information entropy of the human-observer system presented in this article make it possible to identify and predict critical situations in the state of various functional subsystems of the body in everyday life and/or conditions of professional activity and are aimed at use in preventive and telemedicine.

6.9. The developed solutions can be used for the general assessment of a person's psychophysical state, as well as for the detection of stresses and other critical conditions, as well as for the development of individual recommendations for treatment and prevention. Thus, the possibilities of implementing the basic requirement of modern medicine are being realized: the fight against consequences should be combined as much as possible with the prediction of potential threats (.

*The work was funded by the authors and Soft Master Ltd (Israel)*

## References

1. Винер Н. Кибернетика или управление и связь в животном и машине. // Пер. с англ.— 2-е издание.—М.: Наука, 1983.— 344 с.
2. Горелик С. Способ контроля веса пациента. Пат. РФ 2209582. <https://findpatent.ru/patent/220/2209582.html>
3. Дуплик С.В. Модель адаптивного тестирования на нечеткой математике. Текст. / С.В. Дуплик // Информатика и образование.— 2004.— № 11.— С. 57–65
4. Каплан, Роберт С. Нортон Дэвид П., *Сбалансированная система показателей*, Organizational Publishing, 2003.
5. Коротаев С. Энтропия и информация—универсальные естественнонаучные понятия [Электронный ресурс] // Институт исследований природы времени. URL: [http://www.chronos.msu.ru/old/RREPORTS/korotaev\\_entropy/korotaev\\_entropy.htm](http://www.chronos.msu.ru/old/RREPORTS/korotaev_entropy/korotaev_entropy.htm) (дата обращения 03.10.2022)
6. Мартин Н., Ингленд Дж. *Математическая теория энтропии*. М.: Мир, 1988.
7. Осипов А.И. Термодинамика вчера, сегодня, завтра. Часть 1. Равновесная термодинамика // Соросовский Образовательный Журнал.— 1999.— С. 79–85.
8. Пригожин И., Стенгерс И. Порядок из хаоса. Новый диалог человека с природой. М.: УРСС, 2003.
9. Хакен Г. Принципы работы головного мозга: Синергетический подход к активности мозга, поведению и когнитивной деятельности. М.: Изд-во Per Se, 2001.— 353 с. Н.
10. Adrian, E., and Matthews, B (1934). “The interpretation of potential waves in the cortex,”
11. Haken G. Prediction Theory. // Proceedings of I.R.E., Volume 38. Issue: 4.—P. 417–425, Synergetik. Springer-Verlag Berlin Heidelberg New York, 1982.
12. Chiba T, Kanazawa T, Koizumi A, et al. Current status of neurofeedback for post-traumatic stress disorder: a systematic review and the possibility of decoded neurofeedback. Front Hum

- Neurosci. 2019.
13. Gorelik S, Grudin V, Khaskelberg E. Method for assessing the influence of psychophysical state of drivers on control safety based on monitoring of vehicle movement parameters. *Transportation Research Procedia*. Volume 50, 2020, Pages 152–159. <https://www.semanticscholar.org/paper/Method-for-assessing-the-influence-of-state-of-on-Gorelik-Grudin/e23d204245a80b8c4605d4f96c15c31695f17797>
  14. S. Levine, Malone E, Lekiachvili A. Briss P. Health Care Industry Insights: Why the Use of Preventive Services Is Still Low // *Prev. Chronic Dis.*– 2019. Vol. 16.–E30.
  15. Razzak M. I., Imran M., Xu G. Big data analytics for preventive medicine // *Neural Computing & Applications.*– 2020. Vol. 32.–P. 4417–4451.
  16. Rabasa C., Dickson S. “Impact of Stress on Metabolism and Energy Balance”. *Current Opinion in Behavioral Sciences*, vol. 9, pp. 71–77, 2016.
  17. Lawrie, G. and Cobbald, I. (2004), “Third generation balanced scorecard: evolution of an effective strategic control tool”, *International Journal of Productivity and Performance Management*, Vol. 53 No. 7, pp. 611–623. <https://doi.org/10.1108/17410400410561231>

## Chapter 8. Visual perception of microscopic objects

*Gerasimov A. P.<sup>1,2</sup>, Shevtsov M. A.<sup>3</sup>, Bobkov D. E.<sup>3</sup>, Zabrodskaya Yu. M.<sup>1</sup>, Nazaralieva E.<sup>1</sup>, Kolotii A. D.<sup>4,5</sup>, Yurov Yu. B.<sup>4,5</sup>, Vorsanova S. G.<sup>4,5</sup>, Iourov I. Yu.<sup>4,5,6</sup>*

<sup>1</sup>*Almazov National Medical Research Centre, Saint-Petersburg;*

<sup>2</sup>*Pavlov First St. Petersburg State Medical University, Saint-Petersburg;*

<sup>3</sup>*Institute of Cytology RAS, Saint-Petersburg;*

<sup>4</sup>*Mental Health Research Center, Moscow;*

<sup>5</sup>*Veltischev Research and Clinical Institute for Pediatrics of the Pirogov Russian National Research Medical University, Moscow;*

<sup>6</sup>*Russian Medical Academy of Continuous Postgraduate Education, Moscow, Russia*

Discussion of the mechanisms of visual perception is usually devoted the perception of macroscopic objects. Publications about perceptions of microscopic objects are near absent. But recognition of such object is not common, but very important task in medicine.

Examples of methods in medicine are:

- Classical microscopy (morphology)
- Cytogenetic
- Microsurgery
- Fluorescent microscopy (including FISH and intraoperative 5-ALA microscopy)

The same time (according our opinion) basic mechanisms of perception are typical: recognition of size, form, structure, moving (sometimes) and color. Crowding effect and effect of previous information may be important.

Our search started from discussing about results of confocal fluorescent microscopy of malignant tumors, performed in Institute of Cytology RAS.

### **Immunofluorescence staining and confocal microscopy.**

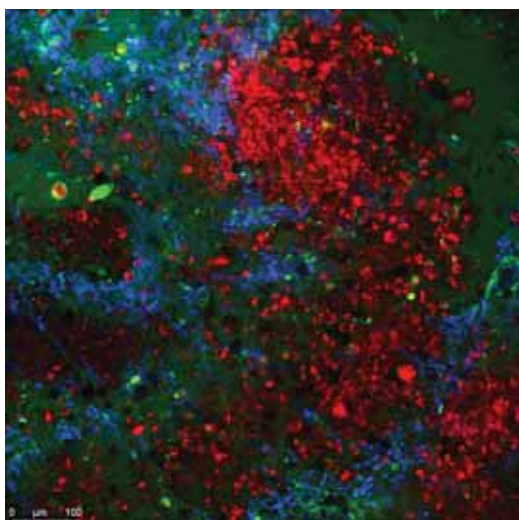
To assess the expression levels of the Hsp70 protein in tumors, we used our method of in vivo staining of tumor samples followed by image recording using laser scanning confocal microscopy. The technique makes it possible to study the features of tumor organization at the cellular and subcellular levels, to analyze cell motility without preliminary enzymatic dissociation, fixation and permeabilization of the material. Material for microscopy was obtained during surgical operations; if possible, the material was divided into zones: contrast-accumulating, necrotic, and perifocal. Samples were dissected into small pieces (2–5 mm in diameter), placed in a cooled culture medium and transported to the microscopy site within 1–2 hours. Then, samples were stained without prior fixation and permeabilization for 1 h, to do this, the material was washed three times with cold PBS and then incubated, gently shaking, with dyes diluted in accordance with the manufacturer's recommendations for the immunofluorescent staining method.

Anti-Hsp70 antibodies fused with FITC were used for in vivo detection of membrane-bound Hsp70, as well as Hoechst 33342 dye to detect DNA of cell nuclei (Bucevičius, J., Lukinavičius, G. at al., 2018). To confirm cell viability, the material was stained with TMRM dye (Thermo Fisher Scientific, USA), which reveals the membrane potential of mitochondria (Monteith, A., Marszalec, W. at al., 2013). Some of the material known to have been taken by patients before surgery with 5-ALA was used to record the fluorescent signal from protoporphyrin IX (pPIX), which accumulates mainly in mitochondria (Ji, Z., Yang, G. at al., 2006).

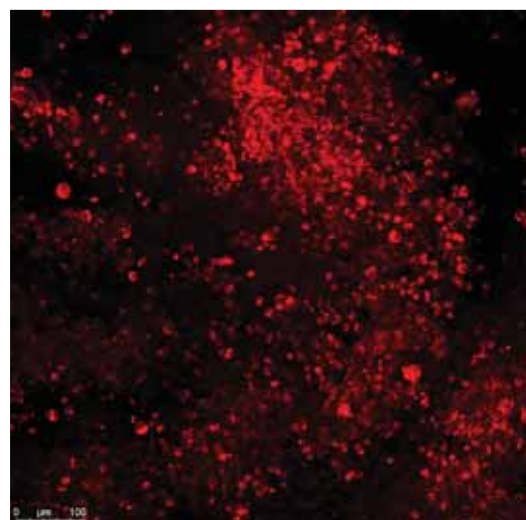
After staining, the material was washed with cold PBS and then examined by confocal laser scanning microscopy using a Leica TCS SP8 inverted microscope equipped with argon and helium-neon lasers. Microscopy was performed for 2–4 h: for this, small pieces of material were placed in Ibidi  $\mu$ -Dish 35 mm (Ibidi, USA), then covered with a cover slip on top to ensure a better fit to the bottom of the dish. Fluorescence of Hoechst 33342 and pPIX was excited with a 405 nm laser and recorded in the range of 415–500 nm for Hoechst 33342 and 650–750 nm for pPIX; Hsp70 fluorescence was excited by a 488 nm laser and recorded in the range of 495–560 nm; TMRM fluorescence was excited by a 561 nm laser and recorded in the range of 565–650 nm. Separate scanning was carried out in three channels, the width of the confocal diaphragm was set to 100  $\mu$ m, which approximately corresponds to the optical section thickness of 1  $\mu$ m. Received images at a resolution of 1024 x 1024 pixels with an average of 7 for each scan line. To obtain three-dimensional models, packs of 35 images (z-stacks) were obtained at a resolution of 512  $\times$  512 pixels with a step along the z axis of 0.35  $\mu$ m and averaging 3 for each scan line. We used 20x and 60x oil immersion objectives.

The signal from a tissue sample obtained during a surgical operation to remove an epilepsy focus was taken as the norm. Tumor tissue stained with fluorescently labeled secondary antibodies (Alexa Fluor 488) was used as a negative control without the use of any specific primary antibodies. The background was subtracted from the resulting images in the ImageJ program using the Subtract Background function with the Rolling ball radius = 50 px parameter. Next, the images were transferred to the Rstudio program, where, using the EBImage functions package, the average and median values of the pixel intensities in the red (TMRM), green (Hsp70), and blue (cell nuclei) channels were obtained, and the images from samples represented by several zones were averaged. Then, statistical analysis was carried out in the GraphPad Prism program: the mean and median values of green pixel intensities (Hsp70) were analyzed on intravital confocal images.

Discussion of slides (Fig.1, Fig.2, Appendix) discovered unexpected results. Clinical genetic (informed about mitochondrial diseases) described difference of size and form of mitochondria. Cytologists (not informed) send no attention this phenomenon. Heteroplasmy is the presence of more than one type of organellar genome (mitochondrial DNA) within a cell or individual. This situation illustrated importance of previous knowledge very well.



**Pic. 1.** Anaplastic astrocytoma. Confocal microscopy. Red–TMRM. Green–Anti-Hsp70 Abs. Blue–Hoechst 33342



**Pic. 2.** The same field, but with only red – TMRM. Significant heteroplasmy.

## Microscopic morphology and pathomorphology of central nervous system

To understand next mechanisms of perception we need to discuss classical microscopy and FISH.

Central nervous system is well-ordered structure. This discovery of Santiago Ramón y Cajal was awarded the Nobel prize in 1906. Now we have combination of very different colouring with different morphology phenomena.

Histology refers to the study of the morphology of cells within their natural tissue environment. As a bio-medical discipline, it dates back to the development of first microscopes which allowed to override the physical visual limitation of the human eye. Since the first observations, it was understood that cell shape predicts function and, therefore, shape alterations can identify and explain dysfunction and diseases. The advancements in morphological investigation techniques have allowed to extend our understanding of the shape-function relationships close to the molecular level of organization of tissues. (Mazzarini M., Falchi M. et al., 2021; Hoffmann L., Blümcke I., 2022)

The traditional study of sections of biological tissues in transmitted light under a microscope implies the identification of various features among the overall picture, which is visible through the eyepiece.

**Identification.** When stained with hematoxylin and eosin, the main stain for studying tissue sections, tissue and cellular structures is differentiated. Hematoxylin paints kernels, lime, microbes, mucus, bones in a blue-purple color. This makes it possible to navigate “on the ground”, where the cell is, and where the stroma and other formations are. Eosin paints the cytoplasm of cells, connective tissue fibers in pink. Coloring is needed to perceive the pattern of the fabric, because without coloring it is colorless, except for natural pigments (hemosiderin, lipofuscin, melanin).

**Topography.** Under the magnifying glass  $\times 5$ , histotopographic pictures of the structure of the tissue are revealed to us, where it is possible to determine the distribution and mutual application of tissues of different structures on a “vast” area.

**Ornament, fabric pattern.** A small magnification of  $\times 10$  gives us an idea of the architectonics of the fabric being studied—the “ornament” built by the fabric elements. We compare the ornament-drawing we have seen with the paintings we know, which allows us to find out similar patterns of norm or pathology.

The shape and size of the cells are estimated at high magnification ( $\times 20$ ,  $\times 40$ ). In addition, we analyze the intracellular structure—the state of the cytoplasm of the nucleus—whether or not there are inclusions, density, color intensity. We evaluate and identify specific changes for mitosis, viral inclusions, metabolic disorders, apoptosis, necrosis, and other pathological phenomena.

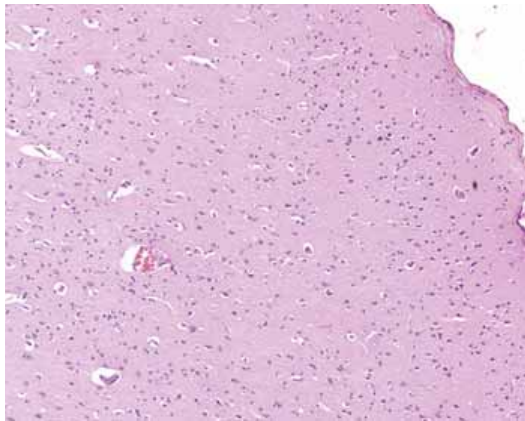
**Colour.** The color highlighting of known structures is important for the perception and analysis of the histological picture. To visualize more differentiated structures, to identify intracellular and extracellular structures, colorings are used to identify more differentiated tissues with the color of known tissue structures. For example, coloration with toluidine blue by the Nissl method is detected in the cytoplasm of neurons, the so-called tigroid substance. We can assess its condition, where it is located in the cytoplasm. The collagen fibers are stained red with picrofuxin according to Van Gieson. Elective coloring with staining of known biochemical elements, give clarity in the presence or absence of a particular component. For example,

Shpilmeir staining reveals myelin. The analysis is whether there is (colored black) or there is no myelin (not colored black) (Taates DJ, Roth J., 2015; Pellicciari C., 2018).

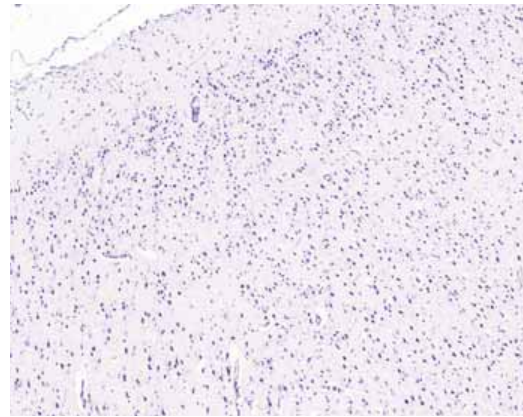
Color contrast. I see—I don't see (yes-no approach). It is more clearly visually possible to determine the presence of specific proteins in the tissue using the immunohistochemical method. The staining of a known protein is visually determined—the algorithm works—is there or not, and where (is stained). Quantitative analysis is used (how many cells) and intensity is estimated (for example, progesterone in the nuclei) (Mori H., Cardiff RD., 2016; Im K., Mareninov S. et al., 2019; Antimonova O.I., Grudinina N.A. et al., 2020).

We may illustrate this using example of focal cortical dysplasia.

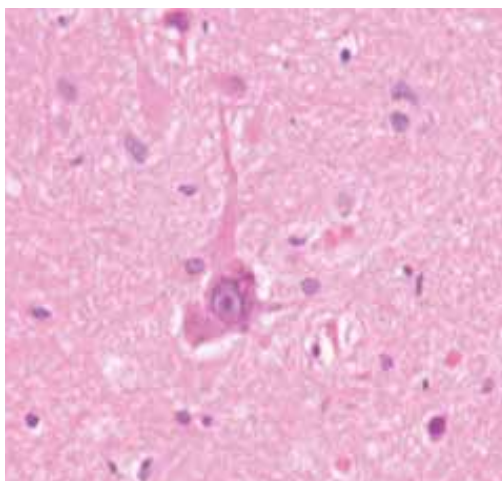
In the case of focal cortical dysplasia, I type we may see disordered tissue picture: zones of neuronal loss, cortical atrophy (pseudosulcus surface of gyrus) and lamination disorder (Fig.3, Fig. 4, Appendix). In the case of focal cortical dysplasia II type, we may see change of size (Fig.5, Appendix), form (Fig.6, Appendix) and, using immunohistochemical search, color heterogeneity (Fig.7, Appendix). In the case of hippocampal sclerosis density of elements is changed (Fig. 8 a, b, Appendix).



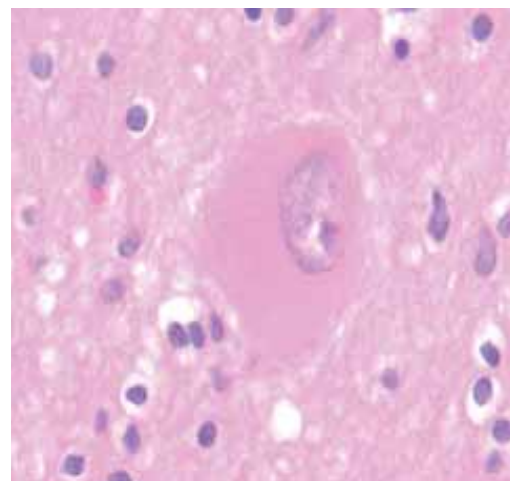
**Pic. 3.** Focal cortical dysplasia I type. Zones of neuronal loss, cortical atrophy (pseudosulcus surface of gyrus). Col. H&E, x100



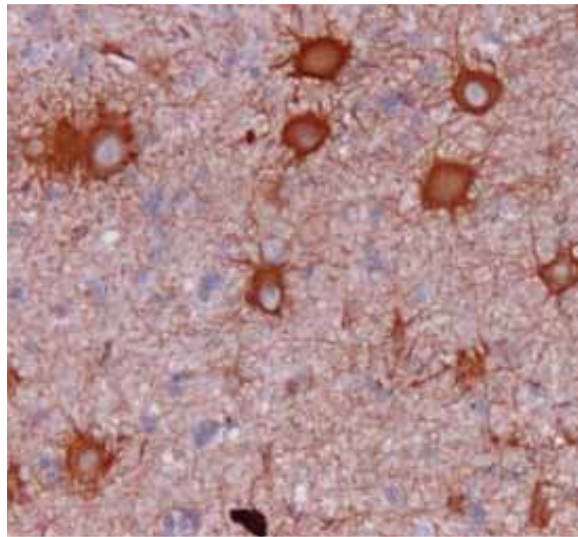
**Pic. 4.** Focal cortical dysplasia I type. Lamination disorder. Col. by Nissl, x100.



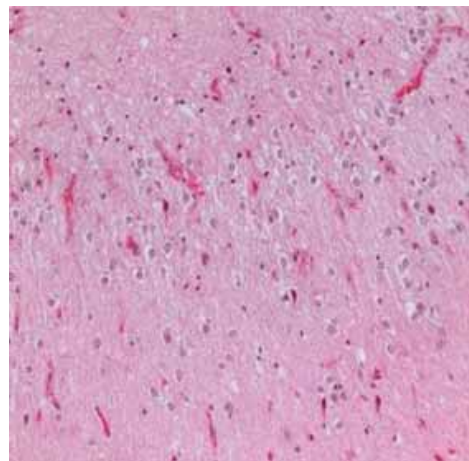
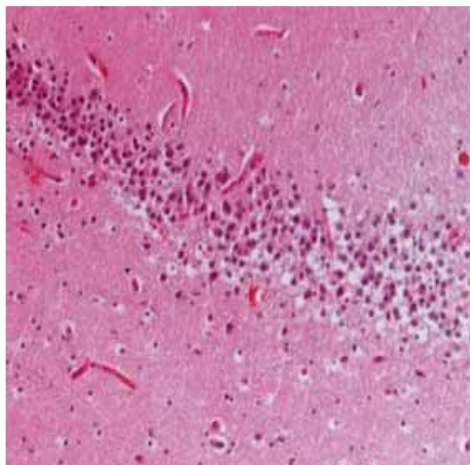
**Pic.5.** FCD IIa. Dysmorphic neurons. Col. H&E



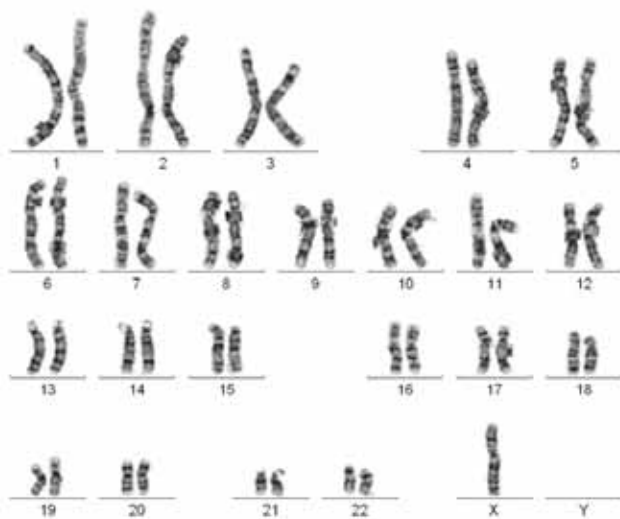
**Pic. 6.** FCD IIb. Ballooned cells Col. H&E



Pic. 7. Focal cortical dysplasia II type. Immunohistochemical search with GFAP—big ballooned cells



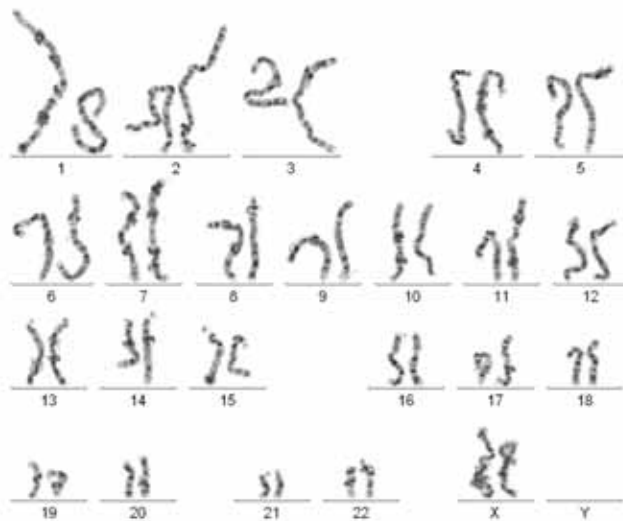
Pic. 8. Hypocampal sclerosis. Gyrus dentatus, col. H&E,  $\times 200$ . A—norma. B—sclerosis. Significant atrophic changes—zones of neuronal loss, nuclear emptiness and dispersion of granular layer of nucleus dentatus.



left



B Right



left



B Right

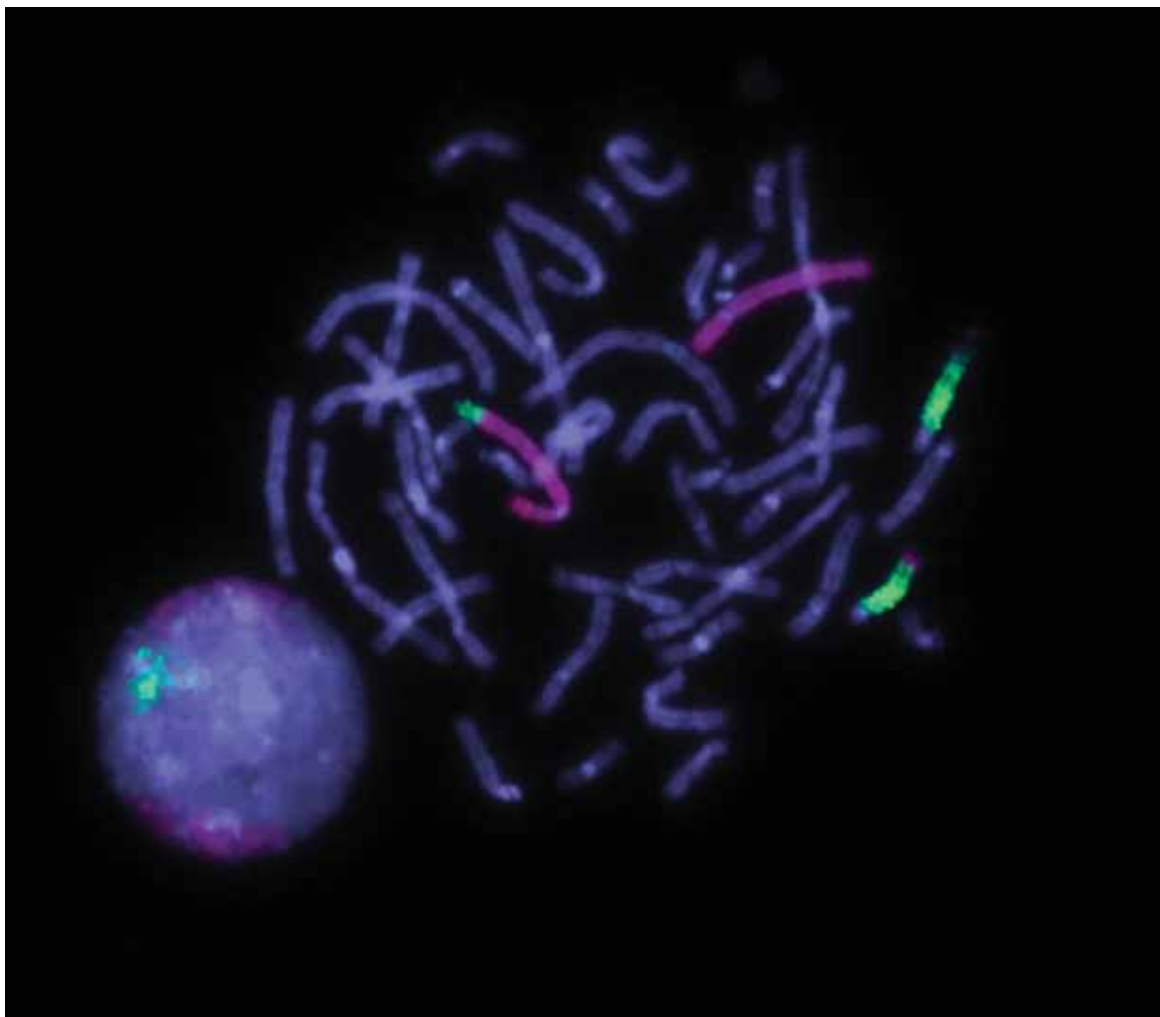
Thus, the morphological analysis of microscopic sections involves complex mechanisms of visual analysis based on our memory of visual images and accumulated experience. This type of analysis determines the expert level. It takes a long experience in order to see–differentiate the seen–recognition in the unknown. The use of techniques using color and contrast images with a known elective value makes it easy to determine the presence or absence of the desired element, where long-term development of visual analysis skills is not required–detection of a known color.

## Cytogenetics

Cytogenetics is essentially a branch of genetics, but is also a part of cell biology/cytology (a subdivision of human anatomy), that is concerned with how the chromosomes relate to cell behaviour, particularly to their behaviour during mitosis and meiosis. It is possible to see the whole cell genome using classical cytogenetic methods (Iourov IY, Vorsanova SG at al., 2012; Iourov IY, Yurov YB at al., 2020).

Standard cytogenetic method is the microscopy of metaphase spread with using G-banding. Standard resolution >550 bands (Fig. 9A, Appendix), but may be higher (Pic. 9B–>800 bands). First task is identification of chromosomes as the objects and excluding the artifacts. Next step is grouping of chromosomes according size and morphology. Then couples of chromosomes must be identified and ordered using structure evaluation. Finally, it is necessary to check number and structure of chromosomes to detect abnormalities (Vorsanova SG, Iourov IY at al., 2021; Iourov IY, Vorsanova SG at al., 2008). Routine karyotyping needs search of 3–5 metaphase spreads, but diagnostic of the mosaic mutations may use up to 50 spreads. (Iourov IY, Vorsanova SG at al., 2019); It is interest fact, that unlike previous methods, classical cytogenetic used only monochromic scale.

But multicolor marks are useful in FISH method. Single color point will demonstrate number of copies of binding site (in norm usually two in one cell). Combination of colors may show chromosomal rearrangements (Fig. 9c, Appendix) (Iourov IY, Liehr T. at al., 2006; Iourov IY, Vorsanova SG at al., 2013).



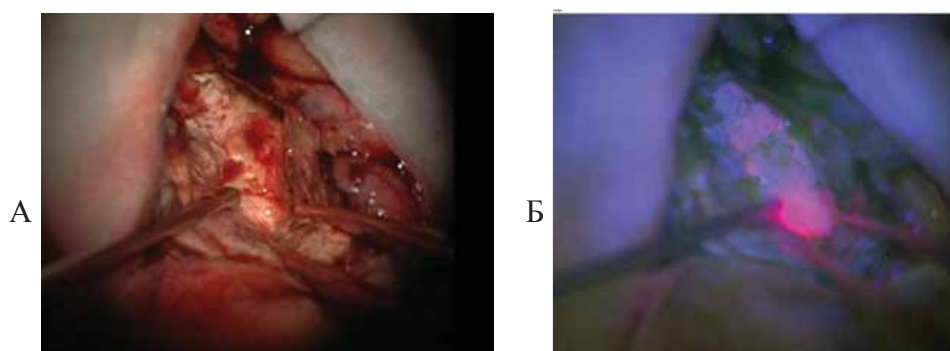
C

**Pic. 9.** Cytogenetics and molecular cytogenetics microscopy. (A) Metaphase spread with a resolution >550 bands (right–microscopic image; left–chromosomes assigned or cellular karyotype). (B) Metaphase spread with a resolution >800 bands (right–microscopic image; left–chromosomes assigned or cellular karyotype). (C) FISH with DNA probes painting whole chromosomes 4 (pink) and 14 (green): note the translocation, i. e. interchange of chromosomal parts between chromosomes 4 and 14.

### **Intraoperative fluorescent microscopy.**

Intraoperative fluorescent microscopy is a variant of intraoperative microscopy with using the 5-aminolevulinic acid as fluorescent marker of tumor. It is method of differentiation the normal tissue and the tumor tissue (especially high grade) (Suero Molina E, Wölfer J at al., 2017). Intraoperative fluorescent microscopy is useful both in gliomas and in meningiomas, with cranial and spinal localization (Millesi M, Kiesel B at al., 2014; Kiesel B, Mischkulnig M at al., 2018; Byvaltsev V.A., Stepanov I.A. at al., 2018; Maragkos GA, Schüpfer AJ at al., 2021)

This method needs switching the microscope from white light (10a) to violet-blue excitation light (10 b). This way normal tissue has deep blue color, but tumor nodes have bright colorization from pink to violet (Della Puppa A, Ciccarino P at al., 2014). At least one publication demonstrated quantitative measurement (Valdes PA, Bekelis K at al., 2014).



**Pic. 10.** Intraoperative visualization: tumor resection under white (A) and blue (B) light using 5-Aminolevulinic Acid fluorescence in frontal high grade glioma. Microscope OPMI Pentero 700, Carl Zeiss Meditec AG, Deutschland. x 36.

## Conclusion.

Perception of microscopic objects bases on estimation the size, form, color, structure. Previous experience (including education) is important.

Influence of crowding effect is not clear, and it is not clear whether recognition strategy individual or standard. It may be task of the future search.

Finally, pictures of chromosomes may be discussed as visual test stimulus.

*Search was supported by the Government Assignment of the Russian Ministry of Health, Assignment no. 121031000359–3.*

## References

1. Antimonova O.I., Grudinina N.A., Egorov V.V. et al. Time machine: Can a dye from 1928 be re-purposed for modern, fluorescence-based detection of amyloid-like fibrils? *Dyes and Pigments*. Volume 172,2020,107863, ISSN0143–7208, <https://doi.org/10.1016/j.dyepig.2019.107863>.
2. Bray MA, Singh S, Han H, Davis CT, Borgeson B, Hartland C, Kost-Alimova M, Gustafsdottir SM, Gibson CC, Carpenter AE. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc*. 2016 Sep;11(9):1757–74. doi: 10.1038/nprot.2016.105. Epub 2016 Aug 25. PMID: 27560178; PMCID: PMC5223290.
3. Bucevičius, J., Lukinavičius, G., & Gerasimaitė, R. (2018). The use of hoechst dyes for DNA staining and beyond. *Chemosensors*, 6(2), 18.
4. Byvaltsev V.A., Stepanov I.A., Kichigin A.I. Outcomes of 5-ALA fluorescence-guided surgery for high grade gliomas. *Siberian Journal of Oncology*. 2018; 17 (2): 18–26.–doi: 10.21294/1814–4861–2018–17–2–18–26.
5. Della Puppa A, Ciccarino P, Lombardi G, Rolma G, Cecchin D, Rossetto M. 5-Aminolevulinic acid fluorescence in high grade glioma surgery: surgical outcome, intraoperative findings, and fluorescence patterns. *Biomed Res Int*. 2014;2014:232561. doi: 10.1155/2014/232561. Epub 2014 Apr 8. PMID: 24804203; PMCID: PMC3997860.
6. Hoffmann L, Blümcke I. Neuropathology and epilepsy surgery. *Curr Opin Neurol*. 2022 Apr 1;35(2):202–207. doi: 10.1097/WCO.0000000000001030. PMID: 35067500.
7. Im K, Mareninov S, Diaz MFP, Yong WH. An Introduction to Performing Immunofluorescence Staining. *Methods Mol Biol*. 2019;1897:299–311. doi: 10.1007/978–1–4939–8935–5\_26. PMID: 30539454; PMCID: PMC6918834.
8. Iourov IY, Liehr T, Vorsanova SG, Kolotii AD, Yurov YB. Visualization of interphase

- chromosomes in postmitotic cells of the human brain by multicolour banding (MCB). *Chromosome Res.* 2006;14(3):223–9. doi: 10.1007/s10577-006-1037-6.
9. Iourov IY, Vorsanova SG, Yurov YB. Recent patents on molecular cytogenetics. *Recent Pat DNA Gene Seq.* 2008;2(1):6–15. doi: 10.2174/187221508783406585.
  10. Iourov IY, Vorsanova SG, Yurov YB. Single cell genomics of the brain: focus on neuronal diversity and neuropsychiatric diseases. *Curr Genomics.* 2012;13(6):477–88. doi: 10.2174/138920212802510439.
  11. Iourov IY, Vorsanova SG, Yurov YB, Kutsev SI. Ontogenetic and Pathogenetic Views on Somatic Chromosomal Mosaicism. *Genes (Basel).* 2019;10(5):379. doi: 10.3390/genes10050379.
  12. Iourov IY, Yurov YB, Vorsanova SG. Chromosome-centric look at the genome. In: Iourov I, Vorsanova S, Yurov Y, editors. *Human interphase chromosomes—Biomedical aspects.* Springer; 2020. pp. 157–170.
  13. Ji, Z., Yang, G., Vasovic, V., Cunderlikova, B., Suo, Z., Nesland, J. M., & Peng, Q. (2006). Subcellular localization pattern of protoporphyrin IX is an important determinant for its photodynamic efficiency of human carcinoma and normal cell lines. *Journal of Photochemistry and Photobiology B: Biology*, 84(3), 213–220.
  14. Kiesel B, Mischkulnig M, Woehrer A, Martinez-Moreno M, Millesi M, Mallouhi A, Czech T, Preusser M, Hainfellner JA, Wolfsberger S, Knosp E, Widhalm G. Systematic histopathological analysis of different 5-aminolevulinic acid-induced fluorescence levels in newly diagnosed glioblastomas. *J Neurosurg.* 2018 Aug;129(2):341–353. doi: 10.3171/2017.4.JNS162991. Epub 2017 Oct 27. PMID: 29076783.
  15. Maragos GA, Schüpfer AJ, Lakomkin N, Sideras P, Price G, Baron R, Hamilton T, Haider S, Lee IY, Hadjipanayis CG and Robin AM (2021) Fluorescence-Guided High-Grade Glioma Surgery More Than Four Hours After 5-Aminolevulinic Acid Administration. *Front. Neurol.* 12:644804. doi: 10.3389/fneur.2021.644804
  16. Mazzarini M, Falchi M, Bani D, Migliaccio AR. Evolution and new frontiers of histology in bio-medical research. *Microsc Res Tech.* 2021 Feb;84(2):217–237. doi: 10.1002/jemt.23579. Epub 2020 Sep 11. PMID: 32915487; PMCID: PMC8103384.
  17. Millesi M, Kiesel B, Woehrer A, Hainfellner JA, Novak K, Martínez-Moreno M, Wolfsberger S, Knosp E, Widhalm G. Analysis of 5-aminolevulinic acid-induced fluorescence in 55 different spinal tumors. *Neurosurg Focus.* 2014 Feb;36(2): E11. doi: 10.3171/2013.12.FOCUS13485. PMID: 24484249.
  18. Monteith, A., Marszalec, W., Chan, P., Logan, J., Yu, W., Schwarz, N., ... & Hockberger, P. (2013). Imaging of mitochondrial and non-mitochondrial responses in cultured rat hippocampal neurons exposed to micromolar concentrations of TMRM. *PLoS One*, 8(3), e58059.
  19. Mori H, Cardiff RD. Methods of Immunohistochemistry and Immunofluorescence: Converting Invisible to Visible. *Methods Mol Biol.* 2016;1458:1–12. doi: 10.1007/978-1-4939-3801-8\_1. PMID: 27581010.
  20. Pellicciari C. Histochemistry as a versatile research toolkit in biological research, not only an applied discipline in pathology. *Eur J Histochem.* 2018 Dec 21;62(4):3006. doi: 10.4081/ejh.2018.3006. PMID: 30572698; PMCID: PMC6317132.
  21. Suero Molina E, Wölfer J, Ewelt C, Ehrhardt A, Brokinkel B, Stummer W. Dual-labeling with 5-aminolevulinic acid and fluorescein for fluorescence-guided resection of high-grade gliomas: technical note. *J Neurosurg.* 2018 Feb;128(2):399–405. doi: 10.3171/2016.11.JNS161072. Epub 2017 Mar 24. PMID: 28338432.
  22. Taatjes DJ, Roth J. The Histochemistry and Cell Biology pandect: the year 2014 in review. *Histochem Cell Biol.* 2015 Apr;143(4):339–68. doi: 10.1007/s00418-015-1313-7. Epub 2015 Mar 6. PMID: 25744491
  23. Valdes PA, Bekelis K, Harris BT, Wilson BC, Leblond F, Kim A, Simmons NE, Erkmén K, Paulsen KD, Roberts DW. 5-Aminolevulinic acid-induced protoporphyrin IX fluorescence

- in meningioma: qualitative and quantitative measurements in vivo. *Neurosurgery*. 2014 Mar;10 Suppl 1(0 1):74–82; discussion 82–3. doi: 10.1227/NEU.0000000000000117. PMID: 23887194; PMCID: PMC4237006.
24. Vorsanova SG, Iourov IY, Yurov YB. Classical clinical cytogenetics (textbook) Moscow: The publishing house of The Russian Academy of Natural History; 2021.
  25. Yurov YB, Vorsanova SG, Iourov IY. Human Interphase Chromosomes: Biomedical Aspects. New York: Springer; 2013.

## Chapter 9. Assessment of color discrimination thresholds by the strict substitution method: a pilot study

Gracheva M.A.<sup>1,2</sup>, Belokopytov A. V.<sup>1</sup>, Timofeev V.A.<sup>1</sup>, Basova O.A.<sup>1</sup>, Maximov P. V.<sup>1</sup>, Nikolaev D. P.<sup>1</sup>

<sup>1</sup>*Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia;*

<sup>2</sup>*Institute for Biomedical Problems, Russian Academy of Sciences, Moscow, Russia*

### Abstract

**Introduction.** Color discrimination of human visual system has particular thresholds, which are usually described by color discrimination ellipses (MacAdam, 1942). Shape of the ellipses may depend on a number of experimental method parameters. Although the “strict substitution” method is considered to be the most accurate (Wyszecki, Stiles, 2000), the ellipses are usually obtained by comparison of two adjacent stimuli.

**The goal:** To estimate color discrimination thresholds with strict substitution method for five CIE-recommended color centers and to validate a new device for color threshold assessment.

**Methods.** After adaptation (1 min), the participant observes a 2° self-luminous stimulus. The initial color of the stimulus was set in xyY coordinates and corresponded to one of the five color centers (Robertson, 1978). The reference stimulus was replaced by a test one at a frequency of 0.33 Hz. The subject had to change the chromaticity coordinates of a test stimulus until the flicker became distinguishable. Test stimulus chromaticity was changed along eight directions (separated by 45°) on the xy plane. The resulting thresholds were compared with the predictions of the CIEDE2000 color difference formula.

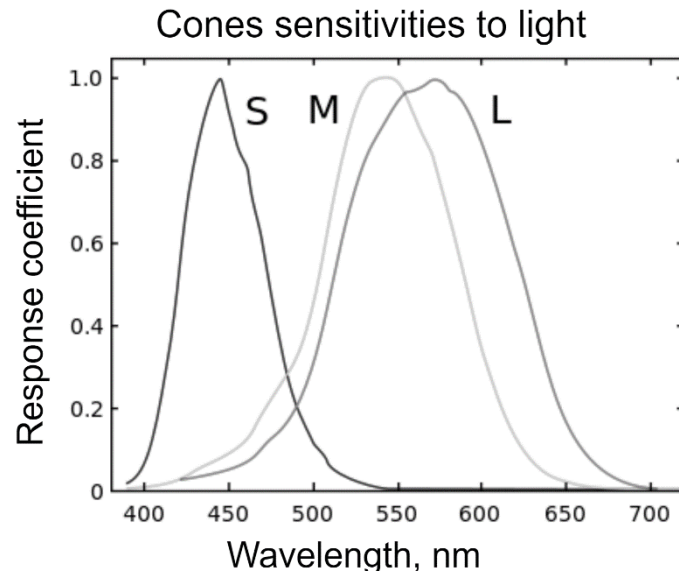
**Results.** STRESS metric results are (color center, STRESS value): red, 0.337; yellow, 0.119; green, 0.170; blue, 0.199; gray, 0.158.

**Conclusion.** The new device tested allowed us to obtain color vision thresholds by strict substitution method. The data obtained for all color centers except for the red one agree well with the CIEDE2000 formula and with other researchers’ data.

### Introduction

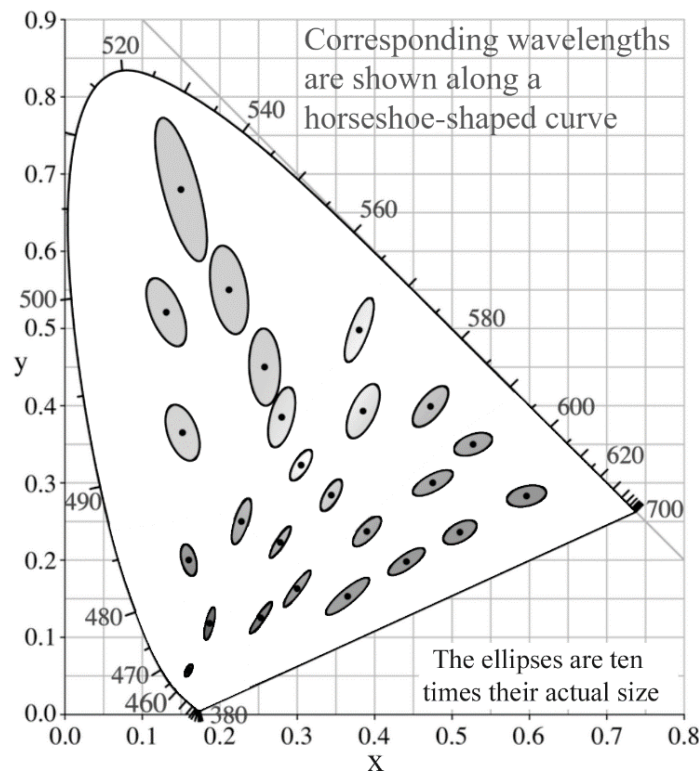
Human vision has an ability to discriminate colors, and that ability is based on several types of photoreceptors in our retina. For a long time, it was mainly accepted that color vision is based on three cone receptors only. Even though there is some evidence that both rods and intrinsically photosensitive retinal ganglion cells may contribute to the perceived colors (Graham, 2014; Graham et al., 2007; Zele et al., 2018), nevertheless most research and application for color vision still work with the three cones paradigm. Three cone based color vision models are also fully acceptable for central vision, since in the central part of the retina (foveola) humans do not have neither rods, nor ganglion cells.

Three types of cones usually discussed for color vision research are S-, M- and L-cones, differing in the position of their spectrum sensitivity peaks. Short wavelength sensitive cones (S-cones) have peaks at about 420–440 nm, middle wavelength sensitive ones have peaks at 534–545 nm, and long wavelength sensitive cones have peaks at about 564–580 nm. Example of sensitivity spectra is presented in Figure 1.



**Figure 1.** Cone sensitivity spectra. S, M and L are referring to the S-, M-, and L-cones, respectively. Image source: [https://en.wikipedia.org/wiki/Spectral\\_sensitivity#/media/File:Cones\\_SMJ2\\_E.svg](https://en.wikipedia.org/wiki/Spectral_sensitivity#/media/File:Cones_SMJ2_E.svg), CC BY3.0

To operate colors in practical needs many attempts were made to create color vision space and color vision models, and most of them are based on data for central color vision and three cones types. One of the most popular spaces are CIE1931, CIE1964, CIELUV, CIE Lab, and many others. An example of the entire range of possible chromaticities in CIE1931 color space is presented in Figure 2 by a horseshoe shaped outer curve.

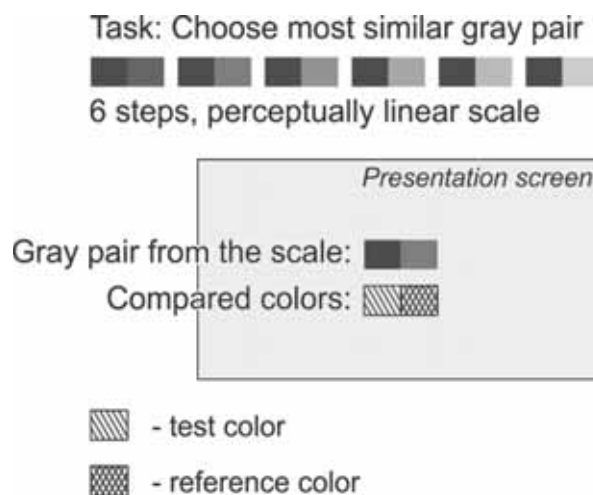


**Figure 2.** Entire range of possible chromaticities in CIE1931 xy chromaticity diagram and MacAdam ellipses of indistinguishable colors. The ellipses are ten times their actual size, as depicted in MacAdam's paper. Along a horseshoe-shaped outline of all possible colors, referring wavelengths are marked.

Intuitively easy to understand, that human color vision has some color resolution thresholds. If a person with normal trichromatic color vision looks at two colors, for example, yellow and violet, he/she could easily see the difference. If different colors are chosen, for example, violet and slightly more purplish violet, the difference between colors is not so obvious. At some point, such colors may be chosen, that even though they have different coordinates in color space, human participants are unable to see the difference.

In 1942 it was shown that in the CIE1931 xy chromaticity diagram the areas of indistinguishable colors have the form of ellipses of different size, axes angles and ratios (see Figure 2). The work was conducted by David MacAdam (MacAdam, 1942), and the color threshold ellipses are often called after him as MacAdam ellipses. For people with normal (trichromatic) color vision those differences in ellipse forms do not reflect any properties of human color vision. The ellipse's different shapes only reflect that the color space is not perfect, or rather it is not uniform in color contrast.

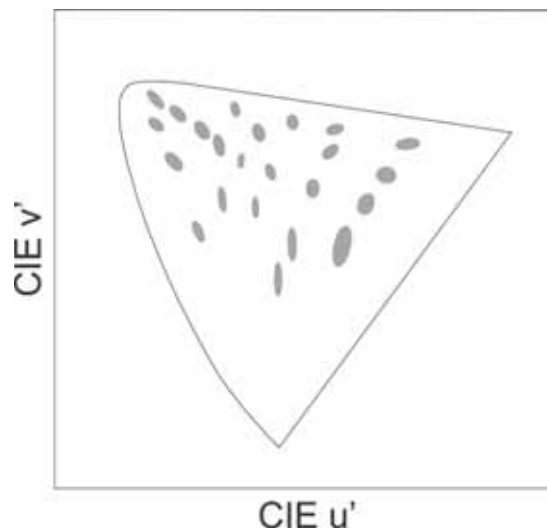
Since the famous MacAdam paper in 1942, a lot of work has been devoted to assessment of color thresholds and perceived color differences. There are several most famous papers on threshold assessment (Silberstein, MacAdam, 1945; Brown, Macadam, 1949; Brown, 1957; Wyszecki, Fielder, 1971; etc.), a concise and nicely written summary of which could be found in (Wyszecki, Stiles, 2000, ch. 5). Another type of research is devoted to assessment of suprathreshold differences: in that case a participant could clearly see the difference between two colors and have to assess the amplitude of such difference. One of the most common experimental designs for such assessments includes comparison with gray scale (see Figure 3 for an example). Some recent examples of suprathreshold studies may be found in (Xu et al., 2021; Zhao et al., 2020; Mirjalili et al., 2019; Xu et al., 2019). Both types of those studies, threshold and suprathreshold, are essential for further progress of color science.



**Figure 3.** An example of experimental design for small color differences assessment (suprathreshold differences). A pair of colored samples are presented (test color and reference color) on a presentation screen or a presentation surface. A set of gray pairs are presented as a scale. Each pair may be chosen and positioned near the colored samples. A participant's task is to choose a gray pair that is the most similar in difference to the colored pair.

After CIE1931, many attempts were made to develop uniform color space. In such a uniform space all MacAdam ellipses should look like circles of equal sizes. In Figure 4, MacAdam ellipses are presented in CIELUV 1976 chromaticity space. It is obvious that ellipses in Figure 4 are more similar to each other than in Figure 2 that presents them in CIE xyY1931.

Nevertheless, even in CIELUV there are some strong disparities in size, angle and axis ratios of ellipses.



**Figure 4.** MacAdam ellipses plotted in CIELUV 1976 chromaticity space. Comparing those ellipses to the ones in CIE  $xyY1931$  (Figure 2) one could easily see that in CIELUV ellipses are much more similar to each other. Nevertheless, there are still some strong disparities in size, angle and axis ratios of ellipses, meaning the space is not purely uniform.

Understanding color differences have quite clear significance for industry and practice: to operate with colors (which is essential in polygraphy, textile industry, design, in image processing, etc.) one has to understand what colors and their combinations would be perceived by a real human observer. That is why a field of uniform color spaces is still blossoming with new color models, new color spaces and new color difference formulas.

Verification of results of all these scientific developments requires a big amount of data obtained in real human participants. There are some datasets available, for example (Luo, Rigg, 1986; Alman et al., 1989; Berns et al., 1991; Witt, 1995; Witt, 1999; etc.). Most of the existing datasets have limited coverage in color space. Nowadays, with industry progress and with the emergence of a new generation of displays, new datasets are required. New experiments have to cover a wide color gamut and a high dynamic range of luminance.

In this paper, we present pilot experimental results of color threshold assessment by a new developed device with a wide color gamut. The device itself is described in another chapter of this book.

#### **The aims of the research:**

- To estimate color discrimination thresholds with strict substitution method for five CIE-recommended color centers.
- To validate the developed device.
- To validate the experimental protocol.

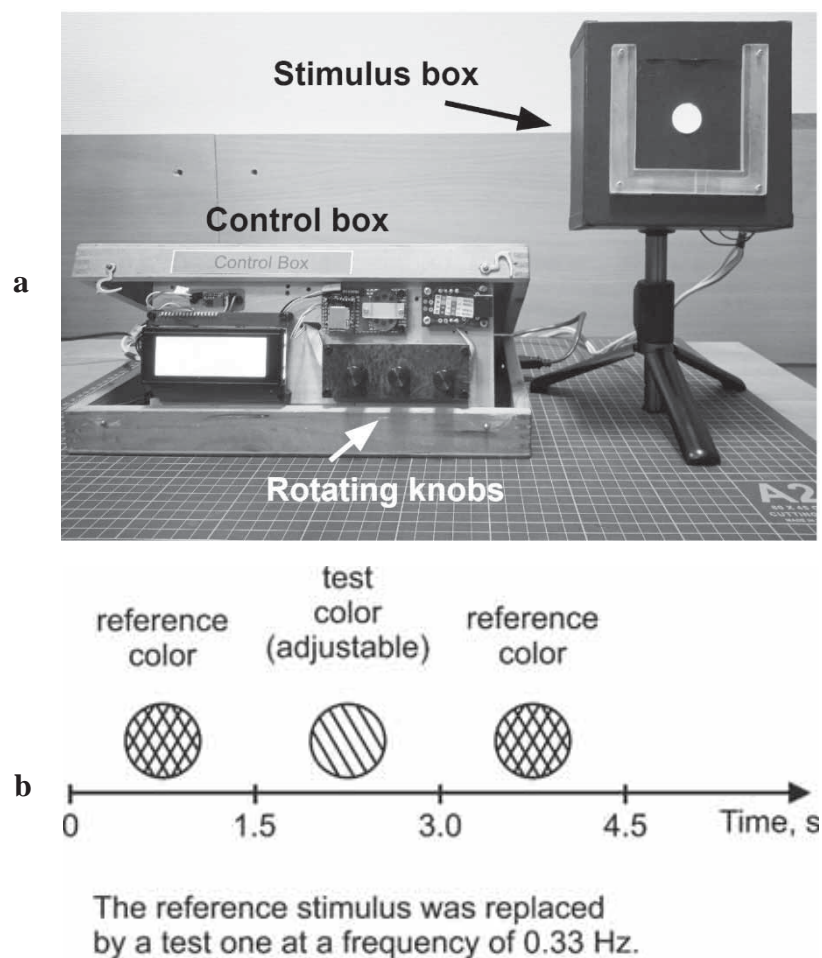
## **Materials and Methods**

### ***The device***

The device is developed in IITP RAS and is fully described in another chapter of this book by Alexander V. Belokopytov.

The device consists of a control box and a stimulus box (Figure 5, a). The control box has an auxiliary screen to display color coordinates of presented colors and rotating knobs (encoders) to change the stimuli. The stimulus box has a display and adjustable apertures to regulate the size of the stimulus. For this study, we used an aperture that provided  $2^\circ$  stimulus from 60 cm viewing distance.

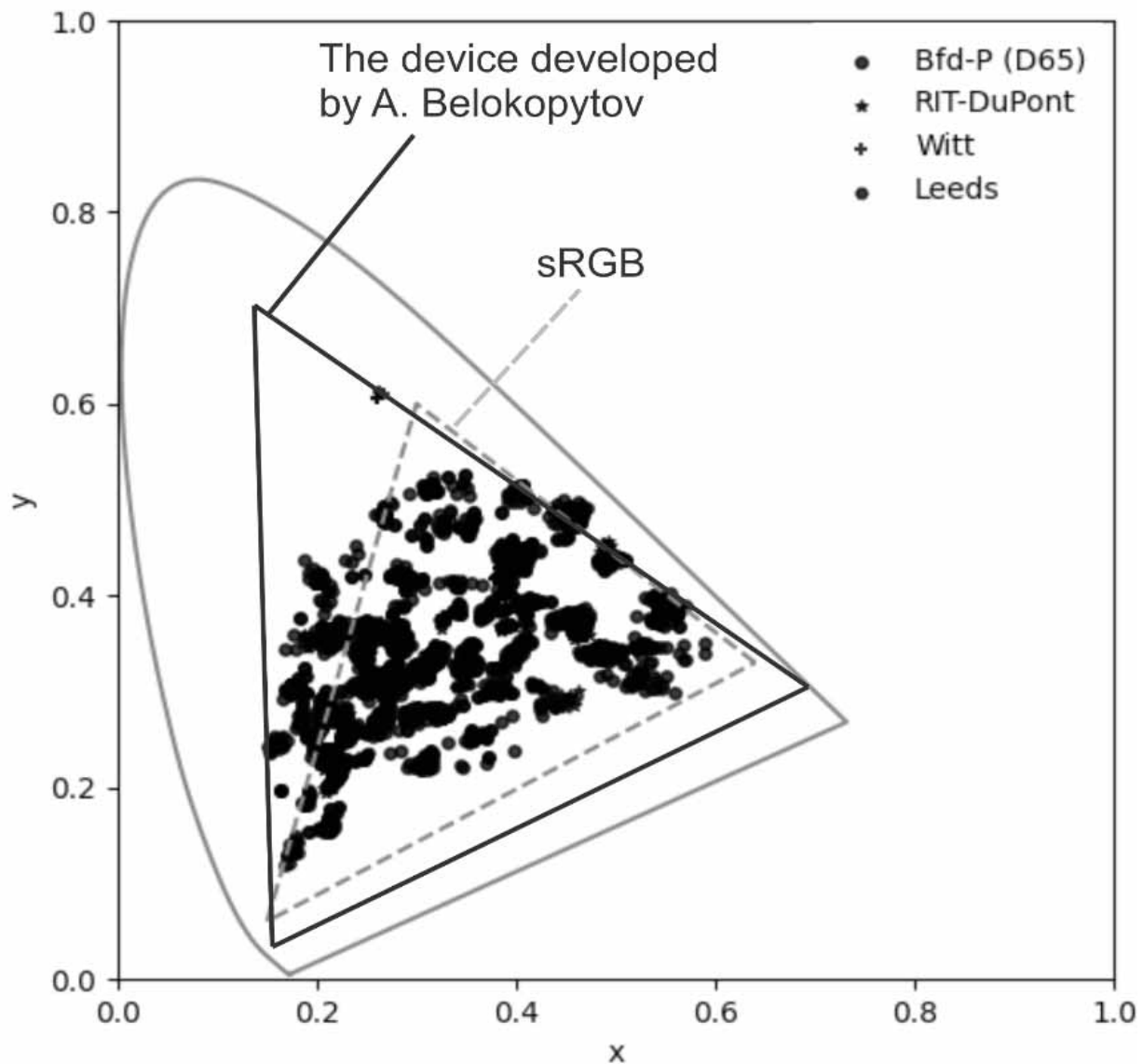
The stimuli presentation was made according to a strict substitution method (Wysecki, Stiles, 2000, pp. 278–293). In strict substitution procedure, test and reference colors are positioned in the same location (in contrast to bipartite procedures) and interchange in time (Figure 5, b). In our experiment, the colors change with 0.33 Hz frequency. The strict substitution method is sometimes considered as the best method for color comparison, since in such a procedure the same retinal area is stimulated by both stimuli.



**Figure 5.** The apparatus used and its description.

- a) The device was developed in IITP by Alexander V. Belokopytov. The device consists of two main parts: control box and stimulus box. Control box has a screen to output the color coordinates (in CIE1931 xyY color coordinate system) and rotating knobs to change the color. Additionally, a separate part with rotating knobs may be added for better participant experience. Stimulus box consists of a LED matrix, cooling fan, temperature sensor and a changeable aperture.
- b) The scheme of a stimulus presentation in a strict substitution method. The test and reference stimuli are interchangeable in time; the change frequency is 0.33 Hz. After 1.5 s of reference stimulus, the test stimulus is presented for 1.5 sec, and so on.

The main benefit of the device is its wide color gamut, providing opportunity to collect the data in a wide color range. Figure 6 presents several datasets and distribution of the samples in color range, standard sRGB gamut (dashed gray line), and color gamut of the device used in this study (solid black line) in a CIE1931 chromaticity diagram. All color datasets presented do not cover the full range of possible colors. New device covers much larger area in the whole color range and could provide an opportunity to extend our data significantly, especially in the area of green-cyan (upper left part of the CIE diagram) and red-purple colors (lower right part of the CIE diagram, bordered by the straight line).



**Figure 6.** The diagram presents the CIE  $xy$  chromaticity range with constant brightness and color ranges for devices and datasets. The color gamut of the new device (solid line), the standard sRGB gamut (dashed line), and the range of samples in various color difference datasets (the datasets presented are: Bfd-P (D65), RIT-DuPont, Witt, Leeds). The developed device provides great opportunities for color threshold assessment in green-cyan (upper left part of the CIE diagram) and red-purple areas (lower right part of the CIE diagram, bordered by the straight line).

## The procedure

The whole procedure tested was as follows:

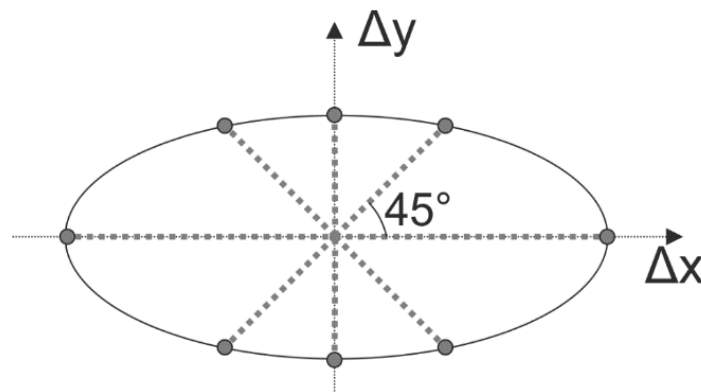
1. Participant's color vision was tested by CAD test (City Occupational Ltd, Great Britain). Only participants with normal (trichromatic without abnormalities) color vision may participate in such studies. We use full binocular medical protocol (approximately 20 min).
2. Adaptation to the dark surround (at least 1 min) . The experiment was conducted in the dark surround. According to (Fairchild, 1993), the adaptation of 1 min is enough for such conditions.
3. The initial color of the stimulus was set in xyY coordinates and corresponded to one of the five color centers (Robertson, 1978), see Table 1.
4. The participant changed the chromaticity coordinates of a test stimulus (rotating the knob) until the flicker of test and reference stimuli became distinguishable. The knob rotation changes color in one of eight directions around the initial color center, see Figure 7. Each direction is separated by  $45^\circ$  in the xy diagram. Measuring directions were randomized inside 5 blocks (blocks corresponded to the color centers).

On this stage of research, only one participant was tested.

The whole procedure duration was about 2 hours (breaks are excluded). The sessions were separated into 30-min intervals to lower participant's fatigue and stress.

**Table 1.** The coordinates of five CIE recommended color centers used in the study to assess threshold values (Robertson, 1978).

Color center	Y	x	y
Yellow	69.3	0.338	0.4228
Green	24.0	0.248	0.362
Blue	8.8	0.219	0.216
Gray	30	0.314	0.331
Red	14.1	0.484	0.342



*Figure 7. Directions for threshold assessments (around each color center).*

Test stimulus chromaticity was changed along eight directions (separated by  $45^\circ$ ) on the xy plane.

## Results

### *Results: comparison with CIEDE2000 by STRESS metric*

The resulting thresholds were compared with the predictions of the CIEDE2000 color difference formula (Luo et al., 2001). For an assessment of the proximity of perceived and computed color differences, the Standardized Residual Sum of Squares (STRESS) metric is often used (Garcia et al., 2007), see equation 1.

$$STRESS(\vec{g}, \vec{h}) = \frac{\|k^*\vec{g} - \vec{h}\|_2}{\|\vec{h}\|_2} = \frac{\sqrt{\sum_{i=1}^n (k^*g_i - h_i)^2}}{\sqrt{\sum_{i=1}^n h_i^2}}, k^* = \frac{\vec{g}^T \vec{h}}{\vec{g}^T \vec{g}}, \quad (1)$$

where  $\vec{g}, \vec{h}$  – vectors of color differences to be compared.

STRESS metric can be computed for two vectors of color differences obtained by arbitrary means. The metric values are always between 0 and 1, where 0 indicates the best agreement, and  $STRESS < 0.3$  may be considered as a good agreement.

The results of STRESS metric for all five color centers are presented in Table 2. We obtained good agreement of our threshold results and CIEDE2000 formula for four color centers out of five (yellow, gray, green, and blue), and marginal agreement for red color center.

**Table 2.** The comparison of the obtained thresholds with the CIEDE2000 by STRESS metric for five color centers.

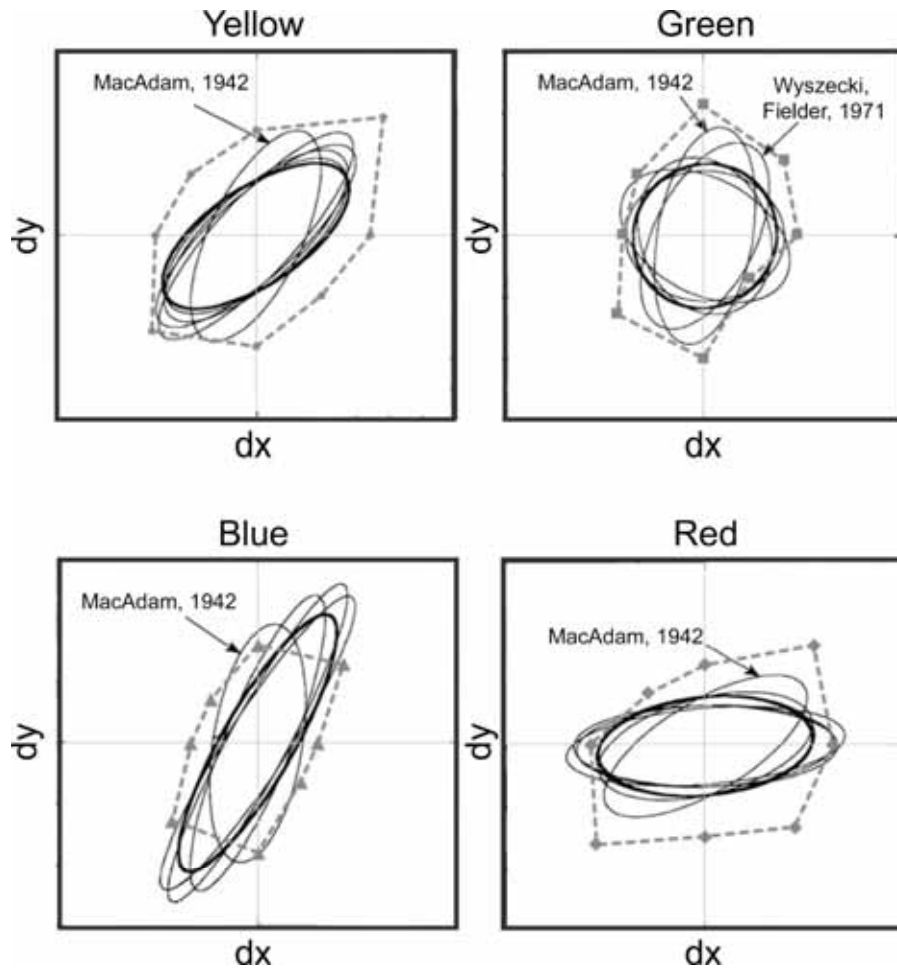
Color center	STRESS value
Yellow	0.119
Gray	0.158
Green	0.170
Blue	0.199
Red	0.337

### *Results: graphical comparison with other researchers' data*

Another way to compare obtained results is to plot them against the results of other researchers. To do so, we took summary data from (Berns et al., 1991) and from (Mirjalili et al. 2019). Both studies summarize several other researchers.

Figure 8 presents graphical comparison of results by different researchers for four color centers. The summary results from Berns and coauthors (Berns et al., 1991) are presented as black ellipses. The authors joined various experimental data obtained around particular color centers, and the data contains both threshold and suprathreshold measurements. Obviously, the sizes of ellipses should be different for different procedures, that is why authors normalized all ellipses to equal area. Our data are presented as 8 dots for each color center joined with a dashed line to assess the whole form of the obtained figure. Since the sizes of black ellipses are normalized, there is no sense in size comparison. The whole form and the tilt of the figure is to be compared.

Berns and coauthors noted that some ellipses were significantly different in comparison with other results. These are MacAdam data (for all four color centers) and Wyszecki and Fielder data (for green center).

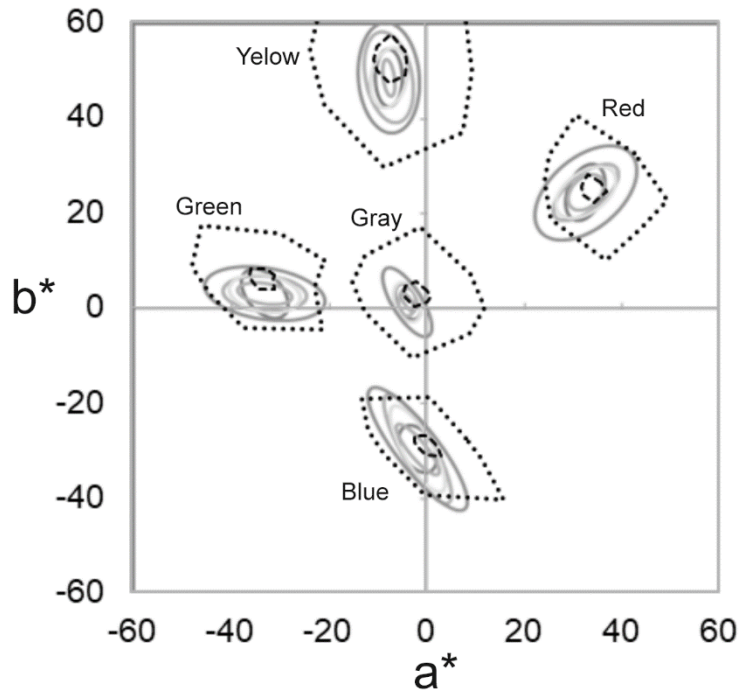


**Figure 8.** Comparison of the obtained threshold results with those of other researchers (Berns et al., 1991), plotted in CIE1931  $xyY$  coordinates. Our results are presented as dots (8 dots for each color center), joined by dashed lines for approximation of the whole form of the figure obtained. The results of previous researchers are presented by ellipses (the figures are taken from (Berns et al., 1991)).

In the plot, different threshold and suprathreshold measurements were joined, that is why the size of the figures is not to be compared: all black ellipses are normalized to equal areas, so the scales are relative and have no numbers. The whole form of figures and the tilt direction are to be compared.

Berns and coauthors marked the ellipses that were different from all other authors, that are MacAdam results for all four color centers (MacAdam, 1942) and Wyszecki and Fielder results for the green color center (Wyszecki, Fielder, 1971).

Figure 9 presents graphical comparison of the results obtained by different researchers for five color centers. The summarized results from the work of Mirjalili and coauthors are presented as gray ellipses (Mirjalili et al., 2019). As in the previous plot, the authors joined various experimental data obtained around particular color centers, and the data contains both threshold and suprathreshold measurements. Our data are presented as 8 dots for each color center joined with a dashed line to assess the whole form of the obtained figure. Since the ellipses from (Mirjalili et al., 2019) paper summarize both threshold and suprathreshold results, the size of the shapes should not be compared, only the form and the tilt of the figure is under consideration. For better visibility, our data were enlarged 5 times and presented by dots joined with black dotted lines.



**Figure 9.** Comparison of the obtained threshold results with those of other researchers (Mirjalili et al., 2019) plotted in Lab coordinates. Our results (transferred to Lab through D65) are presented as dots, joined by black dashed lines for approximation of the whole form of the figure obtained. For illustration, the same results were enlarged 5 times and were plotted by dotted lines. The results of previous researchers are presented by gray ellipses (the figures are taken from (Mirjalili et al., 2019)). In this plot, different threshold and suprathreshold measurements were joined, that is why the size of the figure is not to be compared. The whole form and the tilt direction are to be compared.

## Discussion

### Results of the pilot color thresholds assessment

In this paper we present pilot results of color threshold assessment by a new device developed at the IITP RAS by Alexander V. Belokopytov. The device provides a wide color gamut and is designed for a strict substitution comparison method, in which both stimuli (test and reference) stimulate the same retinal area. This paper provides first experimental results by this new promising device.

The results of STRESS metric for comparison of CIEDE2000 formulas shows reasonable results for all color centers. While for red color center results are slightly over the notional good quality margin, it is not much worse than for other centers (while  $\text{STRESS} < 0.3$  is considered as good quality, we obtained 0.337 for red color center). These results show that the new device could give reasonable color threshold assessments.

The results of graphical comparison with the data summarized by Berns and colleagues (Berns et al., 1991) show that even with a scarce amount of pilot data the whole form of threshold ellipses is similar to the ones of previous researchers. It is noteworthy that our threshold figures seem to tend to the ones obtained by MacAdam, especially for red and green color centers. MacAdam used bipartite field (in which one half of the field is filled with test color, and the other half is filled with reference color) and had big white surround of half brightness of the stimulus (the surround was of  $40^\circ$  of visual field and had brightness of  $24 \text{ cd/m}^2$  while

brightness of stimuli was 48 cd/m<sup>2</sup>). In that sense, neither our stimulus structure, nor our dark surround were similar. Nevertheless, in comparison to many other researchers (Alman et al., 1989; Berns et al., 1991; Witt, 1995; Witt, 1999; Mirjalili et al., 2019), both in our work and in the work of MacAdam the test field had 2° size. Since the inhomogeneity of retina is well-known, and may affect the color assessment results significantly, it is possible that the stimulus size was the crucial parameter for threshold ellipses form.

The results of graphical comparison with summary made by Mirjalili and colleagues (Mirjalili et al., 2019) show a similar tendency. The whole form and the tilt of the threshold figures is quite similar for our data and for data of other researchers. The less similarity could be seen for the red color center, that is in agreement with our STRESS metric results, where the red color center also showed less quality. One of the possible explanations has already been discussed above: it is possible that for red stimuli the size of a test field is of critical importance.

### ***Limitations of the study***

In this study we conducted only pilot experiments with the main goal in mind to be a validation of the new device. Obviously, the data of one participant is far from being enough to assess reliable color thresholds by strict substitution method.

In this study we assessed thresholds only in direction from equal stimuli to the point where the difference is seen (see Figure 7 and the subsection “The procedure” of “Materials and Methods” section). The thresholds may be slightly different if measured in the other direction (from obviously different stimuli to the point of indistinguishable difference). Even though we are aware of the possible discrepancy of various results, the procedure was chosen since it was much easier to conduct, much simpler for participants to understand, and invoked much less stress and fatigue in participants.

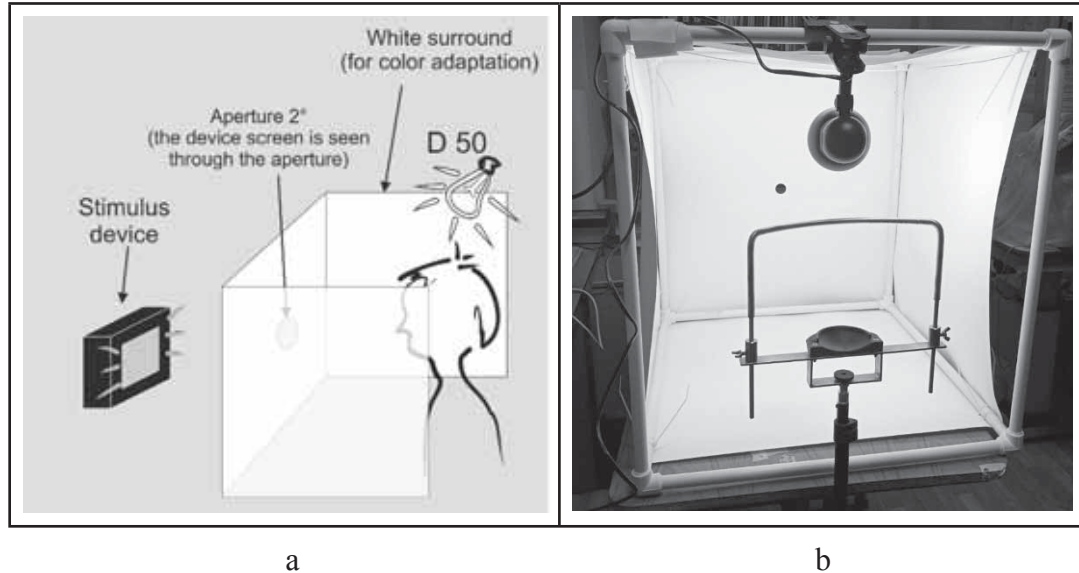
Another limitation of the study concerns the transformation of our results to the Lab coordinates. Since we conducted our studies in a darkened room, there is an uncertainty about the light source that should be used for transformation from CIE1931 xyY to Lab. We used D65 for this purpose, though the choice is not supported by our physical conditions. The transformation to Lab coordinates was used for CIEDE2000 results calculation (see subsection “Results: comparison with CIEDE2000 by STRESS metric”) and for Figure 9. Even though those results may be different if another light source were used, we strongly believe that the difference would not change the main conclusions of our study.

### ***Further work***

The goal of the study was to estimate color discrimination thresholds with strict substitution method for five CIE-recommended color centers and to validate a new device for color threshold assessment. To validate the new device, we conducted threshold assessment around five CIE recommended color centers (Robertson, 1978) since in these areas there are enough data available for comparison. In further research, areas over sRGB gamut would be of most interest.

To improve the experimental conditions in future, we developed an adaptation cabin (Figure 10). The main benefit of the cabin is standardized adaptation conditions for all participants. The cabin has size of about 80\*80\*80 cm. In the front side, there is a chinrest for better fixation of the participant’s head and for stable viewing distance (see Figure 10, b). On the opposite side, there is an aperture providing 2° of visual stimuli. To eliminate any incident light

to be reflected by the stimulus display, the display is positioned at some distance from the aperture (see Figure 10, a). The cabin provides a white surrounding field and a homogenous lighting condition by D50 light sources with regulated brightness. The cabin is placed in a dark room with no windows, so the only light source is D50 lamps. In our further research, we are planning to conduct all our experiments on threshold assessment with this adaptation cabin.



**Figure 10.** A scheme (a) and a photo (b) of an adaptation cabin developed. To better control adaptation conditions, the adaptation cabin is installed in a dark room (with no windows). Five sides of the cabin are covered with white paper, at the other side there is a chinrest and a forehead support for better fixation of a viewing distance. The opposite side has an aperture of  $2^\circ$  to see the device display. The device display is positioned at some distance from the backside of a cabin to ensure that no incident light is reflected by the stimulus screen.

To improve the procedure, we are planning to conduct measurements of threshold points in both directions: from indistinguishable difference to distinguishable flickering, and vice versa. This change may be more tiresome for participants, so additional studies are needed to validate the procedure change.

To improve the data obtained and conclusions drawn from it, the main goal is to increase the data set by recruiting more participants to the study.

## Conclusions

- The device developed allowed us to obtain color vision threshold assessments by strict substitution method for five CIE recommended color centers.
- The data obtained is in good agreement with CIEDE2000 (by STRESS metric) and with the data from other researchers (by graphical comparison).
- However, agreement for the CIE Red color center was marginal.
- CIE Red color center in our data shows more similarity to the MacAdam 1942 results, than to further investigators.
- The device and experimental procedure with described improvements seems promising for collecting dataset on human color vision threshold in a wide range of colors.

## References

1. Alman D. H., Berns R. S., Snyder G. D., Larsen W. A. Performance testing of color difference metrics using a color tolerance dataset. *Color Research & Application*. 1989. 14(3): 139–151. <https://doi.org/10.1002/col.5080140308>
2. Berns R. S., Alman D. H., Reniff L., Snyder G. D., Balonon Rosen M. R. Visual determination of suprathreshold color difference tolerances using probit analysis. *Color Research & Application*. 1991. 16(5): 297–316. <https://doi.org/10.1002/col.5080160505>
3. Brown W. R. J., MacAdam D. L. Visual sensitivities to combined chromaticity and luminance difference. *JOSA*. 1949. 39(10):808–834. <https://doi.org/10.1364/JOSA.39.000808>
4. Brown W. R. J. Color discrimination of twelve observers. *JOSA*. 1957. 47(2):137–143. <https://doi.org/10.1364/JOSA.47.000137>
5. Fairchild M. Chromatic adaptation in hardcopy/softcopy comparisons, in *Color Hard Copy Graphic Arts II*. J. Bares, Ed., Proc. SPIE. 1993. 1912.
6. Garcia P. A., Huertas R., Melgosa M., Cui G. Measurement of the relationship between perceived and computed color differences. *JOSA A*. 2007. 24(7):1823–1829. <https://doi.org/10.1364/JOSAA.24.001823>
7. Graham D. Melanopsin ganglion cells: a bit of fly in the mammalian eye. 2014. Retrieved 2022, December 12 from: <http://webvision.umh.es/webvision/Melanopsin.html>
8. Graham D., Wong K. Y., Pattabiraman K., Berson D. M. A Bit of fly in the mammalian eye. *Investigative Ophthalmology & Visual Science*. 2007. 48(13): 2850.
9. Luo M. R., Rigg B. Chromaticity discrimination ellipses for surface colours. *Color Research & Application*. 1986. 11(1):25–42. <https://doi.org/10.1002/col.5080110107>
10. Luo M. R., Cui G., Rigg B. The development of the CIE2000 colour difference formula: CIEDE2000. *Color Research & Application*. 2001. 26(5):340–350. <https://doi.org/10.1002/col.1049>
11. MacAdam D. L. Visual sensitivities to color differences in daylight. *JOSA*. 1942. 32(5):247–274. <https://doi.org/10.1364/JOSA.32.000247>
12. Mirjalili F., Luo M. R., Cui G., Morovic J. Color-difference formula for evaluating color pairs with no separation:  $\Delta E_{NS}$ . *Journal of the Optical Society of America A*, 36(5):789–799. <https://doi.org/10.1364/JOSAA.36.000789>
13. Robertson A. CIE Guidelines for coordinated research on colour difference evaluation. *Color Research & Application*. 1978. 3(3):149–151.
14. Silberstein L., MacAdam D. L. The distribution of color matchings around a color center. *JOSA*. 1945. 35(1):32–39. <https://doi.org/10.1364/JOSA.35.000032>
15. Xu Q., Cui G., Safdar M., Xu L., Luo M. R., Zhao B. Assessing colour differences under a wide range of luminance levels using surface and display colours. *Final Program and Proceedings–IS and T/SID Color Imaging Conference*. 2019.:355–359. DOI: 10.2352/issn.2169–2629.2019.27.64
16. Xu Q., Zhao B., Cui G., Luo M. R. Testing uniform colour spaces using colour differences of a wide colour gamut. *Optics Express*. 2021. 29(5):7778–7793. <https://doi.org/10.1364/oe.413985>
17. Witt K. Linearity and additivity of small color differences. *Color Research & Application*. 1995. 20(1):36–43. <https://doi.org/10.1002/col.5080200107>
18. Witt K. Geometric relations between scales of small colour differences. *Color Research and Application*. 1999. 24(2):78–92. [https://doi.org/10.1002/\(SICI\)1520-6378\(199904\)24:2<78::AID-COL3>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1520-6378(199904)24:2<78::AID-COL3>3.0.CO;2-M)
19. Wyszecki G., Fielder G. H. New color matching ellipses. *JOSA*. 1971. 61:1135–1152. <https://doi.org/10.1364/JOSA.61.001135>
20. Wyszecki G., Stiles W. S. *Color science. Concepts and methods, quantitative data and formulae*. John Wiley and Sons, 2000. 976 p.
21. Zele A. J., Feigl B., Adhikari P., Maynard M. L., Cao D. Melanopsin photoreception

- contributes to human visual detection, temporal and colour processing. *Scientific Reports*. 2018. 8(1):3842. <https://doi.org/10.1038/s41598-018-22197-w>
22. Zhao B., Xu Q., Luo M.R. Color difference evaluation for wide-color-gamut displays. *JOSA A*. 2020. 37(8):1257–1265. <https://doi.org/10.1364/JOSAA.394132>

## **Chapter 10. Non-visual eye movements during memorization, long-term and short-term memory retrieval on the blank display: a pilot study**

*Lekhnikskaya P.*

*Kazan Federal University, Russia*

People's eyes are constantly moving, whether for visual processing (Henderson, J., 2003; Damiano, C., 2019) or for thinking (Liversedge, S. P. et al., 2000; Rayner, K., 1998, Rayner, K., 2009; Theeuwes, J. et al., 2009, Yarbush, A. L. 1967; Damiano, C., 2019). If eye movements during scene perception may be explained by the need to understand the world around us, then thinking is a rather complicated process. To the best of our knowledge, thinking encompasses a wide range of components, including imagining, task solving, memory, and so on; all of them are in a strong bound.

The pioneer work in eye movements during thinking emerged in the 1960s and 1970s, it was revealed that while people were answering questions, they typically shifted their gaze away from the questioner (Ehrlichman, H. et al., 2012). Ehrlichman, H., & Micic, D. named such eye movements “non-visual.” These eye movements are independent of visual processing (Ehrlichman, H. et al., 2012). Non-visual eye movements can be explored in the “blank screen paradigm” (Altmann, G. T., 2004), in which we show a participant the task, after that he solves this mentally by looking at the blank screen. In this way, the role of spatial-numerical associations in number categorization tasks in mental arithmetic has been established (Hartmann, M. et al., 2015), the difference between convergent thinking and divergent thinking has been shown (Maheshwari, S et al., 2021), it has been revealed that eye movements are used to coordinate elements of a mental model with elements of the visual field (Spivey, M. J et al., 2001), a novel approach to understanding the potentially causal role of visual imagery on music-induced emotion has been represented (Hashim S et al., 2020). Mental imagery is the phenomenon in which someone imagines an object or a visual scene in his or her “mind’s eye” in order to retrieve information from that mental image or to transform it so as to generate needed information (Yoon, D., 2004; Shepard, R.N. et al., 1986). Laeng and Teodorescu research proved that eye movements are stored as a spatial index that is used to arrange the component parts correctly when mental images are generated (Mast, F., 2002; Laeng, B. t al., 2002). Therefore, the blank screen paradigm contributes to research on mental imagery for which the role of memory is of high importance.

Based on the idea that eye movements are somehow connected with memory retrieval, Hebb explored mental imagery: “If the image is a reinstatement of the perceptual process, it should include the eye movements.” Here oculomotor system connects perception and imagery (Hebb, D. O., 1986; Sheehan, P.W. et al., 1969), this principle was formalized by Noton and Stark’s (Noton, D. et al., 1971; Noton, D. et al., 1971) as scanpath theory (Wynn, J. S., 2019), recognition occurs when the same scanning pattern enacted during initial viewings is recapitulated during subsequent viewing and recognition of a stimulus (Hannula, D. E., 2010). Scanpath theory’s legacy lies in its proposal that eye movements play a functional role in retrieval, an idea that has been critical in shaping current models of memory and visual attention (Wynn, J. S., 2019). In conjunction with this theory, research of Johansson R. et al. suggests that distinct scanpath properties differentially contribute depending on the nature of the goal-relevant memory (Johansson R et al., 2022). Recent articles also indicate the positive correlation between spontaneous eye movements and episodic recollections (Johansson R et al., 2022; Brandt, S. A. et al. 1997; Laeng, B. et al, 2002; Gbadamosi, J. et al, 2001; Johans-

son, R. et al, 2006; Johansson, R. et al., 2012, Richardson, D. C. et al, 2000; Ryan, J. D. et al, 2000). Eye movements may be stored in memory along with the images that were encoded at each fixation (Mast, F., 2002; Kosslyn, S.M., 1994). It has been proven that such eye movements with selective attention serve to facilitate processing in memory encoding and retrieval (Hannula, D. E., 2010; Henderson, J. M. et al, 2005; Noton, D. et al, 1971; Damiano, C. et al, 2019), oculomotor movements may serve as self-generated cues, pointing to stored memories (Johansson R. et al, 2022).

Another sustained interest in interrelation between eye movements and memory emerged in 1976, when Yarbus (1976) showed that eye movements with memory are strongly correlated (Hannula, D. E., 2010). Thus, a sufficient base of research on the relationship between memory and eye movements was accumulated. We can assess memory by using eye movements under circumstances in which behavioral reports may not be reliably obtained. Hannula, D. E. made a review in which came to the conclusion that memory for different aspects of experience (e.g., individual items, spatial and non-spatial relationships, temporal order) guide eye movement behavior (Hannula, D. E., 2010). Eye movement during past and future thinking (El Haj, M et al., 2018), the role of eye movements in retrieval (Wynn, J. S., 2019), eye movements during long-term pictorial recall (Martarelli, C. S. et al., 2012) were investigated. Eye Movement Desensitization and Reprocessing was created in which the patient is urged to retrieve memories of the traumatic event while laterally moving the eyes (Hout, M. et al., 2001, Shapiro, F., 1995), French participants were asked to mentally visualize a map of France while their vocal response times and eye movements were recorded (Bourlon, C., 2011). Thus, taken together, findings from behavioral research converge on a key role for eye movements in memory.

However, research has yet not reached consensus on whether and how eye movements are connected with long-term and short-term memory retrieval of non-pictorial information and its memorization. Similar research was conducted by Ehrlichman, H., & Micic, D.. Their hypothesis was that searching for information in long-term memory generates saccadic activity, whereas maintenance of information in working memory inhibits saccadic activity (Ehrlichman, H. et al, 2012). But they used facial expressions as visual stimulus; we interested in how people move their eyes during retrieval of the poems.

## **Method**

The accumulated experience formed the basis of the forthcoming study and hypotheses:

H1: There are differences in eye movements when retrieving information from short-term memory, long-term memory, during memorization. Eye movements can be automatically divided according to their nature: short-term memory, long-term memory, memorization.

H2: Those eye movements during memorization and reproduction that do not differ are interconnected.

H3: It is possible to automatically identify groups according to eye movements during memorization in memorization, short-term and long-term memory.

In our study, we considered memory as long-term and short-term. Two possible ways in which a difference between short- and long-term memory stores may differ are in duration, and in capacity (Cowan, N., 2008). Atkinson and Shiffrin (1968) assessed short term memory like the faculties of the human mind that can hold a limited amount of information in a very accessible state temporarily (Cowan, N., 2008; Atkinson, R.C. et al, 1968). In contrary, long-term

memory is a vast store of knowledge and a record of prior events, and it exists according to all theoretical views; it would be difficult to deny that each normal person has a rich set of long-term memories (Cowan, N., 2008).

## **2.1. Participant**

A total of 35 participants were recruited from university students. Data from one participant had to be discarded due to technical problems with eye-tracking. Therefore, the eye-tracking results are based on data collected from 34 subjects (29 female) with normal or corrected-to-normal vision. All of the participants are students at the Kazan (Volga Region) Federal University, pedagogical (with major in languages), philological areas of training (mean age = 20 years, SD = 2,4). Participants were given the following instructions: In this experiment you will be asked to try to retell, looking at the screen, the first 7 lines of “On seashore far a green oak towers...”. Next, a small poem will be given, try to learn it and retell it, also looking at the screen.

## **2.2. Procedures**

The experiment consisted of three stages: reproduction of the first 7 lines of the poem “On seashore far a green oak towers...” by A. S. Pushkin, memorization of the poem “A wave-kissed land is known to me...” by A. S. Pushkin (the version with 7 lines from the website “Культура.рф”) and reproduction of the second poem.

The choice of the poem “On seashore far a green oak towers...” was due to the fact that it is usually studied in elementary school and most people remember it, i.e. its retrieval occurs from long-term memory. “A wave-kissed land is known to me...” is usually not recited by heart in schools, the participants did not know this poem and taught it during the experiment, i.e. the retrieval of this poem occurred from short-term memory. This study was carried out basing on blank screen paradigm. A white background was displayed on the screen for retelling the poem, which is why there were no eye fixations caused by the presence of any extraneous image. This allowed tracking eye movements that were directly related to the reproduction of information.

## **2.3. Materials and methods**

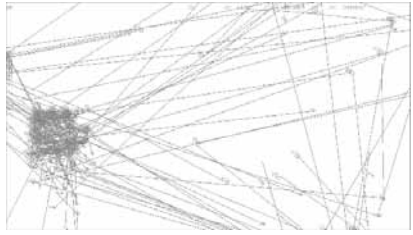
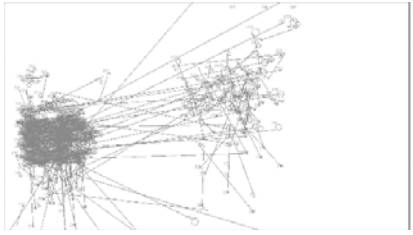

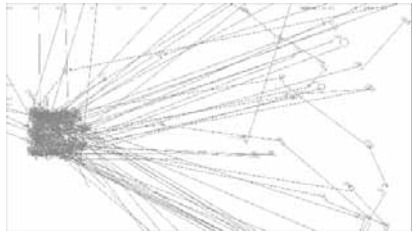
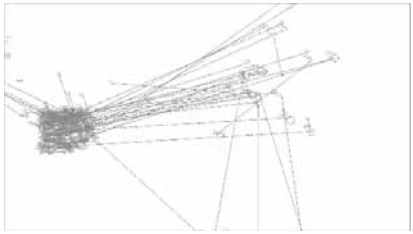
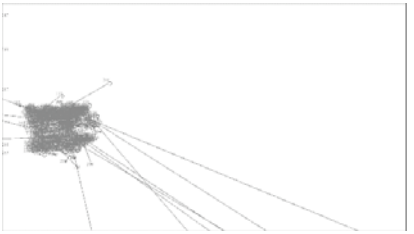
Eye tracking (SR Research EyeLink 1000 Plus, sampling frequency 1000 Hz) was chosen as a research method. The resolution of the monitor screen is 1920×1080, a chin rest was used to fix the position of the head. Calibration was carried out at the beginning of the experiment. Nine points were involved in the calibration process, calibration was accepted if the average error was <0.30° of the viewing angle, and the maximum error was <0.60°. Before proceeding to the next stage, the drift was checked, the entire course of the experiment was controlled by the participant. During the experiment, only the subject and the researcher were present in the laboratory.

### **2.3.1 Preprocessing**

Before conducting statistical analysis, we processed the eye-tracking data as follows: saccades and fixations associated with the reading of the poem were removed. Thus, those eye movements that were associated with the mental reproduction of the poem, looking at the free part of the monitor, were preserved. We also took a look at Gaze Plots in memorization and divided all movements into conditional three groups. Eye movements were considered

grouped if the concentration of fixations in a certain part of the screen was visually traced. For scattered eye movements we will consider a chaotic eye movement that is not concentrated in any particular part of the monitor. Without derivation from the main stimulus, we will consider such eye movements that are not concentrated in a certain part of the monitor, which fixation go beyond the monitor, and are not scattered over it (tab.1). Thus, groups were selected based on the method of memorization, the groups dataset included data on short-term and long-term memory.

**Table 1.** *Classification of eye movements based on memorization group*

Scattered EM	Grouped EM	Without derivation
		
		

To avoid overfitting, we applied feature standardization using StandardScaler<sup>1</sup> to achieve equal scaling among input features. The standardization was performed separately for the training and testing dataset within each fold. All real numbers were reduced to integers for training model, in cases where there was a null value, we put 0. We also set a fixed random state to make the result reproducible<sup>2</sup>.

### 2.3.2 Statistical Analysis

Fixations and saccades were segmented in the DataViewer with the standard EyeLink algorithm using velocity and acceleration thresholds. The obtained data were analyzed in the “Statistica” program. In order to assess differences in eye movements when retrieving information from short-term memory, long-term memory, during memorization the data was collected and an analysis of variance (ANOVA) with repeated measures was conducted. This was followed by post-hoc pairwise comparisons using Mann-Whitney U Test. We also calculated correlation between eye movements in short-term memory and retrieval using Spearman Rank Order Correlations.

In machine learning classification we used following features: current fixation duration, current fixation end, current fixation index, current fixation pupil, current fixation x, current fixation y, current saccade amplitude, current saccade angle, current saccade average velocity, current saccade blink duration, current saccade duration, current saccade peak velocity, current saccade peak velocity time. In statistical analysis—average blink duration, average fix-

<sup>1</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

<sup>2</sup> [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)

ation duration, average saccade amplitude, average x resolution average y resolution, blink count, duration, fixation count fixation duration max fixation duration maximum time, fixation duration minimum, fixation duration minimum time, median fixation duration, median saccade amplitude, pupil size maximum, pupil size maximum time, pupil size mean, pupil size minimum, saccade count, standard deviation, fixation duration, standard deviation, saccade amplitude.

### 2.3.3 Classification Approach

We used many different algorithms of machine learning such as DecisionTreeClassifier, RandomForestClassifier, GaussianNB, LogisticRegression, KNeighborsClassifier, MLPClassifier, GradientBoostingClassifier. Classifier performance was evaluated in terms of precision, recall, f1-score, accuracy.

## Results

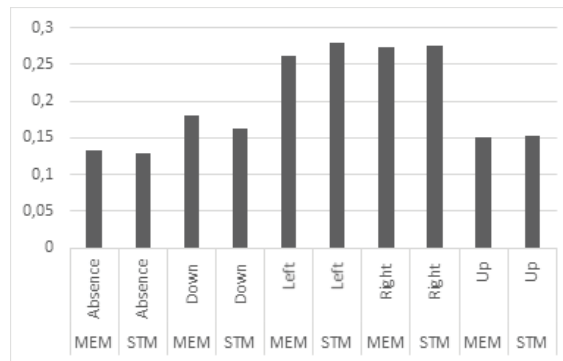
*H1: There are differences in eye movements when retrieving information from short-term memory, long-term memory, during memorization. Eye movements can be automatically divided according to their nature: short-term memory, long-term memory, memorization.*

The EyeLink software has a built-in algorithm which show averages eye movements parameters. First, we analyzed this averaged eye movements parameters. For statistical analyses, we chose repeated measures ANOVA for comparisons across types of memory and memorization, pos-hoc comparisons were performed through Mann-Whitney U test. First, the results of the multivariate tests showed that average fixation duration, average saccade amplitude, fixation count, fixation duration max, fixation duration max time, fixation duration min, fixation duration min time, median fixation duration, median saccade amplitude, pupil size max, pupil size max time, pupil size mean, pupil size min, saccade count had all a significant impact on the type of memory and memorization (Wilks lambda=0,124,  $F(42,158)=6,913$ ,  $p=0,00$ ).

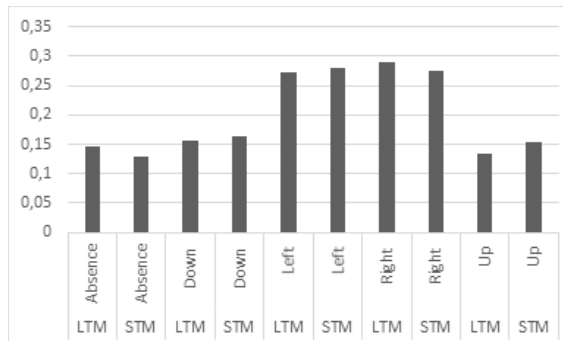
Second, post-hoc comparisons showed that the differences between short-term memory and long-term memory, between short-term memory and memorization were significant in average fixation duration, average saccade amplitude, average x resolution, average y resolution, blink count, duration, fixation duration max, fixation duration min, median fixation duration, median saccade amplitude ( $p=0,00$ ).

Next, we analyzed the direction of every single saccade. Since the sample is not the same ( $n(\text{STL})=1520$ ,  $n(\text{LTM})=1102$ ,  $n(\text{MEM})=2144$ ), we divided the number of eye movements down, up, right, left and absence of eye movements by the total number of observations of a particular type of memory or memorization. We compared memorization with short-term memory retrieval (fig.1) and long-term memory with short-term memory retrieval (fig.2). Since the number of eye movements is different when memorizing and reproducing information, we divided the number of eye directions in a certain direction by the total number of eye movements when memorizing or the type of memory.

When memorizing, a tendency was expressed to look down or to one point in comparison with the reproduction of a poem from short-term memory. When retelling a poem from long-term memory, study participants also looked more focused at one point, or tended to look to the right, compared to short-term memory.



**Figure 1.** Saccade direction of memorization and short-term memory retrieval



**Figure 2.** Saccade direction of long-term memory and short-term memory retrieval

We also tried to automatically divide short-term memory (STM) and long-term memory (LTM) eye movements basing on the raw data. Dataset consists of the characteristics of every single saccade and fixation, results are represented in table 2 and table 3. For saccades in LTM  $n=1102$ , in STM  $n=1520$ , for fixations in LTM  $n=1162$ , in STM  $n=1527$ .

**Table 2.** Evaluation of the classification of eye movements (saccades) based on the type of memory

Classifier	Type of memory/ memorization	Precision	Recall	F1-score	Accuracy
DecisionTreeClassifier	LTM	0.66	0.64	0.65	0.72
	STM	0.76	0.77	0.77	
RandomForestClassifier	LTM	0.71	0.72	0.72	0.77
	STM	0.81	0.80	0.81	
GaussianNB	LTM	0.48	0.80	0.60	0.57
	STM	0.75	0.41	0.53	
LogisticRegression	LTM	0.56	0.55	0.56	0.64
	STM	0.60	0.70	0.70	
KNeighborsClassifier	LTM	0.57	0.75	0.71	0.75
	STM	0.81	0.75	0.78	
MLPClassifier	LTM	0.66	0.64	0.65	0.72
	STM	0.76	0.77	0.77	
GradientBoostingClassifier	LTM	0.65	0.69	0.66	0.72
	STM	0.78	0.74	0.76	

The data was best classified on short-term and long-term memory by RandomForestClassifier

(accuracy=0,76), KNeighborsClassifier (accuracy=0.75) in saccades and RandomForestClassifier (accuracy=0,77), KNeighborsClassifier (accuracy=0.75) in fixations. GaussianNB (accuracy=0.57) and LogisticRegression (accuracy=0.64) coped with this task worse

*H2: Those eye movements during memorization and reproduction that do not differ are interconnected.*

Since the distribution is not normal, we studied correlation between eye movements in short-term memory and retrieval using Spearman Rank Order Correlations. Significant correlations were found in fixation duration min time ( $r=1,00$ ,  $p<0,05$ ), pupil size max ( $r=0,49$ ,  $p<0,05$ ), pupil size max time ( $r=1,00$ ,  $p<0,05$ ), pupil size mean ( $r=0,81$ ,  $p<0,05$ ), pupil size min ( $r=0,51$ ,  $p<0,05$ ). Fixation count ( $r=0,17$ ,  $p>0,05$ ) and saccade count ( $r=0,17$ ,  $p>0,05$ ).

*H3: It is possible to automatically identify groups according to eye movements during memorization in memorization, short-term and long-term memory.*

According to our classification we tried to use machine learning algorithms in identifying eye movements group. For saccades in Grouped EM  $n=1079$ , in Scattered EM  $n=2596$ , Without derivation  $n=1028$  (tab.4). For fixations in Grouped EM  $n=1079$ , in Scattered EM  $n=2596$ , Without derivation  $n=1028$  (tab.5).

**Table 4.** Evaluation of the classification of eye movements (saccades) based on the type of memory/memorization

Classifier	Type of memory/ memorization	Precision	Recall	F1- score	Accuracy
DecisionTreeClassifier	Grouped EM	0.64	0.61	0.62	0.72
	Scattered EM	0.80	0.80	0.80	
	Without derivation	0.55	0.58	0.57	
RandomForestClassifier	Grouped EM	0.81	0.70	0.75	0.82
	Scattered EM	0.84	0.93	0.88	
	Without derivation	0.75	0.63	0.78	
GaussianNB	Grouped EM	0.48	0.15	0.23	0,51
	Scattered EM	0.74	0.56	0.64	
	Without derivation	0.29	0.82	0.43	
LogisticRegression	Grouped EM	0.40	0.07	0.12	0.57
	Scattered EM	0.59	0.93	0.72	
	Without derivation	0.38	0.08	0.13	
KNeighborsClassifier	Grouped EM	0.70	0.67	0.69	0.76
	Scattered EM	0.81	0.88	0.85	
	Without derivation	0.63	0.49	0.55	
MLPClassifier	Grouped EM	0.64	0.61	0.62	0.72
	Scattered EM	0.80	0.80	0.80	
	Without derivation	0.55	0.58	0.57	
GradientBoostingClassifier	Grouped EM	0.69	0.60	0.64	0.74
	Scattered EM	0.78	0.87	0.82	
	Without derivation	0.62	0.50	0.56	

RandomForestClassifier (accuracy=0,82) and KNeighborsClassifier (accuracy=0,76) best identified groups, GaussianNB (accuracy=0,51) and LogisticRegression (accuracy=0,57)

worst identified groups in both saccades and fixations measurements.

**Table 5.** *Evaluation of the classification of eye movements (fixations) based on the type of memory/memorization*

Classifier	Type of memory/ memorization	Precision	Recall	F1- score	Accuracy
DecisionTreeClassifier	Grouped EM	0.67	0.64	0.66	0.73
	Scattered EM	0.80	0.81	0.81	
	Without derivation	0.55	0.57	0.56	
RandomForestClassifier	Grouped EM	0.82	0.71	0.76	0.82
	Scattered EM	0.83	0.93	0.88	
	Without derivation	0.73	0.60	0.66	
GaussianNB	Grouped EM	0.48	0.15	0.23	0.51
	Scattered EM	0.74	0.56	0.64	
	Without derivation	0.29	0.82	0.43	
LogisticRegression	Grouped EM	0.40	0.07	0.12	0.57
	Scattered EM	0.59	0.93	0.72	
	Without derivation	0.38	0.08	0.13	
KNeighborsClassifier	Grouped EM	0.70	0.67	0.69	0.76
	Scattered EM	0.81	0.88	0.85	
	Without derivation	0.63	0.49	0.55	
MLPClassifier	Grouped EM	0.67	0.64	0.66	0.73
	Scattered EM	0.80	0.81	0.81	
	Without derivation	0.55	0.57	0.56	
GradientBoostingClassifier	Grouped EM	0.70	0.60	0.65	0.74
	Scattered EM	0.78	0.87	0.82	
	Without derivation	0.62	0.50	0.55	

## Discussion

Based on the reproduction of information, interesting conclusions have already been obtained in the field of variation in eye movements. So, variations in visual imagery during autobiographical retrieval seem to be related physiologically with variations in the number of fixations and saccades (El Haj, M. et al, 2018). In this pilot study, we found that in the retrieval of poems from different types of memory, there may be variations in that average blink duration, average fixation duration, average saccade amplitude, fixation count, fixation duration max, fixation duration max time, fixation duration min, fixation duration min time, median fixation duration, median saccade amplitude, pupil size max, pupil size max time, pupil size mean, pupil size min, saccade count had all a significant impact on the type of memory and memorization. Based on the random forest algorithm, it is possible to determine with relatively high accuracy what type of memory the information is reproduced from.

Our study confirmed the empirical research suggesting that looking at a “now-empty” location facilitates retrieval of information that was previously exposed in that location (El Haj, M. et al, 2018; Althoff, R. R. et al, 1999; Henderson, J. M. et al, 1999; Hollingworth, A., 2006; Loftus, G. R. et al, 1978; Ryan, J. D. et al, 2000). All participants averted their gaze and as the histogram showed, the participants more often tended to look to the left when retelling

the poem from short-term memory, which can be explained by the position of the poem at the stage of its memorization. There were also objects that grouped the view into an empty part of the screen, perhaps this is related to the spatial properties of absent stimuli which can influence eye movements when those stimuli are imagined (Altmann, G. T. et al, 2004; Brandt, S. A. et al, 1997; Laeng, B. et al, 2002; Richardson, D. C. et al, 2000; Spivey, M. J. et al, 2001; Spivey, M. J. et al, 2004). Histogram results complete the finding that people make multiple eye movements when they “scan” for information in long-term memory and very few eye movements when they “focus” on information in the working-memory “buffer” (Ehrlichman, H. et al, 2012). When retelling from short-term memory, participants were more likely to look up and down, whereas when retelling from long-term memory, the direction of gaze was often absent.

Such results can form the basis for the differentiation of cognitive styles according to the strategy for memorization and the reproduction of information.

### **Limitations and future work**

This work proposes the first steps towards classifying memory during non-visual eye movements on the blank screen paradigm. Despite this promising result, our work has several limitations that we plan to address in future work. First, we have a rather limited sample, for serious conclusions it is necessary to expand it. Second, it is needed to evaluate the quality of the retelling the poems, how the rhyme, the words of the poem are preserved, the meaning is conveyed, whether the person told the poem with expression or not. Based on this assessment, we will be able to judge the relationship of eye movements with more complex aspects.

### **Conclusion**

Despite the identified limitations, a pilot study showed that it is possible to quite successfully distinguish between eye movements during some cognitive processes, such as memorization, retrieval of information from short-term and long-term memory. Also, the results of the differentiation of eye movements according to the memorization strategy, their relationship when memorizing and reproducing a poem from short-term memory can be used to explain the cognitive style.

### **References:**

1. Althoff, R. R., & Cohen, N.J. Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1999, 25(4), 997e1010.
2. Altmann, G. T.. Language-mediated eye movements in the absence of a visual world: the “blank screen paradigm.” *Cognition*, 2004, 93(2), B79–B87. doi: 10.1016/j.cognition.2004.02.005
3. Atkinson, R.C. and Shiffrin, R.M. Human memory: a proposed system and its control processes. In: Spence K. W. and Spence J.T. (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, 1968, Vol. 2. Academic Press, New York, pp. 89–195
4. Bourlon, C., Oliviero, B., Wattiez, N., Pouget, P., & Bartolomeo, P.. Visual mental imagery: What the head’s eye tells the mind’s eye. *Brain Research*, 2011, 1367, 287–297. doi:10.1016/j.brainres.2010.10.03
5. Brandt, S. A., & Stark, L. W.. Spontaneous Eye Movements During Visual Imagery Reflect the Content of the Visual Scene. *Journal of Cognitive Neuroscience*, 1997, 9(1), 27–38. doi:10.1162/jocn.1997.9.1.27
6. Cowan, N.. Chapter 20 What are the differences between long-term, short-term, and working

- memory? *Essence of Memory*, 2008, 323–338. doi:10.1016/s0079–6123(07)00020–9
7. Damiano, C., & Walther, D. B. . Distinct roles of eye movements during memory encoding and retrieval. *Cognition*, 2019, 184, 119–129. doi:10.1016/j.cognition.2018.12.014
8. Duff, M. C., Covington, N. V., Hilverman, C., & Cohen, N. J. . Semantic Memory and the Hippocampus: Revisiting, Reaffirming, and Extending the Reach of Their Critical Relationship. *Frontiers in Human Neuroscience*, 2020, 13. doi:10.3389/fnhum.2019.00471
9. Ehrlichman, H., & Micic, D.. Why Do People Move Their Eyes When They Think? *Current Directions in Psychological Science*, 2012, 21(2), 96–100. doi:10.1177/0963721412436810
10. Ehrlichman, H., Micic, D., Sousa, A., & Zhu, J.. Looking for answers: Eye movements in non-visual cognitive tasks. *Brain and Cognition*, 2007, 64(1), 7–20. doi:10.1016/j.bandc.2006.10.001
11. El Haj, M., & Lenoble, Q.. Eying the future: Eye movement in past and future thinking. *Cortex*, 2018, 105, 97–103. doi:10.1016/j.cortex.2017.05.024
12. Gbadamosi, J., & Zangemeister, W. H.. Visual Imagery in Hemianopic Patients. *Journal of Cognitive Neuroscience*, 2001, 13(7), 855–866. doi:10.1162/089892901753165782
13. Hannula, D. E.. Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, 2010, 4. doi:10.3389/fnhum.2010.00166
14. Hartmann, M., Mast, F. W., & Fischer, M. H.. Spatial biases during mental arithmetic: evidence from eye movements on a blank screen. *Frontiers in Psychology*, 2015, 6. doi:10.3389/fpsyg.2015.00012
15. Hashim S, Stewart L, Küssner MB. Saccadic Eye-Movements Suppress Visual Mental Imagery and Partly Reduce Emotional Response During Music Listening. *Music & Science*. January 2020. doi:10.1177/2059204320959580
16. Hebb, D. O. Concerning imagery. *Psychological Review*, 1968, 75(6), 466–477. doi:10.1037/h0026771
17. Henderson, J.. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 2003, 498–504. doi:10.1016/j.tics.2003.09.006
18. Henderson, J. M., & Hollingworth, A.. The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 1999, 10(5), 438e443.
19. Henderson, J. M., Williams, C. C., & Falk, R. J.. Eye movements are functional during face learning. *Memory & Cognition*, 2005, 33(1), 98–106. doi:10.3758/bf03195300
20. Hollingworth, A.. Constructing Visual Representations of Natural Scenes: The Roles of Short- and Long-Term Visual Memory. *Journal of Experimental Psychology: Human Perception and Performance*, 2004, 30(3), 519–537. doi:10.1037/0096–1523.30.3.519
21. Hollingworth, A.. Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2006, 32(1), 58e69. <http://dx.doi.org/10.1037/0278–7393.32.1.58>.
22. Hout, M., Muris, P., Salemink, E., & Kindt, M.. Autobiographical memories become less vivid and emotional after eye movements. *British Journal of Clinical Psychology*, 2001, 40(2), 121–130. doi:10.1348/014466501163571
23. Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K.. Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*, 2012, 38(5), 1289–1314. doi:10.1037/a0026585
24. Johansson, R., Holsanova, J., & Holmqvist, K. Pictures and Spoken Descriptions Elicit Similar Eye Movements During Mental Imagery, Both in Light and in Complete Darkness. *Cognitive Science*, 2006, 30(6), 1053–1079. doi:10.1207/s15516709cog0000\_86
25. Johansson R, Nyström M, Dewhurst R, Johansson M. 2022 Eye-movement replay supports episodic remembering. *Proc. R. Soc. B*289: 20220964. <https://doi.org/10.1098/rspb.2022.0964>
26. Kosslyn, S. M. *Image and Brain*, 1994, MIT Press
27. Laeng, B., & Teodorescu, D.–S.. Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, 2002, 26(2), 207–231. doi:10.1207/

28. Liversedge, S. P., & Findlay, J. M.. Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 2000, 4(1), 6–14. doi:10.1016/s1364-6613(99)01418-7
29. Loftus, G. R., & Mackworth, N. H.. Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4(4), 565e572.
30. Maheshwari, S., Tuladhar, V., Thargay, T., Sarmah, P., Sarmah, P., & Rai, K.. Do our eyes mirror our thought patterns? A study on the influence of convergent and divergent thinking on eye movement. *Psychological Research*, 2021. doi:10.1007/s00426-021-01520-7
31. Martarelli, C. S., & Mast, F. W.. Eye movements during long-term pictorial recall. *Psychological Research*, 2012, 77(3), 303–309. doi:10.1007/s00426-012-0439-7
32. Mast, F.. Eye movements during visual mental imagery. *Trends in Cognitive Sciences*, 2002, 6(7), 271–272. doi:10.1016/s1364-6613(02)01931-9
33. Noton, D., & Stark, L.. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 1971, 11(9), 929–IN8. doi:10.1016/0042-6989(71)90213-6
34. Noton, D., & Stark, L.. Scanpaths in Eye Movements during Pattern Perception. *Science*, 1971, 171(3968), 308–311. doi:10.1126/science.171.3968.308
35. Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J.. Amnesia is a Deficit in Relational Memory. *Psychological Science*, 2000, 11(6), 454–461. doi:10.1111/1467-9280.00288
36. Rayner, K.. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 1998, 124(3), 372–422. doi:10.1037/0033-2909.124.3.372
37. Rayner, K.. The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 2009, 62(8), 1457–1506. doi:10.1080/17470210902816461
38. Richardson, D. C., & Spivey, M. J.. Representation, space and Hollywood Squares: looking at things that aren't there anymore. *Cognition*, 2000, 76(3), 269–295. doi:10.1016/s0010-0277(00)00084-6
39. Spivey, M. J., & Geng, J. J.. Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, 2001, 65, 235–241.
40. Shapiro, F.. *Eye movement desensitisation and reprocessing. Basic principles, protocols, and procedures*, 1995. New York: Guilford Press
41. Sheehan, P.W.; Neisser, U. Some variables affecting the vividness of imagery in recall. *British Journal of Psychology*, 1969, 60(1), 71–80. doi:10.1111/j.2044-8295.1969.tb01178.x
42. Shepard, R.N., Cooper L. A. *Mental Images and Their Transformations*, 1986. Cambridge, MA: The MIT Press.
43. Spivey, M. J., & Geng, J. J.. Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychological Research*, 2001, 65(4), 235–241. doi:10.1007/s004260100059
44. Spivey, M. J., Richardson, D. C., & Fitneva, S. A.. Thinking outside the brain: Spatial indices to visual and linguistic information. In J.M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*, 2004. New York: Psychology Press
45. Theeuwes, J., Belopolsky, A., & Olivers, C. N. L.. Interactions between working memory, attention and eye movements. *Acta Psychologica*, 2009, 132(2), 106–114. doi:10.1016/j.actpsy.2009.01.005
46. Wynn, J. S., Shen, K., & Ryan, J. D.. Eye Movements Actively Reinstatement Spatiotemporal Mnemonic Content. *Vision*, 2019, 3(2), 21. doi:10.3390/vision3020021
47. Yarbus, A. L.. *Eye Movements During Perception of Complex Objects*. *Eye Movements and Vision*, 1967, 171–211. doi:10.1007/978-1-4899-5379-7\_8
48. Yoon, D., & Narayanan, N. H.. Mental imagery in problem solving. *Proceedings of the Eye Tracking Research & Applications Symposium on Eye Tracking Research & Applications–ETRA'2004*, 2004. doi:10.1145/968363.968382

## Chapter 11. Least action classifiers

*Malashin R., Boiko A.*

*Pavlov Institute of Physiology RAS, State University of Aerospace Instrumentation,  
Saint-Petersburg, Russia*

### Introduction

Least Action principle is a fundamental law in physics. According to this principle real objects follow trajectories that satisfy stationary points of a functional called Action: maximum potential and minimum kinetic energy.

Shelepin et al. (Shelepin et al., 2003) draw analogy between least action principle and human thinking. Indeed, opportunities for optimal decision-making are limited—two opposing factors arise:

- the desire to analyze the situation and get the maximum information so that the decision is the most accurate (by analogy, it can be compared with the maximum potential energy in physics);
- the desire to reduce the time and resources that are needed to obtain additional information (analogous to the minimum kinetic energy).

At a certain level of abstraction, the process of information processing becomes sequential, and the decision to continue the calculations is made based on the understanding of the current quality of the solution already achieved. In this case, the system is dynamically configurable because it handles different data differently.

The principle of least action can be used in the development of advanced artificial intelligence systems. In this interpretation the principle of least action says that if it is possible to obtain a reliable solution in a simpler way, then this should be done. More concretely, we want our dynamically configurable system conditioned on the data instance to unfold computational graph in the way that satisfies maximum expected accuracy and at the same minimum computational resources that we need to spend to get an answer. Such systems can save resources on easy examples, while keep spending enough computations on hard examples to achieve desirable accuracies. Note, that these not only assume termination of computation adaptively, but also adaptive selection of the next computational module, given the current state of solution.

Dynamic systems of image analysis based on this interpretation of least action principle was proposed in (Malashin, 2019). It was shown that building such a system can be reduced to training an agent to perceive images through the set of pretrained classifiers with the objective to minimize number of classifiers used for every image. In (Malashin, 2021) concrete architecture called Least action classifier was proposed. In (Malashin, 2022) and (Malashin, Boiko, 2022) the idea was further developed. In this chapter we summarize main achievements, current problems and further development directions in this area.

### Dynamic networks

Dynamically configurable systems attract a lot of research interest. For example, Viola and Jones (Viola, Jones, 2004) used cascades of classifiers for face detection: the image is processed sequentially by classifiers and if any of the classifier responses negatively then the recognition process stops and negative response is produced by the whole cascade. Dynamically

cally configurable systems based on neural networks are also known as dynamic networks. In (Bolukbasi et al, 2017), an approach to the sequential application of networks of various well-known architectures to the classification problem is described. The paper describes, among other things, a cascade model using an early exit from the network. In (Aria et al, 2019) the authors consider options for early exit from computations by training a fully connected classifier on top of the features of each of the network levels.

Scaling up neural networks (provided with enough data) boosts performance qualitatively, and this is especially true when using transformers (Vaswani et al, 2017) in the problem of language modeling. Large transformers tend to use a static calculation structure—different data instances cause firing of the same neurons (Delvin et al, 2018), (Brown et al 2020). Same is true for big convolutional neural networks (CNNs) and most of recurrent networks that perform a complete chain of all transformations in any case, whether the input is very trivial or very complex. Dynamically configurable computing systems are those in which different data is processed differently (dynamic network is another term frequently used when talking about deep learning).

Sparse models (or switch transformers) (Zoph, 2022), (Fedus et al 2021) complement the architecture of the transformer with additional modules—controllers whose task is to select fully connected blocks, where to send data. As a result, different expert blocks process different examples of data. This approach (called a mixture-of-experts or switch transformers) was first used back in 2017 for convolutional neural networks (Shazeer et al, 2017), where it proved to be efficient for very large networks. Today sparse models allow further increase of number of free parameters, the number which already reaches trillions (Fedus et al, 2021).

Noteworthy the mixture-of-experts uses only a part of the capabilities of the dynamic configuration systems:

- the number of operations is always fixed for both hard and easy examples;
- there is no possibility of obtaining intermediate result;
- dynamic configuration is carried out at the level of experts of the same size, while an expert who is responsible for solving a more complex problem should have more free parameters.

Another important aspect that is not touched upon in the giant sparse transformers is a matter of reusing experts in solving other problems. This can be extremely useful when creating retrainable systems. Although some experiments show, for example, that using information about which expert processed the data does not improve the result of the entire model.

The authors of (Gururangan et al, 2022) build a modular dynamically configurable system based on the fact that there are different domains for a language with different corpora (technical literature, scientific, medical, etc.), and for each domain, a separate expert model is trained. As authors show this approach outperforms baseline. During the inference, each expert processes queries of its own domain; metadata provides information for routing. An important feature of such approach is that you can add experts to it, as well as retrain already existing. The disadvantages include that the approach requires significant human intervention in setting up and adapting the system to new tasks.

Researchers applied a lot of efforts to extend different neural network architectures with the ability to dynamically adapt computations for different image examples. McGill and Perona (McGill, Perona, 2017) proposed usage of “stop” and “go” signals in neural networks that allows

adaptive termination of computations. (Zhou et al, 2021) considers the problem of constructing ensembles from experts, where each expert specializes in its domain. The authors experiment with different image domains (sketch, photography, etc.) though different domains are treated differently only during training, so this approach cannot be called truly dynamic.

Graves (Graves, 2016) proposed recurrent neural network with adaptive computational time (ACT). Figurnov et al. applied ACT in residual blocks of convolutional neural networks.

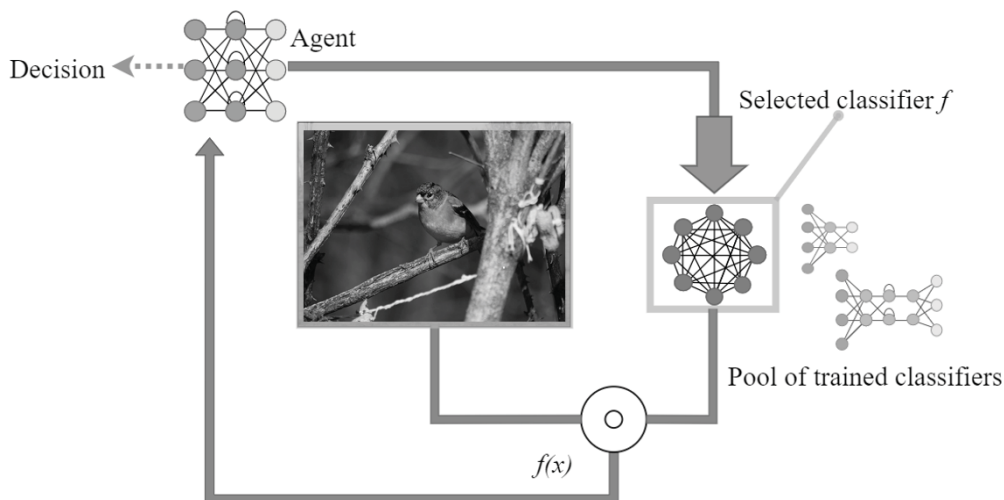
Models that use hard visual attention (Mnih et al, 2014; Li et al, 2017; Wang et al, 2020) can be considered as partially dynamically configurable: they adaptively select regions of images, though every time the same neurons fire.

Good systematization of dynamically configurable systems can be seen from the perspective of different variants of algorithm configuration (Adriaensen et al 2022). Term algorithm configuration (AC) refers to the process of determining a policy for setting algorithm parameters; the goal is to maximize performance across entire population. Adriaensen et al specify three variants of AC: *per-distribution*, *per-instance* and *dynamic*. Per-distribution AC specifies policy that gives parameter settings for entire distribution, per-instance AC policy sets parameters individually to every example in the dataset, while dynamic AC (DAC) exploits iterative nature of AI task: parameters are adjusted after every iteration. InstaNAS (Cheng et al, 2019) can be thought as example of per-instance AC. In this case for every image an individual network is constructed. Least action classifier (LAC) (Malashin, 2021) discussed in the next chapters directly exploits iterative nature of image recognition task and is a special case of Dynamic Algorithm Configuration.

## General approach

Malashin (Malashin, 2019; Malashin 2021) proposed to train the algorithm for automatic image analysis according to the problem of dynamic algorithm configuration (Biedencamp, 2020). In this case after each step of the analysis, it is required to select the next branch of calculations in accordance with the information already received about the analyzed image (context). Such strategies are called context-sensitive (unique for each particular image).

In practice the agent is taught to perceive images through the set of pre-trained classifiers and the goal of the agent is to unfold the computational graph with the trajectory that refers to the minimal number of operations and maximal expected accuracy (Figure 1)



**Figure 1.** General scheme of agent-image interactions in Least Action Classifier (LAC)

The proposed computation structure includes three key elements (Malashin, 2019): a set of classifiers  $A$ , a function for choosing the next classifier  $\Phi_1$ , and a function for updating the state based on the results of the classifier  $\Phi_2$ .

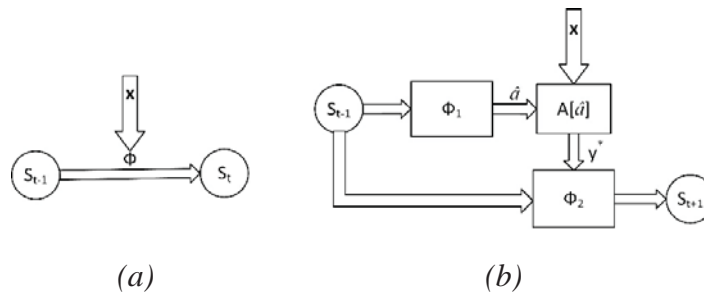
The question of how to combine the results of the classifiers, first of all, is how to obtain effective functions  $\Phi_1$ ,  $\Phi_2$ , which should minimize the computational resources and maximize the expected accuracy of the solution. The selection function  $\Phi_1(s_t, x) \rightarrow \hat{a}$  maps the current state  $s_t$  and the data  $x$  (image under consideration) to the space of “keys” of classifiers  $\hat{A} = \{\hat{a}_i\} \cup \{\varepsilon\}$ , where  $\varepsilon$  is a symbol meaning that the calculation needs to be stopped.

The term “key” in this case is used to emphasize that  $\Phi_1$  does not generate the algorithm  $a$ , but only selects it from the list, which in python corresponds to accessing the dictionary  $D$  by the “key”  $a = D[\hat{a}]$ . The classifier  $a \in A$  takes data  $x$  as input and returns a decision vector  $y^*$ :  $a(x) \rightarrow y^*$ . For convenience, we can assume that the “key”  $\hat{a}$  also contains information about the “complexity” of the algorithm  $a$ . The state update function  $\Phi_2(s_t, \hat{a}, y^*) \rightarrow s_{t+1}, y_{t+1}^{\text{cur}}$  maps the current state and the result of the selected classifier to the state  $s_{t+1}$  and a new generalized solution  $y_{t+1}^{\text{cur}}$ , taking into account the previous answers of the classifiers.

It should be noted that the considered decision process in the most general case is not a Markov decision process (MDP), since the optimal  $\Phi_1$  and  $\Phi_2$  must take into account all previous key-decision pairs, i.e.  $s_t = \{\hat{a}_1, y_{t-1}^*, \dots, \hat{a}_{t-1}, y_{t-1}^*\}$ . If the data  $x$  does not change in time (which is not necessary in the general case when observing the environment in real time), then the order in which the classifiers are called is not important. If the data changes, then for the optimal decision-making algorithm  $s_t = \{x_0, \hat{a}_1, y_{t-1}^*, x_1, \dots, \hat{a}_{t-1}, y_{t-1}^*, x_{t-1}\}$ , which leads to an excessively large state space  $S$ . In addition, it can be noted that, without imposing restrictions, the optimal selection function  $\Phi_1$ , which receives data  $x$  as input, must a) solve the classification problem inside itself, b) encapsulate the entire set of classifiers  $A$ , c) and then choose the order of classifiers that will give the correct solution by manipulating classifiers, without using the knowledge embedded in them. Obviously, this is not a solution that could be useful in practice when implementing the principle of least action, since in this case the burden of all calculations simply lies on the  $\Phi_1$  function.

In order to avoid such behavior and partially reduce the size of the state space, we will consider the indirect interaction of  $\Phi_1$  with data only through classifiers, in this case the function  $\Phi_1$  takes the form  $\Phi_1(s_t) \rightarrow \hat{a}$ . Additional complexity is caused by the “dependence” of the required functions  $\Phi_1(s_t, x) \rightarrow \hat{a}$  and  $\Phi_2(s_t, \hat{a}, y^*) \rightarrow s_{t+1}, y_{t+1}^{\text{cur}}$ .

In fact, this is the same transition function  $\Phi$  (Figure 2a), which provides the transition from the state  $s_{t-1}$  to  $s_t$ , but it is more convenient for us to divide it into two due to the presence of an external set of fixed classifiers  $A$  (Figure 2b).



**Figure 2.** The state diagram of the MDP for the problem of sequential classification of image  $x$ . General case of sequential classification using function  $\Phi$  (a), the case of using a pool  $A$  of external classifiers (b) (Malashin, 2019)

First, let's consider an approximation of the selection and update functions for solving a simplified problem in the form of an MDP based on a naive Bayes classifier and minimizing the expected entropy (Malashin, 2019).

### Naive Bayes classifier and expected entropy minimization

For simplicity, let's assume:

- the entire image is fed to the input of the classifiers;
- the classifiers' answers are conditionally independent of each other;
- responses of classifiers are deterministic and discrete (for example, they issue a label of a recognized class);
- the “cost” (computational/time costs) of running all classifiers is the same.

The use of these assumptions allows us to use the current probability distribution of the image belonging to one of the given classes as a state, the naive Bayes classifier as the update function  $\Phi_2$ , and the expected entropy of distributions as a criterion for the selection function  $\Phi_1$ .

Let  $\{P_k^{(y_i)}\}$ ,  $y_i \in [1, N]$  be the probabilities that image  $x$  belongs to class  $y_i$  after running  $k$  classifiers, where  $N$  is the number of recognizable image classes. The classifier  $A_i \in A$  takes the data  $x$  and returns the label  $c_i$ :  $A_i(x)=c_i$ , where  $c_i \in [1, N]$ . For simplicity, the simplified notation  $A_i=c_i$  will be used below, which by default implies that the data was submitted to the input of the classifier.

We will use the vector  $P_k$  as the current intermediate solution:  $y_k^{\text{cur}}=P_k$ , and as the state  $S_k$  we will also store the “keys” of all called classifiers in order to prevent the same algorithm from being called several times:  $S_k \{y_k^{\text{cur}}, \{\hat{a}_i: i \in [1, k]\}\}$ . At the beginning of the calculations, the probability vector of the image belonging to each of the classes  $P_0$  can be selected from a uniform (or other a priori) probability distribution.

Let  $k$  classifiers  $\{A_i\}$ ,  $i \in [1, k]$  be launched and corresponding responses  $\{c_i\}$ ,  $i \in [1, k]$  were received from them. Since the classifiers are conditionally independent of each other, then:

$$P(y_i | A_1=c_1, A_2=c_2, \dots, A_k=c_k) = P(y_i | A_1=c_1) \cdot P(y_i | A_2=c_2) \cdot \dots \cdot P(y_i | A_k=c_k), \quad (1)$$

where  $P(y_i | A_j=c_j)$  is the probability that the image belongs to class  $y_i$ , provided that the  $j$ -th classifier recognized the image as belonging to class  $c_j$ .

Using the Bayes formula, we get an iterative classification procedure, when after running each classifier, the probability distribution is refined. For example, if the classifier  $A_j$  is selected at step  $k$  and the label  $c_j$  is obtained:

$$P_k(y_i | A_j = c_j) = \frac{P(A_j=c_j | y_i) \cdot P_{k-1}(y_i)}{P(A_j=c_j)}. \quad (2)$$

$P(A_j=c_j | y_i)$ ,  $P(A_j=c_j)$  can be calculated at the training stage. For example,  $P(A_j=c_j | y_i)$ —is the response frequency  $c_j$  on the images of class  $y_i$  in the training set. The classification algorithm using the assumption of conditional feature independence when observing a class label is called a naive Bayesian classifier (Murphy, 2012). Thus, we have determined  $\Phi_2$ .

According to the principle of least action, the system is required not only to get the correct answer by running all available classifiers but to use only a small subset of them, while also obtaining a high-quality solution. In this regard, a very important function is the  $\Phi_1$  function,

which intelligently selects the sequence of launching classifiers, as well as the criterion for stopping the solution refinement procedure. From the point of view of information theory and without taking into account the complexity of the algorithm, we want to choose at each iteration such a classifier  $A^*$  that provides a minimum of the expected entropy of the distribution of classes  $H(P|A^*=c^*)$ , where  $c^*$  is any of the possible “answers” of the classifier  $A^*$ . In other words, the expected entropy in this case is the entropy expectation after the classifier has been launched, taking into account all possible classifier responses:

$$l_{all}(x, y) = l_{BCE}(x, y) - kD(x) \quad (3)$$

where  $N$  is number of classes,  $P(A_j=c_j | y_i)$  is a ratio of answers  $c_j$  for the classifier to the sum of all possible answers of  $A_i$  (on the training set).

Selection function  $\Phi_1$  calculates the set of expected posterior distributions after calling each classifier, and uses them to calculate the expected entropy after calling each classifier.  $k$ -th classifier, which corresponds to the minimal expected entropy, is selected for execution:

$$k = \operatorname{argmin}_k H(P|A_k). \quad (4)$$

The stopping criterion in this case can be a decrease in entropy below a predetermined threshold or an excess of the number of called classifiers. Thus, we have defined the second desired function  $\Phi_1$ .

In this case, sequential classification resembles the game “guess the animal” (in some countries a similar game is called twenty questions”). In this game, one player thinks of an animal (for example, an elephant or a turtle), and the task of the second player is to ask questions, the answers to which are either “yes” or “no” (for example, does this animal live in water? Is it domestic?).

By the rules of the game, the number of questions that can be asked is limited, so very specific questions (such as “Does this animal eat bamboo?”) are best asked after more general ones (“Does this animal walk on four legs?”). Questions are an analogue of classifiers, the player asking questions, in fact, implements the functions  $\Phi_1$  and  $\Phi_2$ . The role of the player who answers the questions is played by the environment. If we imagine that the classifiers recognize only two classes without mistakes, and all the classes are equally likely, then the decision-making process is described by a decision tree, and in the most optimal case, the average number of runs should be  $2\log_2 N$ .

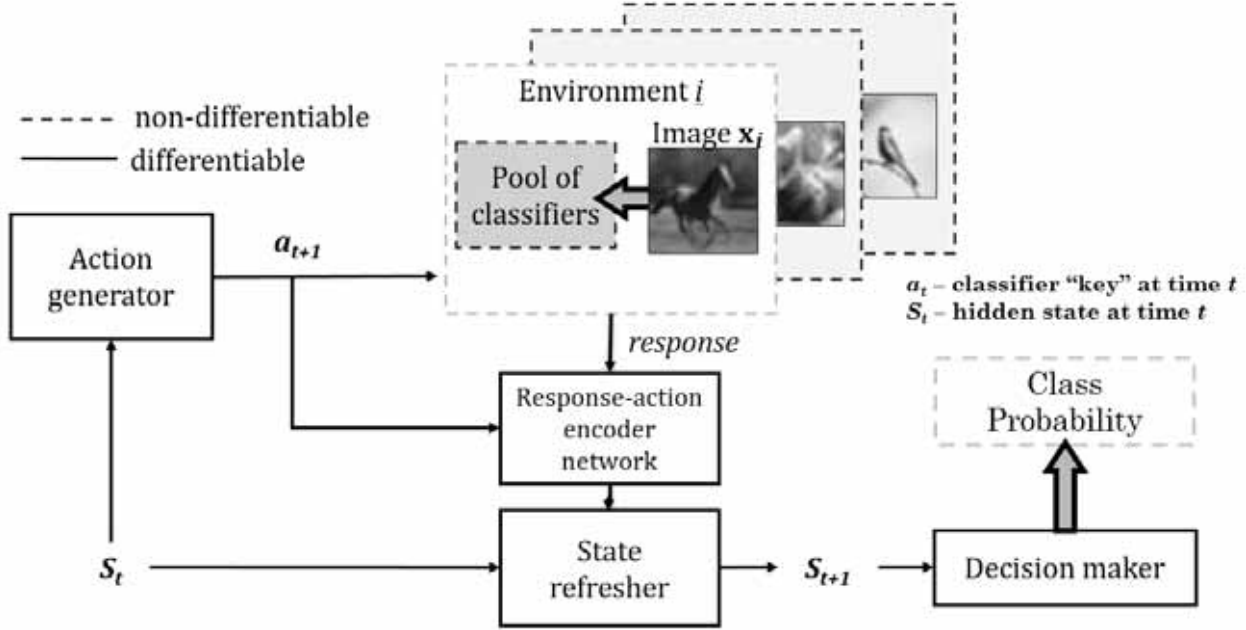
In the general case, however, the above four assumptions are not satisfied: solution of a more complex problem with “entangled” parameters is required, for which the proposed functions  $\Phi_1$  and  $\Phi_2$  are poorly suited. Modern image classifiers do not output a label, but a probability distribution, and different classifiers have different complexity. The most “problematic” is the assumption of the conditional independence of classifiers, which is not feasible in the case of the same response space of two classifiers, because otherwise, one should expect that the responses of the classifiers are not related to the image at all and do not carry any information about the object in principle.

The conditional dependence of classifiers greatly complicates the decision rule, but potentially allows you to make the right decisions when the probability table  $P(A_j=c_j|x=j)$  is very sparse (the expected case when the classifier’s answer is not a label, but a larger discrete or real vector). One example of the successful application of a neural network to overcome the curse of dimensionality that occurs in tabular graph models is neural autoregressive networks

(Larochell, Murray, 2011).

### Least action classifier

Neural networks seem to be reasonable choice to approximate optimal functions  $\Phi_1$  and  $\Phi_2$  (Malashin, 2019). In (Malashin, 2021) least action classifier (LAC) architecture was proposed (Figure 3).



**Figure 3.** Least action classifier architecture (Malashin, 2021)

LAC consists of the following five main components (Malashin, 2021):

1. Environment response generator, a non-differentiable element that takes an image and the index of the requested classifier, and returns response of the classifier.
2. State refresher that implements  $\Phi_2$  function; at step  $t$  it inputs hidden state vector and encodes classifier response; returns new hidden state vector.
3. Action generator that implements  $\Phi_1$  function; it inputs hidden state vector and returns the “key” of the classifier.
4. Decision maker that inputs hidden state vector and outputs current solution.
5. Response-action encoder that encodes action and classifier response in the format that is appropriate for state refresher.

Classifiers in the pool are learned on different subsets of classes (Malashin, 2021).

Let the dataset  $D$  consist of  $N$  images  $x_i$  with appropriate labels  $y_i$ :

$$D = \{(x_i, y_i), i \in [1, N], x \in X, y \in Y\} \quad (5)$$

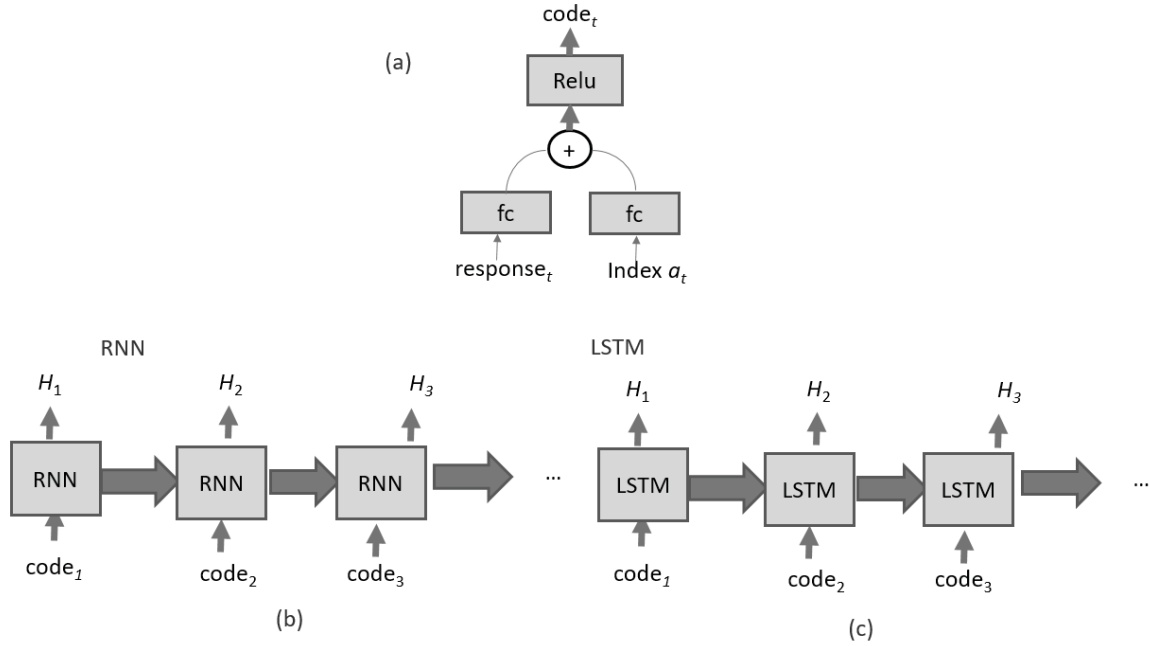
Subset of classes  $Y^k \subset Y$  split  $D$  into overlapping datasets  $D_k$ :

$$D_k = \{(x_k, y_k) \in D, y_k \in Y^k\} \quad (6)$$

Classifiers in the pool are learned on different  $D_k$ .

In (Malashin, 2021) response-action encoder works as follows:

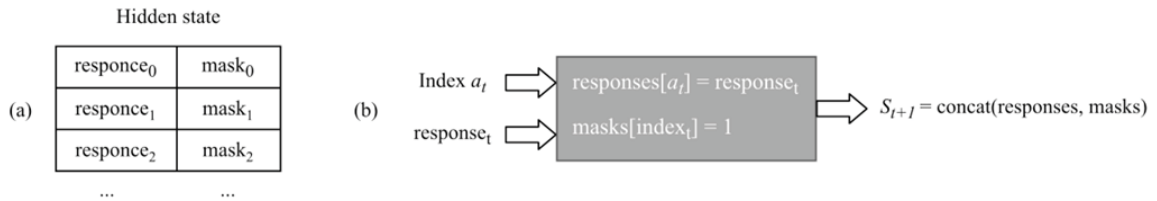
- Classifier response and classifier index are linearly transformed by fully connected layer;
- Then two obtained codes are summed;
- Final code is obtained by applying Relu activation function (Fig.4)



**Figure 4.** Response-action encoder (a) and state-refresher in the form of RNN (b) or LSTM (c)

State-refresher can be represented by RNN or LSTM (Figure 4b and Figure 4c), but it appeared that more robust dynamically configured computations are learned with state refresher with a short memory.

Figure 5 shows its hidden state structure and diagram of the update algorithm.



**Figure 5.** State refresher with short memory. Hidden state structure (a) and the diagram of its update algorithm (b) (Malashin, 2021)

Instead of encoding the state with hidden units, responses of every classifier are stored in a table (short memory). The state representation is extended with a masks that are set to ones when saving the corresponding classifier response into the response table. The masks provide a clear marker of what classifiers had been already called; they help to avoid duplicate actions. The size of a state-vector  $N \times C \times 2$ , where  $N$  is a number of responses to memorize and  $C$  is the size of the response vector. At the beginning of an “episode”, both tables are filled with zeros.

In (Malashin, 2021) LAC is learned by the hybrid function (Malashin, 2021):

$$Loss = \gamma Loss_{RL} + Loss_S, \quad (7)$$

where  $Loss_{RL}$  is reinforcement loss, and  $Loss_S$  is standard cross-entropy loss with ground truth label,  $\gamma$  is

hyperparameter. Reinforcement loss is a sum of action loss  $L_{action}$  and entropy-bonus  $L_H$ :

$$Loss_{RL} = L_{action} + \alpha L_H, \quad (8)$$

where  $\alpha$  is another hyperparameter and  $L_{action}$  has the following form (Malashin, 2021):

$$L_{action} = \sum_k^K \sum_t^T A_{k,t} \log[\pi(a_{k,t} | s_{k,t-1}; \theta_a)], \quad (9)$$

where  $K$  is the number of images in the batch,  $T$  is the number of actions taken in each “episode”,  $\pi$  is action policy,  $\theta_a$  is vector of action generator’s weights,  $s_{k,t-1}$  is hidden state vector,  $A_{k,t} = R_{k,t} - b(s_{k,t-1})$  is advantage: an extra reward  $R$  over prediction of the baseline network  $b$  agnostic to the action taken.

Formula (9) refer to A2C loss.

Entropy bonus has the following form (Malashin, 2021):

$$L_H = \sum_{i=1, t=2} \log[P(a_{i,t})]P(a_{i,t}) + \beta \sum_{k=1, i=1, t=1} \log[P(a_{k,i,t})]P(a_{k,i,t}) \quad (10)$$

where  $\beta$  is hyperparameter,  $P(a_{i,t})$  is probability of selecting classifier  $i$  on step  $t$  averaged across all  $K$  images in a batch. First term of (10) forces the agent to use different classifiers on different steps in course of every episode, except the first step, because first step is context-free and it’s ok to always start from the same action. Second term softens predicted action distribution avoiding non-alternative decisions during training.

The reward for every episode  $r$  equals 1 if the image is classified correctly and 0, otherwise. In the experiments (Malashin, 2021) number of allowed actions is less than a number of all classifiers in the pool and, therefore, direct penalty for resource consumption is not necessary.

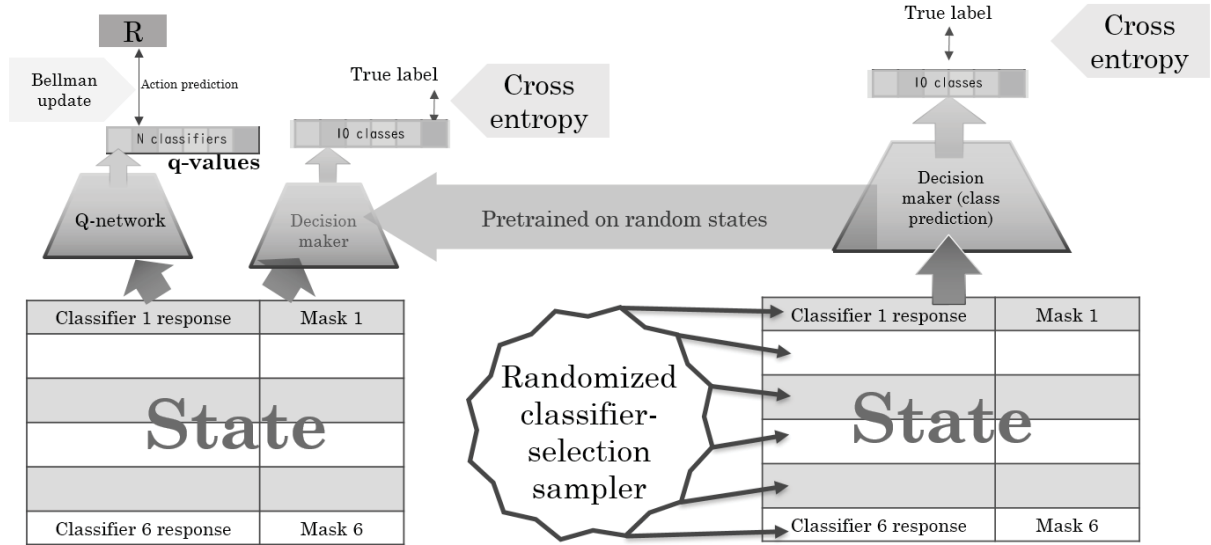
Malashin and Boiko (Malashin, Boiko, 2022) apply DQN for the sparse ensemble problem. The problem has a straightforward formulation in terms of DQN. Single Q-network is responsible for predicting values of actions two types:

- Index of the classifier to launch next;
- Index of predicted class image.

When Q-network selects action of the second type the recognition process stops. It appeared that without applying additional heuristics DQN approach very inefficient, therefore authors proposed hybrid DQN. The resulting modification of least action classifier is called QLAC.

For this, two modules are distinguished—a classifier selection module (Q-network) and a decision maker, which are taught independently to solve individual problems. The decision maker is responsible for predicting the label of the image given the state  $s_t$ , and the classifier selection module is responsible for selecting classifiers from the pool to update the state. This is very similar to the scheme depicted in Figure 3, but selection module is learned with bellman update instead of actor-critic loss.

To train the decision maker, which directly allows you to extract information about the class label, the cross-entropy loss function is used. Prediction is formed from states formed by random selection policy; these states can potentially occur with more advanced agent’s policy. Random policy consists of choosing a random set of  $L$  responses to an image. The value  $L$  is also chosen either randomly in the range from 1 to  $C$ , where  $C$  is the number of classifiers in the pool. The module is trained to produce true image label from the states obtained from the above algorithm. Figure 6 illustrates process of training hybrid DQN.



**Figure 6.** Hybrid DQN scheme (QLAC) (Malashin, Boiko, 2022)

LAC и QLAC were tested on CIFAR-10 dataset. Pool of classifiers was obtained by randomly sampling subsets of class indexes for training six individual classifiers according to formula (6). Table 1 depicts information about trained classifiers.

**Table 1.** Classifiers in the pool (Malashin, 2021)

index	Classes of images that were used for training	Test set accuracy (10 classes),%
0	0,1,8,4	35.6
1	1,2,3,5,6,7,9	57.1
2	3,2,4	24.6
3	7,2	18.3
4	0,1,6,7,8,9	51.0
5	0,2,3,5	29.5

Results of training LAC and QLAC in the pool are depicted in table 2.

**Table 2.** LAC and QLAC performance (Malashin, 2021)

Number of steps	LAC (Malashin, 2021)	QLAC (Malashin, Boiko, 2022)
1	67.8	67
2	75.81	76.15
3	77.81	77.74
4	78.62	78.01
5	79.1	77.71
6	79.29	77.17

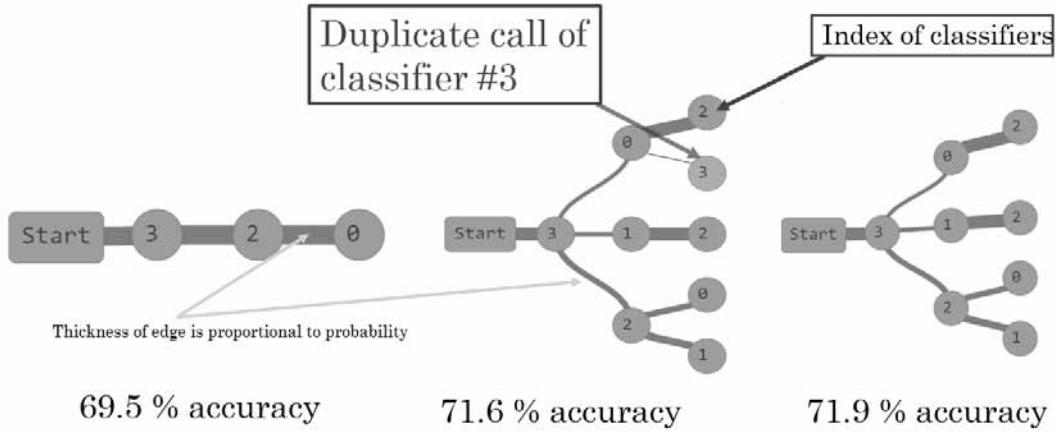
Table shows that the agent can incorporate information from multiple classifiers. We can also see that LAC outperforms QLAC. Better performance of LAC comes at the cost of heavily designed loss function which is much more complex and has more free parameters. QLAC performance starts to drop when number of steps exceeds 3. This can be caused by overfitting:

larger state space makes the model larger as a consequence.

To further analyze LAC results let's consider the following setup:

- use smaller pool of 4 classifiers (for details refer to (Malashin, 2021));
- restrict number of actions available for the agent to 3.

In this case, LAC outperforms neural network trained on responses of any 3 of 4 classifiers in the pool (by 2 %). That shows ability of LAC to exploit dynamic structure of computation. Figure 7 shows visualization of the computational graphs of LAC trained with different parameters and respective accuracies on the test set.



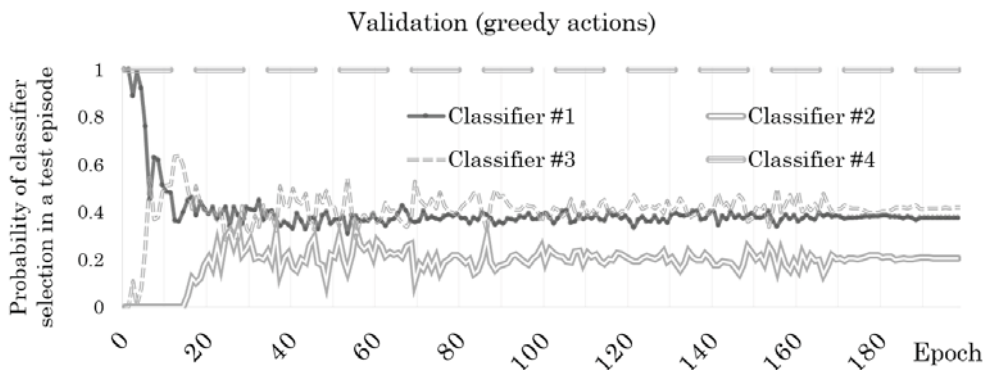
**Figure 7.** Computational graphs of trained Least Action Classifier. (a) LAC is trained with inappropriate training parameters, (b) LAC is trained with proper hyperparameters, (c) LAC with maximum accuracy on validation set

The computational graphs were constructed as follows:

- nodes are indexes of classifiers (actions);
- thickness of edge is proportional to the probability of the taken action.

As can be seen from the figure if the hyperparameters are chosen in such a way that final computational graph is unfolded dynamically than it leads to accuracy boost. Maximal accuracy corresponds to the dynamic computations without any duplicate actions.

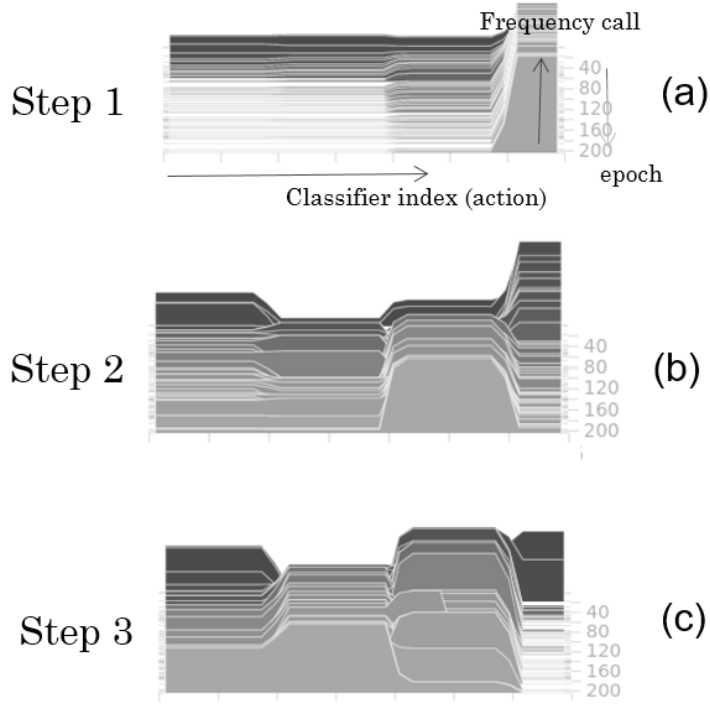
Figure 8 demonstrates change in the policy during the course of training.



**Figure 8.** Classifiers' call frequency on test set during course of LAC training (2 actions) (Malasin, 2021)

Agent finds the best combination of classifiers easily but it takes a lot of epochs to reveal good context-dependent dynamic configuration. On Figure 8 classifier 2 is ignored until the twentieth epoch.

The ability to find dynamic context-dependent policies depends on the loss function and entropy bonus (10) in particular. Figure 9 visualizes probability distribution of actions on each of three steps.



**Figure 9.** Distribution of actions for each of three steps of LAC

First part of entropy bonus (10) ignores first step, and distribution stabilizes around classifier #4 (Figure 9a). It looks reasonable to do so because agent knows nothing about the image and in this case wants to use the most informative classifier all the time. This helps good initialization. In the begging of training action distribution on 2<sup>nd</sup> and 3<sup>rd</sup> step stabilizes around the same classifier (Figure 9b and Figure 9c). This is related to the fact that decision maker is also learned during training and at first epochs better accuracy can be obtained if the hidden state is more stable. Then entropy bonus forces agent to use different classifiers than classifier #4 and exploration of better policies continues. Later action distributions on step 2 and step 3 changes so that at the end of the training the agent totally ignores classifier #4.

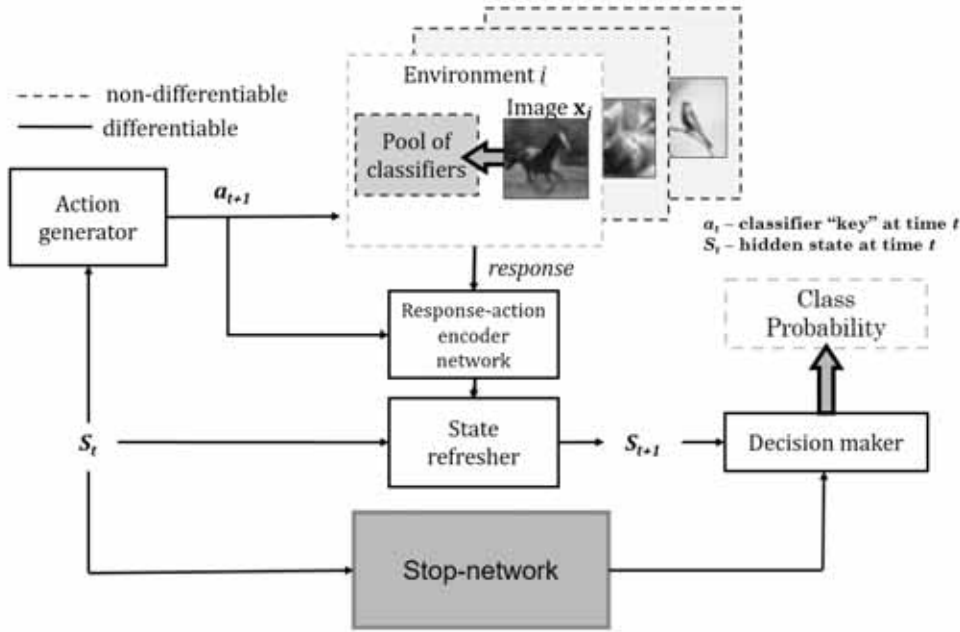
### Dynamic stop

There are two aspects to the dynamic configuration. The first aspect is the choice of suitable modules for a particular example, and the second is the ability to stop further calculations when a solution with acceptable quality is found (Bidencamp, 2020).

The second aspect can be especially useful for complex systems operating at a high level of abstraction, with multi-modal data, operating in real time. LAC and QLAC work with a fixed number of actions, so the following modification was proposed in (Malashin, 2022).

An additional module stop network was added to the architecture of least action. The stop network takes a hidden state vector as input, and at its output there are two neurons, one of which

generates a stop signal, and the second—a signal to continue calculations. Based on which signal value is greater, a decision is made either to continue calculations or to stop them and use the decision generated by the decision module in the current step. Architecture of LAC with stop network is depicted in Figure 10.



**Figure 10.** Least action classifier with stop network

The action selection network and the stop control network are trained with reinforcement, while the decision network is trained using cross entropy (and using a reference label).

To encourage short trajectories, when image is classified correctly episode reward  $R = a^n$ , where  $a \in [0,1]$  is a hyperparameter and  $n$  is the number of steps it took to recognize the image. Different coefficient  $a$  provides different average length of episodes; in experiments,  $a$  was chosen in the range  $[0.98, 0.99]$ .

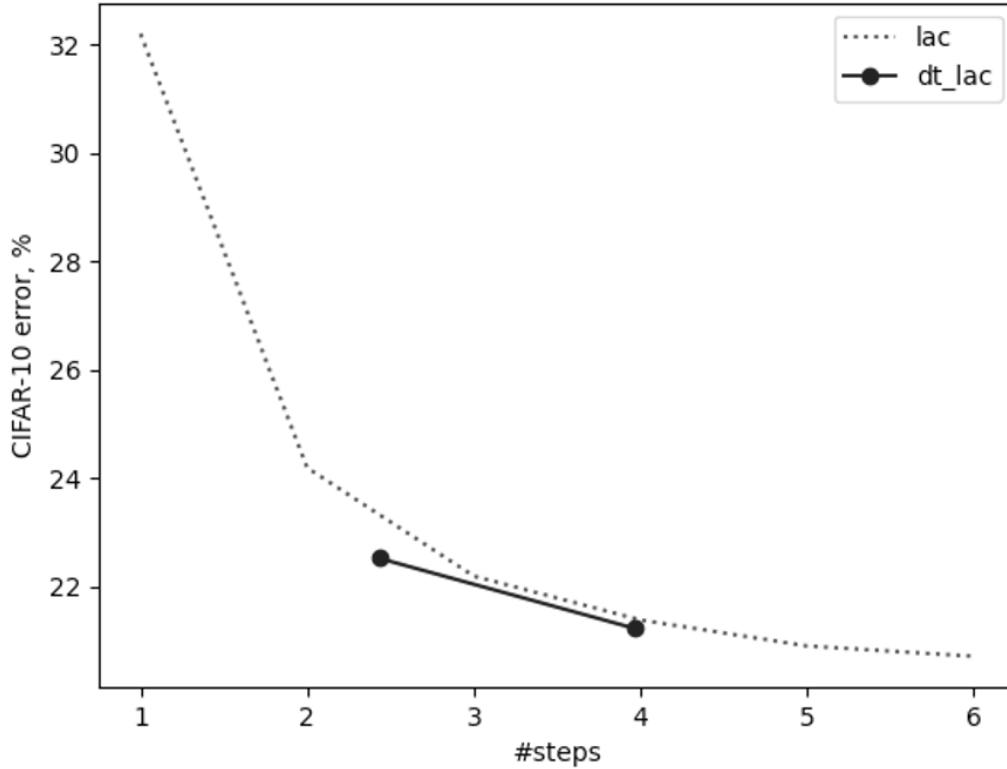
During the experiments, the use of curriculum learning turned out to be extremely important. The model is trained in three stages. First, Least Action Classifier is trained without the stop network, similar to how it was described in the previous section. At the second stage, the obtained weights are frozen, and a stop network is added and is trained to stop the calculations, provided that the action trajectories are fixed and set by the state of the rest of the model. At the third stage, the entire model is fine-tuned. Three-stage learning of the attention model was proposed in DT-RAM (Li et al, 2017) and a lot of modern complex dynamically configured systems use similar training scheme (Cheng, 2019), (Wang et al, 2020). By analogy with DT-RAM LAC with stop network can be called DT-LAC (LAC with dynamic time).

Pool with 6 classifiers is used for experiments. Table 4 shows performance of two instances of DT-LAC trained with different hyperparameter  $a$ .

**Table 4.** Performance of the least action classifier with dynamic computational time

Method	Test accuracy	Average number of steps
DT-LAC-1 ( $a=0.98$ )	77.48	2.43
DT-LAC-2 ( $a=0.99$ )	78.78	3.97

DT-LAC spends 2.43 steps in average and reaches 77.48 accuracy on CIFAR-10 dataset. DT-LAC-2 spends extra 1.5 steps for 1.3 % of accuracy. Figure 11 gives a visual comparison with the original LAC.



*Figure 11. Comparison of LAC and DT-LAC (Malashin, 2022)*

As can be seen from the figure, the dynamic termination of calculations increases the recognition accuracy for a given average episode length, however, in this case, the difference between calculations with a fixed duration and dynamic ones is not significant.

## Conclusion

Least action classifier (LAC) is an agent that perceives image through the set or pretrained classifiers. It has been shown (Malashin, 2021) that carefully trained LAC can get benefit by exploiting dynamic configuration. Still there is a lot of ways to improve the architecture and in conclusion we would like to mention several modifications we consider to be promising.

**Loss function with oracle bonus.** Entropy bonus (10) of the loss function encourages the entropy of the distribution of actions has a number of drawbacks. First of all, it is necessary to use large data packages (256 images were used in the current work). The second drawback is more important. In the learning process a random action is selected in accordance with the distribution predicted by the action generator. In the case the increase in the entropy bonus can be achieved by predicting a uniform distribution with respect to all classifiers that have not been selected before. This gives the agent to “cheat” and get entropy bonus without usage of dynamic configuration. This undesirable “trick” may let network to avoid binding classifier selection to context. Manifestation of this effect can be detected by visualizing probabilistic computation graphs on a test sample when a greedy action is selected (similarly to graphs depicted in Figure 7). LAC at the beginning of training often learns just such strategies, however, this avoids falling into the local minimum of the objective function, which, with a sufficient number of iterations, allows obtaining the desired context-sensitive strategies.

To overcome the indicated shortcomings, one could consider an approach with the introduction of penalties for repeating trajectories, however, it is worth noting that there is an exponentially large (to the number of classifiers) symmetric trajectories, in practice this approach cannot be implemented explicitly. For its approximation, one can consider using oracle neural network.

This network attempts to predict the output of the action generator from the “history” of previous actions in the current episode. If the predictions are systematically accurate, then this indicates that the context is not used by the action generator and such strategies can be “penalized”. It is important that this method can be fully applied during the stochastic learning procedure, since the “oracle” predicts the distribution of actions, and not a specific action (the distribution itself must be context-sensitive).

**Hierarchical agents.** If the final output of the agent in the classification problem, is the label of a recognized class, which means that it is in the response space of the agent itself (Malashin, 2019). Thus, a trained agent can itself be a module used by a higher-level agent. In this case, additional uncertainty is introduced by the fact that, unlike CNN classifiers, the agent itself is a dynamically configurable system, which means that its operational time is different for different input data. However, in the paradigm of reinforcement learning and approximation of the desired function using a neural network, this does not introduce fundamental difficulties. Training higher-level agents enables gradual training of increasingly complex and efficient models. Calling another agent is essentially a transfer of control to it. For example, in a “guess the animal” game, a low-level agent might have a list of questions ready to guess any bird or any animal from Australia.

The training of a high-level agent has common features with the algorithm (Rosenbloom, 2006), which was created on the basis of the observation that a person perceives and remembers information in fragments. So, for example, a professional chess player uses “fragments” consisting not of one position, but immediately of a chain of the following actions. To solve a problem, a person decomposes it into subproblems. The authors of the algorithm (Rosenbloom, 2006) assumed that if a solution to a certain subproblem exists, then this allows you to run an algorithm for solving the problem without further splitting into subtasks.

**Unsupervised learning.** LAC is trained with reinforcement, but this requires labeled data: reinforcement depends on the coincidence or mismatch of the predicted and ground truth labels. In addition, the decision-making module is trained directly by supervised signal (cross entropy of the reference and predicted class distributions). This significantly reduces the amount of data available for training and leads to bright pronounced effect of retraining.

It is possible to suggest methods to use unlabeled data for LAC training. For example, one can learn useful features by forcing agent to predict full table of classifiers responses based on only few responses.

Applying unsupervised learning might be promising direction for future research as it can make scaling LAC.

## References

1. Adriaensen, S., Biedenkapp, A., Shala, G., Awad, N., Eimer, T., Lindauer, M., and Hutter, F. Automated Dynamic Algorithm Configuration. 2022. arXiv preprint arXiv:2205.13881.
2. Adrià Ruiz, Jakob Verbeek. Adaptive Inference Cost with Convolutional Neural Mixture Models. ICCV. 2019. 1872–1881.

3. Biedenkapp A., Furkan Bozkurt H., Eimer T., Hutter F., Lindauer M. Dynamic Algorithm Configuration: Foundation of a New Meta-Algorithmic Framework. ECAI. 2020.
4. Bolukbasi, T., Wang, J., Dekel, O., Saligrama, V. Adaptive neural networks for efficient inference. In International Conference on Machine Learning. 2017. 527–536.
5. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., et al. Language models are few-shot learners. Advances in neural information processing systems. 2020. 33:1877–1901.
6. Cheng A., Lin C. H., Juan D., Sun M. InstaNAS: Instance-aware Neural Architecture Search. 2019. arXiv preprint arXiv: 1811.10201
7. Devlin J., Chang M., Lee K., and Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2018. arXiv preprint arXiv: 1810.04805
8. Fedus W., Zoph B., Shazeer N., Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. 2021. arXiv preprint arXiv: 2101.03961
9. Gururangan S., Lewis M., Holtzman A., Smith N., and Zettlemoyer L. Demix layers: Disentangling domains for modular language modeling. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022.
10. Larochelle H. and Murray I. The Neural Autoregressive Distribution Estimator. AISTATS. 2011.
11. Li Z., Yi Y., Liu X., Zhou F., Wen S., Xu W. Dynamic Computational Time for Visual Attention. 2017. arXiv preprint arXiv: 1703.10332.
12. Malashin R. Dynamic termination of computations in computer vision systems. J. Opt. Technol. 2022. 89(8): 469–475.
13. Malashin R. Principle of least action in dynamically configured image analysis systems. J. Opt. Technol. 2019. 86(11): 678–685.
14. Malashin R. Sparsely Ensembled Convolutional Neural Network Classifiers via Reinforcement Learning. Proceedings of the 2021 6th International Conference on Machine Learning Technologies. 2021.
15. Malashin R., Boiko A. Training a dynamically configurable classifier with deep Q-learning. J. Opt. Technol. 2022. 89(8): 437–447.
16. McGill M. Perona P. Deciding How to Decide: Dynamic Routing in Artificial Neural Networks. In Proceedings of the 34th International Conference on Machine Learning. 2017.
17. Murphy K. Machine Learning: A Probabilistic Perspective. Cambridge, Massachusetts: MIT Press. 2012. 1098 P.
18. Rosenbloom P. A cognitive odyssey: From the power law of practice to a general learning mechanism and beyond. 2006.
19. Shazeer N., Mirhoseini A., Maziarz K., Davis A., Le Q., Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. 2017. arXiv preprint arXiv: 1701.06538
20. Shelepin Y. E., Krasil'nikov N. N. Principle of least action, physiology of vision, and conditioned reflexes theory. Rossiiskii fiziologicheskii zhurnal imeni IM Sechenova 89.6 (2003): 725–730.
21. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser Ł., and Polosukhin I. Attention is all you need. NeurIPS. 2017. 30: 5998–6008
22. Viola P. and Jones M. J. Robust Real-Time Face Detection. International Journal of Computer Vision. 2004. 2(57): 137–154.
23. Wang Y., Lv K., Huang R., Song S., Le Yang, Gao Huang. Glance and Focus: a Dynamic Approach to Reducing Spatial Redundancy in Image Classification // NeurIPS. 2020. P. 2429–2441.
24. Zhou K., Yang Y., Qiao Y., and Xiang T. Domain adaptive ensemble learning. IEEE Transactions on Image Processing. 2021. 30:8008–8018.
25. Zoph B., Bello I., Kumar S., Du N., Huang Y., Dean J., Shazeer N., Fedus W. Designing effective sparse expert models. 2022. arXiv preprint arXiv: 2202.08906.

## **Chapter 12. The effect of transcranial Direct Current Stimulation on the electrical activity of the humans` cerebral cortex during facial expression emotion recognition**

*Meleshenko E.A., Egorova V.A., Ermakov P.N., Gorbenkova O.A.*

*Southern Federal University, Academy of Psychology and Educational Sciences, Rostov-on-Don, Russia*

The work aims to study the influence of transcranial Direct Current Stimulation on the facial expression emotion recognition in humans. In the work, the authors used the methods of transcranial Direct Current Stimulation (tDCS) and electroencephalography.

The authors attempt to show the possibility to affect the facial expression recognition by stimulating the specific parts of the cerebral cortex with a constant electric current of small quantities. As the literature review has shown, right lateral occipital sulcus, right prefrontal cortex, right lingual gyrus are responsible for the recognition of faces and emotions expressed, so the areas described were exposed to electrical stimulation. The study`s novelty is that in contrast to the experiments described by a variety of authors, where there is direct exposure of the brain structures to an electric current, the stimulation used in our research does not violate the anatomical integrity of the body.

Due to the results obtained, the non-invasive influence of electrical stimulation can change the time of facial expression recognition that is required to determine the emotion being demonstrated, as well as the quality of this recognition. As a result of exposure, a significant decrease in recognition time is detected in comparison with a placebo. In turn, the quality of facial expression emotion recognition decreases. The authors conclude that there is some disruption in the process of facial expression recognition. The analysis of the event-related potentials during facial expression recognition before and after electric stimulation or without it is also presented. The authors show the differences in the pattern of event-related potentials for three types of demonstrated emotions: negative, neutral and positive. These results confirm the assumption that the mechanism of facial expression emotion recognition at the physiological level has changed.

**Keywords:** EEG, ERP, tDCS, emotion recognition, visual stimulation, reaction rates.

### **List of abbreviations**

- EEG – electroencephalogram
- ERP – event-related potential
- MRI – magnetic resonance imaging
- fMRI – functional magnetic resonance imaging
- PET – positron emission tomography
- TMS – Transcranial magnetic stimulation
- tDCS – Transcranial Direct Current Stimulation

### **Introduction**

Currently, researchers actively study how different non-invasive methods of neuromodulation, namely tDCS, affect the functioning in various parts of the cerebral cortex. This article

aims to study the effect of tDCS on the cerebral cortex areas, which are responsible for visual recognition and evaluation of emotions. The study discovers whether it is possible to change the quality and speed of facial expression emotion recognition using tDCS and compares the patterns of the electroencephalography observed before and after stimulation. Similar studies were conducted mainly by invasive methods. This study evaluates the results of tDCS affecting the structures of the cerebral cortex involved in the recognition of emotional expressions.

## **Emotion recognition**

The perception of another person's emotions is an important factor in the interpersonal communication and interaction between humans. The evolutionary role of emotion perception is (1) to adapt to environmental conditions, (2) to identify and to differentiate positive and negative stimuli, and (3) to develop a system of how to respond to different stimuli. Moreover, emotions have a role of a signing system that allows individuals to interact (Сериков А. Е., 2018). To improve activities closely related to the need for interpersonal interaction, it becomes valuable to search for methods of influencing the emotions' perception.

In neuropsychology, the process of emotions' perception and their categorization is described sufficiently. There is a close relationship between the perception of one's own emotions and the ability to perceive the emotions of another person (Terasawa et al. 2013). For the classification of emotions, we used the circumplex theory of emotions (the Russell's model), based on the convenience of its application in psychophysiological research. The Russell's model (1980) assumes that valence and activation are independent, bipolar dimensions of emotions. Within the framework of the theory, Russell and co-authors identified 7 main emotions—neutral, fear, surprise, anger, disgust, sadness, joy. We used these emotions as proposed objects for categorization.

The perception of facial emotions' expressions itself seems to be a complex phenomenon, covering various parallel processes—perception of the face itself as well as the processes of facial expressions categorization by their emotional and semantic categories (Terasawa et al. 2013).

As it was mentioned earlier, there is a lack of works devoted to the possibility of influencing the process of perception and categorization of another persons' emotions. Among the non-invasive methods that influence the functioning of the cerebral cortex, transcranial magnetic stimulation (TMS) attracts much attention. The method has proven itself in clinical practice as a useful tool to help people with developmental delays and depression. Despite the possibility to influence a person's own emotional state with the help of TMS, there is not much data on the possibility to use TMS in order to study its influence on the perception of another persons' emotions (Маслеников Н. В., Э. Э. Цукарзи и др. 2013; Шорохова М. В. 2017).

Similar evidence we can find in the works devoted to the method of transcranial direct current stimulation. Various methods of influencing the functional state with the use of tDCS, and tDCS usage in assistance during the treatment of depressive disorders, developmental delays, etc. are well described. (Филимонова А. Е., Борсуков А. В., 2009; Корзенев А. В., Абриталин Е. Ю., 2010). However, despite the possibility of both methods to modulate the activity of individual brain structures, evidence about the possibility of tDCS to affect human perception is insufficient, because the implementation of tDCS was limited to specific conditions of clinical practice (Lewis et al. 2007).

In order to study the possibility of tDCS to influence the perception of human facial expressions, it is necessary to determine the intervention targets, which are associated with the process of recognition and categorization of emotions. These targets are specific parts of the brain involved in the process described above.

Transcranial Direct Current Stimulation (tDCS) is a method used in medicine and physiological research; a type of neuromodulation that uses a weak direct current supplied through electrodes on the scalp. It was originally developed to help patients with brain injuries or neuropsychiatric conditions such as major depressive disorder. There is evidence of the effect of tDCS on the functional state of various cerebral cortex areas (Rosa, MA; Lisanby, SH., 2012; Пушкин А. А., Сухов А. Г., 2020).

The mechanism of impact on the nervous tissue of tDCS is in the change of the neurons' membrane potential towards hyperpolarization or depolarization. This change depends on the direction of charged particles movement created by the tDCS itself. The consequence of such a change in the membrane potential is either an increase in the activity of neurons, or a decrease in activity, due to an increase or decrease in the threshold of their sensitivity. For example, exposure to an anode current, considering the location of neurons in the layers of the cerebral cortex with axons towards the skull, causes depolarization of the axon membrane, and the soma is slightly hyperpolarized due to its location somewhat further from the scalp surface, as through the surface the impact occurs. Thus, anodic stimulation causes greater activity of neurons exposed to it, cathodic stimulation, respectively, involves the contrary processes (Michael A Nitsche, Leonardo G Cohen, 2008; M. A. Nitsche and W. Paulus, 2000).

### **Cerebral cortex areas responsible for facial emotion expression recognition.**

Physiologically, the recognition of human facial emotions is carried out by a number of formations related to both cortical and subcortical structures of the brain. Many authors note that neural substrates of the right ventral prefrontal cortex, left fusiform gyrus and right lingual gyrus, and anterior insular cortex play an important function in this process (Cacioppo J. T., Gardner W. L. 1999; Steimer, 2002; MacLean, P.D., 1952).

The right ventral prefrontal cortex, according to the data available, is activated when a person is solving tasks related to facial expression emotion recognition, while ignoring other tasks related to facial recognition and verbal factors. The authors agree that this area of the brain takes part in the nonverbal emotional processing of incoming visual information. Researchers also state that the right ventral prefrontal cortex is activated at the moment of facial emotions recognition, during the process of assessing the attractiveness of faces, while it does not show any activity above the resting level during the tasks related to gender recognition. The lack of response in the area to verbal tasks is explained by the need for active interaction of both hemispheres in such tasks, which is consistent with the ideas of Bowers (6st). Despite the lack of correlation in the activity of the right ventral prefrontal cortex with the tasks of verbal recognition of emotions, many authors believe that right ventral prefrontal cortex is associated with the conscious evaluation of emotions. These findings are confirmed by the fact that neural activity in this area negatively correlates with neural activity in the amygdala, thus, the right ventral prefrontal cortex suppresses emotional reactions generated by the amygdala, and because of this fact right ventral prefrontal cortex is being associated with a conscious assessment of emotions. The connection with conscious processing of emotions is also confirmed in studies on categorization and comparison of human faces emotions, results of which

show activation in this area of the cerebral cortex (Cordaro D.T. et al., 2018; Cacioppo J.T., Gardner W.L. 1999; Gardner W.L. 1999; MacLean, P.D., 1952).

The left fusiform gyrus and the right lingual gyrus are also significantly activated during facial emotion recognition tasks, but the authors indicate that there is less activation of them for gender recognition tasks and verbal tasks. The authors interpret these results as meaning that these visual associative zones of the cerebral cortex participate in facial recognition but are not specific for processing emotions themselves. Thus, the left fusiform gyrus and the right lingual gyrus participate in the process of matching facial expressions with their verbal descriptions, which requires more resources than comparing facial expressions with each other without the mediation of verbal descriptions. At the same time, both in the case of positive and negative emotions, there is a relatively low activity of the amygdala, which is explained by the weakening of the unconscious processing of facial emotions, occurring at a certain level of tension in conscious processing (Tetsuya I., 2014).

The anterior insular cortex, as the authors show, plays a significant role in the perception of internal changes in the body (interoception) and the interpretation of information about the environment, which is important in the process of emotion recognition. This opinion is confirmed by fMRI studies that have shown that the right insular cortex is activated when a person is aware of his own emotions and bodily states (Terasawa et al., 2013). Also, the results of neuropsychological studies show that a damage to this area of the brain leads to deterioration in the recognition of emotions by facial expression, especially the recognition of disgust. Thus, the authors state that the anterior insular cortex plays a significant role in recognizing a person's own emotions and the emotions of other people, regardless of the emotions' category (Narumoto J., Yamada H., 2000; Kurth F., Zilles K., 2010).

It is noted that the anterior insular cortex has reciprocal connections with the frontal and anterior temporal cortex, and the posterior insular cortex has similar connections with the parietal and posterior temporal cortex. The anterior insular cortex is associated with cognitive and socio-emotional processing, and the posterior insular cortex is associated with sensorimotor processing (Kurth F., Zilles K., 2010). Thus, these data may indicate that the insular cortex is important for the integration of multimodal sensory information with high-level cognitive processing (Narumoto J., Yamada H., 2000).

Studies devoted to the electrical stimulation of the anterior insular cortex show that the data obtained in these studies can be well described and explained by the circumplex model of emotions. As this model describes, emotions can be grouped on a graph by indicators such as valence and arousal. In studies with electrical stimulation, it turned out that the insular cortex can fairly be considered a neural substrate that determines the level of arousal reflected in the circumplex model of emotions (Lewis et al. 2007). Similar results were obtained by the authors investigating the effect of changes in blood flow in the insular cortex—in the case of a decrease in blood flow, there is a blunting of interoception, which is the basis to recognize levels of arousal. Similarly, with electrical stimulation of the anterior insular cortex, the interpretation of emotions shifts towards greater excitement than this emotion really is. However, this area of the brain is associated with the assessment of the arousal level of a particular emotion and, apparently, does not affect the actual categorization of an emotion by its valence (Narumoto J., Yamada H., 2000; Kurth F., Zilles K., 2010; Hariri AR et al., 2000).

## Materials and methods

The study involved 30 healthy volunteers of both sexes (12 men, 18 women) which were students, postgraduates and employees of the Southern Federal University aged 17 to 30 years (the average age was 18.6 years). All volunteers were instructed and familiarized with the principles of how the devices operate. The meaning and mechanisms of the procedures for registering EEG and tDCS stimulation were also explained to them. Each subject was familiarized with the conditions of the study; the signed document on voluntary informed consent to participate in the study indicated the risks of the experiment, the right to interrupt the experiment at any time without explanation, as well as the obligation of the experimenter to anonymize the data obtained. The volunteers were divided into 2 groups—control and experimental, in which, respectively, no exposure was performed and was performed using micropolarization.

The electroencephalogram (EEG) was recorded using an encephalograph analyzer Neuro-image BMM-40. The EEG electrode placement scheme is a modified “10–20” scheme, from which the FP1 and FP2 electrodes were excluded due to the alleged large number of recorded artifacts.

Transcranial Direct Current Stimulation (tDCS) was performed using a Magnon-SLIP micropolarization device for 20 minutes; the current power was determined individually for each volunteer at a level below the pain threshold (no higher than 0.70 mA) based on the oral report of the volunteer. Anodes for tDCS were placed on the scalp in the area of the right lateral sulcus near T4, the right prefrontal cortex near F8 and the right lingual cortex near T6. Cathodes were placed in the area of the left fusiform gyrus, in the area of the Brock and Wernicke zones near F7 and T5.

Visual stimulation was performed using a Python script using the procedures of the PsychoPy application programming library. Registration of the response to the stimulus was carried out using a computer mouse. To minimize the error in determining and rendering the visual stimuli on the display, a photosynchronization sensor connected to the trigger channel of the encephalograph was used. The stimuli were a set of photographs of faces formed from the collections of KDEF (Lundqvist E., 1998) and WSEFEP (Olszanowski M., 2014), standardized in size, brightness and contrast; 2 photo banks of 320 images each were created. The banks included the following emotions: anger, neutral, sadness, joy, fear, surprise, disgust.

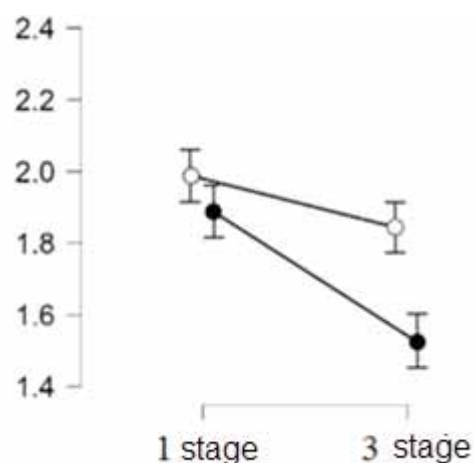
The empirical part of the study was divided into 3 stages:

1. The stage of stimuli presentation—registration of EEG, measurement of the speed of interpretation and the result of the presented stimuli interpretation (photos with an image of a face expressing a certain emotion) from the first photo bank of 320 images of emotion expression. The task of the volunteer was to identify negative (100 photos), neutral (120 photos) and positive (100 photos) facial expressions;
2. The stage of exposure—application of tDCS electrodes. Depending on whether the volunteer was in the experimental or control group, micropolarization was either carried out or simulated; during the tDCS, the EEG registration procedure was not carried out, and no stimuli were given—the volunteer was given time to rest;
3. The re-presentation stage—EEG registration, measurement of the speed of interpretation and the result of interpretation of the presented stimuli (photos with a face image) from the second photo bank of 320 images expressing emotions.

The data obtained were cleared of artifacts and outliers, after which they were subjected to statistical analysis with subsequent interpretation of the results of the study. To analyze the data obtained on the time and quality of responses in the JASP program, ANOVA with repeated measurements (before and after exposure) was performed for two groups (control and experimental). The analysis of event-related potentials (ERP) was carried out in the MatLab R2018b environment using the EEGLab 2022.0 tools; intra-group and inter-group comparison was carried out; variance analysis was used with repeated ANOVA measurements at a significance level of 0.05. The time interval from –200 ms before the presentation of the stimulus to 500 ms after the presentation was used to average the ERP.

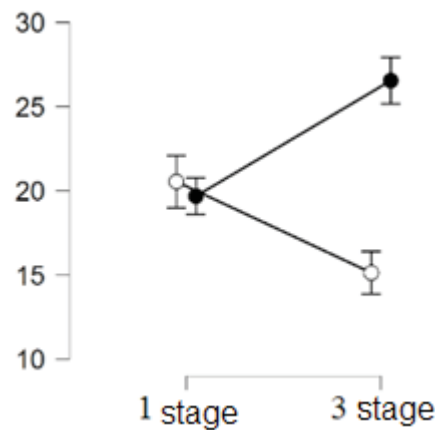
## Results

At the first stage of the study, the average response time to stimuli in the representatives of the control group was 2.01 seconds, in the representatives of the experimental group – 1.89 seconds. Differences in the reaction time of representatives in the control and experimental groups at the first stage of the study do not show significant differences ( $p=0.05$ ), which indicates the homogeneity of both groups of volunteers. At the third stage of the study, the representatives of the control group had an average reaction time to stimuli of 1.90 seconds. The average reaction time of the representatives in the experimental group at the 3rd stage was 1.57 seconds. These differences in the response time to the stimulus of representatives in the control and experimental groups are statistically significant ( $p=0.05$ ). The differences are presented in the Figure 1.



**Figure 1.** The reaction time to the presented stimulus in the control and experimental groups at the 1st and 3rd stages of the study. Note: white dots stands for the control group, black—for the experimental group; vertical scale is the average time (sec.)

The average number of errors in determining the facial emotion expression at the 1st stage of the study in representatives of the control group was 21.2; the average number of errors in determining the facial emotion expression in representatives of the experimental group was 19.2. These indicators do not show significant differences ( $p=0.05$ ), which also indicates the homogeneity of both groups. The average number of errors in determining the facial emotion expression at the 3rd stage of the study in representatives of the control group was 16.5; the average number of errors in representatives of the experimental group is 22.3. These differences in the number of errors in determining facial emotion expression in representatives of the control and experimental groups are statistically significant ( $p=0.05$ ). The differences are presented in the Figure 2.



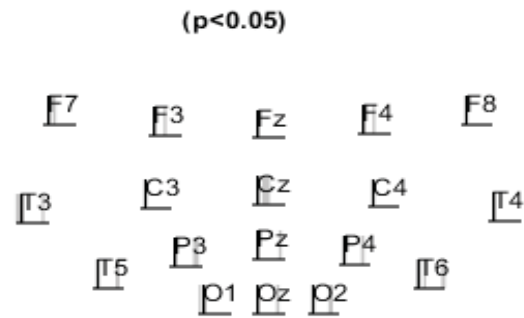
**Figure 2.** The number of errors in determining the facial emotion expression in the control and experimental groups at stages 1 and 3 of the study. Note: white dot stands for the control group, black—for the experimental group; vertical scale—the average number of errors.

When comparing the ERP in representatives of the control and experimental groups before and after placebo, or tDCS respectively, no statistically significant differences were found. Differences were obtained at the level of statistical tendency ( $p=0.10$ ) on the leads T4, O2, concerning the time period from  $\sim 280$  to  $\sim 300$  ms after the presentation of the stimulus. Statistically significant differences ( $p=0.05$ ) according to the leads P3, P4, O2, C3, C4, Cz, T3, concerning the time period from  $\sim 250$  to  $\sim 300$  ms after the presentation of the stimulus, were found. These results indicate that there are no significant differences between the representatives of the control and experimental groups at the 1st stage of the study, which mean some homogeneity of these samples. The differences noted after placebo and tDCS may indicate a significant presence of the tDCS effect on representatives in the experimental group. When comparing the ERP of representatives in the control group before and after placebo, differences were obtained at the level of statistical tendency ( $p=0.10$ ) for leads F8, T6, O1, O2, concerning two time periods  $\sim 80$ – $100$  ms after the presentation, as well as after  $\sim 330$  ms after the presentation of the stimulus. These differences affect the trace potentials of the ERP and the ERP's component N1, however, they are not statistically significant ( $p=0.10$ ). These differences may indicate the presence of a placebo effect, traceable at the level of a statistical trend, but no statistically significant differences were found. The differences are presented in the Figures 3,4,5.

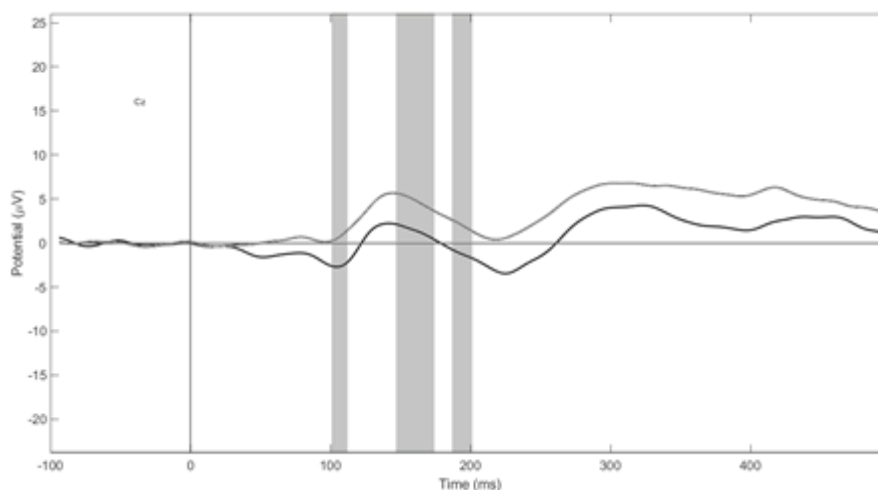
When comparing the ERP of representatives in the experimental group before and after micropolarization, depending on the nature of the stimuli presented statistically significant differences were found ( $p=0.05$ ). When perceiving negative facial expressions, differences were observed in leads F3, Cz, F7, concerning time periods  $\sim 100$ – $130$  ms and  $\sim 150$ – $210$  ms after the presentation of stimuli.

According to the P2 component, there are no differences in the time of appearance in the representatives of the experimental group before and after the tDCS, but in amplitude, there are differences in the representatives in the experimental group before and after the tDCS. It can be assumed that these changes in the P2 component are related to the orientation of attention to the semantic category of the stimulus. When attention focused on semantic categories is activated, there is a decrease in the latency of other potentials. The P2 component can be observed in cases where the processing of stimuli affects different levels of perceptual analysis.

Changes in component N1 may indicate changes in pre-attentive processing, allowing an individual to process a certain number of pieces of information and evaluate the number of potential targets for directing attention, choose the most adequate of them. Also, it can be assumed that it allows to effectively switch to adequate goals in the conditions when a variety of stimuli is presented, which is accompanied by the choice of a certain reaction; it has a connection with the mobility of nervous processes during perception, as well as with the speed of information processing in neural networks.

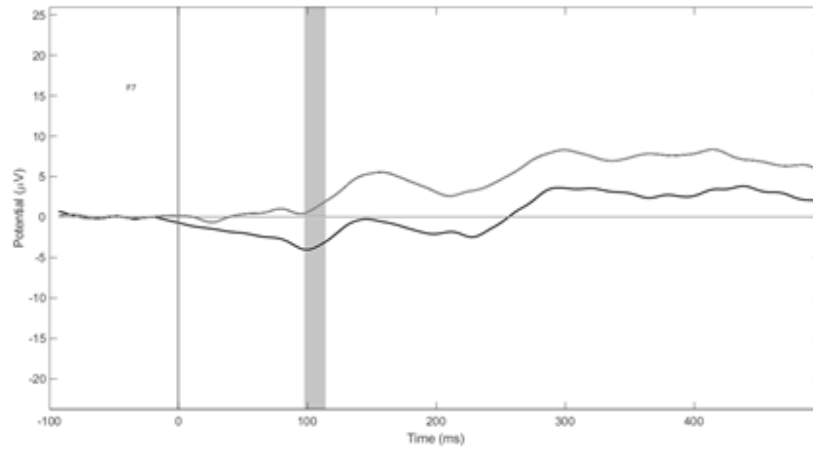


**Figure 3.** The epochs comparison of the experimental group before and after the intervention when presenting photographs of people expressing negative emotions.

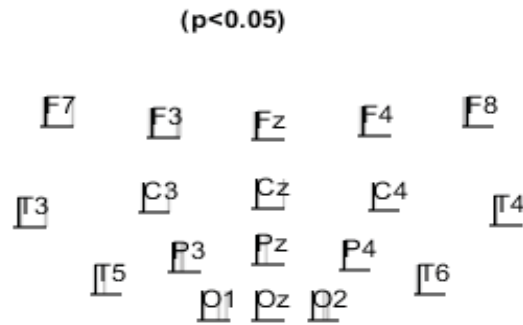


**Figure 4.** Comparison of ERP by the Cz withdrawal of the experimental group before and after micropolarization when interpreting photographs of people expressing negative emotions. Note: the average ERP after micropolarization is displayed in grey, before micropolarization in black, the vertical line series shows significant differences at a significance level of 0.05.

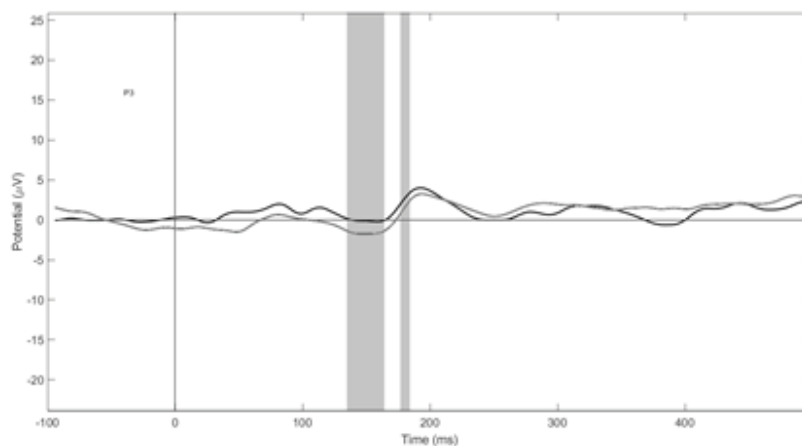
In the perception of neutral facial expressions, differences were observed in the leads T5, P3, concerning the time period ~150–200 ms after the presentation of stimuli. For this component, there is a relative increase in the negativity of the N2 component in the representatives in the experimental group after the tDCS of small current in comparison with the representatives in the experimental group before the tDCS. At the time of the component occurrence, representatives of the experimental group before and after micropolarization have no differences. The N2 component manifests itself when activating the processes of working memory, correcting the process of complex sensorimotor reaction. The differences are presented in the Figures 6,7.



**Figure 5.** Comparison of ERP by the F7 lead of the experimental group before and after micropolarization when interpreting photographs of people expressing negative emotions. Note: the average ERP after micropolarization is displayed in grey, before micropolarization in black, the vertical line series shows significant differences at a significance level of 0.05.

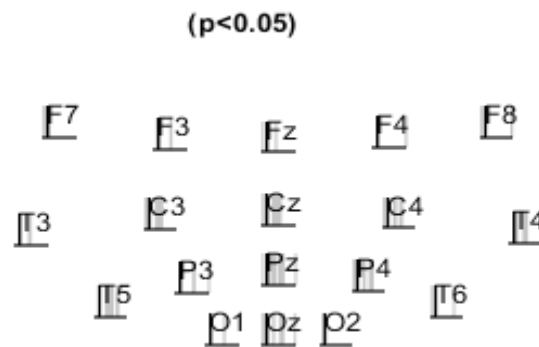


**Figure 6.** The epochs comparison of the experimental group before and after the intervention when presenting photographs of people expressing neutral emotions.

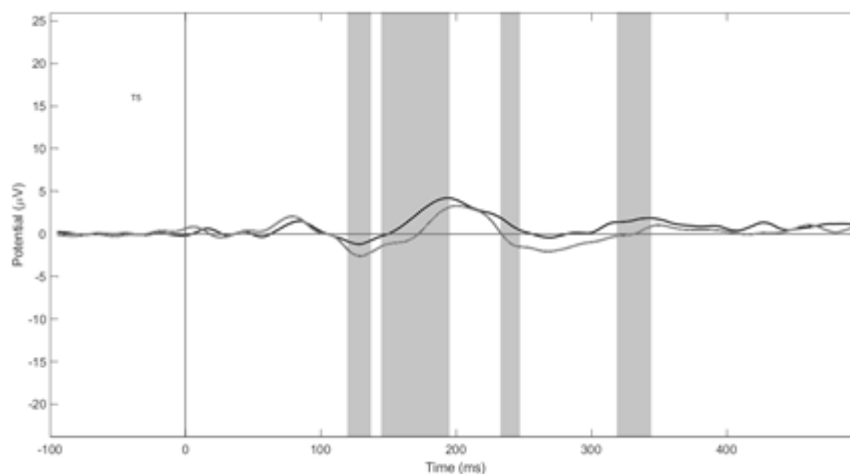


**Figure 7.** Comparison of ERP by the P3 lead of the experimental group before and after micropolarization when interpreting photographs of people expressing negative emotions. Note: the average ERP after micropolarization is displayed in grey, before micropolarization in black, the vertical line series shows significant differences at a significance level of 0.05.

When perceiving positive facial expressions, differences were observed in leads C3, C4, Cz, Pz, T5, concerning time periods ~140–230 ms, ~ 250–300 ms after presentation of stimuli. According to the P2 component, there are no differences in the time of appearance in the representatives of the experimental group before and after the tDCS, the differences manifest themselves in a slight increase in negativity. It can be assumed that these changes in the P2 component are related to the orientation of attention to the semantic category of the stimulus. When attention focused on semantic categories is activated, there is a decrease in the latency of other potentials. The P2 component can be observed in cases where the processing of stimuli affects different levels of perceptual analysis. The differences are presented in the Figures 8,9.



**Figure 8.** The epochs comparison in the experimental group before and after the intervention when presenting photographs of people expressing positive emotions.



**Figure 9.** Comparison of ERP by the T5 lead in the experimental group before and after micropolarization when interpreting photographs of people expressing negative emotions. Note: the average ERP after micropolarization is displayed in grey, before micropolarization in black, the vertical line series shows significant differences at a significance level of 0.05.

For this component, there is a relative increase in the negativity of the N2 component among the representatives of the experimental group after the tDCS in comparison with the representatives of the experimental group before the tDCS. At the time of the component occurrence, representatives of the experimental group before and after micropolarization have no differences. The N2 component manifests itself when activating the processes of working memory, correcting the process of complex sensorimotor reaction.

## Discussion

Differences in the speed and quality of facial emotion expression recognition show a decrease in the time required for this determination, while reducing its quality during tDCS placed on the areas of the scalp. These results can be explained by the fact that the electric current passing through the motor areas of the cerebral cortex had a stimulating effect on them, reducing the necessary time for a complex visual-motor reaction. At the same time, the impact on the areas of the cortex, perceived as involved in the processing of negative facial expressions by the anode current, introduced some disturbances in the process of perception and processing of information itself, due to, probably, their greater activation, which could result in the indicated decrease in the quality of facial expression recognition.

The data obtained by electroencephalography can be interpreted as tDCS leads to significant changes in the following components of the ERP—N1, N2 and P2. There are differences in the time of the components' occurrence, as well as changes in their amplitude.

Changes in the ERP's component N1 were observed when analyzing the perception of negative facial expressions by leads Cz and F7 and were accompanied by a slightly earlier appearance of this component and a slight increase in its amplitude relative to the data obtained before the tDCS. Observations of such changes exclusively in the case of perception of negative stimuli, may indicate an increase in the role of assessing the physical characteristics of stimuli and attention processes in connection with anodic stimulation of the right lateral sulcus, right prefrontal cortex and right lingual cortex. It can be assumed that this stimulation could provoke excessive inhibition of these areas of the cortex, which requires additional research.

Significant changes in the ERP's component N2 were observed when presenting all variants of stimuli: F3, Cz when presenting negative facial expressions; T5, P3 when presenting neutral facial expressions; C3, C4, Cz, Pz, T5 when presenting positive faces. Taking into account the equal representation of the diversity of the stimulus material at the 1st and 3rd stages of the study, it is likely that changes in the amplitude of the ERP's component N2 are associated with the activation of attention processes and an increase in cognitive load. It can be assumed that the anodic stimulation and a decrease in the quality of facial expression recognition contribute to the tension and deterioration of the entire mechanism of facial expression emotion recognition, since these changes were recorded in all groups of presented stimuli.

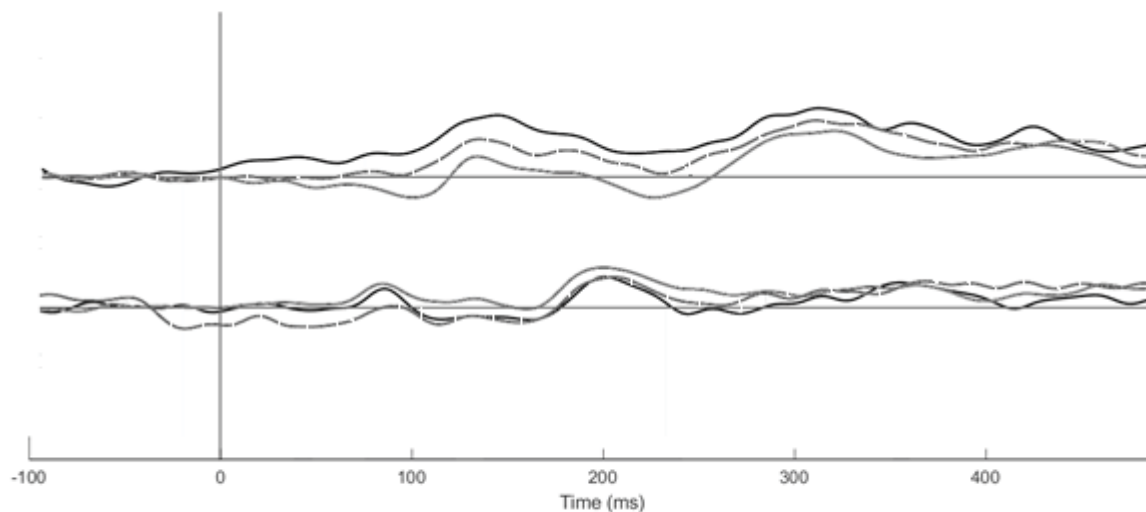
Changes in the amplitude of the P2 component were also observed in the perception of all groups of stimuli: by the Cz lead in the perception of negative stimuli; by the T5 lead in the perception of neutral stimuli; by the C3, C4, Cz, T5 leads in the perception of positive stimuli. These results may possibly indicate an increase in the activity of the neural network responsible for the perception and semantic categorization of emotionally significant heterogeneous stimuli during the anodic stimulation. Probably, these results may also indicate the increase in the intensity of the recognition system and categorization of visual emotionally significant stimuli.

It is interesting to record statistically significant differences in electroencephalogram leads: in response to the demonstration of negative facial expressions, a change in bioelectric activity is observed mainly in the frontal leads; neutral—in the left temporal region; positive—in the central sulcus and left temporal region.

When comparing the electroencephalogram data before and after micropolarization, the absence or at least weak intensity of the ERP's component P3 is detected simultaneously with

the intensity of the ERP's component P2. These results may indicate that there is no need for an emergency categorization of perceived stimuli because these stimuli before and after the tDCS were familiar and easily categorized, but at the same time, the categorization processes themselves were present. A significant change in the P2 component was observed in response to the assessment of positive facial expressions in photographs.

When comparing the amplitude of the ERP components, it can be noticed before the tDCS that when perceiving positive facial expressions, positivity is higher compared to the perception of neutral facial expressions, in turn, the perception of neutral facial expressions shows greater positivity compared to the perception of negative facial expressions.; after micropolarization, these indicators change: positivity in the perception of negative facial expressions is higher in comparison with the perception of neutral facial expressions, and the perception of neutral facial expressions shows positivity higher in comparison with the perception of positive facial expressions. These indicators can probably indicate an increase in the activity of the recognition system and categorization of emotionally significant stimuli when perceiving negative facial expressions, simultaneously accompanied by a decrease in similar activity when perceiving positive facial expressions after the tDCS. Comparison of the ERPs is presented in the Figure 10.



**Figure 10.** Comparison of the ERP in the experimental group before and after micropolarization when interpreting photographs of people expressing negative, neutral, positive emotions. Note: the average ERP is displayed in grey when perceiving negative emotions, grey dashed line – neutral emotions, black – positive emotions; ERP is displayed on top before micropolarization, below – after.

It can be assumed that the impact of the described tDCS scheme can lead to changes in the work of the perception system and categorization of the facial emotion expression when presenting heterogeneous stimuli. Considering the electrode placement scheme and the results obtained, it can be assumed that an increase in activation or, if there was, an excessive inhibition of the right lateral sulcus, the right prefrontal cortex and the right lingual gyrus can lead to a deterioration in the recognition of negative facial expressions of a person, and also, probably, these changes introduce errors in the operation of this mechanism during perception and categorization of other expressions of human faces.

The results obtained demonstrate the complex operation of the entire system of facial expression emotion recognition and, accordingly, the influence of the areas primarily responsible for

recognizing negative facial expressions on the operation of areas or a system for categorizing other human facial expressions.

*Research was financially supported by Southern Federal University, 2022, 07/2020–02-AP.*

## References

1. Корзенев А. В., Абриталин Е. Ю. Преодоление резистентности депрессивных расстройств методом нейроэлектростимуляции. Сибирский научный медицинский журнал. 2010. 30 (5): 30–34
2. Маслеников Н. В., Э. Э. Цукарзи и др. Депрессии при шизофрении: оценка когнитивных функций в динамике при лечении транскраниальной магнитной стимуляцией. Социальная и клиническая психиатрия. 2013. 23 (1): 5–11
3. Пушкин А. А., Сухов А. Г., Лысенко Л. В., Попов Д. А., Руденко В. В., Мелещенко Е. А. Об особенностях влияния транскраниальной микрополяризации на пространственно-временную организацию биоэлектрической активности мозга при коррекции эпилептиформной пароксизмальной активности у детей. Вопросы практической педиатрии. 2020. 15 (1): 104–109
4. Сериков А. Е. Ч. Дарвин о выражении эмоций: судьба основных идей. Вестник Самарской гуманитарной академии. 2018. 2 (24): 30–40
5. Филимонова А. Е., Борсуков А. В. Использование транскраниальной электростимуляции в клинике (обзор литературы). Вестник Смоленской государственной медицинской академии. 2009. 1: 130–133
6. Шорохова М. В. Коррекционно-логопедическое сопровождение детей с нарушениями речи в учреждении здравоохранения (на примере психоневрологического диспансера). Концепт. 2017. 511: 1–7
7. Cacioppo J. T., Gardner W. L. Emotion. Annual Review of Psychology. Oxford Handbook of Political Psychology. 1999. 50: 191–214.
8. Cordaro Daniel T., Sun, Rui, Keltner, Dacher, Kamble, Shanmukh, Huddar, Niranjana, McNeil Galen Universals and cultural variations in 22 emotional expressions across five cultures. Emotion. 2018. 18 (1): 75–93.
9. Hariri AR, Bookheimer SY, Mazziotta JC. Modulating emotional responses: effects of a neocortical network on the limbic system. Neuroreport, 2000, 11: 43–48.
10. KDEF: E. Lundqvist, D., Flykt, A., & Öhman, A. The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.
11. Kurth F, Zilles K, Fox PT, Laird AR, Eickhof SB A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. Brain Struct Funct. 2010. 214 (5–6): 519–534
12. Lewis DA, Curley AA, Glausier JR, Volk DW. Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. Trends Neurosci. 2012. 35(1): 57–67.
13. MacLean, P. D. Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). Electroencephalography and Clinical Neurophysiology. 1952. 4 (4): 407–18.
14. Narumoto, Jin; Yamada, Hiroki; Iidaka, Tetsuya; Sadato, Norihiro; Fukui, Kenji; Itoh, Harumi; Yonekura, Yoshiharu. Brain regions involved in verbal or non-verbal aspects of facial emotion recognition. NeuroReport. 2000. 11 (11): 2571–2574
15. Nitsche M. A., Cohen L. G., Wassermann E., Priori A. Transcranial direct current stimulation: State of the art 2008. Brain Stimulation. 2008. 1 (3): 206–223
16. Nitsche M. A., Paulus W. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. The Journal of Physiology. 2000. 527 (3): 633–639
17. Rosa MA, Lisanby SH. Somatic treatments for mood disorders. Neuropsychopharmacology. 2012. 37: 102–116.

18. Steimer, Thierry The biology of fear-and anxiety-related behaviors. *Dialogues in Clinical Neuroscience*. 2002. 4 (3): 231–49.
19. Terasawa, Y., Fukushima, H., and Umeda, S., How does interoceptive awareness interact with the subjective experience of emotion? An fMRI Study. *Hum. Brain Mapp*. 2013. 34 (3), 598–612
20. Tetsuya I. Role of the fusiform gyrus and superior temporal sulcus in face perception and recognition: An empirical review. *Japanese Psychological Research*. 2014. 56 (1): 33–45
21. WSEFEP: Olszanowski M., Pochwatko G., Kuklinski K., et al. Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Front Psychol*. 2015. 5

## Chapter 13. Features of Visual Information Processing in Patients with Schizophrenia

*Murav'eva S. V., Schemeleva O. V., Lebedev V.S., Vershinina E.A.*

*Pavlov Institute of Physiology, Academy of Sciences, St. Petersburg, Russia*

We performed the analysis of electrophysiological markers of visual information processing in schizophrenia. We studied the nature of changes in the amplitudes of the components of evoked potentials in response to the presentation of a combination of stimuli with different spatial-frequency and semantic characteristics (objects of animate and inanimate nature) in patients with schizophrenia in the early stages. The obtained data indicated a predominant decrease in the activity of the “high-frequency” parvo system, which manifests itself in a perception disorder and the abnormality of processing of small images and their details. Also, we obtained data in patients with schizophrenia that signifies an abnormality of the involuntary classification of images of objects of animate and inanimate nature. The obtained result is important for the understanding of the features of visual information processing in patients with schizophrenia in the early stages of the disease and the development of methods of cognitive impairments measuring.

**Keywords:** cognitive visual evoked potentials, low and high spatial frequency, stimuli of animate and inanimate nature, schizophrenia

### Introduction

Disturbances in the processing of visual information in schizophrenia have been known for a long time, yet they still require more in-depth study. It is known that multiple structural and functional disorders of eyes and various brain structures can be observed in schizophrenia (Козуб и др., 2020; Муравьева, Козуб, 2021; Silverstein S.M., Rosen, 2015). On the one hand, they can be associated with the disease, and on the other hand, with medication or concomitant diseases. Visual processing disorders are well known in schizophrenia, including contrast sensitivity disorders (Kelemen et al., 2013); various excitatory and inhibitory functions, shape and movement processing (Robol et al., 2013). Despite this growing body of evidence, the extent to which the observed problems are connected with the changes in the brain versus changes in the structure of the eye remains unanswered. Abnormalities (i.e., hypo- or hyperactivation) have already been shown to exist in the occipital, temporal, parietal, and prefrontal regions during various cognitive tasks in schizophrenia (Cruz-Martinez, Reyes-Garcia, 2022).

The relevance of this work is determined by the advantages of combining of the method of cognitive visual evoked potentials and the method of spatial-frequency filtering of images with different semantics in order to detect disorders. This method allows to assess the functional state of the visual system in the early stages of cognitive impairment, based on the objective electrophysiological methods. To assess cognitive impairment in patients with neuropsychiatric pathology, the method of cognitive visual evoked potentials has long been used. Based on this method, a subtle research method was developed and tested in the clinic—with the perception of images of objects processed by the wavelet filtration, for the selective assessment of the system of “object” vision (parvocellular system) or “spatial” vision (magnocellular system).

This work is devoted to the analysis of the data obtained to assess the role of the brain in

visual information processing disorders in schizophrenia. The electrophysiological method of image classification was chosen as a research method. The purpose of this work is to analyze electrophysiological markers of visual information processing in schizophrenia with a disease duration of 1 to 7 years. The objectives of the study were to analyze the ratio of the amplitudes of the early and late components of visual evoked potentials obtained using the image classification method.

## **Materials and Methods**

The study involved 27 patients with a paranoid form of schizophrenia (F20 according to the ICD10) with a disease duration of 1 to 7 years, including 15 men and 12 women. All observers were without ophthalmic pathology and had visual acuity of at least 0.9. All subjects had given written informed consent before the study. All patients who took part in the studies were in hospital and received antipsychotic therapy. Emotional uniformity, hypomimicity, tension and alertness, and anxiety should be noted among the symptoms that were observed in most patients with schizophrenia who participated in the study. Auditory hallucinations (voices) were also characteristic. The patients were characterized by the dominance of productive symptoms over negative ones.

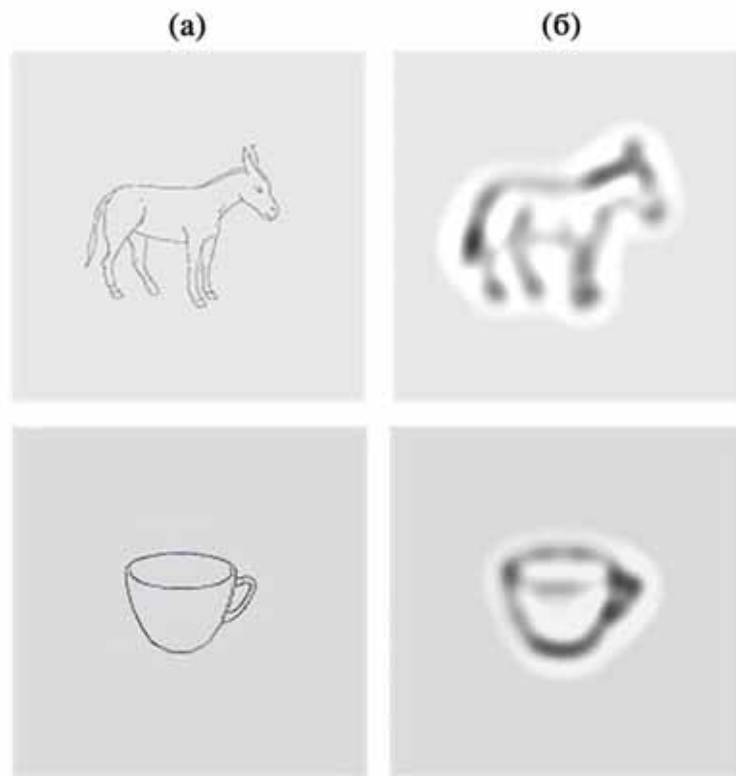
Following disturbances of the brain functions were noted: disturbances in the structuring of thinking, a decrease in the ability to concentrate.

Disturbances of behavior did not determine the clinical picture of the examined patients.

Subjects sat in an acoustically treated darkened room. Stimuli were presented on a monitor screen (17 inches) at a distance of 1 m. The effectiveness of the technique and the control of exposure were evaluated using the electrophysiological method (cognitive visual evoked potentials), described previously [4,6,7]. We used comparative analysis of the amplitude of the components of the evoked potentials during perception of images of objects that were filtered by spatial frequencies using digital wavelet filtering to selectively affect the magno (low-frequency) and parvo (high-frequency) channels of the visual system. The stimuli were digitally filtered images so that half of them contained only the low-frequency part of the spatial frequency spectrum, and the other high-frequency. To isolate high spatial frequencies, a filter was used with a maximum transmittance at a frequency of 10 cycles/deg; and low frequencies, 1 cycle/deg. Filtration was performed by convolution of images with the DoG-function (abbr. from Difference of Gaussians), a wavelet representing the difference of two two-dimensional Gauss functions with different half-widths. We used a set of 90 monochrome contour images in shades of gray (45 images of animate, 45 images of inanimate nature). The average brightness and contrast of all images were the same. Images were presented in random order for 100 ms with an interval of 1 s. The observer's task was to press the mouse button as quickly as possible: to the left when he/she saw an object of animate nature for half cases and to the right when he/she saw an object of inanimate nature (distinguishing objects according to semantic attributes) for half cases.

Recording of the electroencephalogram was performed using an electroencephalographic cap (ElectroCap, International Inc., United States) with 19 electrodes located on the surface of the head in accordance with the 10–20 International System in the leads Fp1; Fp2; F7; F3; Fz; F4; F8; T3; C3; Cz; C4; T4; T5; P3; Pz; P4; T6; O1; O2. Reference electrodes were placed on the earlobes, and a ground electrode was placed in the frontal region. To record evoked potentials, an encephalograph (Mitsar, Russia) with a sampling frequency of 250 Hz and the WinEEG software were used. To process the EEG recording, the method described in [5,6]

was used. The analysis period of 700 ms was divided into time intervals according to components: N60; P100 (N100); N170 (P170); N200; N250 (P250); P300; P500. For each lead in each time window, the amplitude values were found corresponding to the peaks of the main components of the evoked potential (maximum for positive waves and minimum for negative waves). After that, statistical comparison of the amplitude values of the components obtained in response to the presentation of various types of stimuli was performed. The data of evoked potentials were averaged in all patients included in the study group. Statistical analysis of the data was performed using two-way analysis of variance for the dependent variables ANOVA, with such factors as Frequency (High vs Low) and Animacy (Animate vs Inanimate). Statistical decisions were made at the 5 % significance level, data are presented as mean  $\pm$  standard deviations Mean  $\pm$  SD (the graph shows the means with errors of the mean SE). The calculations were carried out using the software package in the SPSS Inc 13 program [1].



**Figure. 1.** Examples of two categories of black and white images, animate and inanimate category, filtered by the wavelet transform in the (a) high and (b) low spatial frequencies.

Average evoked potentials of patients with schizophrenia in the central vertex lead (Cz), in the occipitotemporal lead on the left (T5) and on the right (T6), in the central parietal lead (Pz) and in the occipital leads on the left (O1, O2), on imaging stimuli objects of animate (A) and inanimate (B) nature, filtered by low (blue line) and high (black line) spatial frequencies. Asterisks show significant differences depending on the level of significance when comparing the ratio of the amplitudes of one component within each semantic group. Significance level: “\*” –  $p < 0.05$ ; “\*\*” –  $p < 0.01$ ;

There is a significant decrease in the amplitude of the early components of evoked potentials–N60 and P100 to stimuli of high spatial frequencies compared to low ones in the occipital and occipital-temporal regions on the left. These results are consistent with those of the control group. Late VEP components (N170 (P170); N200; N250 (P250); P300 and P500) are characterized

either by a decrease in the amplitude of the components of visual evoked potentials, or there is no significant difference in response to stimuli of high spatial frequency compared to low. In our early studies, the opposite ratio of the amplitudes of evoked potentials was obtained in the control group.

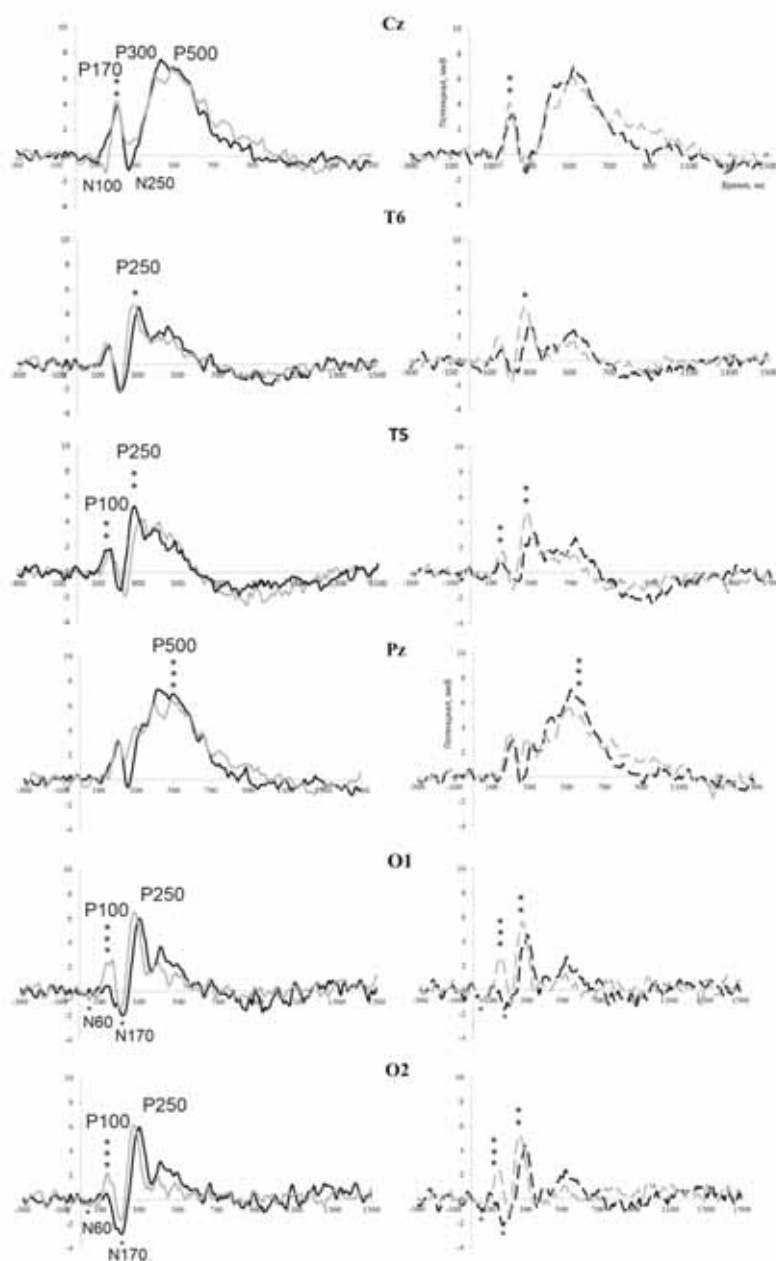
## Results

As a result of statistical analysis of patient data, when assessing the ratio of the amplitudes of the components of evoked potentials, depending on the spatial-frequency and semantic characteristics of the images, the following significant changes were revealed. Upon presentation of images containing the high-frequency part of the spectrum, a significant decrease in the amplitude of the components is observed compared with the response to images containing the low-frequency part of the spectrum, regardless of the semantics of the stimulus in the following areas of the brain (Table 2). In the occipital region (in leads O1 and O2), a significant decrease in the amplitude of N60, P100 and N170 was obtained, in the occipital-temporal region on the left (T5)–P100 and P250; in the central region (Cz)–P170; in the central parietal region (Pz)–P500. In the rest of the analysis zones, no significant difference between the amplitudes in response to images of high and low spatial frequencies was obtained. A significant difference in the amplitudes of the P250 (N250) component depending on the change in the semantics of the stimulus was observed in the occipital (O1 and O2), posterior temporal (T5, T6), central parietal (Pz) and central vertex (Cz) leads. The average evoked potentials with the level of statistical significance are shown in Figure 2.

**Table 2.** Significances of the main effects  $p$  ANOVA,  $df1=1$ ,  $df2=24$  ( $p < 0.05$ ) by leads for the components of evoked potentials depending on the characteristics of the stimulus: spatial frequency (low and high). Significance level: «\*»– $p < 0,05$ ; «\*\*»– $p < 0,01$ ; «\*\*\*»– $p < 0,001$ .

Lead	Component	Time	Spatial frequency of an image (LF and HF)
O1	N	60	LF > HF: 0,020 *
O2	N	60	LF > HF: 0,030 *
T5	N	60	0,184
T6	N	60	0,384
Cz	N	100	0,333
O1	P	100	LF > HF: 0,000 ***
O2	P	100	LF > HF: 0,000 ***
T5	P	100	LF > HF: 0,003 **
T6	P	100	0,181
Cz	P	170	LF > HF: 0,005 **
O1	N	170	LF > HF: 0,017 *
O2	N	170	LF > HF: 0,003 **
Pz	P	170	0,059
T5	N	200	0,184
T6	N	200	0,862
Cz	N	250	0,916
O1	P	250	0,395
O2	P	250	0,386
Pz	N	250	0,828
T5	P	250	LF > HF: 0,002**

T6	P	250	LF > HF: 0,010*
Cz	P	300	0,181
Pz	P	300	0,158
Cz	P	500	0,258
Pz	P	500	LF > HF: 0,000***



**Figure. 2.** Average evoked potentials of patients with schizophrenia in the central vertex lead (Cz), in the occipital-temporal lead on the left (T5) and on the right (T6), in the central parietal lead (Pz) and in the occipital leads on the left (O1, O2), on stimuli of images of objects of living (A) and inanimate (B) nature, filtered by low (gray line) and high (black line) spatial frequencies. Asterisks show significant differences depending on the level of significance when comparing the ratio of the amplitudes of one component within each semantic group. Significance level: “\*” –  $p < 0.05$ ; “\*\*” –  $p < 0.01$ ; “\*\*\*” –  $p < 0.001$ .

Thus, according to the data obtained, in patients with schizophrenia in these areas, upon pres-

entation of images filtered by high spatial frequencies, there is a significant decrease in amplitudes for the vast majority of VEP components, regardless of the semantics of the stimulus, and a decrease in amplitude in response to images of wildlife compared to with the non-living. At the same time, the average probability of correct answers in patients is at least 80 %, which corresponds to the data of the control group described earlier (Silverstein S. M., Rosen R. 2015).

## **Discussion**

According to our data, there is a significant decrease in the amplitude of the early components of evoked potentials—N60 and P100 to stimuli of high spatial frequencies compared to low ones in the occipital and occipital-temporal regions in the left hemisphere. These results are consistent with those of the control group, obtained in the course of early studies in our laboratory using the same technique (Мойсеевко и др., 2015; Муравьева и др., 2017). In the occipital and occipital-temporal regions, there is a decrease in the amplitude of the components responsible for primary perception (N60 and P100) in response to the presentation of images of high spatial frequency compared to low. This is because at the stage of primary perception, analysis of low frequencies comes first, while the analysis of high frequencies comes after with global analysis occurring before the detailed one (Муравьева и др., 2008). Thus, it can be said that in the occipital region and in the occipital-temporal region, the ratio of amplitudes does not change at the stage of primary perception compared with the data of the control group.

In this work, data were also obtained that the later components of VEP (N170 (P170); N200; N250 (P250); P300 and P500) are characterized by either a decrease in the amplitude of the components of visual evoked potentials, or there is no significant difference in response to stimuli of high spatial frequency compared to low. In our earlier studies, the opposite ratio of the amplitudes of evoked potentials was obtained in the control group (Муравьева и др., 2017). These components are responsible for the stages of primary differentiation, comparison with known images and primary recognition, translation into short-term working memory and decision making. Previous data from a group of healthy subjects are consistent with recent magnetic encephalography and fMRI studies showing that both dorsal and ventral regions are involved in both low and high spatial frequency imaging. In the primary visual area and ventral areas (topographic regions V1-V4, lateral and ventral occipitotemporal regions involved in visual processing), as well as in dorsal regions (topographic regions V3A, V3B, IPS01-IPS04, superior and inferior regions of the intraparietal sulcus) high-frequency images elicit a higher response amplitude than low-frequency images (Vitali P., 2005). Thus, our data on patients testify in favor of the fact that patients with schizophrenia with a disease duration of 1 to 7 years have a dominant decrease in the activity of the high-frequency parvo system at the late stages of visual information processing. This system, according to studies in newborns, is formed later (Vaziri-Pashkam M. et al., 2019) and can be developed through additional visual experience (Bosworth R. G., 2013).

## **Conclusion**

In the work, an electrophysiological method for classifying images with different spatial and semantic characteristics was used in patients with schizophrenia with a disease duration of 1 to 7 years. An analysis was made of electrophysiological markers of visual information processing in schizophrenia. Evidence has been obtained in favor of impaired visual information processing at later stages. Thus, the stages of primary differentiation, comparison with known images and primary recognition, transfer to short-term working memory and decision making, transfer to working memory are affected. This testifies in favor of brain dysfunction at the

level of deep processing of visual information. But in order to assess the contribution of the retina to these disorders, a deeper analysis is needed using the method of optical coherence tomographic angiography to investigate possible vascular changes in the retina. It is possible that in the later stages of the development of schizophrenia, changes in the retina will increase due to the pathological process or long-term drug therapy. However, at the early stages of the development of neuropsychiatric pathology, the described method of cognitive visual evoked potentials can be recommended for use in the clinic for an objective assessment of cognitive impairment and therapy monitoring.

*The work was supported by the State Program 47 of the State Enterprise “Scientific and Technological Development of the Russian Federation” (2019–2030), topic 0134–2019–0006 (section 63.3).*

## References

1. Козуб К. Е., Шелепин Ю. Е., Чомский А. Н., Шарыбин Е. А., Иванова Е. А. Структурно-функциональные исследования сетчатки при шизофрении // Офтальмологический журнал. 2020 № 4 С. 38–43.
2. Моисеенко Г. А., Шелепин Ю. Е., Хараузов А. К., Пронин С. В., Чихман В. Н., Вахрамеева О. А. Классификация и распознавание изображений живой и неживой природы // Оптический журнал. 2015. Том 82. № 10. С. 53–64.
3. Муравьева С. В., Дешкович А. А., Шелепин Ю. Е. Магно- и парвосистемы человека и избирательные нарушения их работы // Российский физиологический журнал им. И. М. Сеченова. 2008. Т. 94. № 6. С. 637–649.
4. Муравьева С. В., Козуб К. Е., Пронин С. В., Оптические и электрофизиологические методы оценки функционального состояния нейронных сетей зрительной системы // Оптический журнал. 2021. Том 82. № 12. С. 42–49.
5. Муравьева С. В., Пронина М. В., Моисеенко Г. А., Пневская А. Н., Поляков Ю. И., Кропотов Ю. Д., Пронин С. В., Шелепин Е. Ю., Шелепин Ю. Е. Исследование зрительных когнитивных вызванных потенциалов при шизофрении на ранних стадиях заболевания и их коррекция при помощи интерактивных виртуальных сред // Физиология человека. 2017. Том 43. № 6. С. 24–36.
6. Bosworth R. G., Dobkins K. R. Effects of prematurity on the development of contrast sensitivity: testing the visual experience hypothesis // Vision Research. 2013. Vol. 82. P. 31–41.
7. Cruz-Martinez C., C.A. Reyes-Garcia, N. Vanello A novel event-related fMRI supervoxels-based representation and its application to schizophrenia diagnosis Comput. Methods Prog. // Biomed., 2022. P. 213.
8. Kelemen O., Kiss I., Benedek G., Keri S., Perceptual and cognitive effects of antipsychotics in first-episode schizophrenia: the potential impact of GABA concentration in the visual cortex // Prog. Neuropsychopharmacol. Biol. Psychiatry 2013. Vol. 47, P. 13–19.
9. Robol V., Tibber M. S., Anderson E. J., Bobin T., Carlin P., Shergill S. S., et al., Reduced crowding and poor contour detection in schizophrenia are consistent with weak surround inhibition // PLoS One 8, 2013. e60951.
10. Silverstein S. M., Rosen R. Schizophrenia and the eye // Schizophr Res Cogn. 2015. Vol. 2. № 2. P. 46–55.
11. Vaziri-Pashkam M., Taylor J., Xu Y. Spatial frequency tolerant visual object representations in the human ventral and dorsal visual processing pathways // Journ. of Cognitive Neuroscience. 2019. Vol. 31. № 1. P. 49–63.
12. Vitali P., Abutalebi J., Tettamanti M., Rowe J., Scifo P., Fazio F., Cappa S. F., Perani D. Generating animal and tool names: An fMRI study of effective connectivity // Brain and Language. 2005. Vol. 93. P. 32–45.

## Chapter 14. Functional near-infrared brain spectroscopy for ergonomics and clinical practice.

*Rodionova I.<sup>A</sup>, Korotaev V.<sup>A</sup>, Shebut D.<sup>A</sup>, Ryzhova V.<sup>A</sup>, Khlynov R.<sup>A</sup>, Mikheev S.<sup>A</sup>, Shelepin Yu.<sup>B</sup>, Morozov S.<sup>B</sup>, Vasiliev P.<sup>B</sup>*

*<sup>A</sup>—ITMO University, Saint-Petersburg, Russia;*

*<sup>B</sup>—Pavlov Institute of Physiology RAS, Saint-Petersburg, Russia*

### Stating the Problem

Human brain is one of the body organs with the highest metabolic activity, which requires energy obtained with the help of oxygen consumption for its high physicochemical performance (Watts et al., 2018). Although the brain only makes up 2 % of the body weight, it consumes approximately 20 % of the total oxygen utilized by the body as a whole (Watts et al., 2018; Zhong et al., 2021). Therefore, to support the proper operation of the brain, continual supply of oxygen is needed.

It is critical in neuroscience to understand which part of the brain is activated during a specific activity.

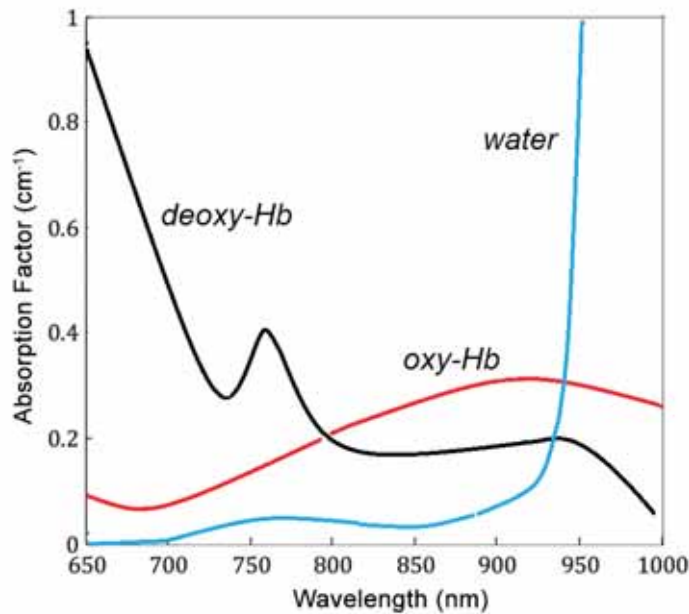
Functional magnetic resonance imaging (fMRI) produces colored images of the brain (Scarapicchia et al., 2017). A specific color denotes brain regions with increased or decreased oxygen levels (Scarapicchia et al., 2017). Since neurons require oxygen to function, fMRI can literally show where neurons are active.

Another important reason for the need to control brain oxygenation is the possibility of a wide range of diseases, such as stroke, traumatic brain injury, tumors, arterial stenosis, Alzheimer's and Parkinson's diseases, etc. (Watts et al., 2018). Also, oxygen saturation monitoring is conducive to the early diagnosis and treatment of cerebral ischemia and hypoxia (Zhong et al., 2021). The development of such diseases correlates with oxygenation disruption of the brain. Each of the disease is characterized by the diverse level of oxygen disruption severity. Thus, understanding the role of brain oxygenation study and control is critical for disease progression diagnostic and is of great interest for both theoretical neuroscience and clinical practice.

Since the fMRI technology is expensive, moreover, and the patient needs to lie still on the device table during the scanning procedure, these limitations make screening difficult in some cases. For instance, when working with children, the elderly, and patients with psycho-neurological issues who are unable to lie still for an extended period of time during the examination. Also, some studies need to be carried out precisely in motion, which cannot be done using fMRI (Scarapicchia et al., 2017).

### Working Principles of fNIRS

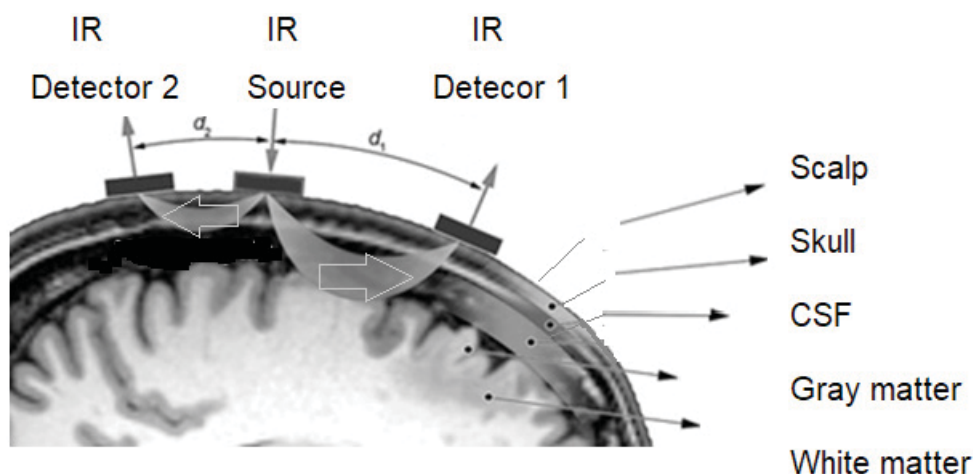
Like fMRI, functional near-infrared spectroscopy (fNIRS) measures oxygen levels in the brain (Scarapicchia et al. 2017). fNIRS is a relatively new, non-invasive, portable neuroimaging method that allows to visualizing and evaluating changes in blood oxygenation levels in the human brain (Chen et al., 2020). Such technology operates on low levels of non-ionizing light in the near-infrared range (700–1000 nm), which can penetrate biological tissues and be absorbed by chromophores, such as oxyhemoglobin (oxy-Hb) and deoxyhemoglobin (deoxy-Hb) (Chen et al., 2020). Optical window of infrared light is shown in Figure 1.



**Figure 1.** Absorption spectrum of infrared light by forms of Hb and water. For near-infrared spectroscopy (NIRS) there is the window from 675 up to 925 nm “open” by water property in this diapason. Izzetoglu et al., 2007)

The fNIRS device emits light into the tissue with two wavelengths—760 nm and 850 nm. Such wavelengths are chosen because they give the best resolution to identify concentration change of both chromophores separately.

In commercially available NIRS monitors, oxy-Hb and total hemoglobin (HbT) are measured using at least two different wavelengths that are emitted from one NIR-diode and received by another two photodiodes spaced apart (Fig. 2) (Skrifvars et al., 2021). The latter property makes it possible to separate light that has passed through tissues that are superficial (extracranial) and deeper (intracranial, at a depth of 2–3 cm) (Skrifvars et al., 2021).



**Figure 2.** NIR photons’ “path” from the light source to the detector through the different layers of the head, CSF—cerebrospinal fluid (Pinti et al., 2020)

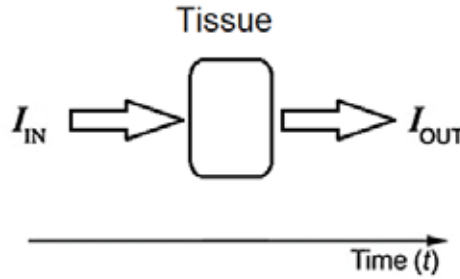
When NIR light passes through biological tissue, it is not only absorbed but also dispersed. Attenuation of light results from scattering, which occurs 100 times more frequently than absorption (Skrifvars et al., 2021). The likelihood of getting absorbed increases with the amount

of scattering that a photon experiences (Skrifvars et al., 2021). The light will be diluted, dispersed, and able to pass through the tissue several centimeters into the skull (Pinti et al., 2020).

The penetration depth of the light is proportional to the source-detector distance ( $d_1$ : deeper channel;  $d_2$ : superficial channel) (Pinti et al., 2020). A channel is made up of the source-detector pair and is located at the midpoint between the source and the detector and at a depth that is approximately equal to the half of the source-detector separation (Pinti et al., 2020).

As a result, backscattered light can be gathered and changes in light attenuation can be tracked through positioning a light detector at a specific distance from the source of NIR light. Since oxy-Hb and deoxy-Hb are primarily responsible for absorption inside the NIR optical window, changes in light attenuation at a certain wavelength can be defined by a linear combination of changes in oxy-Hb and deoxy-Hb concentrations (Pinti et al., 2020).

The majority of commercially available systems, referred to as continuous wave (CW) fNIRS devices, measure the ratio of the injected ( $I_{in}$ ) to the output ( $I_{out}$ ) NIR-light to assess light attenuation ( $A$ ) caused by tissue scattering and absorption (Pinti et al., 2020). The scheme of measurements is shown in Figure 3.



**Figure 3.** Graphic representation of input and output NIR-light in the tissue (Pinti, Tachtsidis et al., 2020)

By calculating the logarithm of the ratio of  $I_{in}$  to  $I_{out}$  that is connected to variations in hemoglobin concentrations, absolute attenuation can be assessed:

$$A = -\log_{10} \left( \frac{I_{in}}{I_{out}} \right) \quad (1)$$

By subtracting the first attenuation measurement from the subsequent attenuation measurements, changes in attenuation ( $\Delta A$ ) are computed and used to calculate changes in concentration of oxy-Hb and deoxy-Hb (Pinti et al., 2020). This technique (2), which is commonly used in fNIRS, is also known as differential spectroscopy or the modified Beer-Lambert law (mBLL) (Pinti et al., 2020).

$$\Delta A = \varepsilon(\lambda) \cdot \Delta c \cdot d \cdot DPF(\lambda) \quad (2)$$

To reduce the effect of scattering, melanin, and water concentrations, the first attenuation measurement is subtracted from the succeeding attenuation results, which is the result of differential spectroscopy (Pinti et al., 2020). The mBLL relates changes in attenuation  $\Delta A$  to changes in chromophore concentrations ( $\Delta c$ , either oxy-Hb or deoxy-Hb) (Pinti et al., 2020). In (2)  $d$  denotes the source-detector distance,  $\varepsilon$ —chromophore's extinction coefficient at a certain wavelength  $\lambda$ ; the differential pathlength ( $DPF$ ) represents the increase in the photon path caused by scattering.

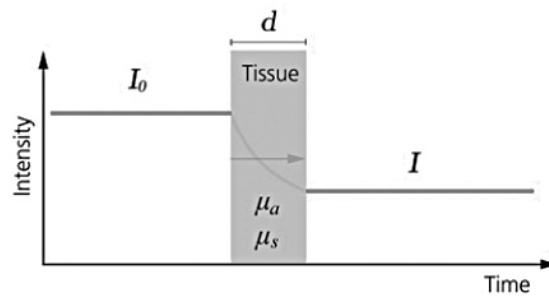
Oxy-Hb and deoxy-Hb concentration variations can be tracked using CW fNIRS sensors, but absolute baseline concentrations cannot be determined since the contributions from absorption and scattering cannot be distinguished and quantified (Scholkmann et al., 2014). Because of this, measurements of oxy-Hb and deoxy-Hb begin at zero (Scholkmann et al., 2014). However, because absolute concentrations are not required and functional activity is typically assessed relative to a baseline, these systems are useful for cognitive neuroscience applications (Pinti et al., 2020).

The downside of CW fNIRS is that they cannot fully determine the optical properties of tissue (e.g. light scattering ( $\mu_s'$ ) and absorption ( $\mu_a$ ) coefficients) and hence cannot predict the oxy-Hb and deoxy-Hb absolutely (Scholkmann et al., 2014). However, with a few acceptable assumptions, changes in oxy-Hb and deoxy-Hb can be quantified (Scholkmann et al., 2014). As a result, in the early years, NIRS sensors were largely used as trend monitors to examine various physiological conditions and clinical therapies (Scholkmann et al., 2014).

In addition to fNIRS systems based on CW technology, there are also two more categories of fNIRS instruments: time-domain (TD) and frequency-domain (FD) devices (Scholkmann et al., 2014). The signal differences of three methods are shown in Figures 4.1–4.3.

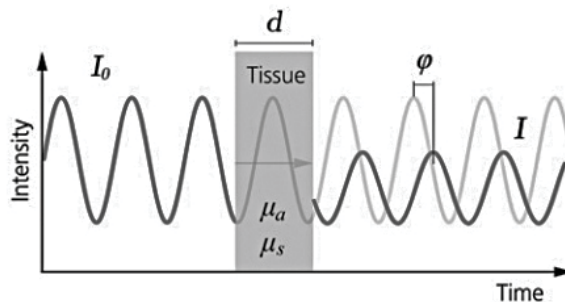
In following Figures,  $I_0$ —incident light signal,  $I$ —transmitted light signal,  $d$ —thickness of the medium,  $\mu_a$ —absorption coefficient,  $\mu_s$ —scattering coefficient,  $\phi$ —phase delay, and  $I(t)$ —temporal point spread function of the transmitted light signal.

With continuous wave technology, light is emitted at a constant intensity, and only changes in the light's intensity as it passes through tissue are then measured (Scholkmann et al., 2014).



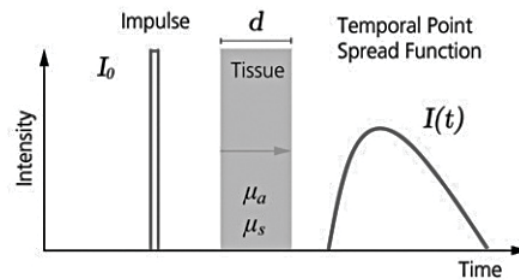
**Figure 4.1.** Illustration of continuous wave NIRS (Scholkmann et al., 2014)

By varying the intensity of the produced light with a sinusoidal function, it is possible to determine absolute values of hemoglobin concentration (van Essen et al., 2020). The frequency domain technology modulates the emitted light intensity and then measures the intensity of the detected light as well as the phase shift, which corresponds to the time of flight (Scholkmann et al., 2014).



**Figure 4.2.** Illustration of frequency domain NIRS (Scholkmann et al., 2014)

The phase shift represents the time difference between the emitted and detected light waves and varies with the optical path length of the light through the tissue. Absolute hemoglobin concentrations and, consequently, regional cerebral oxygen saturation are estimated using the absorption and scattering coefficients (van Essen et al., 2020).



**Figure 4.3.** Illustration of time domain NIRS (Scholkmann et al., 2014)

The time domain technology emits an extremely short pulse of light into the tissue and measures the arrival times of the photons that emerge from the tissue (Scholkmann et al., 2014). Although it produces the most information, this technology is also the most complex (Scholkmann et al., 2014).

### Clinical applications

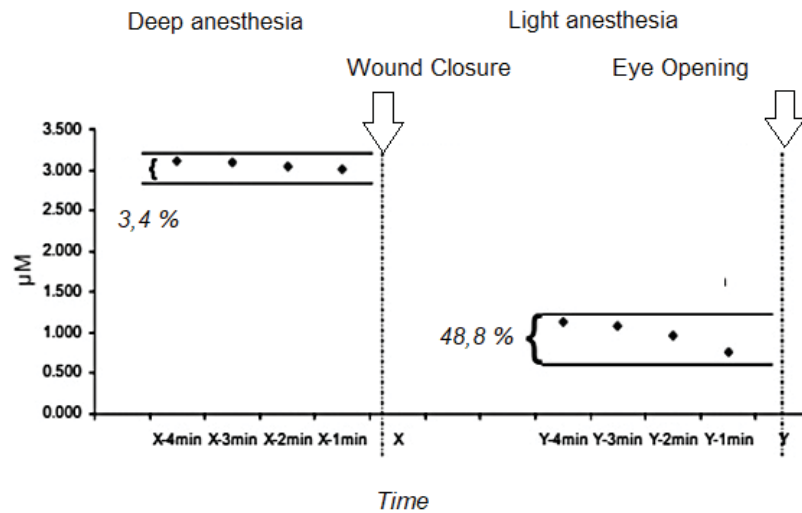
NIRS offers non-invasive online monitoring of tissue oxygenation in a wide range of clinical scenarios (Scheeren et al., 2012).

To begin with, this technology can be used as a clinical tool for monitoring the anesthesia process. To ensure the security of anesthesia treatments, different ongoing measurements, including peripheral oxygenation, circulation, ventilation, and temperature, are needed (Klein et al., 2021). Anesthetics interfere with the destruction of neurons in a dose-dependent way to inhibit the creation of memory and consciousness (Klein et al., 2021). Thus, anesthetics suppress the cerebral metabolic rate and impact cerebral blood flow, since the decrease in brain activity can be explained by a drop in the cerebral metabolic rate of oxygen consumption and the glucose metabolic rate (Klein et al., 2021).

Firstly, such technology can be used as a clinical tool for anesthetic depth monitoring. The first study to look at fNIRS biomarkers for depth of anesthesia (Izzetoglu et al., 2011) was done to confirm fNIRS ability to differentiate between “deep” and “light” anesthesia (Fig. 5). Mean values indicate that light anesthesia is associated with lower Hb level and a high rate of change and includes a 4-minute period before eye opening (Izzetoglu et al., 2011). The effect was observed predominantly in the right hemisphere (Izzetoglu et al., 2011). In contrast, deep anesthesia was associated with a low rate of change in Hb and was classified as a 4-minute period to wound closure (Izzetoglu et al., 2011).

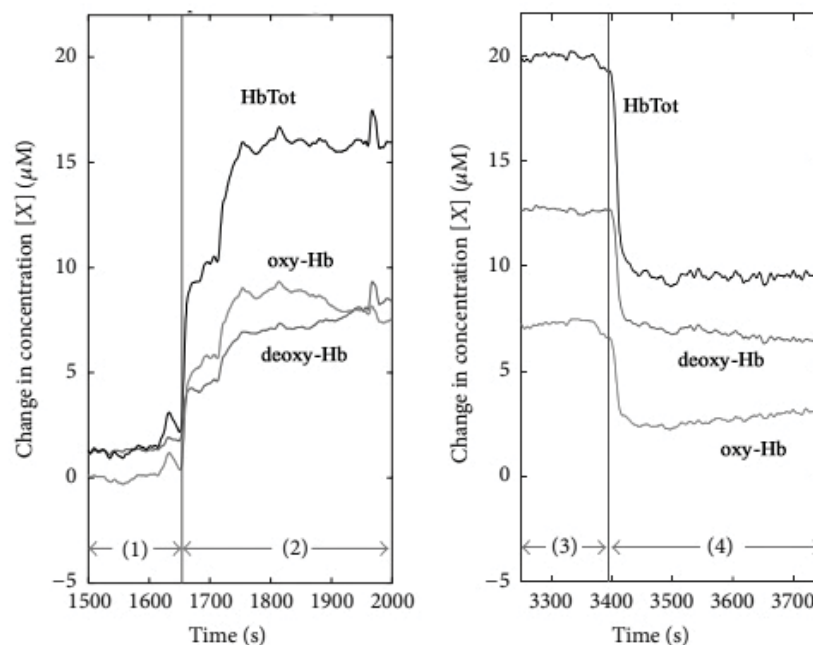
Moreover, this technique can serve as an optimal measure of awareness for use in anesthesia. This would allow some capacity to predict potential awakening with sufficient time to adjust the anesthetic agents to prevent awareness (Hernandez-Meza et al., 2015). During deep anesthesia, deoxy-Hb averages displayed a very slow rate of change (3.4 %) (Izzetoglu et al., 2011). In contrast, as the patient emerges to wakefulness, this rate of change increases drastically (48.8 %) (Izzetoglu et al., 2011).

The technique can predict the patient’s potential awakening with enough time to adjust anesthetic agents to prevent or prolong it, depending on the situation.



**Figure 5.** Rate of hemodynamic changes during surgical manipulation with a patient within deep and light anesthesia stages: The time versus fNIR measurements for deoxy-Hb are shown for deep and light anesthesia stages (Izzetoglu et al., 2011)

Also, using fNIRS, it was demonstrated that changes in patient's position (e.g., different angles of the operating table), as well as the anesthetic type used, affect hemodynamic parameters and can have a significant impact on fNIRS-derived indicators, such as deoxy-Hb, oxy-Hb, total hemoglobin (HbTot), and Hb during anesthesia (Lovell et al., 2000). When moving from a supine position to a Trendelenburg position, the hemoglobin concentration increases significantly, and vice versa (Lovell et al., 2000). fNIRS signal changes as a consequence of position change are shown in Figure 6.

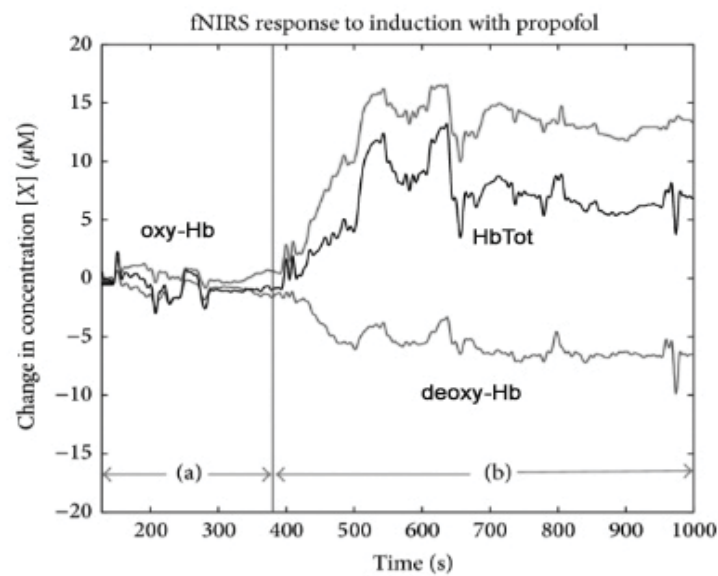


**Figure 6.** The effect of position variations on a 58-years-old female patient during laparoscopic salpingo-oophorectomy under general anesthesia with 2.4 % sevoflurane (Hernandez-Meza et al., 2015)

(a) – oxy-Hb, deoxy-Hb, and HbTot responses during the transition from (1) supine to (2) Trendelenburg positions. (b) – change from (3) Trendelenburg to (4) supine posture.

Furthermore, fNIRS can discriminate between the effects of induction with various medications, such as propofol, etomidate, and thiopental, and monitor changes in anesthetic delivery in real-time (Hernandez-Meza et al., 2015). During induction (Fig. 7), various anesthetic doses of the same medication also generate various fNIRS signal characteristics (Hernandez-Meza et al., 2015). Additionally, using this approach, it is possible to show that various anesthetic medications, such as midazolam, aminophylline, and isoflurane, may impact anesthesia depth but not oxygen saturation at the same time (Hernandez-Meza et al., 2015).

It is only possible to determine the hemoglobin concentration from the hemodynamic response detected by fNIRS, making it an indirect method of determining metabolic rate. Thus, it can be affected by distorting factors affecting the relative amounts of hemoglobin and oxy-Hb without changing the metabolic rate (Hernandez-Meza et al., 2015). Therefore, future research may include the development of algorithms for mitigating the impact of distorting factors.



**Figure 7.** fNIRS response (changes in oxygenation) to induction with propofol (Hernandez-Meza et al., 2015)

Another application fNIRS finds in assessment of the cerebral cortex during motor task behaviors. Safe mobility is essential for daily life and calls for effective motor control mechanisms (Herold et al., 2017). Such assessment can be one of the key factors in gaining a deeper comprehension of neuromotor control that, consequently, may lead to the improvement of rehabilitation strategies (Herold et al., 2017). fNIRS technology is used in studying brain activation during walking on treadmill, overground walking and postural tasks (e.g., balance test) (Herold et al., 2017). With the help of fNIRS, areas, which are being activated during any kind of activity, could be determined. By observing areas behavior, conclusions about the optimal duration and quantity of exercises, needed breaks between exercises could be drawn. fNIRS may also be a valuable addition in prehospital care, including the transport of critically ill patients (Herold et al., 2017).

The fact that fNIRS research is on the rise is largely due to the technique's high tolerance for motion artifacts. During any activity (e.g., walking, dancing, communicating), a properly positioned fNIRS cap will still continue to provide reliable signals.

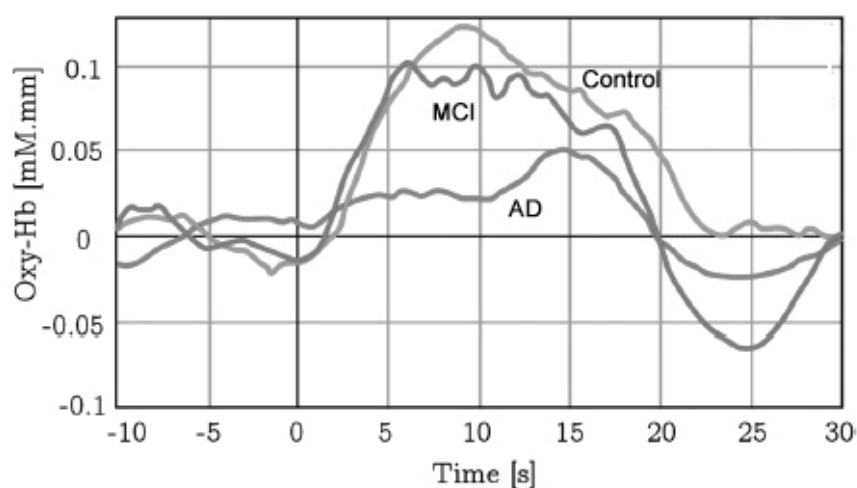
Furthermore, fNIRS finds its application in diagnostics and treatment of neurological disorders. That way, migraine symptoms, depression, autism and other can be studied. For exam-

ple, accurate localization and spatial resolution of this technology are essential for epilepsy treatment and management (Rahman et al., 2020; Monrad et al., 2015; Guevara et al., 2020). By combining multichannel fNIRS with long-term EEG, epileptologists could gain a better understanding of hemodynamic changes that occur during different phases (preictal, ictal, postictal, and interictal) of seizures and, therefore, identify and localize them more accurately (Chen et al., 2020; Guevara et al., 2020).

A number of studies show the possibility of differences between focal and generalized seizure. For instance, in (Monrad et al., 2015) patients with temporal and extratemporal lobe epilepsies showed consistent ictal increases in oxygenation on the epileptic hemisphere in comparison to the non-epileptic hemisphere. Regional frontal oxygenation has been reported to fall during seizures in patients with generalized epilepsies and secondary generalized seizures. (Rahman et al., 2020; Chen et al., 2020). However, fNIRS has not yet been widely employed in epilepsy clinical practice because of the inconsistent seizure detection rate. To the exact moment, it potentially can offer an additional clinical parameter for more precise earlier seizure detection and localization (Monrad et al., 2015).

Alzheimer's disease (AD) may be detectable using fNIRS, even in its early stages (Rahman et al., 2020; Li et al., 2018). It is studied that compared to healthy controls, AD patients show lower levels of activity at particular brain regions during a variety of cognitive tasks (Li et al., 2018). Also, when given a verbal fluency task, a decrease in oxy-Hb and HbT in the parietal cortex of AD patients was registered (Li et al., 2018). Current limitation of the technology in this field is that it is not able to characterize differences across various stages of AD (Li et al., 2018).

Arai et al. (2006) conducted a study on 32 healthy controls, 15 Alzheimer's patients, and 15 people with mild cognitive impairment (MCI). They determined that the activation index was significantly lower in the bilateral parietal and frontal areas in the AD group, while the concentration change of the oxy-Hb in the MCI group was significantly worse only in the right parietal area, based on the activation index computation and the weight of variations in the oxy-Hb concentration in each zone during the activation task. The waveforms in Figure 8 display the representative results for each group.



**Figure 8.** The overall average changes in the concentration of  $HbO_2$  of different groups during the verbal fluency test in the right parietal area of the brain (Rahman et al., 2020)

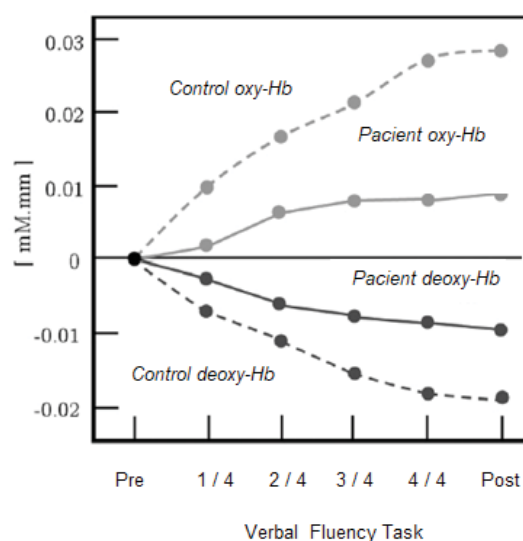
fNIRS has been used in several researches to examine the neural process behind autism spectrum disorder (ASD). This technique has a potential to outstand other neuroimaging techniques

currently used in studying ASD (e.g. EEG, fMRI and magnetoencephalography (MEG)) due to its relative low cost, portability and possibility to use it in naturalistic applications. Another benefit of using fNIRS is that it allows to gain a better understanding of brain functional activation and its organization in infants and young children, mainly when they are hyperactive, which is not possible with fMRI (Siddiqui et al., 2017). The fNIRS is especially useful for investigating the linguistic viewpoint, which is essential for adequate interpersonal relations and frequently deteriorated in ASD due to reduced noise (Rahman et al., 2020).

Despite the fact that attention-deficit/hyperactivity disorder (ADHD) is the most researched illness, few studies have employed fNIRS in this mental disorder in children and adolescents (Lee et al., 2021). Nowadays, neurofeedback is often employed in the treatment of ADHD. fNIRS studies are focused on the difference in hemodynamic activity based on ADHD symptoms or the prediction of treatment response to drugs using various cognitive activities. First, the evaluation of changes in hemodynamic activity depending on symptoms is based on the fundamental symptoms of ADHD (inattention, hyperactivity, and impulsivity) are connected to executive function impairment and primarily reflect prefrontal brain function (PFC) (Lee et al., 2021). Second, the prediction of treatment response was studied in Monden et al. (2012) the ADHD group showed less hemodynamic activity than the typically developing children during the Go/No-Go task in the right inferior frontal gyrus and middle frontal gyrus.

fNIRS technique can be also useful in studying schizophrenia. Schizophrenia is a mental illness characterized by positive and negative symptoms, cognitive impairment, and persistent social difficulties (Koike et al., 2013). Although about 0.7 % of population globally suffer from this mental disorder, efficient therapies and reliable indicators for all symptoms and functioning of the disorder are still lacking. So, there is a need to continue and broaden research in this area.

Watanabe and Kato (2004) discovered that oxy-Hb and deoxy-Hb fluctuations in language fluency test were lower in schizophrenia patients than in healthy controls. In addition, compared to trials in individuals given standard antipsychotic medicines, this study indicated increased activities and corresponding oxy-Hb improvements (Watanabe et al., 2004). Figure 9 depicts the comparable variations in  $\text{HbO}_2$  concentration caused by verbal fluency test (VFT) and letter-number span test (LN) in the control state.



**Figure 9.** Averaged change in oxy-Hb and deoxy-Hb during VFT and LN in control group and patients with schizophrenia disorder (Rahman et al., 2020)

Another application of fNIRS is in cardiac surgery, including the measurement of cerebral oxygenation (rSO<sub>2</sub>) (Scheeren et al., 2012). Tissue hypoxia occurs frequently in the perioperative setting, particularly in cardiac surgery (Scheeren et al., 2012). As a result, measuring and obtaining adequate tissue oxygenation may help to avoid (postoperative) complications and thus be cost-effective. NIRS monitoring can also be used to detect tissue hypoxia in (prehospital) emergency settings, where it has prognostic value and allows therapeutic interventions to be monitored (Scheeren et al., 2012).

## Conclusion

Studies show that fNIRS is a valid and accurate diagnostic tool for task-associated oxygenated blood. The portability, movement tolerance, and safety of fNIRS make it an especially good method for examining how the brain develops in infants and children (Pinti et al., 2020). Although it may still be too early to describe all the uses of this novel technology, it is anticipated that its use in clinical practice may extend much beyond this. However, there are a lot of limitations that are yet to be solved. For the examination of fNIRS data, there are currently no set standards. The question of how much chromatic biomolecules, other than the molecules of interest (such as hemoglobin), affect NIRS measurements arises when working with them (Pinti et al., 2020). For instance, skin pigmentation or hair color and thickness may be relevant in this situation. The inability to access cortical regions that are not adjacent to the scalp is another drawback of fNIRS. The ventral surface of the frontal cortex, the basal ganglia, and much of the cerebellum cannot be easily measured with fNIRS due to the increased light absorption with distance from the head (Scarapicchia et al., 2017). Finally, NIRS monitoring is relatively expensive when using single-use patient sensors. Further research is also required in that regard because the majority of potential therapeutic uses have not undergone a cost-benefit review. However, fNIRS makes the prospects for cognitive and social neuroscience appear more promising.

## References

1. Chen W.L., Wagner J., Heugel N., Sugar J., Lee W., Conant L., Malloy M., Heffernan J., Quirk B., Zinos A., Beardsley S.A., Prost R., Whelan H.T. Functional Near-Infrared Spectroscopy and Its Clinical Application in the Field of Neuroscience: Advances and Future Directions. *Frontiers in Human Neuroscience*. 2020. 14:724. 15 p.
2. Guevara E., Edgar, Flores-Castro J. A., Peng Ke., Nguyen D.K., Lesage F., Pouliot P., Rosas-Romero R. Prediction of Epileptic Seizures Using fNIRS and Machine Learning. *Journal of Intelligent & Fuzzy Systems*. 2020. 38(2): 2055–2068.
3. Hernandez-Meza G., Izzetoglu M., Osbakken M., Green M., Izzetoglu K. Near-Infrared Spectroscopy for the Evaluation of Anesthetic Depth. *BioMed Research International*. 2015. 11 p.
4. Herold F., Wiegel P., Scholkmann F., Thiers A., Hamacher D., Schega L. Functional near-infrared spectroscopy in movement science: a systematic review on cortical activity in postural and walking. *Neurophotonics*. 2017. 4(4):041403. 26 p.
5. Izzetoglu K., Ayaz H., Merzagora A., Izzetoglu M., Shewokis P.A., Bunce S.C., Pourrezaei K., Rosen A., Onaral B. The Evolution of Field Deployable fNIR Spectroscopy from Bench to Clinical Settings. *Journal of Innovative Optical Health Sciences*. 2011. 4(3). pp. 239–250.
6. Izzetoglu M., Bunce S., Izzetoglu K., Onaral B. Functional brain imaging using near-infrared technology. *IEEE Engineering in Medicine and Biology Magazine*. 2007. 26(4). pp. 38–46.
7. Klein A.A., Meek T., Allcock T.E., Cook T.M., Mincher N., Morris C., Nimmo A.F., Nimmo J., Pandit J., Pawa A., Rodney G., Sheraton T., Young P. Recommendations for standards

- of monitoring during anaesthesia and recovery. Guideline from the Association of Anaesthetists. 2021. 12 p.
8. Koike S., Nishimura Y., Takizawa R., Yahata N., Kasai K. Near-infrared spectroscopy in schizophrenia: a possible biomarker for predicting clinical outcome and treatment response. *Front Psychiatry*. 2013. 4:145.
  9. Lee Y.J., Kim M., Kim J.S., Lee Y.S., Shin J.E. Clinical Applications of Functional Near-Infrared Spectroscopy in Children and Adolescents with Psychiatric Disorders. *J Korean Acad Child Adolesc Psychiatry*. 2021. 32(3): 99–103.
  10. Li R., Rui G., Chen W., Li S., Schulz P.E., Zhang Y. Early Detection of Alzheimer's Disease Using Non-invasive Near-Infrared Spectroscopy. *Front Aging Neurosci.* 2018. 9(10):366. 11 p.
  11. Lovell T. Marshall A.C., Elwell C.E., Smith M., Goldstone J.C. Changes in cerebral blood volume with changes in position in awake and anesthetized subjects. *Anesthesia & Analgesia*. 2000. 90(2). pp. 372–376.
  12. Monden Y., Dan H., Nagashima M., Dan I., Tsuzuki D., Kyutoku Y., et al. Right prefrontal activation as a neuro-functional biomarker for monitoring acute effects of methylphenidate in ADHD children: an fNIRS study. *Neuroimage Clin* 2012. 1:131–140.
  13. Monrad P., Sannagowdara K., Bozarth X., Bhosrekar S., Hecox K., Nwosu M. Haemodynamic response associated with both ictal and interictal epileptiform activity using simultaneous video electroencephalography/near infrared spectroscopy in a within-subject study. *J. Near Infrared Spectrosc*. 2015. 23. pp. 209–218.
  14. Pinti P., Tachtsidis I., Hamilton A., Hirsch J., Aichelburg C., Gilbert S., Burgess P.W. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann N Y Acad Sci*. 2020. 1464(1). pp. 5–29.
  15. Rahman A.A., Siddik A.B., Kanti T., Khanam G., Khanam F., Khanam M. Narrative Review on Clinical Applications of fNIRS. *Journal of Digital Imaging*. 2020. 33. pp. 1167–1184.
  16. Scarapicchia V., Brown C., Mayo Ch., Gawryluk J.R. Functional Magnetic Resonance Imaging and Functional Near-Infrared Spectroscopy: Insights from Combined Recording Studies. *Frontiers in Human Neuroscience*. 2017. 11(419). 12 p.
  17. Scheeren T.W.L., Schober P., Schwarte L.A. Monitoring tissue oxygenation by near infrared spectroscopy (NIRS): background and current applications. *J Clin Monit Comput*. 2012. 26. pp. 279–287.
  18. Scholkmann F., Kleiser S., Metz A.J., Zimmermann R., Mata P.J., Wolf U., Wolf M. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage*. 2014. 85(1). pp. 6–27.
  19. Skrifvars M.B., Sekhon M., Aneman E.A. Monitoring and modifying brain oxygenation in patients at risk of hypoxic ischaemic brain injury after cardiac arrest. *Crit Care*. 2021. 25(312). 8 p.
  20. van Essen T., Goos T.G., van Ballegooijen L. Comparison of frequency-domain and continuous-wave near-infrared spectroscopy devices during the immediate transition. *BMC Pediatr*. 2020. 20(94).
  21. Watts M.E., Pocock R., Claudanos Ch. Brain Energy and Oxygen Metabolism: Emerging Role in Normal Function and Disease. *Frontiers in Molecular Neuroscience*. 2018. 11(216): 13 p.
  22. Zhong W.A., Ji Z., Sun C. Review of Monitoring Methods for Cerebral Blood Oxygen Saturation. *Healthcare (Basel)*. 2021. 9(9): 20 p.

## **Chapter 15. Features of the perception of emotionally colored images by adolescents and young people depending on the presence of social scenes on them**

*Skuratova K.<sup>1</sup>, Shelepin E.<sup>1</sup>, Naumova D.<sup>2</sup>*

<sup>1</sup>*Pavlov Institute of Physiology, Academy of Sciences, St. Petersburg, Russia*

<sup>2</sup>*Saint Petersburg State University, Saint-Petersburg, Russia*

### **Introduction**

When scanning the visual environment that surrounds us every day, we are faced with an excessive array of its constituent elements, but the capabilities of the human visual system are limited, which makes it impossible to analyze each of them in detail. Therefore, the human brain is forced to evaluate the priority of incoming visual signals and allocate the largest amount of cognitive resources to process the most significant elements and events. But what is significant? It is assumed that the main method of prioritization is the assessment of the emotional significance of the stimulus or event (Compton, 2003). It was also found that stimuli assessed as emotionally significant are subjected to increased processing. While the emotional significance of stimuli may vary from person to person, there are some stimuli, such as human faces, that are emotionally meaningful to most people. This is due to the fact that humans are inherently social species, whose brains have specialized areas that are sensitive to various visual social cues, especially faces (Landsiedel et al., 2022).

Neuroimaging studies have shown that the orbitofrontal cortex and ventral striatum respond to socially reinforcing stimuli such as beautiful or smiling faces, while lesions in the orbitofrontal cortex impair interpersonal behavior (Hornak et al., 2003). Research has also shown that both the ventral and dorsal striatum respond to more complex social information such as cooperation or the opportunity to punish a traitor (Sanfey, 2007). In addition, the amygdala plays an important role in calculating and updating the value of social information. The perception of emotional faces has also been shown to increase amygdala activity because they are biologically relevant stimuli that predict critical events in our environment (Davis et al., 2010). Thus, the study of this topic contributes to the understanding of the mechanisms by which the psyche analyzes objects in the environment and prioritizes social incentives.

Neuroimaging has also shown that images of faces preferentially activate visual cortical areas corresponding to the central visual field, while other images, such as buildings, cause stronger activation in areas responsible for peripheral vision (Wang et al., 2013). It is important to note that face recognition and discrimination are impaired when using peripheral vision (Mäkelä et al., 2001). This is due to the fact that the perception of faces is a complex cognitive task, since its constituent structures are small in size, as well as a low signal-to-noise ratio (Behrmann, 2022).

There is no doubt that faces are very effective at attracting attention if they express one emotion or another. For example, researchers have found that fearful and angry expressions are recognized faster than neutral and happy expressions (Lanfranco et al., 2023). In the Stroop emotional test, the rate of naming facial color is lower if the facial expression is angry rather than neutral, suggesting that the processing of angry faces requires more cognitive resources (van Honk et al., 2001).

Features of the perception of faces change with some mental and neurological disorders; for example, people with depressive disorder pay more attention to sad faces, anxiety increases

attention to fearful and angry faces, and autism spectrum disorders in general affect visual processing of faces, causing avoidance of eye contact. Thus, the study of the perception of people's faces, including those expressing a particular emotion, can shed light on the cognitive mechanisms of the perception of social information in various psychopathological conditions (Lanfranco et al., 2023).

Social interaction plays a central role in human life. People watch and participate in numerous social scenes every day. Visual processing of human faces underlies the detection, recognition and identification of relatives, and is also an important aspect of social interaction. Faces contain a wealth of information (e.g., emotional and physical state, intentions) that facilitates communication. The ability to process such information is a highly developed visual skill (Haxby, Hoffman., Gobbini, 2000; Geringswald et al., 2020).

The purpose of this study is to compare the parameters of oculomotor activity during the perception of social stimuli containing human faces and stimuli containing images of animals or inanimate objects in a choice situation.

## **Procedure and research methods**

The study was carried out using the eye tracking method using the Neurobureau hardware and software system.

The stimulus material consisted of 25 stimuli, of which 20 were experimental and 5 were controls. The experimental stimuli contained 4 images of different emotional coloring (positive, dysphoric, threatening, neutral), and the control stimuli contained 4 neutral images. One part of the stimuli contained social scenes with images of people's faces, the other part contained scenes with animals or inanimate objects. All images used were licensed CC0 and were selected and emotionally categorized by peer review.

Between stimuli, a fixation cross was presented for 3000 ms in order to fix the gaze position in the center of the screen. Stimuli were presented in random order.

The task of the subjects was to carefully look through the images and choose the most attractive ones. The time of presentation of stimuli was not limited.

The study involved 57 people from 15 to 45 years old. The subjects were combined into three age groups: from 15 to 17 years old (9 people), from 18 to 30 years old (40 people), from 31 to 45 years old (9 people). The sample is normalized by sex. All subjects had normal and corrected-to-normal vision, and were free of psychiatric and neurological illnesses.

## **Results**

### ***Time to the first fixation***

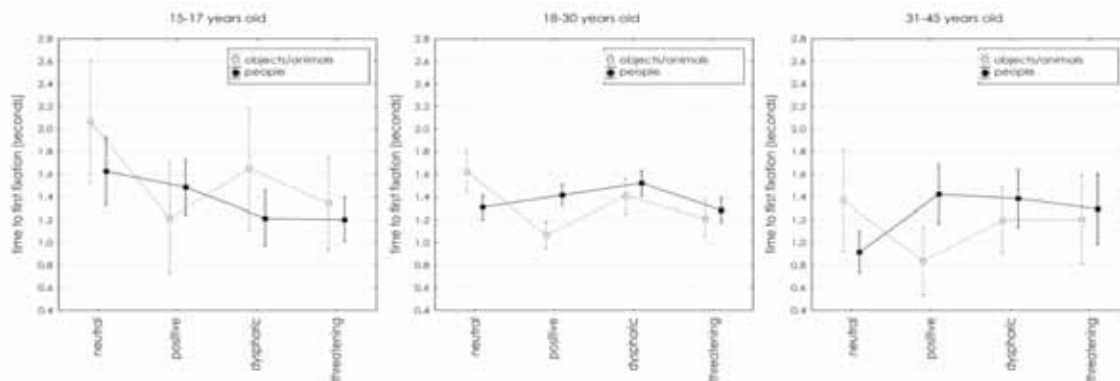
Consider such an indicator as the time to the first fixation on an emotionally colored image

On stimuli containing social scenes, adolescents most quickly notice threatening images, on stimuli containing only objects or animals—positive images. Neutral images are the last thing on the mind of teenagers.

Subjects between the ages of 18 and 30 pay the fastest attention to positive images of objects or animals. Among the social scenes, no significant differences in time to the first fixation were found.

Subjects over the age of 30 notice neutral social scenes and positive images of animals and objects the fastest.

Figure 1 shows the average values and confidence intervals for three age groups for the time to first fixation parameter, depending on the emotional coloring of the image and the presence of a social scene on it.



**Figure 1.** Time to first fixation on an emotionally colored image

### Number of fixations

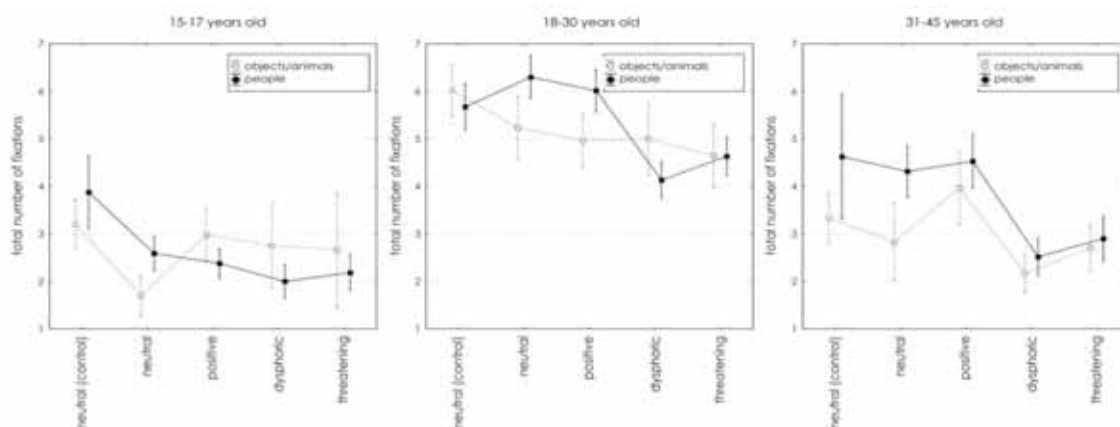
How many fixations do the subjects make on each image?

Adolescents make the fewest fixations, which makes their choices more impulsive. They need the fewest fixations to analyze neutral scenes containing images of animals or objects. At the same time, they make the most fixations on control stimuli, since the same neutral emotional coloring makes it difficult to choose the image they like. Among emotionally colored images, more fixations are made on images with animals and objects than on social scenes.

People between the ages of 18 and 30 require the most fixations to select an image they like, which may be due to higher levels of anxiety in this age group. The graph shows that they try to avoid dysphoric and threatening social scenes, giving more visual preference to neutral and positive images of people. We can assume that defense mechanisms of the psyche, in particular avoidance, may play a role here.

In people over 30 years of age, the pattern is generally the same, with avoidance of dysphoric and threatening scenes.

Figure 2 shows the average values and confidence intervals for three age groups for the number of fixations parameter, depending on the emotional coloring of the image and the presence of a social scene on it.



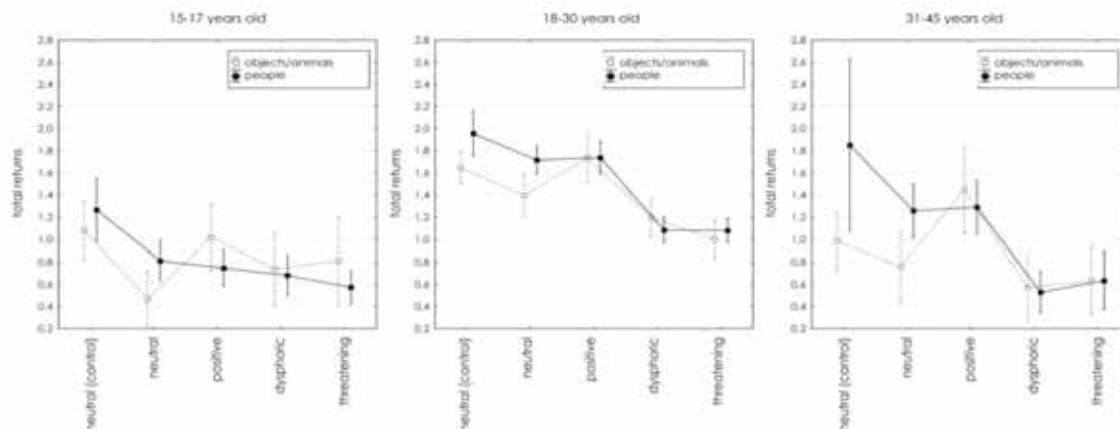
**Figure 2.** Number of fixations on an emotionally colored image

## Number of returns

Adolescents are less likely to return to image reanalysis

Subjects aged 18 to 45 most often return to positive images. Neutral images on control stimuli also require a greater number of repeated returns to analysis.

Figure 3 shows the average values and confidence intervals for three age groups for the number of returns parameter, depending on the emotional coloring of the image and the presence of a social scene on it.



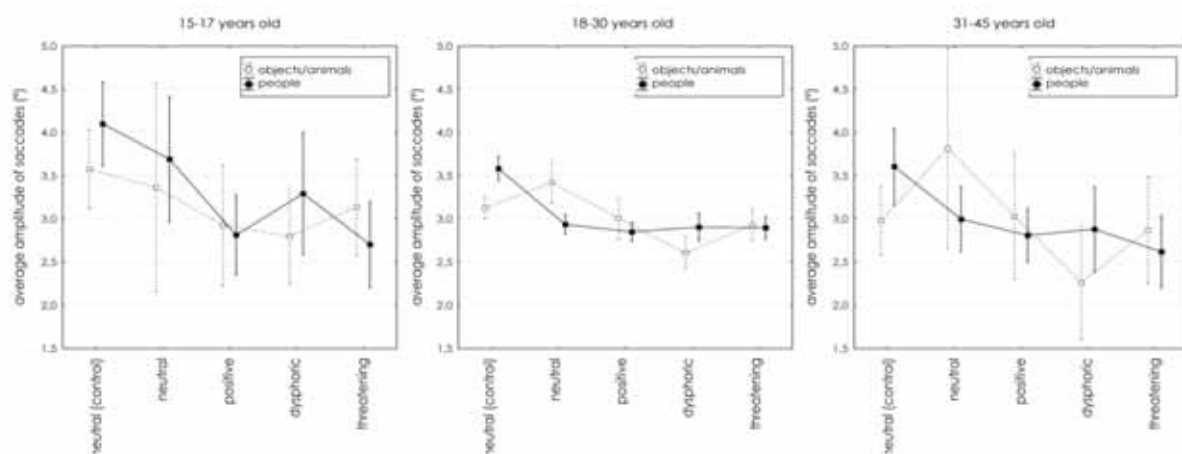
**Figure 3.** Number of returns to an emotionally colored image

## Saccade amplitude

No age differences in saccade amplitude were found

The hypothesis of a decrease in the amplitude of saccades in the analysis of social stimuli due to a decrease in the effectiveness of their analysis by peripheral vision was not confirmed.

Figure 4 shows the average values and confidence intervals for three age groups for the average saccade amplitude parameter depending on the emotional coloring of the image and the presence of a social scene on it.



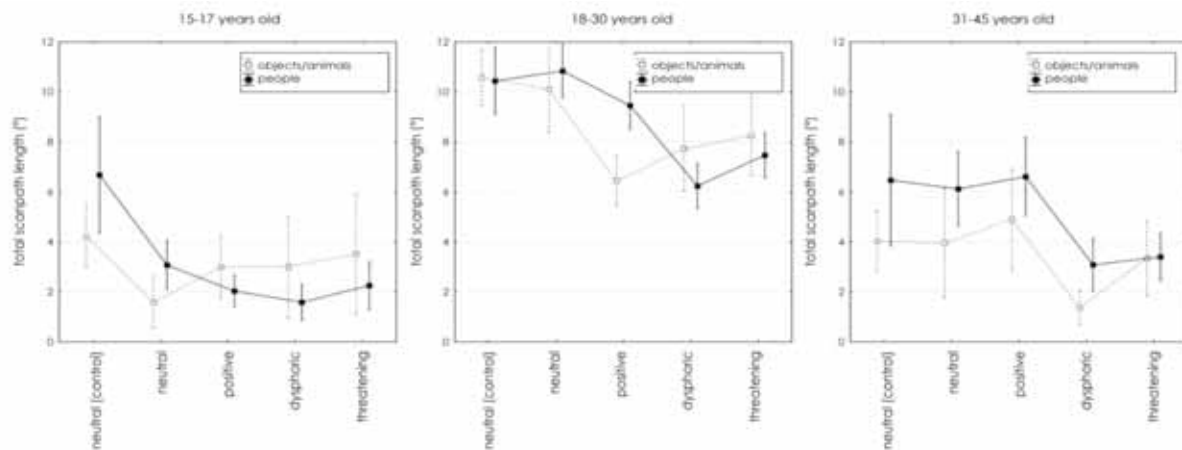
**Figure 4.** Average amplitude of saccades on an emotionally colored image

## ***Total scanpath length***

The total scanpath length is the sum of the amplitudes of all saccades.

We see that the greatest length of the scanning path is typical for subjects from 18 to 30 years old. This suggests that they analyze the proposed images in the most holistic way.

Figure 5 shows the average values and confidence intervals for three age groups for the total scanpath length parameter depending on the emotional coloring of the image and the presence of a social scene on it.



**Figure 5.** *The total length of the scanpath on an emotionally colored image*

## ***Attention Index***

It was calculated as the ratio of the time spent viewing an emotionally charged image to the total time spent viewing the stimulus.

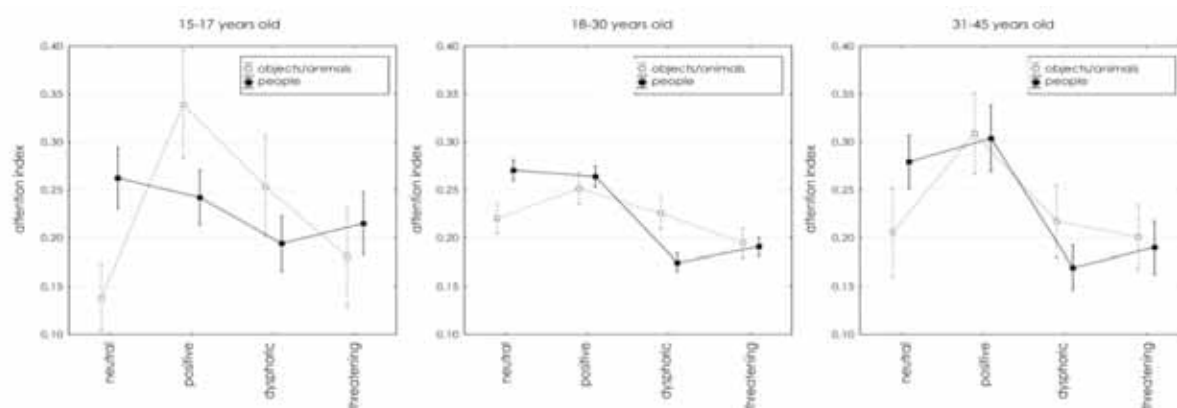
Among social scenes, teenagers prefer neutral images and avoid dysphoric ones. Among the images of animals and objects, positive ones are preferred, and neutral ones are avoided.

Among social scenes, people aged 18 to 30 prefer neutral and positive images, while avoiding dysphoric and threatening ones. Among the images of animals and objects, positive ones are preferred, and threatening ones are avoided.

Among social scenes, people over 30 prefer positive images and avoid dysphoric and threatening ones. Among the images of animals and objects, positive ones are preferred, while all the rest are given much less attention.

Figure 6 shows the average values and confidence intervals for three age groups for the attention index parameter, depending on the emotional coloring of the image and the presence of a social scene on it.

We would like to note that despite the fact that the sample of subjects from 18 to 30 years old is the largest, they have more consistent data with less variation in them. We hope that further analysis of the psycho-emotional state of the subjects of this group will help to understand why their perception patterns are so similar. A cursory analysis of the results of clinical questionnaires showed that the subjects of this group have high rates of anxiety, hopelessness and more pronounced depressive symptoms. Among subjects over 30 years of age, the results lie within the normal range or deviate minimally from it.



**Figure 6.** Index of attention to an emotionally colored image

## Conclusions

- Positive images that do not contain people attract attention the fastest.
- The fewest fixations in the analysis of emotionally colored images are made by teenagers, the most by people from 18 to 30 years old. Adolescents are also less likely to return to image reanalysis
- The most holistic analysis of the proposed images is typical for people from 18 to 30 years old.
- Adolescents make more fixations on images with animals and objects, people over 30—on social stimuli
- People aged 18 to 45 avoid dysphoric and threatening social scenes and most often return to positive images

## Practical significance

The developed stimulus material is planned to be used for the analysis of oculomotor patterns in the perception of emotionally colored scenes by people with various psycho-emotional disorders.

This will allow us to consider the cognitive mechanisms characteristic of various psychoemotional disorders and develop objective methods for their diagnosis.

At the moment, we have already discovered the relationship between hopelessness and oculomotor patterns in the perception of emotionally colored social scenes.

*This work was supported by the State Program 47 GP “Scientific and Technological Development of the Russian Federation” (2019–2030), subject 0134–2019–0006.*

## References

1. Behrmann M, Avidan G. Face perception: Computational insights from phylogeny // Trends in Cognitive Sciences–2022–№ 26(4)–P. 350–363.
2. Compton R.J. The interface between emotion and attention: A review of evidence from psychology and neuroscience // Behavioral and cognitive neuroscience reviews–2003–№ 2(2)–P. 115–29.
3. Davis F.C., Johnstone T., Mazzulla E.C., Oler J.A., Whalen P.J.. Regional response differences across the human amygdaloid complex during social conditioning // Cerebral

Cortex–2010–№ 20(3)–P. 612–621

4. Geringswald F., Afyouni A., Noblet C., Grosbras M.–. Inhibiting saccades to a social stimulus: A developmental study // *Scientific Reports*–2020–№ 10(1).
5. Haxby J. V., Hoffman E. A., Gobbini M. I. The distributed human neural system for face perception // *Trends in Cognitive Sciences*–2000–№ 4(6)–P. 223–233.
6. Hornak J, Bramham J, Rolls ET, Morris RG, O’Doherty J, Bullock PR, Polkey CE. Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices // *Brain*–2003–№ 126–P. 1691–1712.
7. Landsiedel J., Daughters K., Downing P. E., Koldewyn K. The role of motion in the neural representation of social interactions in the posterior temporal cortex // *Neuroimage*–2022–№ 262.
8. Lanfranco R. C., Rabagliati H., Carmel D. The importance of awareness in face processing: A critical review of interocular suppression studies // *Behavioural Brain Research*–2023–№ 437.
9. Mäkelä P., Nasanen R., Rovamo J., Melmoth D. Identification of facial images in peripheral vision // *Vision Research*–2001–№ 41–P. 599–610.
10. Sanfey A. G. Social decision-making: insights from game theory and neuroscience // *Science*–2007–№ 318–P. 598–602.
11. Wang B., Yan T., Wu J., Chen K., Imajyo S., Ohno S., Kanazawa S. Regional neural response differences in the determination of faces or houses positioned in a wide visual field // *PLoS One*–2013–№ 8(8).
12. van Honk J., Tuiten A., De Haan E., van den Hout M., Stam H. Attentional biases for angry faces: Relationships to trait anger and anxiety // *Cognition and Emotion*–2001–№ 15(3)–P. 279–297

## **Chapter 16. Cortical mechanisms of speech comprehension: from index sign to symbolic sign**

**Sizova O. B.**

*Psychological and Pedagogical Center for Social Adaptation of Children with Severe Speech Disorders, St. Petersburg, Russia*

Speech comprehension disorders are traditionally associated with primary dysfunction of the temporal cortex, which manifests itself in the form of sensory aphasia syndrome in adults and in sensory alalia or developmental verbal auditory agnosia in case of congenital immaturity of speech comprehension. However, the symptoms of the two syndromes under consideration differ significantly. Common are only such signs as the presence of echolalia, high speech activity and the intactness of the intonational component of speech production of adults and children (Лурия 1947, Трауготт 1994). With sensory aphasia in adults, there are features of speech comprehension and production that are quite consistent with the idea of a basic disorder of the phonological processing of the incoming speech signal. Adults have difficulty distinguishing phonetically similar words and meanings of morphological units. Patients with sensory aphasia find it difficult to differentiate the lexical and the grammatical meanings realized by phonetically similar sequences of sounds. As a result, phonetic and semantic paraphasias are found in their own speech, due to the similarity of phonetic characteristics or meanings of interchangeable in paraphasias words (Лурия 1947). Thus, damage to the functions of perceiving sounding speech in adults with sensory aphasia can reasonably be considered as a disorder of the symbolic level of the speech-language system. This violation concerns the connection between phonetics and linguistic meaning, the processing of speech as a specific acoustic code referring to information transmitted by conventional sound sequences.

Meanwhile, congenital disorders of speech comprehension in children are accompanied by manifestations of a much deeper dysfunction of acoustic signal processing. They reveal selective difficulties in the formation of a conditioned response to sound, while maintaining its formation to signals in the visual modality. The formed conditioned response to an acoustic stimulus is unstable. Children do not understand the meaning of the sounds that they often encounter in their everyday environment, they cannot adequately determine the location of the sound source in space. Familiar sounds (voices of familiar people, the sound of familiar toys), that is, signals that have already been presented to the child repeatedly, are more likely to evoke a reaction in children with sensory alalia than sounds perceived for the first time. Therefore, for successful sound recognition, such children need a greater number of presentations, a greater frequency of perception of the sound signal. Children with sensory alalia do not develop a preference for speech signals over non-speech ones. (Трауготт, Кайданова 1975, Трауготт 1994). All of the above features of auditory perception ultimately led to significant restrictions on the impressive and expressive vocabulary up to the interlocutor's impression of the complete inability of the child to understand the sounds of speech (Rapin&Allen 1983). An interesting feature of children with sensory alalia is greater success and accuracy in self-producing words compared to understanding the same words. This trend gives the impression of a larger volume of expressive vocabulary compared to impressive lexicon (Трауготт, Кайданова 1975).

Our longitudinal observations of the development of 7 children (2 girls and 5 boys) with sensory alalia syndrome make it possible to supplement the existing literature descriptions with some important characteristics of the perception and production of speech by this group of

children. The understanding and use of an extremely limited set of familiar lexemes and not-one-word utterance is strictly determined by the specific situation in which these language units are used and perceived by children.

For example, Masha, in age 5;5, responded stably and adequately to the request of an adult, “Masha, bring me a notebook!”, if this appeal was made at the end of a session with a speech therapist. The girl found a notebook and brought it to the teacher. When the speech therapist once turned to Masha with exactly the same request in the middle of the lesson, the girl turned her head to the speaker, smiled, but did not even try to find the notebook with her eyes. The phrase was heard, but its meaning in this context was not understood by the child.

Similarly, the use of familiar phrases is situationally restricted. Four-year-old Grisha always said confidently when meeting *\*Pivetkakdea!?* (=Hellohowareyou!?). But not one of the words that make up this phrase was reproduced or perceived by Grisha in other situations.

Speech intonations, accurately and adequately reproduced by children with sensory alalia, can give the interlocutor the impression of a child’s understanding of speech. In a standard dialogue, in which an adult tried to find out the name and age of the child and gave verbal clues, facilitating the search for an answer, five-year-old Gulya demonstrated a virtuoso understanding of the meaning of interrogative and affirmative intonations. The girl singled out the last segment in the adult’s question, marked with interrogative intonation, and replaced the interrogative intonation with an affirmative one in the answer.

Adult: What is your name? Gulya?

Gulya, 5;2: *\*Gulya!*

However, when Gulya applied this tactic to a deliberately provocative utterance from an adult, it was found that she probably does not differentiate the phonological features of the words, and, accordingly, does not understand the meaning of the final segments of the adult’s phrases.

Adult: What is your name? Masha?

Gulya: *\*Masha!*

With the most profound dysfunction of understanding, the child’s speech production may be limited precisely to the reproduction of various intonational patterns. Masha, 5 years old, playing with dolls, produced sequences of sounds corresponding to the intonation of a question, exclamation, resentment, joy, depicting a dialogue between toys. At the same time, these completely recognizable intonation contours did not contain not only usual words, but even sounds that distinguish the phonetic system of the Russian language.

Taking linguistic means, children with sensory alalia use them without understanding and without encoding any meanings that go beyond the current situation with their help. This qualitative difference in the functioning of language units in the communicative system of children with sensory alalia is confirmed, in particular, by the fact that the translation of their speech features does not cause any difficulties: there are no proper language meanings, with specific characteristics of the language used, in their communicative system. Thus, sensory aphasia and sensory alalia, despite the commonality of the primary defect (dysfunction of the temporal cortex), are fundamentally different syndromes.

However, in children’s speech pathology, a syndrome has been described that has much more similarities with sensory aphasia than sensory alalia. Symptoms of afferent motor alalia (hereinafter АМА) (Белова-Давид 1969) or primary somatosensory dysfunction (hereinafter PSD)

(Сизова 2017), the mechanism of which is associated with dysfunction of the parietal cortical zones, includes phonetic and semantic paraphasias (Белова-Давид 1969), difficulties in differentiating words and grammatical markers that are similar in sound and meaning (Сизова 2017, 2022). These features of the impressive and expressive speech of children with AMA/PSD are due to the primary defect in controlling the accuracy of articulatory movements, provided by postcentral parts of the parietal cortex. Cortical representations of articulatory postures, necessary for the realization of the sounds of the native language, are formed defectively and cannot be adequately used to control the correspondence of the current pronunciation to the target. When generating an utterance, the target articulation can, in principle, be replaced by any of the articulatory movements available to the child. As a result, any target phoneme in a word can be randomly substituted. That's why the phonetic realization of the word may change with each subsequent production, which leads to phonetic paraphasias. At the same time, the connection between the sound and meaning of lexemes or grammatical markers also turns out to be unreliable and ambiguous. This leads to failures in the elicitation of lexemes and in the perception of lexemes and grammatical markers (Сизова 2017, 2022). due to dysfunction of the proprioceptive component of phonological analysis (Орфинская 2003). Thus, speech comprehension disorders, phonological processing dysfunction in AMA/PSD are secondary due to the primary dysfunction of kinesthetic / proprioceptive control of articulation. The adequacy of the intonational design of the utterance and the deployed complex syntax, that provides high speech activity, are also characteristics of the described group of children, uniting them both with adults with sensory aphasia and with children with sensory alalia.

Let us consider the speech characteristics of children with AMA/PSD that are common to them and patients with sensory aphasia. They are derived from a series of our longitudinal developmental observations of 19 preschoolers with PSD (10 girls and 9 boys). The difficulty of distinguishing between phonetically and semantically similar words is illustrated by the following dialogue taking place at an exhibition of children's crafts:

Adult: Choroshaya poluchilas' podelka! (=Good work!)

Timur, 4;11 (surprised): *\*poddelka?! (=fake?)*

Misunderstandings are also revealed in the perception of morphological units, such a prefixes:

Adult: Xenia, razdevaysya! (=Xenia, undress!)

Xenia, 4;9: *\*Kuda odevatsya? (=Where to dress up?)*

The given examples clearly show that the signifiers, phonetic words are not clearly differentiated, first of all, in the pronunciation of children, that is, not primarily acoustic, but articulatory representations of substituted lexemes and morphemes are not sufficiently differentiated. Similar-pronounced and, accordingly, similar-sounding words are not differentiated in the language system of children with AMA/PSD in speech perception too, which in turn leads to an inadequate understanding of the meaning of erroneously perceived and reproduced words. The fact that the words were not only incorrectly reproduced, but also incorrectly perceived, is evidenced by the presence of a repeat question from child. The child feels the inadequacy of the lexeme identified by him in the existing context, and the way the word is reproduced demonstrates to the listener that it is the pronunciation inaccuracy, proprioceptive, articulatory analysis of the perceived word caused a failure of understanding.

In the utterances produced by children with AMA/PSD, phonetic paraphasias are found (sub-

stitution the word by phonetic similarity):

Christina, 4;11 (fashioned a hill with a hole on the side): *\*Eto beryoza!* (instead berloga) *\*Eto medved'!* (=It's birch! (instead den) This is a bear!)

In Christina's speech, the similarity of the phonetic words «beryoza» and «berloga» is enhanced due to age-related pronunciation features: the substitution of sonors with the sound Y.

The substitutions of the target word with a lexeme close in meaning (semantic paraphasias) are revealed:

Slava, 4;5: *\*Ya vchera na dache upal, byla shishka* (=I fell at the dacha yesterday, there was a bump)

Christina, 4;5: *\*Ya tozhe upala, byl zholud'...shishka* (=I also fell, I have an acorn... a bump (shishka=cone=bump))

It should be noted that in the speech of children with sensory alalia, phonetic paraphasias appear in the process of correctional work, but there are very few semantic paraphasias (Трауготт, Кайданова 1975).

Thus, the general mechanism of impairment—dysfunction of the temporal cortex—in sensory aphasia in adults and sensory alalia/ developmental verbal auditory agnosia in children leads to fundamentally different patterns of impairment. On the contrary, sensory aphasia (temporal cortex) and afferent motor alalia / primary somatosensory dysfunction (parietal cortex) associated with dysfunction of various parts of the cortex are realized in the form of completely identical speech symptoms.

To explain the reasons for the described paradox, it seems reasonable to refer to the concept of the semantic triangle (Ogden & Richards 1936). According to this concept, any sign, including verbal, is a unity of three components: first is the exponent or sign carrier (in relation to a verbal sign realized in oral speech, it is a phonetic word as a sequence of certain sounds realized through a certain sequence of articulatory movements). The second component is the denotate, or object relatedness of the sign (Лүрия 1979), finally, the third component is the significat, or the meaning of the sign. In ontogeny, the object relatedness of the verbal sign is first formed, and only then does its generalized linguistic meaning, the significat, appear (Лүрия 1979).

At the same time, the existing classification of signs divides them into symbols, indices and icons (Peirce 1868, Бразговская 2018). The vast majority of linguistic signs are verbal symbols that have an exponent, a denotate and a significat. However, some means of verbal and non-verbal language, such as personal pronouns and pointing gestures, are index signs; they have an exponent, or sign-carrier (phonetic word-pronoun or pointing movement and posture of the hand), but their object relatedness is strictly connected with the situation of their use: The signified of the pointing gesture is the object in the current situation to which the gesture is directed; the signified of a personal pronoun is one of the participants directly present in the situation or mentioned in the current dialogue taking place here and now. The function of the index sign is an indication, and the index can perform its sign function only due to the joint (spatial and/or temporal) occurrence of the sign carrier and its denotate in the current situation. The index does not have a generalized meaning, a significat, precisely because its meaning is determined by the current specific circumstance, and not by the generalized situation (Бразговская 2018).

Let us consider speech production and features of speech understanding by children with sensory and afferent motor alalia in terms of the production and perception of signs-symbols and signs-indexes.

In the case of phonetic paraphasia in Christina's statement *\*Eto beryoza!* (instead berloga) *\*Eto medved'!* (=It's birch! (instead den) This is a bear!) occasional homonymy arises. The word "birch" in the child's lexical system acquires, along with the usual meaning (tree), one more meaning (bear's dwelling) i.e. one exponent acquires two denotates and two significats instead of one. This metamorphosis occurs due to the phonetic and articulatory similarity of the exponents "beryoza" and "berloga", merging into one diffuse articulation representation due to insufficient accuracy of articulation control. As a result, one signifier is correlated in the child's individual language system with two meanings.

The reverse metamorphosis occurs with the word "acorn", which turned out to be a synonym for the lexeme "bump" in the individual language system of Christina within the framework of the statement: *\*Ya tozhe upala, byl zholud'...shishka* (=I also fell, I have an acorn... a bump). Indeed, the lexemes "cone" and "acorn" have a common significat—"tree fruit". But the failure occurs when Christina tries to use the similarity of the direct meaning and the denotations of two words (cone and acorn) to replace the lexeme used in the derived meaning: "bump is result of trauma" ("shishka" have two means in Russian: cone and bump). The lexeme "acorn" does not have this derived meaning, and it is important that Christina probably notices this semantic inconsistency and corrects her slip of the tongue. Thus, both the failure and its correction indicate that in the language system of the child, not only the direct object relatedness, the denotate of the used linguistic sign, but also its derivative meaning, significat are actualized. This paraphasias in Christina's speech, both phonetic and, especially, semantic, indicate that, despite the ambiguous, unstable relationship between the exponent, denotate and significat in the lexemes used by the child with AMA/PSD, they are signs-symbols, that have a generalized meaning, a significat, i.e., function as a verbal language sign, a symbolic level sign. Difficulties in translating these utterances, related to the specifics of the Russian language, also confirm this statement.

Understanding a formal linguistic sign—the statement "Masha, bring me a notebook!" within the framework of a stably reproducible and new situation—indicates a different way of processing this incoming acoustic signal. The meaning of the statement is actualized only within the framework of a familiar situation due to the frequency of the joint occurrence of the utterance sound and the situation, that necessitates bringing a notebook to an adult. The actual recognizable acoustic signal serves as a trigger for the start of the corresponding action without the need to refer to its linguistic meaning, significat. Masha's misunderstanding of the same speech signal outside of a familiar situation indicates that this utterance has no significat in the child's acoustic system, it's just a sequence of sounds. This sound sequence expected and therefore recognizable in a specific situation and launching the corresponding motor program in relation to a specific fragment of this situation. Thus, the verbal sign "Masha, bring me a notebook" has a standard speech exponent—a sequence of phonemes, but its denotate is the object (and the action associated with it) exclusively in a specific situation, the formally verbal exponent does not have a generalized linguistic meaning, and the sign function of the signal is performed by means of a verbal "pointing" to a certain fragment of the current situation. In fact, a full-fledged sign-symbol transmitted by an adult, in the acoustic system of a child with sensory alalia is reduced to an index sign devoid of signification. Similarly, being perceived and reproduced by the child exclusively in specific current situations,

all linguistic signs that are available to children with sensory alalia function. At the same time, researchers notice that children with sensory alalia do not spontaneously correct speech underdevelopment without special speech therapy intervention (Трайготт, Кайданова 1975). That is, the sound sequences, which are linguistic units that encode information, regardless of the situation of their presentation, in the acoustic system of a child with sensory alalia remain index signs that acquire contextual meaning exclusively within the framework of a specific situation, in which the child learns to recognize them due to the frequency of their use and joint occurrence with certain objects and actions.

The described level of functioning of the acoustic sign is correlated with the mode of functioning of three differentiated sound signals in green monkeys. Indeed, these animals are able, having heard the signal of relatives, indicating the approach of an eagle, to descend from a tree, in response to a signal issued when a leopard appears, on the contrary, rise higher so as not to become a victim of a predator. Finally, a specific signal about the approach of a snake makes the monkeys make noise together since such actions can drive away the snake. Despite the fact that some researchers have described this communication system as a language consisting of three different words with their signifiers (a specific voice signal) and signifieds (three predators dangerous for green monkeys) (Tecumseh Fitch 2010), it is easy to see that the interpretation of the described signals as signs-indexes will be quite plausible. Indeed, they are used in strictly specific situations and cause a rigidly conditioned motor reaction, to which the corresponding vocalization “points”. Vocalization is learned due to the frequency co-occurrence of a specific acoustic signal and the always identical actions following it. This sign does not have a signified, it is not needed, since monkeys can and will report a predator only if they really see it here and now, and the only adaptive response to this sound will be appropriate actions to save one’s own life. In no other situations are these screams issued and, accordingly, do not cause other reactions, i. e., their production and understanding are strictly situational.

But Masha also reacts in a similar way to the phrase of an adult, having limited capabilities for recognizing multicomponent acoustic signals due to the primary defect in the sensory alalia and recognizing this signal insofar as it expects to hear it strictly in defined situation. Outside the situation, the child does not have appropriate acoustic expectations, the phrase, like background insignificant noise, remains unrecognized, and no reaction follows it. The only difference is that all green monkeys live in a communicative system of index signs, while a child with sensory alalia, communicating in the world of signs-symbols, reduces them, however, to the level of indexes due to the limited ability to recognize acoustic signals without relying on the situational context.

At the same time, however, the ability to communicate in a child with sensory alalia is completely intact, and he uses the entire arsenal of communication means at his disposal to maintain contact and translate his intentions. So, Kolya, 11 years old, painted multi-figure drawings and series of pictures like comics (essentially icon signs in which the signifiers are similar to the signifieds) to use them to tell, for example, about a trip with the family to the country or with the father to the gas station. These stories contained a minimum of usual words, but a lot of gestures, pointings and vocalizations. Intact visual modality is thus used as a channel of communication when it is impossible to use verbal communication as a usual means of encoding information.

The described characteristics of auditory perception and ways of communication of children with sensory alalia confirm, that the primary defect in this syndrome does not directly affect

the symbolic system of the language and is not a primary dysfunction of the phonological, semantic processing of the incoming acoustic signal. Due to the instability and diffuseness of representations formed in the auditory cortex, stable and accurate recognition of current acoustic signals is difficult. The signal that often sounds in this situation is not recognized as identical to the one that has already perceived, and that is why a larger number of repetitions or a longer duration of signal presentation is required for its recognition. This is a low-level primary deficit in cortical processing of acoustic signals (not only verbal, but verbal as well) secondarily affects the ability to master speech as a way of encoding and transmitting information. This happens at the earliest stages of development. This deficiency in the processing of incoming signals leads to the inability to fully participate in situations of shared attention (Scaife & Bruner 1975, Tomasello 2008). Within the framework of such situations, a typical developing child at the age of 12 months masters the meaning of an acoustic verbal signal and the sign function of speech. Normally, a child already at 8 months is able to turn his head towards a well-known and often called object. But due to the above reasons, for children with sensory alalia, the inability to assess the identity of the incoming acoustic signal to the one already perceived in the presence of the corresponding object makes it difficult for such children to form this basic function. However, even the described reactions are insufficient for a true understanding of speech. Within the framework of shared triadic attention, three participants in the situation—a child, a caregiver (more often a mother) and the object of their joint attention (more often a toy)—are involved in a common purposeful activity. Mother and child manipulate the toy together, and at the same time, the mother spontaneously performs verbal commentary on the activity, naming the actions performed and pronouncing the name of the toy especially often (repetitions are characteristic of the child-direct speech) (Цейтлин 2000). Within the framework of such interaction, the child learns to follow the direction of the caregiver's gaze, to understand his intentions regarding the object of shared attention (Scaife&Bruner 1975). The child also learns to process a verbal comment, isolating the most frequently sounding patterns in it and correlating them with the object and purpose of the activity (toy) (Tomasello 2008).

It is clear that the first two skills are quite accessible for a child with sensory alalia, which is why the actual ability to communicate develops in him: children with sensory alalia are sociable, participate in common activities with pleasure, are able to track the direction of the interlocutor's gaze and in this way read his intentions. But the third most important function of activity within the framework of triadic shared attention—the understanding of the meaning of the words that sound frequently in the commentary on the activity—for a child with sensory alalia is not available due to the inability to capture the identity of repetitive frequency acoustic patterns. To form a stable connection between an object and its verbal naming, a child with sensory alalia needs a significantly larger number of signal presentations, that is, as mentioned above, special correctional education that is not available within the framework of ordinary communication in the family circle.

Another described strategy for mastering new words is a fast mapping (Carey & Bartlett 1978). The strategy is based on the ability to isolate the first perceived pattern from the speech flow and correlate it with a new object that has appeared in the field of activity. Fast mapping is demonstrated by typically developing children older than 2 years. However, this strategy is also inaccessible to a child with sensory alalia, since the novelty of an acoustic signal perceived for the first time can only be assessed by comparing its dissimilarity with already stored acoustic representations. A child with sensory alalia does not have accurate differentiated representations of already perceived acoustic signals, which is why the assessment of

the novelty of a newly received signal is also inaccessible to him. Thus, a child with sensory alalia, participating, like all children of 12 months, in situations of shared attention, successfully learns interaction and communication within them, but not its verbal form. Even for a situational, contextual understanding of the meaning of a word, his cortical auditory system lacks the resources to adequately process incoming signals.

Therefore, speech understanding in sensory aphasia in adults and afferent motor alalia/primary somatosensory dysfunction in children is built on the basis of fundamentally different basic cortical mechanisms, than the perception of any, including a verbal acoustic signal in sensory alalia / developmental verbal auditory agnosia. It is important to note that the difference is not in the quality of the processing of the actual verbal signal that encodes the linguistic meaning. The differences relate to the formation of acoustic (including verbal) cortical representations. Since the diagnosis of “sensory aphasia” applies only to patients who have already formed understanding of speech is impaired in the presence of well-formed productive speech (Трайготт, Кайданова 1975), therefore, in both adults and children with acquired sensory aphasia, full-fledged cortical representations of verbal signals in the temporal cortex are already formed in the anamnesis. With ongoing verbal signal processing, patients with sensory aphasia may experience central acoustic processing difficulties similar to those described in children with sensory alalia. Thus, in adult patients with aphasia, there is a need for a longer presentation of the signal to respond to it, difficulties in determining the source of the signal in space, and instability of auditory thresholds (Трайготт, Кайданова 1975) ю But in cases of sensory aphasia, the most important idea has already been formed that speech matters, that the verbal signal transmits not only the emotional state of the interlocutor, but information stably encoded by a specific sequence of sounds, regardless of the situation of its presentation, and this information needs to be decoded and understood. Simply put, with sensory aphasia, the idea that speech is necessary and important to understand is preserved. Therefore, the existing low-level features of cortical signal processing only affect the accuracy and reliability of phonological analysis, make it difficult to convert sounds into meanings. But patients with sensory aphasia try, with varying degrees of success, to discover the linguistic meaning of sound sequences.

With afferent motor alalia / primary somatosensory dysfunction, the functioning of the auditory system is also not primarily impaired. Therefore, the full participation of the child in a situation of triadic shared attention is likely, when the name of the object of joint action is pronounced by an adult participant in the situation, and the child only perceives speech and evaluates the frequency or novelty (using the fast mapping strategy) of the acoustic patterns broadcast by the adult. The distortion of the cortical representation probably occurs later, when the child himself tries to reproduce the frequency and new verbal units from the input. In this case, the dysfunction of the kinesthetic control of the current articulation due to defectively formed representations of articulation movements affects the accuracy of reproduction and, secondarily, the already established acoustic representation of the reproduced speech segment. Acoustic representations of sounding speech are formed in the cortex of the developing brain earlier than the corresponding speech-motor representations, and the development of the auditory system precedes and sanctions the formation of pronunciation skills (Бельтюков 1977). Thus, even in motor alalia based on primary somatosensory dysfunction, the idea of speech as an information coding system is formed in ontogenesis. Despite the inaccuracy of perception and even speech understanding errors due to dysfunction of the kinesthetic control system, a child with motor alalia is ready to decode the verbal acoustic stream and decipher the meaning behind the phonetic signifier.

With sensory alalia, the qualitative difference in the processing of verbal signals is due to the basic insufficiency of auditory gnosis and the formation of inaccurate representations of acoustic signals regardless of the general biological or symbolic level of their potential decoding. At the same time, auditory sensitivity in children with sensory alalia may be normal or, with a decrease in certain frequency ranges (often high frequencies), it is still sufficient for mastering speech. One of the differential diagnostic signs of sensory alalia is proposed to be the ability to hear and respond to whispered speech (Трайгорт, Кайданова 1975). However, the need for greater repeatability, duration and intensity of the acoustic signal in order for a child with sensory alalia to respond to it leads to critical difficulties in the formation of auditory gnosis skills.

Of course, the condition for the successful formation of a cortical representation of an acoustic signal is its repeatability, reproducibility, and high adaptive significance. All these conditions correspond to the intonation patterns of the language: their variability is limited, for example, in the Russian language there are only 7 intonation constructions that express the meanings specific for each intonation construction using various patterns of raising and lowering the tone of the voice (Брызгунова 1981). Intonational constructions are regularly reproduced within situations whose emotional meaning is easily perceived by children with sensory alalia, since the basic ability to cooperate and understand the intentions of another is formed in them in a timely and adequate manner in conditions of shared attention. Understanding the emotional characteristics of the situation is extremely important for successful adaptation: it is these characteristics that confirm or refute the appropriateness and adequacy of the child's behavior. The vital emotional characteristics of the situation are associated with the intonational, prosodic characteristics of the speech flow that accompanies the deployment of the situation. Thus, the probability of fixing an intonation pattern and linking it with a certain emotional meaning for a child with sensory alalia is much higher than the ability to isolate and associate with linguistic meaning infinitely variable combinations of phonemes presented without pauses between semantic segments and often emotionally neutral. This is probably why, by preschool age, children with sensory alalia are able to differentiate, reproduce intonation patterns, and even successfully use them to imitate participation in a meaningful dialogue by carrying out emotional interaction with interlocutors.

However, children with sensory alalia until a certain age are not even aware of the presence of a linguistic meaning of a verbal sign, the need to link the exponent not only with the denotate, the objective meaning suggested by the situational context, but also with the signification, the generalized meaning, that does not depend on the situation of use. The connotations of a linguistic sign are most inaccessible to a child with sensory alalia. So, Kolya, at the age of 11, affectionately called his mother "toad", because his mother's big eyes and her physique reminded Kolya of this animal, which he sincerely liked. It is this state of the level of understanding of speech and its role in communication that caused the almost complete absence of negativism in communication situations among younger preschoolers with sensory alalia. Children with different syndromes of motor alalia worry about the imperfection of their own speech or, in extreme cases, the dullness of their interlocutors. Children with sensory alalia simply do not realize that speech is needed for something more than an exchange of emotions, which is quite accessible to them, therefore they do not worry about the success of their communication.

It is the described critical differences in the levels of understanding of the verbal acoustic signal that affect the way it is processed in the cerebral cortex. In sensory aphasia and motor ala-

lia, the verbal signal is processed inaccurately and an erroneous decision can be made about its meaning or about presence its meaning at all. With sensory aphasia, such failures occur due to a disorder of the auditory component of the phonological processing of the incoming signal. In afferent motor alalia/primary somatosensory dysfunction, comprehension failures are facilitated by dysfunction of the proprioceptive component of phonological processing which secondarily affects the accuracy of the auditory component of this process. But in both cases, the incoming signal is processed as a linguistic sign-symbol, the meaning of which the listener needs to make the right decision about.

With sensory alalia/ developmental verbal auditory agnosia, the level of phonological semantic analysis may not participate at all in the processing the verbal acoustic signal. The incoming speech signal is processed as an index sign indicating a certain fragment of the current situation and/or an action associated with it. At the same time, the recognition of this sign itself is seriously difficult due to the instability and diffuseness of acoustic representations in the listener's cortex. The probability of recognition is tightly related to the non-speech components of the situational context, which predict the probability of a certain acoustic signal previously perceived in similar circumstances. This is the basic level of correlation of an acoustic signal with a fragment of the situation, which a typically developing child reaches by 8 months, and this level remains the only available way of processing the incoming verbal signal for a child with sensory alalia without special corrective training.

Thus, the ontogeny of the cortical mechanisms of speech signal processing in normal and most cases of deviant speech development is an evolutionary development from the processing of an index sign to understanding the meaning of a symbolic sign

**Table 1.** *The state of the cortical provision of speech comprehension in different syndromes with speech comprehension dysfunction*

	Sensory aphasia	Afferent motor alalia/ primary somatosensory dysfunction	Sensory alalia
Formation of acoustic (including speech) representations	formed in ontogeny	formed, secondary defect due to dysfunction of kinesthetic control of articulation	instability, diffuseness
Comprehension of meaning of the verbal sign	formed in ontogeny (shared attention)	formed in ontogeny (shared attention)	not formed
Incoming speech signal processing	as a symbol sign, impaired due to dysfunction of phonological processing	as a symbol sign, impaired due to dysfunction of proprioceptive component of phonological processing	as an index sign, instability, context dependent due to dysfunction of formation acoustic representations

## References

1. Белова-Давид Р.А. К вопросу систематизации речевых расстройств у детей // Нарушение речи у дошкольников. М. Просвещение 1969. С. 10–46
2. Бельтюков В.И. Взаимодействие анализаторов в процессе восприятия и усвоения устной речи М. Педагогика 1977 176 с.
3. Бразговская Е. В лабиринтах семиотики. Очерки и этюды по общей семиотике и семиотике искусства М., Екатеринбург Кабинетный ученый 2018 224 с.

4. Брызгунова Е. А. Звуки и интонация русской речи М., Изд-во «Русский язык» 1981 278 с.
5. Лурия А. Р. Травматическая афазия. Клиника, семиотика и восстановительная терапия М., Изд-во Академии медицинских наук 1947 367 с.
6. Лурия А. Р. Язык и сознание / Под редакцией Е. Д. Хомской М. Изд-во Моск. ун-та, 1979 320 с.
7. Орфинская, В. К. Принципы построения дифференцированной методики обучения алаликов на основе лингвистической классификации форм алалии // Логопедия. Методическое наследие. М. Кн. III: Системные нарушения речи: Алалия. Афазия. 2003 С. 81–121
8. Сизова О. Б. К вопросу об иерархии механизмов речи // Модальность в языке взрослых и детей; Механизмы усвоения языка и овладение речевой компетенцией. АСТА LINGUISTICA PETROPOLITANA. Труды Института лингвистических исследований РАН Т. XIII. Ч. 3. СПб. Наука, 2017. С. 741–773
9. Сизова О. Б. «Два по цене одного»: одна (или две?) модели порождения лексических инноваций в детской речи // Проблемы онтолингвистики–2022: речевой мир ребенка (универсальные механизмы и индивидуальные процессы). Материалы ежегодной международной научной конференции / Редакторы Еливанова М. А., Крылова И. А., Краснощекова С. В. Издательство «BBN» с. 166–174
10. Трауготт Н. Н. Как помочь детям, которые плохо говорят СПб Изд-во «СМАРТ» 1994 59 с.
11. Трауготт Н. Н., Кайданова С. И. Нарушение слуха при сенсорной алалии и афазии. Экспериментально-клиническое исследование Л., Изд-во «Наука» 1975 180 с.
12. Цейтлин С. Н. Язык и ребенок: Лингвистика детской речи: Учеб. пособие для студ. высш. учеб. заведений. М: Гуманит. изд. центр ВЛАДОС, 2000. 240 с.
13. Carey S., Bartlett E. Acquiring a single new word Proceedings of the Stanford Child Language Conference. Vol. 15. 1978 pp. 17–29
14. Ogden C. K., Richards I. A. The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism London, K. Paul, Trench, Trubner & Co.; New York, Harcourt, Brace & Co. 1936 363 p.
15. Peirce C. S. On a New List of Categories Proceedings of the American Academy of Arts and Sciences vol. VII. Boston and Cambridge: Welch, Bigelow, and Company 1868 pp. 287–298
16. Rapin I., Allen D. Developmental language disorders: nosologic considerations. In Kirk U. (ed.) Neuropsychology of language, reading, and spelling New York: Academic Press. 1983 pp. 155–184.
17. Scaife, M., Bruner, J. S. Capacity for joint visual attention in the infant Nature, 253, 1975 pp. 265–266.
18. Tecumseh Fitch W. The Evolution of Language Cambridge University Press 2010 611 p.
19. Tomasello M. Origins of human communication A Bradford books. The MIT Press Cambridge Massachusetts, London, England 2008 400 p.

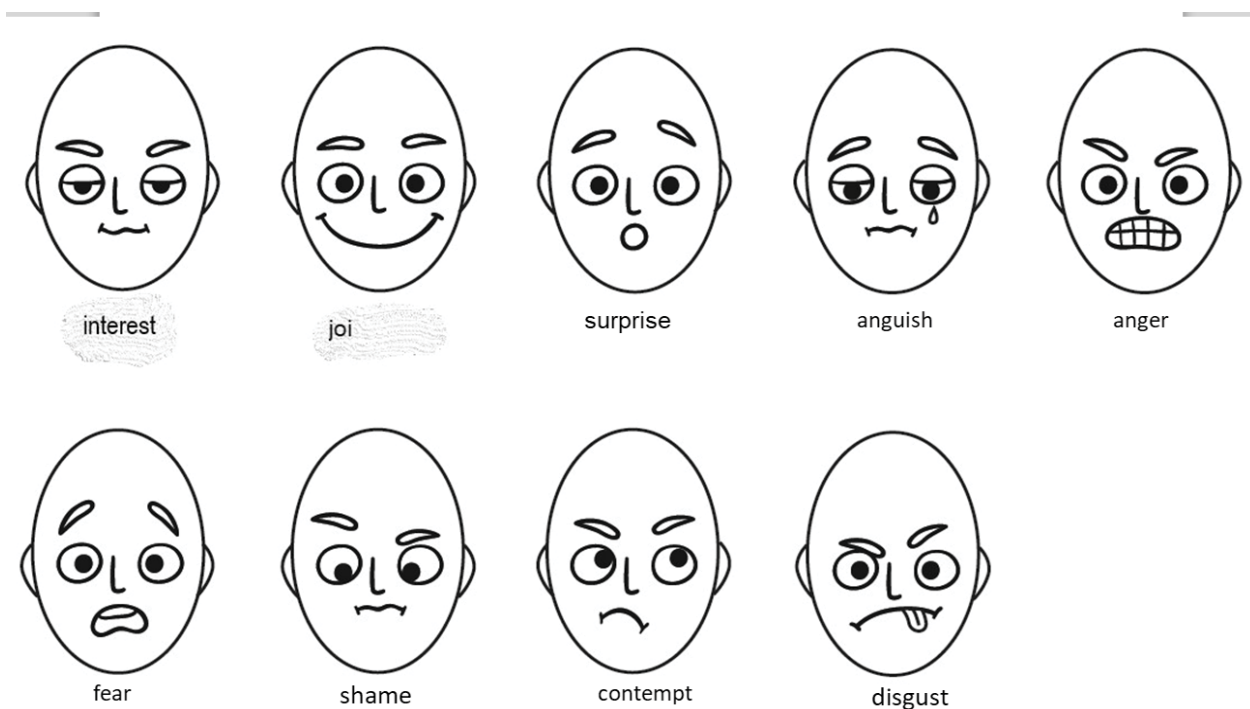
## Chapter 17. Neural mechanisms of emotional memory

*Tsvetkov E. A.<sup>1</sup>, Krasnoschekova E. I.<sup>2</sup>, Kharazova A. D.<sup>2</sup>*

<sup>1</sup>*Medical University of South Carolina, USA*

<sup>2</sup>*St. Petersburg State University, Russia*

Those objects and situations that evoke emotions attract attention. This makes them memorable and influences behavior, as evidenced by the universality of grimaces and facial expressions that reflect basic emotions (fig 1).



**Figure 1.** Mimic grimaces corresponding to the basic emotions according to C. Izard's classification

Being a form of negative emotions, fear accompanies a person from birth and plays an important role in their daily life. Scientific approach for mechanisms of fear, however, has recently begun to be formed and developed within the framework of general ideas about the mechanisms of emotions. Charles Darwin greatly contributed to the development of the biological approach for understanding emotions (Darwin, 1872). In his publications, he explained emotions and their physiological manifestations from the perspective of evolution. He suggested that emotions can be rudiments of reactions developed evolutionary in the struggle for existence, or their modification. Ivan P. Pavlov's theory of the higher nervous activity was a significant milestone in the study of the brain, in general, and in particular of the emotions, which formed a new approach for physiological basis of emotions. According to this theory, emotions are reflexes in nature. They have features of both conditioned and unconditioned reflexes resulting from external and internal irritations, either physiological or cognitive. The central part of the reflex is emotion. A great contribution to the theory of reflexes was also made by another outstanding scientist, V. M. Bekhterev. He defined Pavlov's conditioned reflexes as "associative" or "associative-motor reflexes". According to Bekhterev, this type of reflexes arises both from a slight advance of the neutral stimulus, and from a slight delay. In the first case, the

definition corresponds to the Pavlov's conditioned reflex, while in the second, it is similar to the B. Skinner's concept of "operant behavior". Operant behavior by Skinner differs from the classical conditioned reflex in that the behavior of the animal (or experimental) is modulated, and is not passively trained to be automatic. Within the Pavlov's theory of reflex, the mechanisms of emotions were studied by a number of prominent Soviet scientists. Among them are academicians Anokhin P. K. (1968), Veldman A. V. (1976), Bekhtereva N. P. (1975), Professor Smirnov V. M. (1976), Simonov P. V. (1981, 1987) and many others. Remarkable advances have also been achieved within the cognitive approach. In particular, Schachter-Singer's cognition-arousal theory (Schachter, 1964; Schachter and Singer, 1962), Arnold-Lazarus's cognitive theory (Arnold, 1960; Lazarus, 1991), the theory of cognitive dissonance (Festinger, 1962), appeared under the umbrella of cognitive science. Besides cognitive theories, other concepts were developed such as Freud's psychoanalytic theory systematized by Rapaport (Rapaport, 1960), Simonov's information theory (Simonov, 1981; 1987), Zajonc's vascular theory (Adelman and Zajonc, 1989), and Izard's differential emotions theory (Izard, 1980; 2007). Despite significant progress in the study of emotions, none of the theories mentioned above can bring definitive clarity into the understanding of mechanisms providing the fullness of emotional manifestations and experience. Nowadays Izard's differential emotions theory is probably one of the widest and most universal that combines many hypotheses and describes mechanisms or neural architectonics of emotional experiences (Izard, 1980; 2007). This theory suggests that several independent emotions are fundamental, each of them being defined as a complex process with neurophysiological, neuromuscular and phenomenological aspects. At the neurophysiological level, emotion is identified by the electrochemical activity of the nervous system, in particular the cortex, hypothalamus, basal ganglia, limbic system, facial and trigeminal nerves. At the neuromuscular level, emotion is, above all, mimic activity, and then it is pantomimic, visceral-endocrine and vocal reaction. At the phenomenological level, emotion is either a strongly motivated experience or an experience of direct relevance to the person. This theory is based on the five key assumptions given below (Izard, 1980; 2007): 1. Ten fundamental emotions of human existence are anger, contempt, disgust, distress, fear, interest, joy, shame, surprise/guilt. 2. Each fundamental emotion has unique motivational and phenomenological properties. 3. Fundamental emotions causes specific experiences and external manifestations. 4. One emotion can activate, strengthen or weaken the other. 5. Emotions affect motivation, homeostatic, perceptual, cognitive, and behavior processes.

Thus, the model proposed by C. Izard is convenient due to its flexibility and integration of different theories. One of the important characteristics of this theory is compatibility with the terms and ideas of Pavlov's reflex theory. The latter allows to consider emotions as a reflex in the context of Izard's theory.

Moreover, this theory can be expanded through researches in other fields of study. The theory also enables to examine the interaction of emotional and cognitive processes from the positions, for example, of cognitive theories and, in particular, Simonov's information theory (Simonov, 1981; 1987). The flexibility and high integrative capacity of the Izard's theory make it appropriate as a basis for characterizing the emotion of fear, as well as studying its neural and synaptic mechanisms in experiment.

Based on the Izard's differential emotions theory, fear is a fundamental emotion, which means it is an inherited emotional process with genetically determined physiological component, strictly defined facial expression, and certain subjective experience. It cannot be divided into components, but can be the basis for other emotions (Izard, 2007). Fear may

be caused by real or imaginary (though experienced as real) danger, such as a threat to life, beliefs, values, etc. Fear mobilizes the body to implement protective behavior, including freezing, avoidance, aggression (Izard, 2007). The neural mechanisms provide not only genetically determined fears, but also formation of learned fears, that is important for the adaptation to the environment. Kluver and Bucy laid the foundation to the neurophysiological understanding of processes triggering the fear. They evaluated behavior after medial temporal lobe destruction in monkeys and found a number of dramatic changes in emotional behavior described as “mental blindness”, and later called the Kluver-Bucy syndrome (Kluver and Bucy, 1937, 1997).

Animals with this syndrome displayed no fear which radically changed their behavior (Kluver and Bucy, 1937, 1997; Rosvold et al., 1954). Later it was shown that Kluver-Bucy syndrome results from focal brain lesions including bilateral disturbance of amygdala (Weiskrantz, 1956; Zola-Morgan et al., 1991; Bucher et al., 1970). Similar effects were discovered in other mammals (Goddard, 1964) and led to the assumption that amygdala, being a complex structure of the forebrain and part of the limbic system, plays the main role in the mechanisms of fear.

The modern definition of amygdala is a complex of nuclei consisting of more than thirteen formations identified by specific cytoarchitectonics and histochemistry, as well as by the characteristics of internuclear and intranuclear connections (Pitkanen et al., 1997). At the moment, amygdala has been described in various vertebrates, including monkeys (Amaral and Price, 1984; Russchen et al., 1985), cats (Krettek and Price, 1978; Price et al., 1987), rats (Pitkanen, 2000), turtles (Belekhova, 1992), and birds (Cheng, 1999).

Animal studies of amygdala revealed a considerable similarity of its structure among different species, as well as the afferent and efferent connections of its nuclei (Sah et al. 2003). In the vertebrates from reptiles to humans, however, there is a significant increase in the volume of afferent basolateral nucleus of amygdala (BLA) with respect to its efferent central nucleus (CeA) (Janak and Tye, 2015; Chareyron et al., 2011). This volume increase is associated with the rapid development of the cortical formations related to the BLA through reciprocal connections (Janak and Tye, 2015; Chareyron et al., 2011). Since the subject of this study is rodents (rats and mice), we describe amygdala in relation to the data obtained in rats. Both rats and mice are most common and widely used in experiment.

Classification of the amygdala nuclei in rats has several variations as stated in the literature (Krettek and Price, 1978; Pitkanen et al., 1997; Amunts et al., 2005; Heimer, 2003; Solano-Castiella et al., 2010; Sah, 2003). These differences are insignificant to our study, so we describe amygdala nuclei from the perspective of the most widespread nomenclature proposed by Price (Price et al., 1987) and later modified by Sah (Sah, et al. 2003). In this classification, amygdala nuclei are divided into three groups (Table 1): 1) The deep or basolateral group (sometimes called basolateral nucleus of amygdala, BLA), which includes lateral (LA), basal (B), and accessory basal (AB) nuclei; 2) The superficial or cortical-like group comprises the nucleus of lateral olfactory tract (NLOT), the bed nucleus of accessory olfactory tract (BAOT), the anterior and posterior cortical nucleus (CoA and CoP, respectively), and periamygdaloid cortex (PAC); 3) The centromedial group composed of medial (M), central (CeA), and bed nucleus of stria terminalis (BNST). The final group of nuclei comprising remaining amygdala areas is the anterior amygdala area (AAA), the amygdalo-hippocampal area (AHA), and the intercalated nuclei that do not easily fall into any of these three groups. Almost all the amygdala nuclei are further divided into smaller formations which are called subnuclei and detailed in the Table 1.

*Table 1. Classification of amygdala nuclei (Sah et al., 2003)*

<b>DN - DEEP NUCLEI</b>	<b>SN - SUPERFICIAL NUCLEI</b>	<b>CN - CENTROMEDIAL NUCLEI</b>
<b>BLA - BASOLATERAL NUCLEI</b> <b>LA - Lateral nucleus:</b> LAdl – dorsolateral division; LAvl – ventrolateral division; LAm – medial division.  <b>B – Basal nucleus:</b> Bmc – magnocellular division; Bi – intermediate division; Bpc – parvicellular division.  <b>AB – Accessory basal nucleus:</b> ABmc – magnocellular division; ABpc – parvicellular division.	<b>NLOT - Nucleus of the lateral olfactory tract.</b>  <b>BAOT – Bed nucleus of the accessory olfactory tract.</b>  <b>CoA – Anterior cortical nucleus.</b>  <b>CoP – Posterior cortical nucleus.</b>  <b>PAC - Periamygdaloid cortex:</b> PACm – periamygdaloid cortex, medial division; PACs – periamygdaloid cortex, sulcal division.	<b>CeA – Central nucleus:</b> CeC – capsular division; CeL – lateral division; CeI – intermediate division; CeM – medial division.  <b>M – Medial nucleus:</b> Mr – rostral division; Mcd – dorsal part of central division; Mcv – ventral part of central division; Mc – caudal division.  <b>BNST – bed nucleus of stria terminalis.</b>

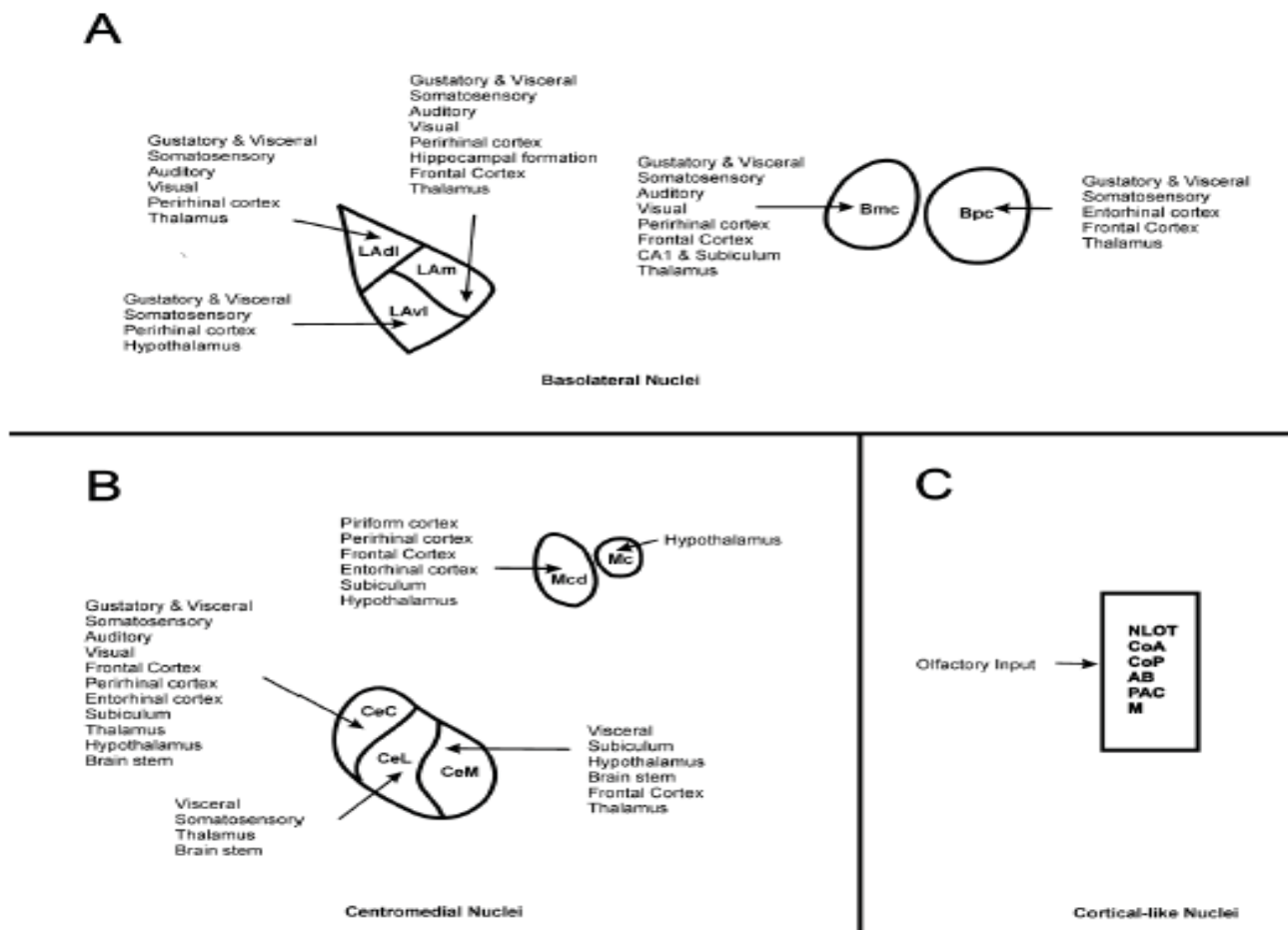
The present data do not allow a definite conclusion, whether the amygdala is a single functional unit or it is a complex of closely located, but multifunctional groups of nuclei. It is suggested, however, that functional systems within the amygdala connect its individual nuclei to provide certain functions. In particular, according to several authors (Davis, 2000; 2001; LeDoux, 2000; Sah, 2003), it is possible to distinguish a set of nuclei responsible for the formation and expression of fear. This functional system is believed to include the basolateral nucleus of amygdala, its central nucleus, and some other subdivisions integrating activity of other units of this structure (Davis, 2000; 2001; LeDoux, 2000; Pitkanen, 2000; Sah, 2003).

The concept of a conditioned reflex is the most appropriate to investigate the mechanisms of fear formation and expression (Sechenov, 1961; Pavlov, 1951). This theory considers fear as a reflex in nature which, therefore, possesses all the properties of conditioned and unconditioned reflexes. The fear can be inherited and learned as any other reflexes. Inherited fear is similar to an unconditioned reflex and involves genetically determined “stationary” neural pathways, whereas learned fear is like a conditioned reflex and realized through the integrated neural connections. Researches of the second half of XX century were focused on the location of associative connections in the learned fear and its molecular mechanisms. In this direction, Dr. Joseph LeDoux, a professor at New York University, is thought to be a pioneer and recognized leader. His research group and some other laboratories showed the amygdala to be not only the central part of the fear as a reflex, but also the center of the associative connections between condition and reaction.

The morphological substrate of these connections was the convergence of amygdala afferents on the cells of the dorsolateral nucleus of basolateral complex, the plasticity of its synapses caused the process. It was demonstrated, in particular, that this process changes the efficiency of synaptic connections in this amygdala nucleus (Huang and Kandel, 1998; Quirk, Repa and LeDoux, 1995; McKernan and Shinnick-Gallagher, 1997; Rogan et al., 1997; Rumpel et al., 2005).

Each of the amygdala nuclei receives specific afferents from several structures of the diencephalon and telencephalon. The most widely presented are sensory inputs from both the thalamus nuclei and the cortical fields. The main sources of polysensory afferents of amygdala are the prefrontal cortex, the hippocampus and the para-hippocampal structures (peririnal,

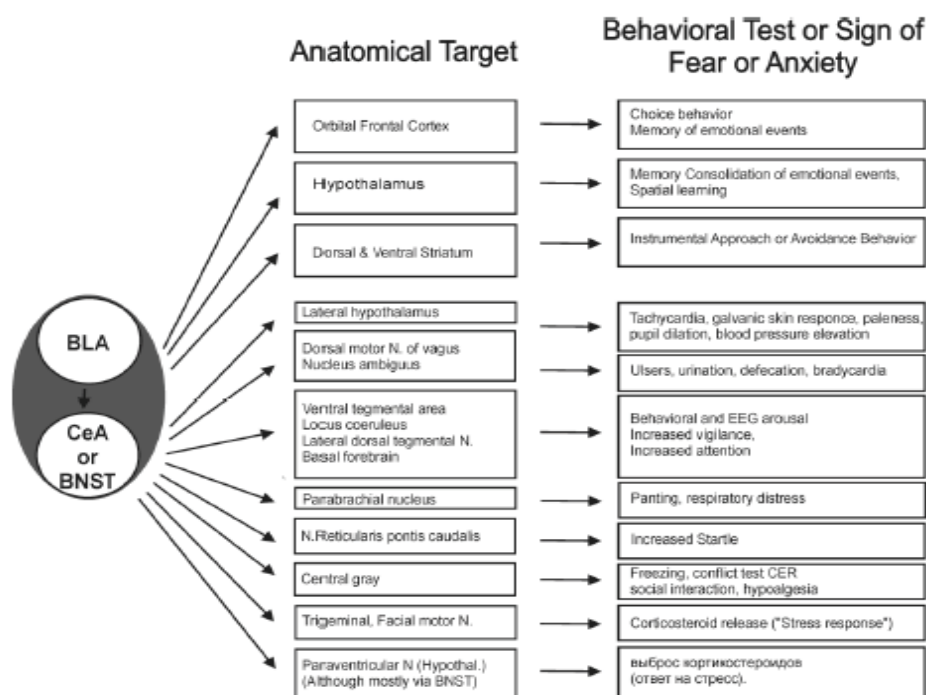
entorhinal cortex, and subiculum). A general scheme of these afferents and innervation peculiarities of the individual amygdala nuclei are shown in the Figure 2.



**Figure 2.** Summary of the inputs to the amygdalar nuclei of the basolateral (A), ventromedial (B) and cortical (B) cell groups. Compiled from the data: (Sah et al., 2003), source: (Tsvetkov i dr., 2015, p. 410). See table 1 for definition

Knowledge of anatomy and pathways is necessary to analyze the principles of the conditional and unconditional stimuli processing in the amygdala. These principles are best described in auditory inputs. Thus, LeDoux and co-authors (LeDoux, 1984; 1985) have shown that the main sources of the auditory afferents of amygdala (fig. 3).

It was established that the auditory information reaches the amygdala through two independent pathways. The first is a direct thalamo-amygdala pathway, and the second is an indirect thalamo-cortico-amygdala pathway (Romanski, 1992); they reach the amygdala through the external (Mascagni et al., 1993) and internal (Sah et al., 2003) capsule respectively and contain glutamatergic fibers. The thalamo-amygdala pathway brings information from the MGB substructures mentioned above, while the thalamo-cortico-amygdala pathway also includes the ventral (MGv) and dorsal (MGd) nuclei of the medial geniculate body and projects the information to amygdala through the auditory cortex (area Te3). Both thalamo-amygdala and thalamo-cortico-amygdala tracts terminate on the neurons of the same amygdala nuclei, more densely in the dorsal (LAdl) and medial (LAm) divisions of the lateral nucleus (LA) (LeDoux et al., 1991; McDonald, 1998; Roger and Arnault, 1989; Turner and Herkenham, 1991).



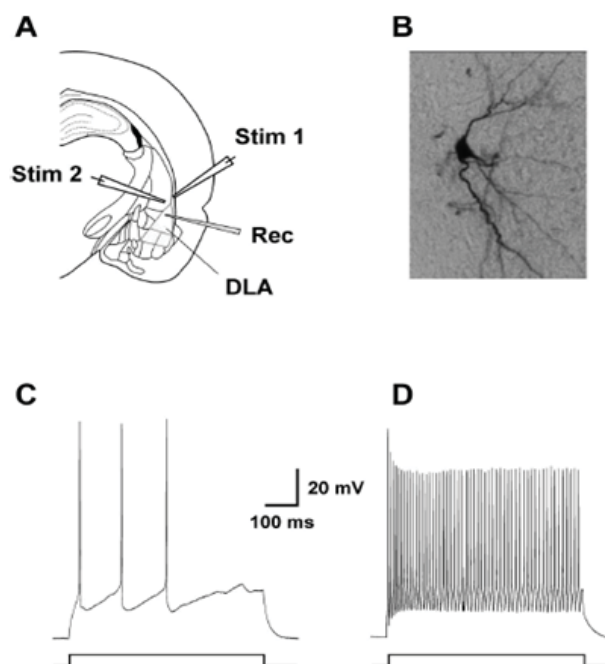
**Figure 3.** Summary of the main outputs from the cell groups of the amygdalar complex and the functions that regulate by these connections. Compiled from the data: (Davis, Whalen, 2001; Repucci, Petrovich, 2016), source: (Tsvetkov i dr., 2015, p. 413). See table 1 for definition

It should be noted that one of these two pathways is sufficient to form a conditioned fear response to a simple sound. Only if both tracts are destroyed, the fear conditioning cannot be produced (Romanski and LeDoux, 1992b). This illustrates that the input of sensory information into the amygdala through at least one of these pathways is necessary to form the conditioned fear (LeDoux et al., 1984; Romanski and LeDoux, 1992a; Romanski, 1992). The main difference between the thalamo-amygdala and the thalamo-cortico-amygdala tracts is that the first is a “quick response” channel, which provides faster, though less detailed information to the amygdala. On the contrary, the thalamo-cortico-amygdala path delivers highly detailed information (Jarrell et al., 1987), but much slower due to the several additional thalamo-cortical connections that are necessary for complex processing of sensory information in the cortex. In this way, two information channels to the amygdala complement each other.

Other types of sensory information travel to the amygdala the same way (Fig.2) (Sah et al., 2003). Thus, the efferents of supragenulate thalamic nucleus deliver to the lateral nucleus of amygdala auditory, somatosensory and visual signals (Aggleton, 1992; Campleu and Davis, 1995). Most information enters the amygdala not from the projection nuclei of the thalamus or the primary somatosensory cortex, but mainly from the thalamic structures of the extralemnisal system and the association areas of the cortex (Sah et al., 2003; Usunoff et al., 2006). The amygdala participates in the fear genesis because of the neurons of its nuclei receive information about both conditional and unconditional stimuli. The afferent pathways delivering the last type of information were scrutinized in the pain system. After delivering through the spinothalamic tracts to the posterior group of thalamus nuclei, nociceptive information is sent to the somatosensory and insular cortex. Then this information comes to amygdala from both the thalamus, in particular the posterior intralaminar nucleus–PIN (LeDoux et al., 1987), and indicated cortical fields (McDonald, 1998; Turner and Herkenham, 1991). It has been shown in the experiment that the destruction of the posterior intralaminar nucleus

(PIN) does not lead to inability of fear conditioning to the pain (LeDoux et al., 1987; Turner and Zimmer, 1984). Only if both PIN and insular cortex are destroyed, complete inability to form this kind of reaction occurs. The latter shows that the information about the conditional and unconditional stimuli comes to the amygdala through the two pathways which are direct thalamo-amygdala and in direct thalamo-cortico-amygdala.

The efferent system of amygdala provides the expression of fear, including its physiological, behavioral, and psychological components, due to the connections with the executive systems of the brain. The nuclei of basolateral complex (BLA) play a certain role in this process. Their direct connections with the dorsal striatum are thought to affect the behavioral strategy, the connections with the hippocampus relate to spatial memory, the connections with the orbitofrontal cortex influence the memory of emotional events (Fig.3) (Davis, 2000; 2001;). An important role in the mechanisms of fear is played also by the nuclei of CeA or BNST, which have numerous inputs from the basolateral complex (BLA) and efferents to a large number of brain structures mediating the fear expression (Davis, 1992; LeDoux et al., 1988; Pitkanen 2000; Sah et al., 2003). These nuclei mediate the processes in the BLA and the activity of the brain structures that cause fear specific behavioral and vegetative reactions. The connections of these nuclei and the effects of their chemical and electrical stimulation according to the M. Davis, is presented in the Figure 4 (Davis, 2000; 2001;). The CeA is believed to process more rapid reactions, which are classified as fear, whereas the BNST deals with longer ones, classified as anxiety states (Davis, 1997).



*Figure 4. Responses of the projection neurons of dorsolateral nucleus of amygdala. Compiled from the data: (Tsvetkov i dr., 2009; Tsvetkov i dr., 2015; Tsvetkov et al., 2002; Tsvetkov et al., 2004).*

*A: I—A slice of the brain at the level of the amygdala and the position of the electrodes (Stim 1 is the electrode stimulating afferents in the external capsule, Stim 2 is the electrode stimulating afferents in the internal capsule, p<sub>3</sub> is the recording electrode); B—Projection neurons in DLA. Horseradish peroxidase enzyme staining; C, D—responses of the projection neuron (A) and interneuron (D) recorded in the current clamp modes during the stimulation of cells by recording electrode. The recorded cell was identified as an interneuron (II) based on the nonaccommodating firing pattern in response to prolonged depolarizing current injection. Stim 1, 2—stimulating electrodes; Rec—recording electrode; 4—projection neurons; DLA—dorsolateral nucleus of amygdala.*

The similar mechanisms of information processing and fear expression are described in humans and, in particular, proved by studies performed *in vivo* with the use of Functional Magnetic Resonance Imaging (fMRI). These studies have shown that expression of fear caused by the presentation of genetically determined (innate) or conditional aversive stimuli (as a spider, snake, or pain) in humans, as well as in animals, leads to bilateral activation of the amygdala and associated structures involved in fear mechanisms (Delgado et al., 2006).

The similarity of the fear mechanisms in humans and animals was also confirmed by the analysis of the visual pathways to the amygdala. In particular, it has been shown with the MRI technologies that in humans, as in animals, visual information enters the amygdala both by fast (thalamo-amygdala) and slow (thalamo-cortico-amygdala) tracts (Sato et al., 2013). Some features of the fear mechanisms in humans have been revealed due to the researches of the interaction of amygdala with cortical structures. A remarkable feature of the human brain is the prefrontal cortex that enables developed cognitive abilities. This part of the brain phylogenetically develops first in primates (Salzman and Fusi, 2010), where it occupies an extensive rostral part of the hemispheres, and does not have any homologs in rodents (Salzman and Fusi, 2010; Batuev, 1987). The prefrontal cortex is usually divided into the dorsolateral, medial and orbitofrontal areas. The orbitofrontal area is closely connected with the amygdala nuclei, where in some of the cortico-amygdala efferents connect with intercalary neurons, that has not been classified into any nuclear structure (Ghashghaei et al., 2007). These neurons exert an inhibitory effect on the cells of centromedial group of nuclei (Ghashghaei and Barbas, 2002; Stefanacci and Amaral, 2002; Stefanacci and Amaral, 2000), which results in the ability of prefrontal cortex to both activate and inhibit the main efferent neurons of amygdala. The connections of the orbitofrontal cortex with amygdala and dorsolateral prefrontal cortex provide close interaction between cognitive and emotional spheres in humans, enabling us to understand and analyze fears (and other emotions), describe and classify them. Apparently, it is the prefrontal cortex network that allows people to experience anxiety and fear in response to physical and psychological stress, and to consciously control emotions and associated with them behavior.

Thus, according to modern data, the complex of amygdala nuclei described above is a central part of the fear mechanisms. The existence of such a center has been suggested by C. Izard, while I. P. Pavlov described it as an element of unconditioned reflex. The center receives stimuli from outer and inner worlds and then activates systems of the brain involved in the fear mechanisms. This process is experienced as a specific feeling associated with somatic and visceral reactions, whereas at the consciousness level it is interpreted either as fear and anxiety, or as a complex of emotions involving them. Since the cause of fear can be both physical or psychological, this scheme supports the cognitive theory and views it as an example of fear being a person's reaction to themselves.

The synaptic plasticity mechanisms were first studied in the early 1970s by T.B. Bliss and T. Lomo who discovered that the exciting synapses in three-synaptic hippocampus pathway show the facilitation ability (Bliss and Lomo, 1973). The latter occurs in response to high-frequency stimulation of presynaptic fibers and can last for hours or days. Due to its length, the authors called this process a long-term potentiation (LTP). The works of T. Bliss and T. Lomo aroused interest of the scientific community, and the plasticity they discovered almost immediately became the main possible neural mechanism of memory and conditioned reflex as a form of learning. The method of surviving brain slices made significant progress in the neural mechanisms of LTP. Because of this method, scientists were able to investigate any synapses

in the brain and develop stimulation protocols for the artificial initiation of LTP. This approach gave the opportunity to investigate biochemical and molecular mechanisms of the synaptic plasticity not only in the hippocampus (Kandel, 1997; Milner et al., 1998; Malenka and Nicoll, 1999; Izquierdo et al., 2008; Lu et al., 1997; Bliss and Collingridge 1993; Collingridge et al., 1983), but also in other brain structures. In particular, the LTP ability has been shown in synapses in various cortex areas (Hirsch and Crepel, 1990; Jay et al., 1996, Heynen and Bear, 2001; Herry and Garcia, 2002), cerebellum (Losi et al., 2002), and amygdala (Huang et al., 2000; Huang and Kandel, 1998; Schafe et al, 2000; Brambilla et al., 1997; Blair et al., 2001).

According to the studies, the LTP can be caused by changes of both presynaptic and postsynaptic membranes. At the presynaptic level, this process presents a change in the mediator secretion, while at the postsynaptic level it changes the membrane sensitivity to the mediator. Moreover, it has been shown that there are several forms of the long-term plasticity. Among them are a LTP induced by modulating neurons described by E. Kandel in the nervous system of mollusks (Kandel, 2000), an NMDA receptor-independent LTP found in certain synapses in the hippocampus (Zalutsky and Nicoll, 1990; 1991), and NMDA receptor-dependent LTP described by T. V. Bliss and T. Lomo (Bliss and Lomo, 1973).

NMDA receptor-dependent is the most common form of LTP (Maren, 2001; Lynch, 2004) that can be seen in synapses of many brain structures, including hippocampus and amygdala. Despite the variability in different structures, the main mechanism of this LTP form is similar in many synapses. The latter allows to describe biochemical and molecular mechanisms of LTP on the basis of the study of synapses in the CA1 field of hippocampus, where this phenomenon was originally discovered and scrutinized.

The process of the LTP development can be divided into two phases (Frey et al., 1993; Schafe et al., 2001; Izquierdo et al., 2008). The first phase lasts 10–20 minutes and does not include synthesis of the protein and mRNA (Schafe et al., 2001). The second phase lasts from tens of minutes to several days, weeks or even months, requires the protein and mRNA synthesis, and results in the structural changes of the synapses (Schafe et al., 2001; Huang et al., 2000; Huang et al., 1994; Nguyen and Kandel, 1996; Nguyen et al., 1994). The length and biochemistry of these phases are comparable to the short-term and long-term phases of memory which suggests that the LTP mechanism can underlie not only the molecular mechanism of associative connections, but also memory and learning in their broad sense.

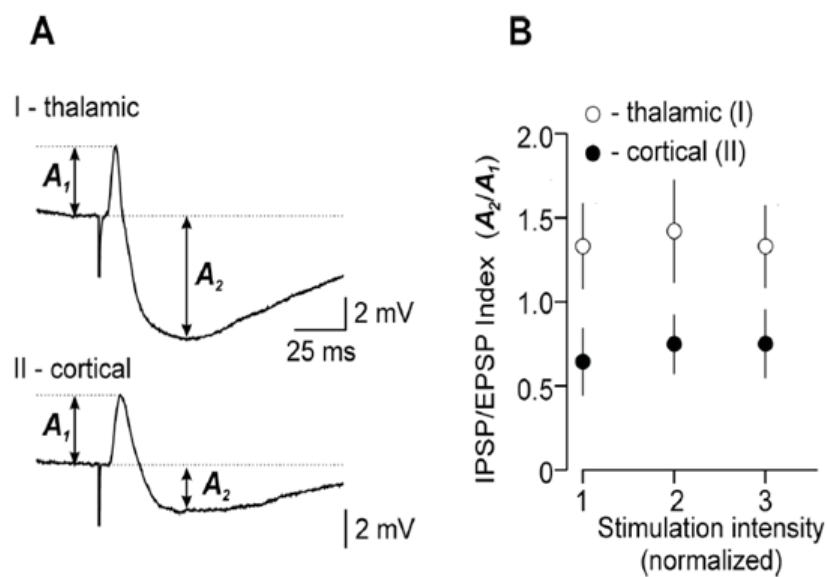
The early phase of LTP of the synapses in the CA1 field of hippocampus begins immediately when induced by any of the stimulation protocols. Most protocols induce both the electrical stimulation of presynaptic fibers and the depolarization of postsynaptic cells. The depolarization includes either tetanization (high-frequency stimulation) of the fibers, or direct cell depolarization through the recording electrode. At the first moment after the LTP initiation, presynapses and postsynaptic cells show a significant increase in the concentration of calcium ions, followed by the massive activation of protein kinases. The source of calcium influx can be either NMDA receptors (Collingridge et al., 1983; Malenka and Nicoll, 1993; Nguyen and Kandel, 1996; Huang and Kandel, 1998), or potential-activated L-type calcium channels (Weisskopf et al. 1999; Grover and Teyler, 1990; Huang and Malenka, 1993), or both (Maggie and Johnston, 1997), depending on stimulation protocol. The calcium can also enter the cell through the specific type of AMPA receptors because their structure does not contain GluR2 subunit. (Geiger et al., 1995) The glutamatergic NMDA and AMPA receptors can control calcium influx at the postsynaptic level, whereas potential dependent  $\text{Ca}^{2+}$  channels of L-type modulate both membranes. A change in the concentration of calcium ions acti-

vates a big amount of protein kinases (Huang and Kandel, 1994; Malinow et al., 1989) at the pre- and postsynaptic levels; their activity maintains the phosphorylation at both levels. As an example, phosphorylation of AMPAR subunits is induced at the postsynaptic level, and phosphorylation of the presynaptic protein GAP-43 activates at presynaptic level. The trigger of the first process is the autophosphorylation of the  $\alpha$ CaMKII (Bevilaqua et al., 2005) associated with NMDARs which leads to higher conductance of postsynaptic AMPA channels (Soderling and Derkach, 2000). The second process is associated with the activity of  $\gamma$ PKC (Routtenberg, 2000) which mobilizes glutamate-containing synaptic vesicles (Camarota et al., 1998; Routtenberg, 2000). All these processes depend on LTP, so they are supposed to be parts of a complex biochemical mechanism of short-term synaptic plasticity, and a base for some forms of short-term memory (Soderling and Derkach, 2000; Sweatt, 1999).

The late phase of LTP involves RNA and protein synthesis and can last for days, weeks or even months (Huang et al., 2000; Huang et al., 1994; Nguyen and Kandel, 1996; Nguyen et al., 1994). The protein synthesis is mediated by the extracellular signal-regulated kinases (ERK) related to the type of mitogen-activated protein kinases (MAPK) (Roberson and Sweatt 1996; English and Sweatt 1996), and cAMP-dependent protein kinases, including type A kinase (PKA-cAMP dependent protein kinase). These, and some other kinases, can be active at the both pre- and post-synaptic levels (see review: Lynch, 2004; Orsini and Maren, 2012). Being activated, phosphorylated ERK/MAPK and/or catalytic subunits of PKA are translocated to the cell nucleus where they activate the transcription initiators (Alberini et al., 1995; Bacsikai et al., 1993; Martin et al., 1997) such as the CREB transcription factor (cAMP response-element binding protein). After phosphorylation, this factor can activate CRE-dependent genes responsible for the structural changes in synapses (Kandel, 1997; Milner et al., 1998; Malenka and Nicoll, 1999; Silva, et al. 1998). In experiments *in vitro*, it has been shown that addition of PKA inhibitors or ERK/MAPK stops the induction of the late stage of LTP completely (Huang et al., 2000; Huang et al., 1994; Nguyen and Kandel, 1996; Nguyen et al., 1994; Schafe et al., 2000; English and Sweatt, 1997; Atkins et al., 1998), while activation of adenylate cyclase by the cAMP- agonist forskolin has the opposite effect (Nguyen and Kandel, 1996). The processes induced by the activation of genes at the late stage of LTP are also blocked by PKA inhibitors and ERK/MAPK (Impey et al., 1996; Impey et al., 1998), though can be activated by high-frequency electric stimulation of presynaptic fibers or by forskolin (Huang et al., 2000; Davis et al., 2000). Nowadays, investigation of the LTP mechanisms of synapses in the brain and spinal cord continues to this day. Nevertheless, there are numerous reviews of the available data (Lynch, 2004; Maren, 2001; Martin et al., 2000; Orsini and Maren, 2012). In our investigations behavioral features of animals experiencing negative emotions were studied in model experiments, producing a conditioned fear reflex. Cellular mechanisms of memory of negative stimuli were investigated on brain slices (fig. 4).

The substrate for the formation of conditioned reflex reactions to aversive stimuli provides a convergence of sensory inputs of basolateral amygdala (BLA) neurons. The mechanism of the realization of conditioned reflexes is provided by the multimodal nature of afferents and synaptic plasticity. This is demonstrated by the responses of the projection neurons of the amygdala caused by the stimulation of the cortico- and/or thalamo-amygdala fibers. Such responses have a combined nature and include monosynaptic excitatory glutamatergic AMPA and NMDA components and a disynaptic inhibitory GABAergic component. At the same time, the areas of stimulation of cortico- and thalamo-amygdala fibers are independent and do not overlap. The electrophysiological parameters of monosynaptic AMPA and NMDA components for the stimulation of thalamic and cortical afferents do not differ. The AMPA

components in both cases are mediated by the receptors that contain the GluR2 subunit. The NMDA component is more complicated. It contains the currents mediated by the receptors with NR2A and NR2B subunits with an approximate ratio of 1:3, respectively. The amplitude of the disynaptic GABAergic component of the response during the stimulation of thalamic fibers is significantly higher than the amplitude of the response provoked by the stimulation of the cortical fibers. The analysis of the provoked unitary miniature currents showed that an increase in the effectiveness of synaptic transmission in the thalamic synapses was associated with a higher amplitude of one-quantum response of interneurons. At the same time, the current-voltage characteristic of the GABAergic component of the response of projection neurons during the activation of interneurons via the thalamic input and probable parameters of GABA release ( $PPF_{70}$ ) do not differ from those observed during the stimulation of cortical fibers (fig. 5).



**Figure 5.** The AMPA and GABA components of biphasic response of projection neurons elicited by (upon) stimulation of cortical and thalamic inputs. Compiled from the data: (Shin et al., 2006).

*A*—the biphasic response of projection neurons elicited by (upon) stimulation of cortical (I) and thalamic (II) inputs:  $A_1$ —amplitude of AMPA component,  $A_2$ —amplitude of GABA component; *B*—ratio of amplitude of AMPA and GABA components of biphasic response at different intensity of stimulation of cortical and thalamic fibers.

Long-term potentiation (LTP), i. e., enhanced activity of the synaptic transmission that remains for a long time, underlies the processes of learning and memory. The cortical and thalamic synapses on the projection cells of LA have the ability for LTP. Its development is explained by the fact that during the stimulation of afferents, the probability of the mediator release from the presynaptic vesicles increases. The concentration of ions of intracellular calcium, required for the initiation of LTP, critically depends on the polarization of the postsynaptic neuron and is provided by the activity of potential-controlled calcium channels of L-type and NMDA receptors of the NR2B group. The activation of other types of calcium-transmitting ligand-dependent channels, which include NMDA receptors of the group NR2A and inositol phosphate ( $IP_3$ ) and ryanodine receptors, is not a factor of initialization of LTP. AMPA receptors, associated with GluR2 subunits, provide the depolarization of the postsynaptic membrane only at the first stage of glutamate release. One of the most important properties of conditioned reflex

reactions is its stimulus specificity, i.e., the recognition of conditioned stimuli associated with aversive stimulus before their manifestations. LTP of the thalamic and cortical synapses on the living brain slides also shows specificity to the stimulus and develops only on the stimulated input. The resources of regulation of the excitatory inputs of the amygdala are inhibitory GABAergic interneurons. These cells, as well as projection cells, receive inputs from the cortex and the thalamus but the highest effectiveness is observed in the thalamic synapses. Such asymmetry enhances the shunting of the projection cell membranes and decreases the level of LTP during the stimulation of the thalamic input (vs cortical). Molecular mechanisms of LTP are verified by the pharmacological blockade of the receptors and channels. Thus, LTP is a mechanism of plastic changes in the amygdala synapses that provide the formation of conditioned reflex amygdala-dependent reactions (fig. 6).

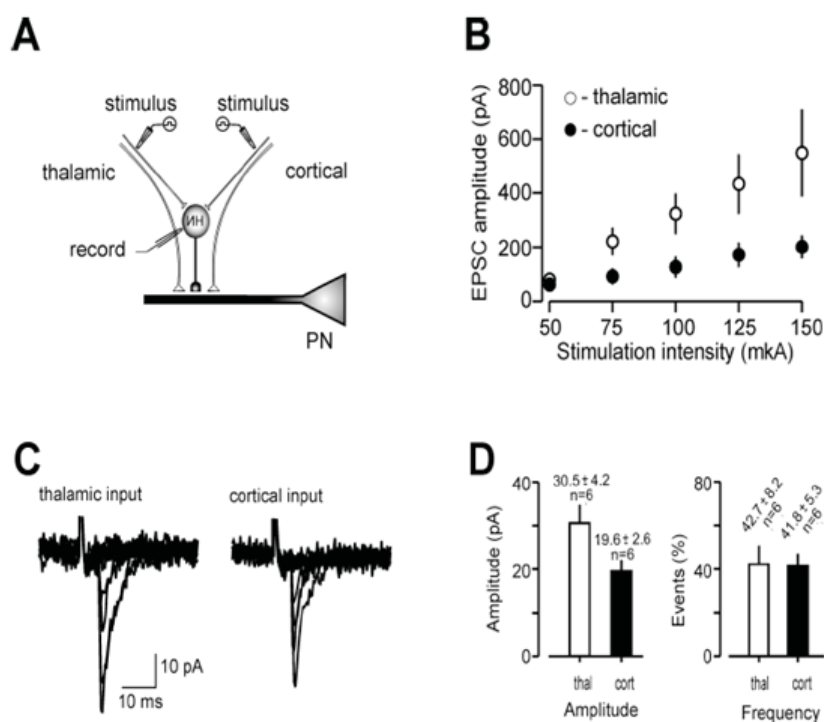
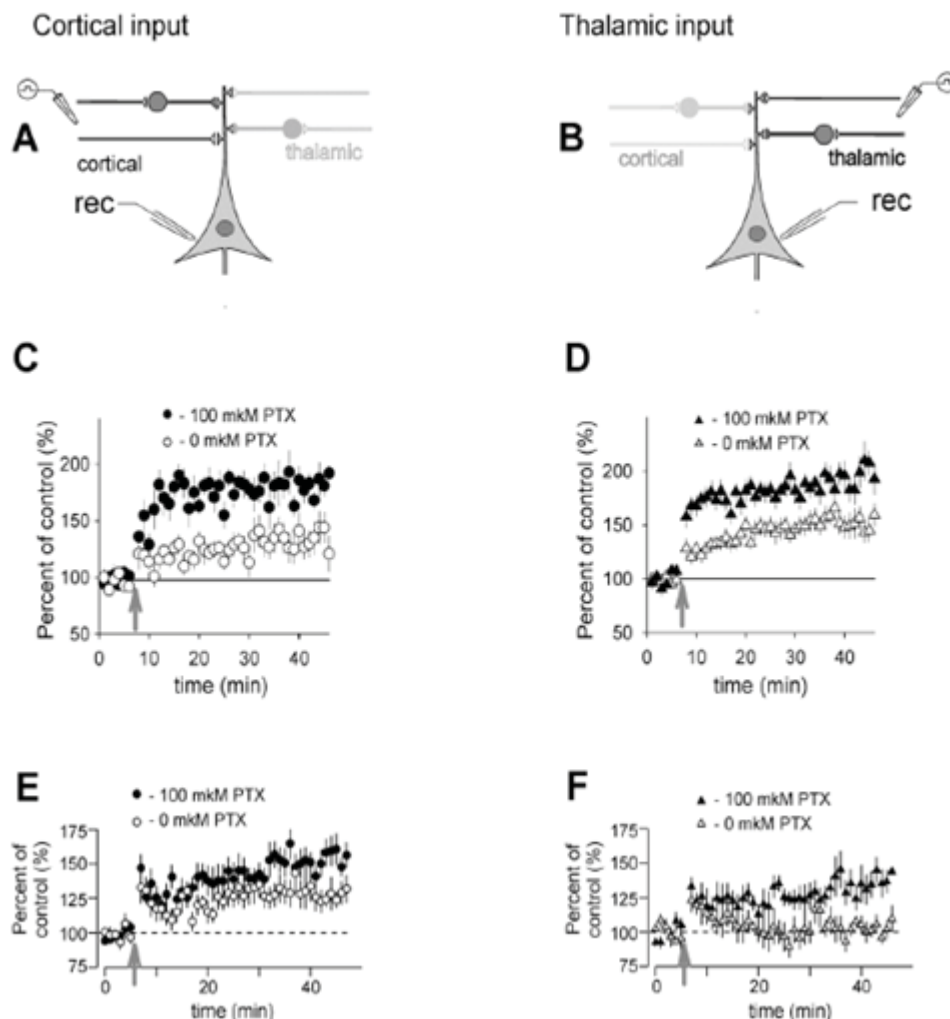


Figure 6. Electrophysiological characteristics of cortico- and thalamo-amygdala synapses on interneurons. Compiled from data: (Shin et al., 2006).

A—a schematic representation of the experimental design when cortical (black) and thalamic (white) inputs to the same interneuron were alternately activated. B—the recorded cell was identified as an interneuron based on the non accommodating firing pattern in response to prolonged depolarizing current injection (bottom trace—supra threshold stimulus); amplitude of AMPA component EPSC recorded in interneurons at convergent cortical and thalamic pathways; C—examples of the AMPA component EPSC pairs induced with a 50 ms interstimulus interval in convergent cortical (I) and thalamic (II) inputs; D—summary plot of mean amplitude and frequency estimates for unitary EPSCs in cortical and thalamic inputs to interneurons.

As a mechanism of plastic changes in the synapses, LTP provides the formation of conditioned reflex amygdala-dependent reactions. The early and late stages of LTP of the synapses of projection neurons in the living brain slides correspond to two types of memory: short-term and long-term. The formation of a conditioned reflex *in vivo* is based on the short-term memory about the coincidence of a conditioned signal and unconditioned traumatizing stimulus, and the consolidation—on the long-term memory about it. The structural changes associated with the development of LTP are the results of complicated biochemical reactions that are induced by  $\text{Ca}^{2+}$  ions that penetrate the cells after the activation of postsynaptic NMDA receptors.

Antagonists of NMDA receptors AP5 and MK-801 that prevent the initiation and development of LTP *in vitro*, inhibit the development of conditioned reflex reaction of fear *in vivo*. The overexpression of these receptors facilitates the amygdala-dependent tasks and development of LTP on living brain sections. Although the similarity of artificially induced LTP and conditioned reflex learning is indirect, the facilitation and habituation are the simplest forms of learning associated with electrophysiological and biochemical changes in the presynaptic and postsynaptic links of a reflex arch. Three main features of LTP (specificity, associativity, and selectivity) indicate that the formation of a conditioned reflex, one of the forms of associative learning, is provided by the potentiation of the synaptic transmission (fig. 7).



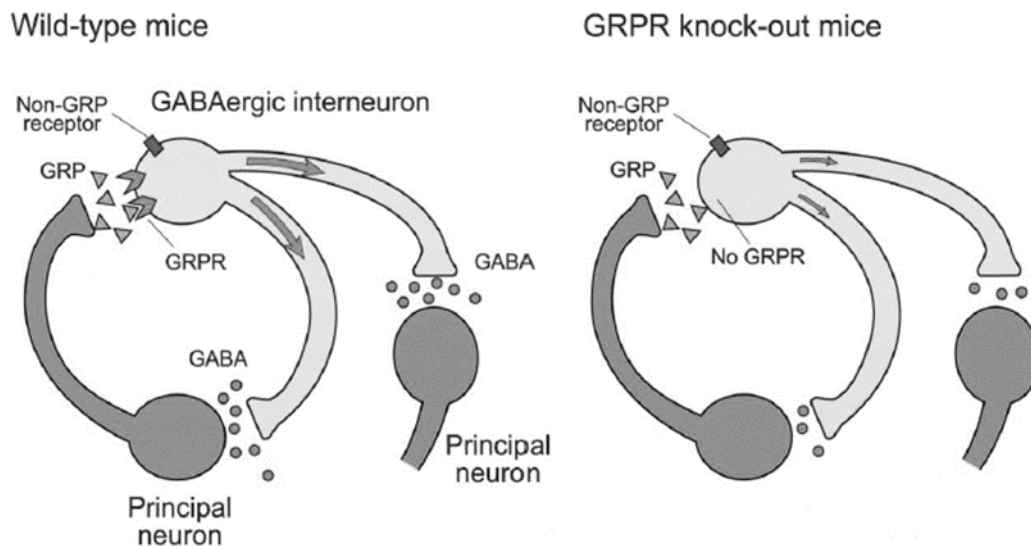
**Figure 7.** The effects of GABAergic receptors on long-term potentiation of cortical amygdala synapses and thalamic amygdala synapses initiated under LTP1 protocol.

Adapted from: (Tsvetkov et al., 2011; Shin et al., 2006).

A–B: diagram of stimulation of cortical amygdala fibers (A) and thalamic amygdala fibers (B); C, E: LTP1 protocol long-term potentiation of cortical amygdala synapses (C) and thalamic amygdala synapses (D) with/without picrotoxin (PTX) in the superfusion solution. The arrow shows the moment of stimulation initiation; D, F: LTP35–40 indices of the cortical input (D) and thalamic input (F) with PTX in the superfusion solution (1) and without it (2).

The association between the formation of conditioned reflex reactions and LTP was studied on animals with genetic proteins modifications that modulate synaptic transmission. The studies on the role of gastrin-releasing peptide (GRP) and its receptor (GRPR) showed that the activa-

tion of receptors led to an increase in the rate of inhibitory neurotransmission in the projection neuron synapses in the control animals and did not cause such an effect in GRPR knockout animals. The pharmacological or genetic inhibition of GRPR led to the facilitation of potentiation and growth of LTP in the living brain slices of experimental animals in comparison with the control. In GRPR knockout mice, the rate of the formation of a conditioned reflex of fear and the context-dependent reaction was higher in comparison with control animals (fig. 8). At the behavioral level, this is manifested as increased anxiety and easy formation of a conditioned reflex reaction to fear. Nociception and the level of anxiety in experimental animals did not differ in comparison with the control. Stathmin (oncoprotein-18) takes part in the regulation of the plasticity of cortico- and thalamo-amygdala synapses of the projection cells of the amygdala. In stathmin knockout animals, a deficit of LTP is observed. At the behavioral level, this is expressed as a decrease in the level of anxiety and ability of animals to the formation of conditioned reflex reactions of fear. At the same time, a decrease in the sensitivity of pain or locomotor activity in these animals is not observed.



**Figure 8.** A model of GRP-dependent negative feedback to lateral amygdala principal neurons.

*Adapted from: (Shumyatsky et al., 2002).*

A: controls; B: GRPR knockout animals.

In conclusion, it is necessary to emphasize the main points of the analysis of our own data and the literature on the neural mechanisms of emotional memory. Two types of amygdala cells, glutamatergic projective and GABAergic interneurons, are under the synaptic influence of the thalamus and cortex. While the synapses on the projective cells are equally effective, the thalamic synapses are more effective on the interneurons. Responses of projection neurons to stimulation of cortical and thalamic afferents include excitatory monosynaptic glutamatergic AMPA and NMDA components and disynaptic inhibitory GABAergic component. The most important resource of excitatory efferent regulation of the amygdala is inhibitory GABAergic interneurons. When the thalamic input is stimulated, interneuron synapses shunt the membrane of projection cells more effectively and decrease the level of LTP than when the cortical one is stimulated. The stages of the development of LTP are similar to those of short-term

and long-term memory. The formation of the conditioned reflex is based on the short-term memory of the coincidence of the conditioned signal with the unconditioned stimulus, and the consolidation is based on the long-term memory of it. The relationship between conditioned reflexes and LTP is confirmed by model experiments with manipulation of genes that encode proteins regulating synaptic transmission and its plasticity. In animals knocked out by the gene for gastrin-releasing peptide (GRP), it is easier to initiate LTP and to form conditioned fear reflex responses. Knockout of the oncoprotein18/statmin gene, leads to a deficiency of LTP and impaired production of amygdala-dependent conditioned reflexes. An analysis of the world literature and the authors' own studies supplement Cannon-Bard's theory of emotions with evidence that the key role in emotion regulation belongs to the amygdala complex, which presents the neural mechanisms of stimulus filtering in relation to stimulus relevance, learning, and stimulus-mediated memory formation. The interdependence of plastic changes in the synapses and the success in producing conditioned-reflex reactions allow us to consider the amygdala as a learning filter, which blocks trivial information, but lets in stimuli associated with emotionally significant signals and forming a memory about them. The efficiency of such a filter depends on the emotional memory of the conditioned stimulus. If the sensory signal is familiar, easily recognized, has an adequate intensity, and is not burdened with aversion, as in the above experiments, postsynaptic potentials are easily bypassed by inhibitory interneurons, and the excitation level of projective neurons is insufficient for generating an action potential. As a result, a stimulus with no aversive load is ignored. If, however, the stimulus is associated with danger based on previous experience, the probability of generating a LTP increases. It is possible that such an "amygdala filter" provides generation of not only negative ones, but also positive stimulus-dependent emotions. According to Cannon-Bard theory, the intermediate brain is the subcortical center of emotions: through the thalamus the ascending flow of sensory impulses is directed to the cortex, providing the conscious experience of emotions; hypothalamic neural networks regulate their visceral component. At the same time, the neurohumoral processes controlled by the hypothalamus and accompanying emotional states, are universal, do not depend on the sign of emotions and do not outrun them. The analysis of the world literature and the authors' own studies do not contradict the Cannon-Bard theory concept and supplement it with evidence that the key role in emotion regulation belongs to the amygdalar complex, which represents neural mechanisms of stimulus filtering depending on its relevance, learning and stimulus-mediated memory formation.

## References

1. Adelman P. K., Zajonc R. B. Facial efference and the experience of emotion // *Annu Rev Psychol.* 1989. Vol. 40. P. 249–280.
2. Adesnik H., Nicoll R. A. Conservation of glutamate receptor 2-containing AMPA receptors during long-term potentiation // *Journal of Neuroscience.* 2007. Vol. 27. P. 4598–4602.
3. Aggleton J. P. The Amygdala: neurobiological aspects of emotion, memory, and mental dysfunction. 1992. New York: Wiley-Liss. Xii. 615 p.
4. Alberini C. M., Ghirardi M., Huang Y. Y., Nguyen P. V., Kandel E. R. A molecular switch for the consolidation of long-term memory: cAMP-inducible gene expression // *Annals of the New York Academy of Sciences.* 1995. Vol. 758. P. 261–286.
5. Batuev A. S. Higher Integrative Systems of the Brain. 1987, New York: Gordon and Breach Science Publisher. P. 296.
6. Bekhtereva N. P., Smirnov V. M. Brain organization of human emotions // *Bulletin of the Academy of Medical Sciences of the USSR.* 1975. Vol. 8. P. 8–19.
7. Belekova M. G., Chkheidze D. D., Veselkin N. P., Kenigfest N. B., Kratskin I. L., P'err Z., Reperan Z. Distribution of GABA-immunoreactive elements in the reptile amygdaloid

- complex // *Neirofiziologiya*. 1992. Vol. 1. P. 68–77.
8. LeDoux, 2000–LeDoux J. E. Emotion circuits in the brain // *Annu. Rev. Neurosci.* 2000. Vol. 23. P. 155–184.
  9. LeDoux et al., 1984–LeDoux J. E., Sakaguchi A., Reis D. J. Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned to acoustic stimuli // *J. Neurosci.* 1984. Vol. 4, no. 3. P. 683–698.
  10. LeDoux et al., 1985–LeDoux J. E., Ruggiero D. A., Reis D. J. Projections to the subcortical forebrain from anatomically defined regions of the medial geniculate body in the rat // *J. Comp. Neurol.* 1985. Vol. 242, no. 2. P. 182–213.
  11. LeDoux et al., 1987–LeDoux J. E., Ruggiero D. A., Forest R., Stornetta R., Reis D. J. Topographic organization of convergent projections to the thalamus from the inferior colliculus and spinal cord in the rat // *J. Comp. Neurol.* 1987. Vol. 264, no. 1. P. 123–146.
  12. LeDoux et al., 1988–LeDoux J. E., Iwata J., Cicchetti P., Reis D. J. Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear // *J. Neurosci.* 1988. Vol. 8, no. 7. P. 2517–2529.
  13. LeDoux et al., 1991–LeDoux J. E., Farb C. R., Romanski L. M. Overlapping projections to the amygdala and striatum from auditory processing areas of the thalamus and cortex // *Neurosci. Lett.* 1991. Vol. 134, no. 1. P. 139–144.
  14. Lee et al., 1980–Lee K. S., Schottler F., Oliver M., Lynch G. Brief bursts of high-frequency stimulation produce two types of structural change in rat hippocampus // *J. Neurophysiol.* 1980. Vol. 44, no. 2. P. 247–258.
  15. Lee et al., 1999–Lee K., Dixon A. K., Gonzalez I., Stevens E. B., McNulty S., Oles R., Richardson P. J., Pinnock R. D., Singh L. Bombesin-like peptides depolarize rat hippocampal interneurons through interaction with subtype 2 bombesin receptors // *J. Physiol.* 1999. Vol. 518, part 3. P. 791–802.
  16. Mahanty, Sah, 1998–Mahanty N. K., Sah P. Calcium-permeable AMPA receptors mediate long-term potentiation in interneurons in the amygdala // *Nature*. 1998. Vol. 394, no. 6694. P. 683–687.
  17. Mahanty, Sah, 1999–Mahanty N. K., Sah P. Excitatory synaptic inputs to pyramidal neurons of the lateral amygdala // *Eur. J. Neurosci.* 1999. Vol. 11, no. 4. P. 1217–1222.
  18. Malenka, Nicoll, 1993–Malenka R. C., Nicoll R. A. NMDA-receptor-dependent synaptic plasticity: Multiple forms and mechanisms // *Trends Neurosci.* 1993. Vol. 12. P. 521–527.
  19. Malenka, Nicoll, 1999–Malenka R. C., Nicoll R. A. Long-term potentiation—a decade of progress? // *Science*. 1999. Vol. 285, no. 5435. P. 1870–1874.
  20. Nicoll, 2017–Nicoll R. A. A brief history of long-term potentiation // *Neuron*. 2017. Vol. 93, no. 2. P. 281–290.
  21. Nishiyama et al., 2000–Nishiyama M., Hong K., Mikoshiba K., Poo M. M., Kato K. Calcium stores regulate the polarity and input specificity of synaptic modification // *Nature*. 2000. Vol. 408, no. 6812. P. 584–588.
  22. Pin, Duvoisin, 1995–Pin J. P., Duvoisin R. The metabotropic glutamate receptors: structure and functions // *Neuropharmacol.* 1995. Vol. 34, no. 1. P. 1–26.
  23. Pitkanen, 2000–Pitkanen A. Connectivity of the rat amygdaloid complex // *The Amygdala* J. P. Aggleton (ed.). New York: Oxford University Press, 2000. P. 31–115.
  24. Pitkanen et al., 1997–Pitkanen A., Savander V., LeDoux J. E. Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala // *Trends Neurosci.* 1997. Vol. 20, no. 11. P. 517–523.
  25. Plant K., Pelkey K. A., Bortolotto Z. A., Morita D., Terashima A., McBain C. J., Collingridge G. L., Isaac J. T. Transient incorporation of native GluR2-lacking AMPA receptors during hippocampal long-term potentiation // *Nat. Neurosci.* 2006. Vol. 9, no. 5. P. 602–604.
  26. Price et al., 1987–Price J. L., Russchen F. T., Amaral D. G. The limbic region. II. The

- amygdaloid complex // Handbook of chemical neuroanatomy / T. Hokfelt, A. Bjorklund, L. W. Swanson (eds.). Amsterdam: Elsevier, 1987. P. 279–388.
27. Puellas, 2001–Puelles L. Thoughts on the development, structure and evolution of the mammalian and avian telencephalic pallium // Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2001. Vol. 356, no. 1414. P. 1583–1598.
  28. Quirk et al., 1995–Quirk G.J., Repa C., LeDoux J.E. Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat // Neuron. 1995. Vol. 15, no. 5. P. 1029–1039.
  29. Sah P. Fear, Anxiety, and the Amygdala // Neuron. 2017. V. 96. P. 1–2.
  30. Sah P., Faber E.S., Lopez D.A., Power J. The amygdaloid complex: anatomy and physiology // Physiol. Rev. 2003. V. 3. P. 803–834.
  31. Schafe et al., 2001–Schafe G.E., Nader K., Blair H.T., LeDoux J.E. Memory consolidation of Pavlovian fear conditioning: a cellular and molecular perspective // Trends Neurosci. 2001. Vol. 24, no. 9. P. 540–546.
  32. Schenberg, Oliveira, 2008–Schenberg E.E., Oliveira M.G. Effects of pre- or posttraining dorsal hippocampus D-AP5 injection on fear conditioning to tone, background, and foreground context // Hippocampus. 2008. Vol. 18, no. 11. P. 1089–1093.
  33. Schienle et al., 2007–Schienle A., Schafer A., Hermann A., Rohrmann S., Vaitl D. Symptom provocation and reduction in patients suffering from spider phobia: an fMRI study on exposure therapy // Eur. Arch. Psychiatry Clin. Neurosci. 2007. Vol. 257, no. 8. P. 486–493.
  34. Selden et al., 1991–Selden N.R., Everitt B.J., Jarrard L.E., Robbins T.W. Complementary roles for the amygdala and hippocampus in aversive conditioning to explicit and contextual cues // Neurosci. 1991. Vol. 42, no. 2. P. 335–350.
  35. Shin R.M., Tsvetkov E., Bolshakov V.Y. Spatiotemporal asymmetry of associative synaptic plasticity in fear conditioning pathways // Neuron. 2006. V. 5. P. 883–896.
  36. Shumyatsky G.P., Malleret G., Shin R.M., Takizawa S., Tully K., Tsvetkov E., Zakharenko S.S., Joseph J., Vronskaya S., Yin D., Schubart U.K., Kandel E.R., Bolshakov V.Y. *stathmin*, a gene enriched in the amygdala, controls both learned and innate fear // Cell. 2005. V. 4. P. 697–709.
  37. Shumyatsky G.P., Tsvetkov E., Malleret G., Vronskaya S., Hatton M., Hampton L., Battey J.F., Dulac C., Kandel E.R., Bolshakov V.Y. Identification of a signaling network in lateral nucleus of amygdala important for inhibiting memory specifically related to learned fear // Cell. 2002. V. 6. P. 905–918.
  38. Tsvetkov E.A., Carlezon W.A., Benes F.M., Kandel E.R., Bolshakov V.Y. Fear conditioning occludes LTP-induced presynaptic enhancement of synaptic transmission in the cortical pathway to the lateral amygdala // Neuron. 2002. V. 2. P. 289–300.
  39. Tsvetkov E.A., Krasnoshchekova E.I., Vesselkin N.P., Kharazova A.D. Amygdala: neuroanatomy and neurophysiology of fear // Zh. EV. Biokhim. Fiziol. 2015. V. 6. P. 406–418.
  40. Tsvetkov E.A., Masalov I.S., Veselkin N.P. The role of inhibitory interneurons in mechanisms of regulation of sensory synapses formed by thalamic and cortical inputs in pyramidal cells of the dorsolateral amygdala nucleus // Zh. EV. Biokhim. Fiziol. 2009. V. 4. P. 403–411.
  41. Tsvetkov E.A., Shin R.M., Bolshakov V.Y. Glutamate uptake determines pathway specificity of long-term potentiation in the neural circuitry of fear conditioning // Neuron. 2004. V. 1. P. 139–151.
  42. Tsvetkov E.A., Suderevskaia E.I., Veselkin N.P. Role of long-term potentiation in mechanism of the conditioned learning // Zh. EV. Biokhim. Fiziol. 2011. V. 3. P. 185–192.
  43. Tully K., Li Y., Tsvetkov E., Bolshakov V.Y. Norepinephrine enables the induction of associative long-term potentiation at thalamo-amygdala synapses // Proc. Natl Acad. Sci. USA. 2007. V. 35. P. 14146–14150.
  44. Zamanillo et al., 1999–Zamanillo D., Sprengel R., Hvalby O., Jensen V., Burnashev N., Rozov A., Kaiser K.M., Koster H.J., Borchardt T., Worley P., Lubke J., Frotscher M., Kelly

45. P. H., Sommer B., Andersen P., Seeburg P.H., Sakmann B. Importance of AMPA receptors for hippocampal synaptic plasticity but not for spatial learning // *Science*. 1999. Vol. 284, no. 5421. P. 1805–1811.
46. Zola-Morgan et al., 1991–Zola-Morgan S., Squire L.R., Alvarez-Royo P., Clower R.P. Independence
47. of memory functions and emotional behavior: separate contributions of the hippocampal formation and the amygdala // *Hippocampus*. 1991. Vol. 1, no. 2. P. 207–220.
48. Zucker, 1989–Zucker R. S. Short-term synaptic plasticity // *Annu. Rev. Neurosci.* 1989. Vol. 12, no. 1. P. 13–31.

## **Chapter 18. Analysis of the spinal neuronal populations in condition of high background neuropile staining using the “Cell Annotation Software”**

*Veshchitskii A. A., Mikhalkin A. A., Merkulyeva N.S*

*Pavlov Institute of Physiology Russian Academy of Sciences, Saint Petersburg, Russia*

The spinal cord is the most elongated structure of the central nervous system, having a non-homogeneous anatomical and physiological feature along its rostrocaudal axis. Thus, the morphological data of neuronal populations in the spinal regions of interest requires an analysis of a large array of slices often from a significant number of segments. These facts demand searching for convenient approaches to automated determination of the neuronal populations. Some software applications can easily realize this task for the well-visualized individual neurons located on a light background. However, many histological protocols led to the labeling for both neuronal somas and neuropile, which significantly complicates an automatic recognition of individual cells. To solve this problem, we tested the algorithm for image processing “Statistical Dominance Algorithm” as part of the program “Cell Annotation Software”, developed specifically for the analysis of brain morphological images, on some “problematic” spinal neuronal populations identified by various molecular markers. These populations were: (1) densely packed calbindin-immunopositive neurons in laminae II, (2) parvalbumin-immunopositive cells within the Clarke’s nuclei masked by the dark neuropile staining, and (3) calretinin immunopositive neurons within the intermediolateral nucleus having a weakly delineating nuclear border. All these areas are characterized by the low level of the signal (labeled neuron) to noise (labeled neuropile) ratio. Among all regions of interest, the “Statistical Dominance Algorithm” was most successful in recognizing the calbindin-immunopositive neurons, since they are located in the neuropile that has a homogeneous appearance. The neuropile of the Clarke’s in opposite is composed of thin parvalbumin-immunopositive processes. But they can be automatically filtered out by size using “Cell Annotation Software” algorithms. The most difficult task was to define the calretinin-immunopositive preganglionic neurons because the intermediolateral nucleus partially invades the white matter, and the brightness ratio between the white and gray matter is always higher than the neurons to neuropile ratio. Therefore, regardless of the algorithm used, the neuropile staining was recognized as a separate “cells-shaped” object. In that case, manual control was required, but this step could be realized using the “Cell Annotation Software” options.

### **Introduction**

The spinal cord is a unique structure of the central nervous system, the heterogeneity of which is represented throughout its length (Watson et al., 2009). Anatomically the spinal cord is divided into five divisions, which differ significantly in length and width (Tan, 2022). The cervical and lumbar parts contain enlargements, due to the presence of interneurons and large motoneuron pools, which are responsible for the work of the muscles of the limbs (Gilerovich et al., 2008). Each division has a segmental structure. For most segments, the rostral part is devoid of roots, what determines the heterogeneity within the segment at the anatomical level (Shkorbatova et al., 2019). At the same time, the borders of divisions do not always correlate with the borders of various spinal nuclei (Philippidou, Dasen, 2013). Thus, in neighboring segments of the same division, a significant difference in the nuclear composition can be observed (Coughlan et al., 2019). All of this together suggests the impossibility of transferring data obtained from one segment to other.

Thus, complex morphological studies require analysis of a large number of slices from many or all segments of the spinal cord of model objects. These facts demand searching for convenient approaches of automated detection, counting, and measurement of morphological characteristics of neuronal populations. The majority of software applications realize this task for the well-visualized individual neurons located on the light background. However, many histological protocols led to the labeling for both neuronal somas and neuropile, which significantly complicates an automatic recognizing of the individual neurons. Thus, this task for these applications is unfeasible.

To solve this problem, we tested the algorithm for image processing “Statistical Dominance Algorithm” (SDA) (Piorkowski, 2016) as part of the “Cell Annotation Software”, the latter was specifically developed for the analysis of brain morphological images (Nurzynska et al., 2017), on some “problematic” spinal neuronal populations identified by various molecular markers. The SDA algorithm statistically determines reference of each pixel to its neighborhood. If the pixel dominates most of the surrounding pixels, it has higher resulting values and *vice versa*. Previously, the SDA algorithm was successfully used for the analysis of the thalamus (Mikhalkin et al., 2021), hippocampus, neocortex (Churilova et al., 2022), and midbrain (Lyubashina et al., 2022), but never for the structures of the spinal cord.

In the spinal cord, we can define some zones of the gray matter, that have a weak ratio between the signal (labeled neurons) and noise (labeled neuropile) after immunohistochemistry to some calcium-binding proteins—markers of the interneurons (Veshchitskii et al., 2022). Interneurons are one of the most complex and heterogeneous population of the spinal cord, and only a few of them are well studied (Jankowska, 2013). A complete study of these populations requires analysis of their morphological features; due to their numerosity, automatic recognition and analysis tools are required.

Thus, the aim of the study was to apply the SDA to the images of the spinal cord with the low level of discernibility between labeled neurons and neuropile and determine the applicability of such an approach for automatic analysis of similar images. The results of the SDA analysis were compared with the results obtained in ImageJ using the “Threshold” function, as the most simple and common tool for the analysis of biological images.

The following zones of the spinal cord gray matter were analyzed in the present study:

1. Calbindin-immunopositive small fusiform and oval neurons and neuropile in the lamina II of the dorsal horn.
2. Parvalbumin-immunopositive medium neurons and neuropile of the Clarke’s nuclei in the intermediate gray matter.
3. Calretinin-immunopositive medium preganglionic sympathetic neurons and neuropile of the intermediolateral nuclei.

## Methods

Experiments were performed in accordance with requirements of Council Directive 2010/63EU of the European Parliament on protection of animals used in experimental and other scientific purposes, and with the approval of the Ethics Commission of the Pavlov Institute of Physiology (Protocol #30/01/2020). Spinal cord slices from the one intact adult domestic female cat (weighing 3.5 kg) were used for the immunohistochemical procedure and subsequent program processing.

**Perfusion and dissection.** The animal was deeply anesthetized with a mixture of Zoletil (Virbac, France; 20 mg/kg) and Xyla (Interchemie werken “De Adelaar” BV, Netherlands; 2 mg/kg), intramuscularly. Heparin (Endopharm, Russia; 0.5 ml/kg) was injected intramuscularly 10 minutes before the start of perfusion. Then the animal was transcardially perfused with 0.9 % NaCl (2 liters, with Heparin, 0.5 ml/l) followed by 4 % paraformaldehyde (2 liters). Then the animal was placed in a prone position, the spinal cord was exposed by removing the vertebral arches and the dorsal part of the dura mater. The distances between the caudal-most parts of the dorsal rootlet attachment zones, connected to the neighboring dorsal root ganglia were measured as spinal cord segments (Shkorbatova et al., 2019). The lengths of all vertebrae and the positions of the segments in relation to the vertebrae were also thoroughly documented. Then, the spinal cord was cut to segments and each of them was replaced in 10 %, 20 % and 30 % sucrose, for cryoprotection, and thereafter cut into 50 µm transverse slices on a freezing microtome (Reichert, Austria). Storage of the slices was in 0.1 M PBS with 0.1 % NaN<sub>3</sub> at +4 °C.

**Immunohistochemical processing.** Slices were processed as free floating. Between all procedures, the slices were washed in 0.01M PBS (pH=7.4). Antigens were unmasked in 1 % NaBH<sub>4</sub>, an endogenous peroxidase activity was blocked by 0.3 % H<sub>2</sub>O<sub>2</sub>, unspecific immunoreactivity was lowered by incubation in 5 % normal goat serum (NGS, Vector Labs). Thereafter slices were incubated for 70 hours in primary antibody (Table 1) with 0.1 % NaN<sub>3</sub> and 3 % of NGS. Then the slices were incubated for 1 day in a biotinylated secondary antibody (Table 1) with 0.1 % NaN<sub>3</sub> and 3 % of NGS. Thereafter slices were subsequently processed using avidin-biotin horseradish-peroxidase complex (ABC Elite system, Vector Laboratories) and diaminobenzidine (DAB)-NiCl<sub>2</sub>-H<sub>2</sub>O<sub>2</sub> reaction. After washing in distilled H<sub>2</sub>O, slices were mounted on slides, dehydrated, cleared and placed under coverslips in Bio Mount HM (Bio-Optica Milano, Italy). To control the specificity of antibodies, several slices were subjected to a standard immunohistochemical protocol with primary or secondary antibodies omitted. In this case, the slices showed no immunopositive labeling.

**Table 1.** The list of the used antibodies.

	Antibody	Host	Clonality	Dilution	Production	RRID
I	Calbindin	Mouse	Monoclonal	1:3000	Sigma	AB_476894
	Calretinin	Rabbit	Polyclonal	1:40000	Sigma-Aldrich	AB_2068506
	Parvalbumin	Rabbit	Polyclonal	1:3000	Abcam	AB_298032
II	Anti-Mouse IgG	Horse		1:600	Vector Lab	AB_2313581
	Anti-Goat IgG	Mouse		1:1000	Vector Lab	AB_2336171

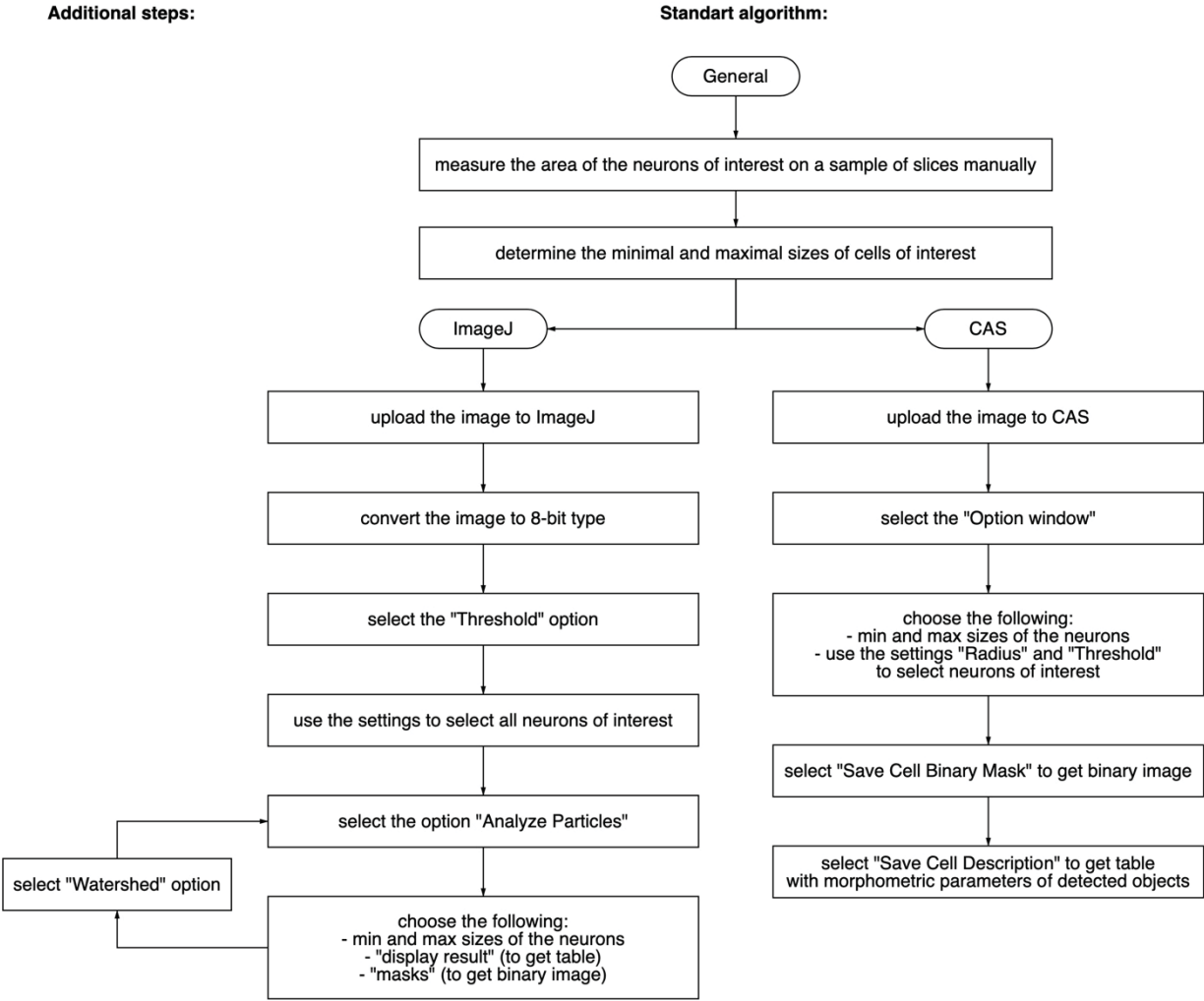
Calbindin antibody (Sigma, Cat.# C9848, RRID: AB\_476894) is mouse monoclonal IgG1, recognizing a single 28 kDa band on western blots of a cat brain (Merkulyeva et al., 2022). The immunostaining of sections through the spinal cord (used at 1:3000) produced a pattern of calbindin labeling that was identical to previously described (Merkulyeva et al., 2016): specific neuronal stripes in the dorsal horns and clusters of cells in intermediate gray matter of spinal enlargements were visualized. The general pattern of immunopositive neurons corresponded to the data on other species using other antibodies to calbindin (Ren, Ruda, 1994).

Parvalbumin antibody (Abcam, Cat.# ab11427, RRID: AB\_298032) is a rabbit polyclonal IgG recognizing 12 kDa band on western blots of a cat brain (Merkulyeva et al., 2022). In the spinal cord, parvalbumin was present in specific pre-motor neural populations and fibers related to the proprioceptive system like in other works (Ren, Ruda, 1994; Anelli, Heckman, 2005).

Calretinin antibody (Sigma-Aldrich, Cat.# AB5054, RRID: AB\_2068506) is a rabbit polyclonal IgG recognizing 29 kDa band on western blots of a cat brain (Merkulyeva et al., 2022).

In the cat spinal cord calretinin was presented by many immunopositive neurons throughout the gray matter in the same structures as in our previous works using the antibody (Veshchitskii et al., 2021) and works of other authors with another calretinin antibody (Anelli, Heckman, 2005).

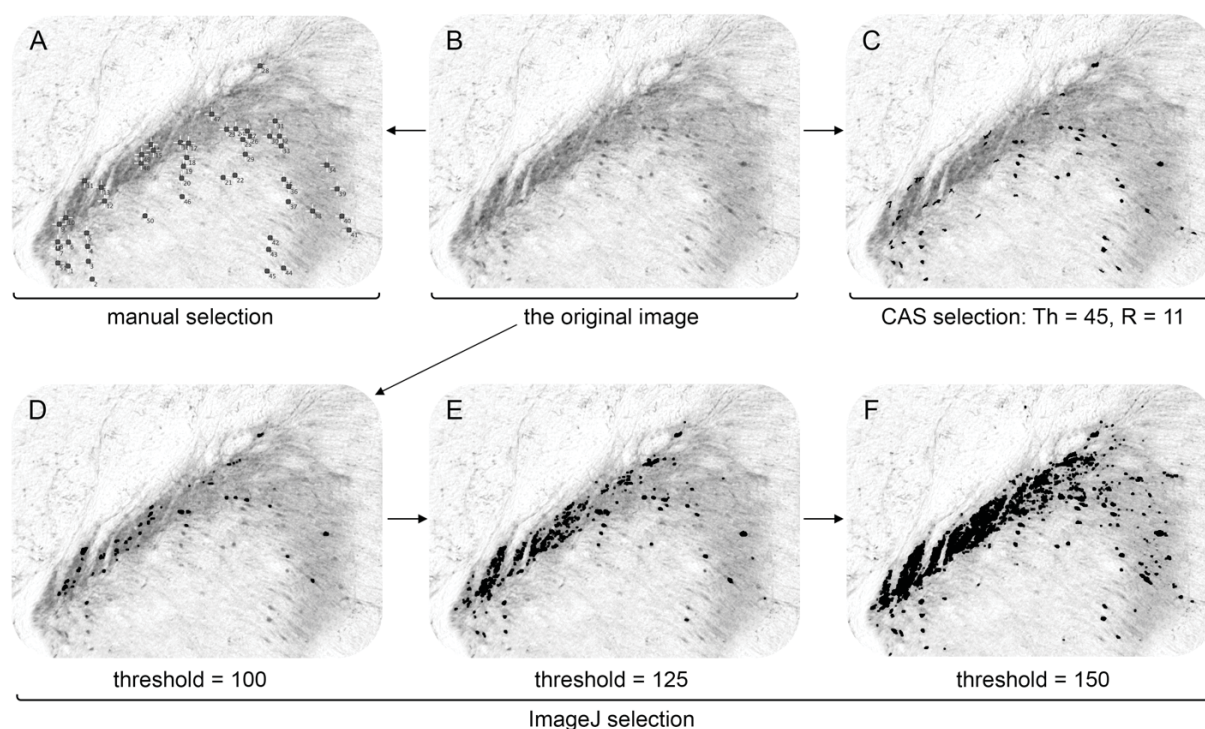
*Image processing.* Digital images of the slices with the identified antigens were obtained using a computer setup equipped with an Olympus CX33 light microscope (Olympus Corporation, Japan; 4× and 10× objectives), digiCamControl software package distributed as an open source, and Nikon camera (D3200, Nikon Corporation, Japan). Selection and counting of the neurons were done in free software ImageJ (Schindelin et al., 2012) and Cell Annotation Software (Nurzynska et al., 2017). The standard order of actions to detect neurons in each application is displayed on Figure 1. Statistical calculations were performed using Prism 9.0 (GraphPad Software, La Jolla, CA, USA).



**Figure 1.** Image processing algorithms in software applications the ImageJ and “Cell annotation software” (CAS).

### Results and Discussion

*Calbindin.* Calbindin is a calcium-binding protein labeling a large number of small fusiform or oval neurons and neuropile of lamina II of the spinal cord dorsal horns (Figure 2B) and a moderate number of neurons in other laminae. Despite the dark immunopositive background in the dorsal horns, these neurons are well recognized by the researcher but their number is too much (for example, in the segments of the cervical and lumbar enlargements hundreds of cells per slice can be detected) for the manual processing.



**Figure 2.** Selection of the calbindin-immunopositive neurons in the dorsal horn of the spinal cord. A—manual selection; B—the original image; C—selection using the “Cell annotation software” (CAS); D, E, F—different selection parameters of the “Threshold” tool in the ImageJ program (100, 125, 150). Th—threshold, R—radius.

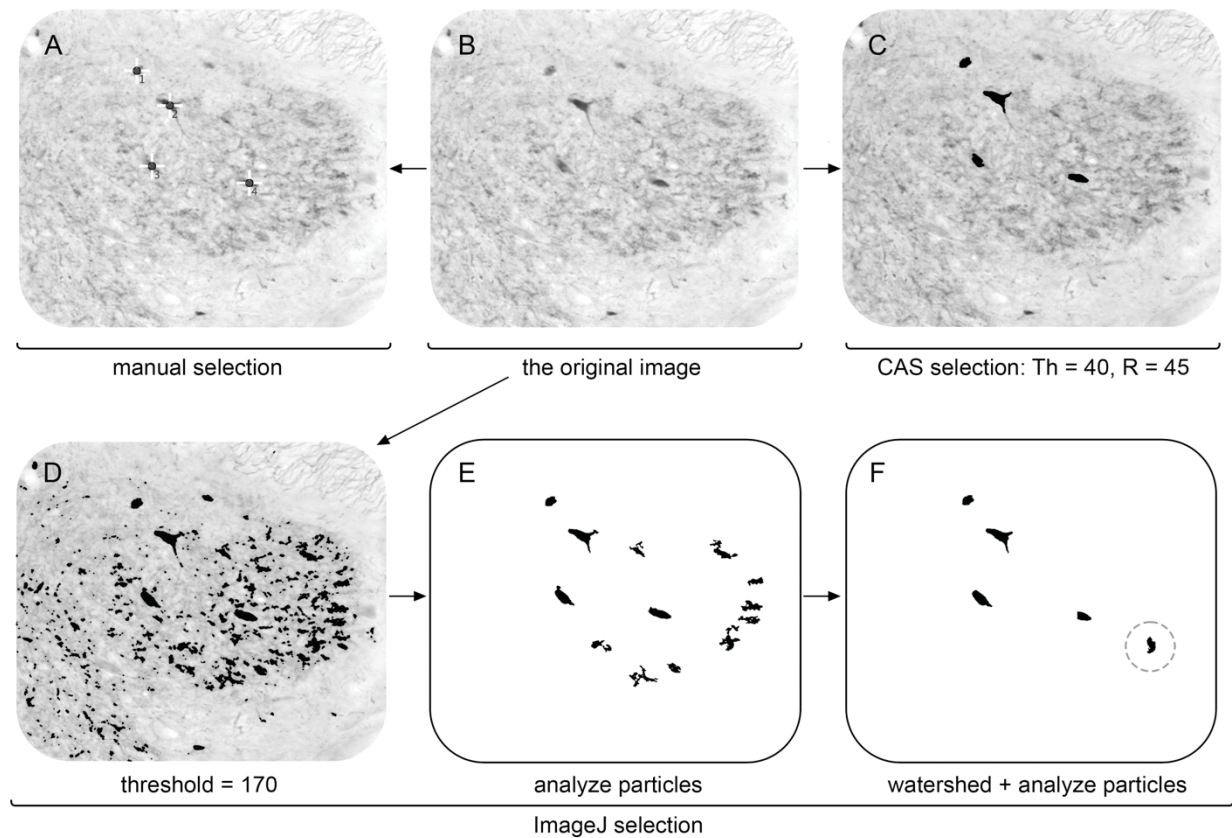
ImageJ analysis using the “Threshold” option fails to define the total population of these neurons (Figure 2D, E, F). In our images, only neurons located within the medial part of lamina II can be distinguished from the dark neuropile by using the low selection threshold; those located in the lateral part of lamina II and lamina III are not (Figure 2D). In this case, an increase in the selection threshold is required. It leads to the total labeling of both neurons of the interest and neuropile (Figure 2E, F); and individual neurons cease to be visible in the medial part of the dorsal horns. No additional manipulations allow detecting these neurons. Such automated analysis is possible only after the finer segmentation of the image into several parts having a different level of background staining. This procedure additionally complicates image processing.

Both CAS and manual recognition of labeled cells lead to similar results (Figure 2A, C). The following CAS settings were used: threshold = 45, radius = 31, area min =  $20 \mu\text{m}^2$ , area max =  $70 \mu\text{m}^2$ . The average difference between CAS and manual counting was  $11 \pm 8\%$  (the CAS algorithm always found more cells).

Throughout the spinal cord, the most serious task for the recognition of the calbindin-stained cells is just the superficial laminae of the dorsal horns (Merkulyeva et al., 2016). Thus, we believe that other labeled cells located in laminae I and V–VIII can be easier recognized. In particular, in the previous study (Merkulyeva et al., 2016) we discovered a clustered organization of the calbindin-positive spinal neurons located at the border of laminae VI and VII. Up today little is known about the function of these clusters. We believe that the full morphological description of this neuronal population can help to understand the presumptive roles of them.

**Parvalbumin.** Parvalbumin is a calcium-binding protein highly expressed within the axon terminals of the proprioceptive afferent neurons of the dorsal root ganglions terminating within

the Clarke's nuclei and in sparse small neurons of the Clarke's nuclei. In many cases, these sparse neurons have the same level of brightness like neuropile; this fact complicated the cells differentiation (Figure 3B). However, due to the low number of cells of interest, in general, this case is significantly simpler than the selection of the calbindin-immunopositive neurons in the dorsal horns.



**Figure 3.** Selection of the parvalbumin-immunopositive neurons in the Clarke's nucleus of the spinal cord. A—manual selection; B—the original image; C—selection using the “Cell annotation software” (CAS); D, E, F—the ImageJ selection steps including the “Threshold” tool (D), “Analyze particles” (E), and combinations of “Watershed” and “Analyze particles” (F). In (F) excessively detected objects in comparison to manual selection are circled. Th—threshold, R—radius.

The multi-step image processing in ImageJ allows recognizing the labeled cells similar to those for manual recognition. This multi-step processing is presented in the left part of Figure 1. At the “Threshold” step, we chose the “170” value allowing to select all cells within the region of interest. But it necessarily leads to the wrong selection of the small immunopositive axonal terminals having the same level of brightness (Figure 3D). The next step is the background filtration realized by the “Analyze Particles” option by the selection of the minimal and the maximal objects areas corresponding to the areas of cells of interest defined. Despite the “optimal” size, several artifact selections of the non-neuronal elements still persisted after this step (Figure 3E). This problem, in general, can be solved using the “Circularity” option in the “Analyze particles” window; but not in the case of the non-round soma shape of cells of interest (like for the Clarke's nuclei). Thus, we used an additional step—the “Watershed” option allowing to cut the touching objects (Figure 1). The resulting image contains plural small objects (Figure 3E) that can be filtrated by the size in the “Analyze Particles” window.

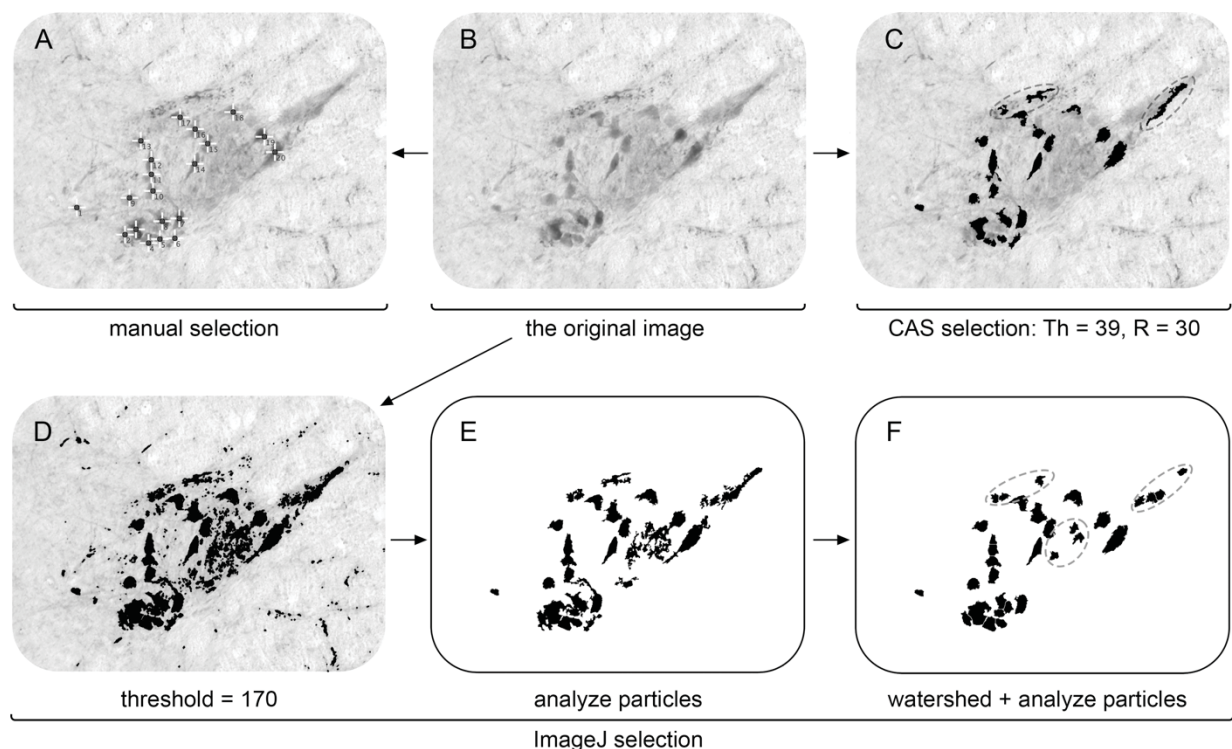
The result of these multi-step manipulations is present in Figure 3F: we can see that all cells of interest are selected and only one selection is an artifact (dashed circle in Figure 3F); we suppose this algorithm to be quite successful.

As for the CAS algorithm, it successfully detects all neurons in one step. For the definition of the Clarke's nuclei neurons, the following settings were used: threshold = 40, radius = 45, area min = 40  $\mu\text{m}^2$ , and area max = 100  $\mu\text{m}^2$ .

Clarke's nuclei illustrate a really difficult task for the automatic selection of the labeled cells since their intensive neuropile staining masks neuronal somas. Other labeled cells located in superficial laminae, in laminae V–VI, and around motor neurons do not have a such background, and their recognizing is a routine task.

In general, antibody to the parvalbumin protein labels the global spinal network responsible for proprioception (Ren, Ruda, 1994; Honda, 1995; Maršala et al., 2007). And the technical tools that allow recognizing this network have significant importance.

*Calretinin.* The most challenging task for the both ImageJ and CAS algorithms was the selection of calretinin-immunopositive preganglionic sympathetic neurons of the intermediolateral nucleus (Figure 4B). This nucleus is located on the border of the gray and white matter and partially invades the white matter. As a result, the labeled neurons and neuropile within this area are much darker compared to the white matter; and regardless of the algorithm, these elements are selected as objects of interest.



**Figure 4.** Selection of the calretinin-immunopositive neurons in the intermediolateral nucleus of the spinal cord. A—manual selection; B—the original image; C—selection using the “Cell annotation software” (CAS); D, E, F—the ImageJ selection steps including the “Threshold” tool (D), “Analyze particles” (E), and combinations of “Watershed” and “Analyze particles” (F). In (F) and (C) excessively detected objects in comparison to manual selection are circled. Th—threshold, R—radius.

In ImageJ, the optimal threshold for the selection of the neurons of interest is 170 (Figure 4D). This value leads to the selection of a large number of neuropile even after “Analyze particles” filtering (Figure 4E). Subsequent using the “Watershed” option leads to the creation of many objects having an area corresponding to the neurons; thus, they are not deleted by the area filtration (dashed ovals in Figure 4F). The average difference between this algorithm and the manual counting was  $76\pm 41\%$  (the algorithm always found more cells).

The CAS algorithm also has difficulties with this task. The following optimal settings for the algorithm were used: threshold = 39, radius = 30, area min =  $60\ \mu\text{m}^2$ , and area max =  $300\ \mu\text{m}^2$ . Like ImageJ, the CAS algorithm distinguishes some large neuropile regions like cells of interest because of their size similar to those of neurons (dashed ovals in Figure 4C). But the average difference between the CAS algorithm and the manual counting is significantly less:  $5\pm 6\%$  (the algorithm always finds more cells).

Calretinin staining in the spinal cord has a particular interest since there are only limited data regarding the morphometry, spatial distribution, and the function of neurons expressing this protein. Only recently the global spatial mapping of the calretinin-immunopositive cells within the spinal cord was documented (Veshchitskii et al., 2021).

In conclusion, both algorithms can be effective in the automatic recognition of neurons of interest, but the CAS algorithm allows obtaining the relevant result with fewer steps. However, we should note that regardless of the selection method, a subsequent inspection of detection results by an operator is an advisable step for obtaining more satisfactory neuron selection.

*Foundation. This study was supported by grants from Russian Science Foundation, grant number: 21–15–00235.*

## References

1. Anelli R., Heckman C.J. The calcium binding proteins calbindin, parvalbumin, and calretinin have specific patterns of expression in the gray matter of cat spinal cord. *J. Neurocytol.* 2005. 34: 369–385.
2. Churilova A., Zachepilo T., Baranova K., Rybnikova E. Differences in the autophagy response to hypoxia in the hippocampus and neocortex of rats. *IJMS.* 2022. 23.
3. Coughlan E., Garside V.C., Wong S.F.L., Liang H., Kraus D., Karmakar K., Maheshwari U., Rijli F.M., Bourne J., and McGlinn E. A Hox code defines spinocerebellar neuron subtype regionalization. *Cell Rep* 2019. 29: 2408–2421.
4. Gilerovich E.G., Moshonkina T.R., Fedorova E.A., Shishko T.T., Pavlova N.V., Gerasimenko Y.P., Otellin V.A. (2008). Morphofunctional characteristics of the lumbar enlargement of the spinal cord in rats. *Neurosci. Behav. Physiol.* 38: 855–860.
5. Honda C.N. Differential distribution of calbindin-D28k and parvalbumin in somatic and visceral sensory neurons. *Neuroscience.* 1995. 68: 883–892.
6. Jankowska E. Spinal Interneurons. In *Neuroscience in the 21st Century*, ed. DW Pfaff, New York, NY: Springer New York. 2013. pp. 1063–99.
7. Lyubashina O.A., Sivachenko I.B., Mikhalkin A.A. Impaired visceral pain-related functions of the midbrain periaqueductal gray in rats with colitis. *Brain Res. Bull.* 2022. 182: 12–25.
8. Marsala J., Lukáčová N., Kolesár D., Sulla I., Gálik J., Marsala M. The distribution of primary nitric oxide synthase- and parvalbumin- immunoreactive afferents in the dorsal funiculus of the lumbosacral spinal cord in a dog. *Cell Mol Neurobiol.* 2007, 27: 475–504.
9. Merkul'yeva N., Veshchitskii A., Makarov F., Gerasimenko Y., Musienko P. Distribution of 28 kDa calbindin-immunopositive neurons in the cat spinal cord. *Front. Neuroanat.* 2016, 9.
10. Merkul'yeva N., Mikhalkin A., Kostareva A., Vavilova T. Transient neurochemical features of the perigeniculate neurons during early postnatal development of the cat. *J. Comp. Neurol.*

2022. 530: 3193–3208.

11. Mikhalkin A., Nikitina N., and Merkulyeva N. Heterochrony of postnatal accumulation of nonphosphorylated heavy-chain neurofilament by neurons of the cat dorsal lateral geniculate nucleus. *J. Comp. Neurol.* 2021. 529: 1430–1441.
12. Nurzynska K., Mikhalkin A., Piorkowski A. CAS: Cell annotation software—research on neuronal tissue has never been so transparent. *Neuroinformatics.* 2017. 15: 365–382.
13. Philippidou P., Dasen J.S. Hox genes: choreographers in neural development, architects of circuit organization. *Neuron.* 2013. 80: 12–34.
14. Piorkowski A. A statistical dominance algorithm for edge detection and segmentation of medical images. In *Information technologies in medicine, advances in intelligent systems and computing.* Springer. 2016. pp. 3–14.
15. Ren K., Ruda M.A. A comparative study of the calcium-binding proteins calbindin-D28K, calretinin, calmodulin and parvalbumin in the rat spinal cord. *Brain Res. Brain Res. Rev.* 1994. 19: 163–179.
16. Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., Preibisch S., Rueden C., Saalfeld S., Schmid B., Tinevez J.–Y., White D.J., Hartenstein V., Eliceiri K., Tomancak P., Cardona A. Fiji: an open-source platform for biological-image analysis. *Nat. Methods.* 2012. 9: 676–682.
17. Shkorbatova P.Y., Lyakhovetskii V.A., Merkulyeva N.S., Veshchitskii A.A., Bazhenova E.Y., Laurens J., Pavlova N.V., Musienko P.E. Prediction algorithm of the cat spinal segments lengths and positions in relation to the vertebrae. *Anat. Rec. (Hoboken).* 2019. 302: 1628–1637.
18. Tan S., Faull R.L.M., Curtis M.A. The tracts, cytoarchitecture, and neurochemistry of the spinal cord. *Anat. Rec. (Hoboken).* 2022.
19. Veshchitskii A.A., Musienko P.E., Merkulyeva N.S. Distribution of calretinin-immunopositive neurons in the cat lumbar spinal cord. *Journal of Evolutionary Biochemistry and Physiology.* 2021. 57(4); 817–834.
20. Veshchitskii A., Shkorbatova P., Merkulyeva N. Neurochemical atlas of the cat spinal cord. *Front. Neuroanat.* 2022. 16: 1034395.
21. Watson C., Kayalioglu G. The Organization of the spinal cord. In: *The Spinal Cord.* Elsevier, 2009. pp 1–7.

## Chapter 19. Hybrid algorithm for optical and SAR image registration based on U-Net neural network and keypoint-based refinement

*Volkov V.*

*Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia*

### Abstracts

The paper considers the problem of satellite multimodal image registration, in particular, optical and SAR (Synthetic Aperture Radar). Such algorithms are used in object detection, change detection, image segmentation and navigation. The paper presents a two-step image registration algorithm: images are pre-aligned by the neural network algorithm based on U-Net, then the result is refined by the image registration algorithm based on keypoints. The paper shows that such pre-alignment increases the final accuracy of image registration.

The neural network algorithm is based on the U-Net-like neural network, which predicts SAR-like image from an optical image for subsequent alignment by an intensity-based (area-based) method. Typically, such networks are trained using BCE (Binary Cross Entropy) loss function. In this paper, we modify the loss function by subtracting the variance of pixel intensities of the predicted SAR-like image. It allows us to get SAR-like images with more contrast and increases the accuracy of image registration at the first step. The second step of the proposed algorithm refines the result of the first step with SIFT algorithm and its modifications.

The paper contains two sets of experiments. The first one considers the image registration accuracy dependency on the value of the variance coefficient in the loss function. In the second experiment, the accuracy of the image registration is compared in cases with and without pre-alignment. Numerical results were obtained on the dataset of 100 images obtained from the Sentinel-1A and Sentinel-2A satellites.

**Keywords:** optical-to-SAR image registration, U-Net-based network, keypoints.

### Introduction

Image registration is the alignment of images of the same scene obtained at different times, from different angles and/or using different sensors. The task of comparing images obtained from sensors of different types is called the multimodal image registration. Optical-to-SAR (Synthetic Aperture Radar) image registration is a particular case of multimodal image registration. Optical-to-SAR image registration is widely used in remote sensing tasks, such as object detection (Errico et al., 2015), change detection (Plank et al., 2016), image segmentation (Wurm et al., 2019), navigation (Yu et al., 2021).

Optical images are well interpreted by humans and do not contain speckle noise, while SAR images are not affected by the atmosphere and day/night illumination (Ye et al., 2020; Hamdi et al., 2021). It is also easier to distinguish built-up areas on SAR images, due to the high intensity of their pixels caused by multiple re-reflection mechanisms (Errico et al., 2015; Сидорчук, Волков, 2018).

The difficulty of comparing optical and SAR images is due to the fact that (i) the images are exposed to different types and strength of noise. For example, speckle noise is present on both optical and SAR images, but the noise is more expressed on SAR images. Moreover (ii) the intensity values of optical and SAR images in some areas of the images may not be correlated

even without the noise. In addition, (iii) the geometric position (for example, the angle of the sensor to the Earth's surface) of the sensor in space may differ at the time of obtaining the image and as a result some lines or shapes of three-dimensional objects may not align (this is especially noticeable in images of mountain terrain).

Optical-SAR image registration methods are divided into two categories: area-based methods (or intensity-based methods) and feature-based methods. The area-based methods are based on the calculation of two corresponding (i.e. "similar" by some metric) subimages. The examples of such methods are normalized cross-correlation (Shi et al., 2012) and mutual information function (Suri, Reinartz, 2009). Such models may have high (subpixel) accuracy, however, they have great computational complexity and often require preliminary approximate alignment of images (Gong et al., 2013), but these models may be ineffective for pairs of images with large geometric differences (Suri, Reinartz, 2009).

Feature-based methods are based on (i) finding some "special" points or image elements (called features) that are easily distinguishable both in the optical and SAR image; (ii) subsequent calculating of a geometric transformation that compares the found features. For example, such methods can use following features: points (Fan et al., 2012; Ma et al., 2016; Paul, Pati, 2018), corners (Xiang et al., 2018; Paul, Pati, 2019; Xiong et al., 2019), lines (Wang et al., 2012) and objects (e.g. roads) (Kunina et al., 2019). Feature-based methods are computationally simpler and often more efficient (Gong et al., 2013) than area-based methods in cases where there are significant geometric distortions between images. However, feature-based methods may be worse than area-based methods in maximum image registration accuracy (Gong et al., 2013).

In literature, algorithms of optical-to-SAR image registration are applied to images of different size from  $256 \times 256$  pixels (Schmitt et al., 2018) to about  $10000 \times 10000$  (Ye et al., 2021). It is worth to note that registration of large-size images has greater computational complexity and are more susceptible to nonlinear geometric distortions. On the other hand, registration of small-sized images may be difficult because of small-sized images may contain insufficient information for successful registration. In this paper, the image registration is considered with images  $1024 \times 1024$  pixels.

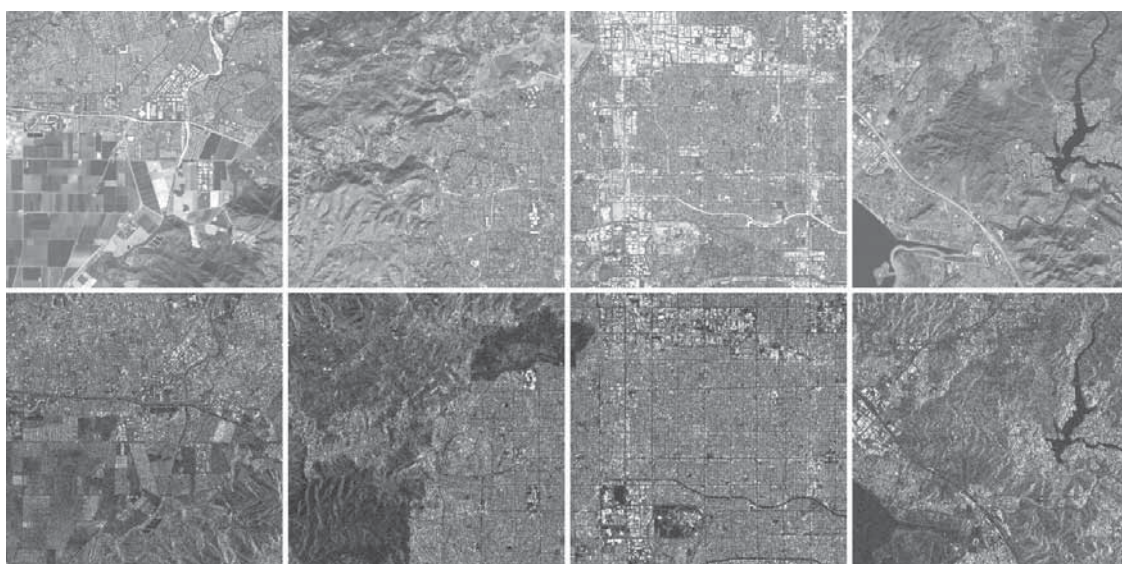
Optical and SAR images, due to the parameters describing the position of the sensor, can be roughly aligned based on georeference. However, due to the inaccuracy of these parameters, the alignment accuracy is tens or even hundreds of pixels (Wang et al., 2022; Hansson N., 2022). In the work (Ye et al., 2021), authors made an estimation of inaccuracy of image registration only with georeference for images obtained from Sentinel-1 (SAR) and Sentinel-2 (optical) satellites. Images from these satellites are used in this work. The authors note that the reason for the significant geometry distortions between these images is that terrain correction process is not conducted for SAR images from the Sentinel-1 satellite, as well as the side-looking geometry for SAR sensor, i.e. the angle of Sentinel-1 sensor is not a perpendicular to the Earth surface. This result in significant geometric distortions relative to optical Sentinel-2 images, which can reach about 100 pixels (at 10 m/pixel). Additionally, it is worth to note the paper in which the authors indicate that georeference still leaves large translation distortions differences between images and small rotation and scale distortions (Wang et al., 2022).

In this paper, we consider an algorithm of optical-to-SAR image registration with high accuracy ( $\pm 1$  pixel) with two limitations. First, the images are assumed to be approximately aligned by georeference with an accuracy of  $\pm 100$  pixels. The number of 100 pixels based on observations with Sentinel-1 and Sentinel-2 image registration (Ye et al., 2021) which were mentioned

before. Secondly, in this study, we considered only translation distortions because of translation distortions have more influence on overall distortions than rotation and scale distortions after georeferenced (Wang et al., 2022). The proposed algorithm consists of two stages: (i) an algorithm based on the U-Net neural network for approximate alignment of optical and SAR images, (ii) the image registration algorithm based on keypoints, which will consider matching keypoints only in the neighborhood of the translation defined on the (i) step. In the second step we use the algorithm based on SIFT (Lowe, 2004) and its modification (Волков, 2022). Numerical results of image registration algorithms were conducted on a dataset of 100 image pairs obtained from Sentinel-1A and Sentinel-2A satellites (Волков, Шве́ц, 2021).

## 1. Dataset description

The dataset used in this work consists of 100 pairs of optical-SAR images with size  $1024 \times 1024$  pixels and is published in (Волков, Шве́ц, 2021). The images used in the dataset are collected from open sources, and licenses allow them to be distributed and modified<sup>3</sup>. The Copernicus Open Access Hub website<sup>4</sup> was used as a source of optical and SAR data. SAR images were obtained from the Sentinel-1A satellite with the following parameters: product type Level-1 Ground Range Detected (GRD), Interferometric Wave (IW) swath mode, VH polarization. Three-channel RGB satellite images from the Sentinel-2A satellite (product type: S2MSI1C) were used as optical images. Images of both types were brought to a spatial resolution of 10 meters/pixel. Each pair of optical-SAR images is aligned relative to each other using georeference. Further, in order to increase the accuracy of alignment, a manual image registration was carried out using a projective transformation. The resulting images are aligned mainly with sub-pixel accuracy, but sometimes there are small areas where the accuracy is lower (an error of up to two pixels) because of non-uniform geometry distortions. The dataset mainly contains images of scenes of cities, mountains and fields. Examples of pairs of images are shown in Fig. 1. The ready-to-use dataset with its metadata are available for download and published in (Волков, Шве́ц, 2021). The data can also be downloaded manually from the resources listed above (Copernicus Open Access Hub website).



**Figure 1.** Examples of images from the dataset. The upper row is optical, the lower one is the corresponding SAR images.

<sup>3</sup> <https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/TermsConditions>

<sup>4</sup> <https://scihub.copernicus.eu/>

## 2. Proposed hybrid image registration algorithm

This section describes a hybrid two-step algorithm for optical-to-SAR image registration. The first stage is an algorithm based on the U-Net neural network for approximate image alignment. The neural network predicts the SAR image with applied Gaussian filter (SAR+gauss) from optical for subsequent registration of SAR-like (neural network result) and SAR+gauss images using pixel-by-pixel comparison. A Gaussian filter is applied to the SAR image to reduce the noise on image. A detailed description of the first stage of the hybrid algorithm is presented in subsection 2.1.

The second stage is an algorithm for optical-to-SAR image registration based on keypoints, which will use the result of the first stage of the hybrid algorithm as a preliminary alignment, i.e., consider candidate keypoints for matching between keypoints only in the surrounding of expected location determined at the first stage of the hybrid algorithm. As such algorithms, we used an algorithm based on SIFT (Lowe, 2004) and its modification (Волков, 2022). A more detailed description of the second stage of the hybrid algorithm is presented in subsection 2.2.

### 2.1. Neural network based on U-Net

This subsection describes the first stage of the hybrid algorithm, which is based on a neural network based on U-Net, which predicts the SAR image with applied Gaussian filter (SAR+gauss) from optical. The neural network is written in Python with using the PyTorch library. Let's call the result of the neural network a SAR-like image. The result of the first stage of the hybrid algorithm for each pair of optical-SAR images is the value of translation distortion between the SAR-like and SAR+gauss images, which will be considered as a pre-alignment at the second stage of the hybrid algorithm. To determine the translation distortion, we consider a set of translation values taken with a certain step, for each of which a metric is calculated based on a pixel-by-pixel comparison. The value of translation distortion with the minimum metric value among the rest of translation values is selected as the best translation for the pair of SAR-like and SAR+gauss images. It should be mentioned that at this stage only a rough image registration is performed, which does not have high accuracy. In this paper we consider two cases in which the translation distortion is determined with an accuracy of  $\pm 5$  and  $\pm 2.5$  pixels. Further refinement of the image registration accuracy is performed at the second stage of the hybrid algorithm described in subsection

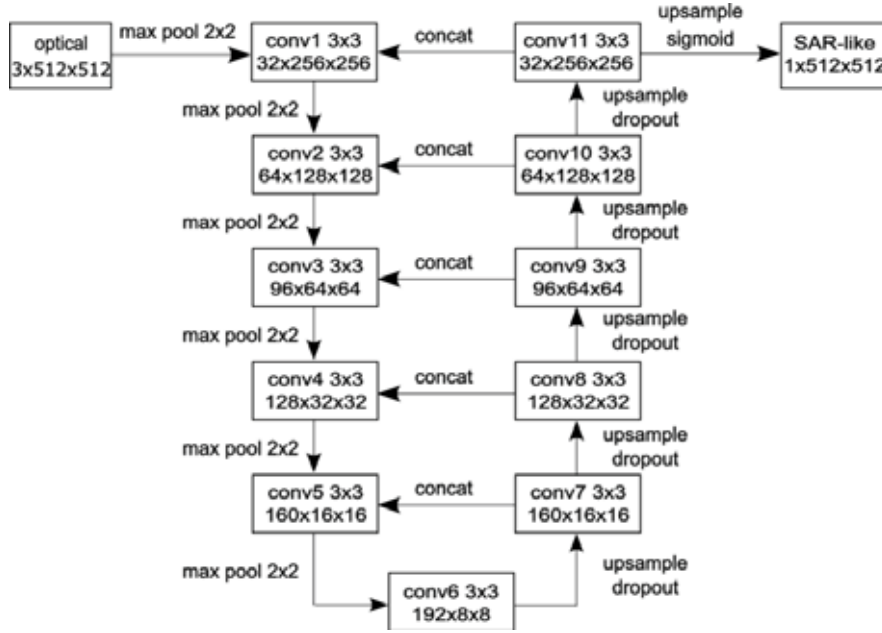
### 2.2. Dataset.

As already mentioned, the dataset contains 100 aligned pairs of optical and SAR images. It was divided into three parts: training (58 pairs), validation (13 pairs) and test (29 pairs) parts. The input of the neural network is an optical image downsampled by 2 times to the size of  $512 \times 512 \times 3$  pixels (HxWxC). The ground truth of the neural network is a downsampled by 2 times SAR image to the size of  $512 \times 512 \times 1$  pixels with applied Gaussian filter with  $\sigma = 2$  (SAR+gauss). A Gaussian filter is used to reduce the noise on SAR image.

### Neural network

The scheme of the neural network used in this work is shown in Fig. 2. The structure of the neural network is based on U-Net. All convolutions use a  $3 \times 3$  pixel kernel and padding equal to 1 pixel. The conv1 cell corresponds to the following sequence: convolution, ReLU activation function and max pooling with size of  $2 \times 2$ . Cells from conv2 to conv6 correspond to a sequence of convolution, BatchNorm layer, ReLU activation function, and max pooling with

size of  $2 \times 2$ . Cells from conv7 to conv11 correspond to a sequence of up sample, convolution, BatchNorm layer, ReLU, dropout with probability 0.2 and concatenation with the result of the left column (Fig. 2). At the end, up sample and the sigmoid function are used to obtain the result of a neural network of the size  $1 \times 512 \times 512$  pixels ( $C \times H \times W$ ).



**Figure 2.** The structure of the neural network predicting a SAR image with an applied Gaussian filter from an optical image.

A function consisting of two terms was used as a loss function. The first term is the BCE loss function (Binary Cross Entropy) which is widely used in U-Net-based neural networks. The formula of the BCE loss function looks as follows (weights were not used):

$$l_{BCE}(x, y) = \text{mean}(l_{BCE}) = \text{mean}(\{l_1, \dots, l_N\}^T), \quad (1)$$

where

$$l_n = -[y_n \log x_n + (1 - y_n) \log(1 - x_n)],$$

$x_n$  – the result of neural network,

$y_n$  – ground truth values (it should be in  $[0, 1]$ ).

For the second term in the loss function variance value of the neural network result was proposed. Thus, the final loss function looks like this:

$$l_{all}(x, y) = l_{BCE}(x, y) - kD(x), \quad (2)$$

where  $k$  – variance value coefficient ( $> 0$ ), which is configurable parameter of the algorithm, ( $D_x$ ) – variance value of the result of neural network.

Adding variance to the loss function made the pixels in the image constructed by the neural network have more different values, which increase the contrast on the SAR-like image.

### Alignment SAR-like and SAR.

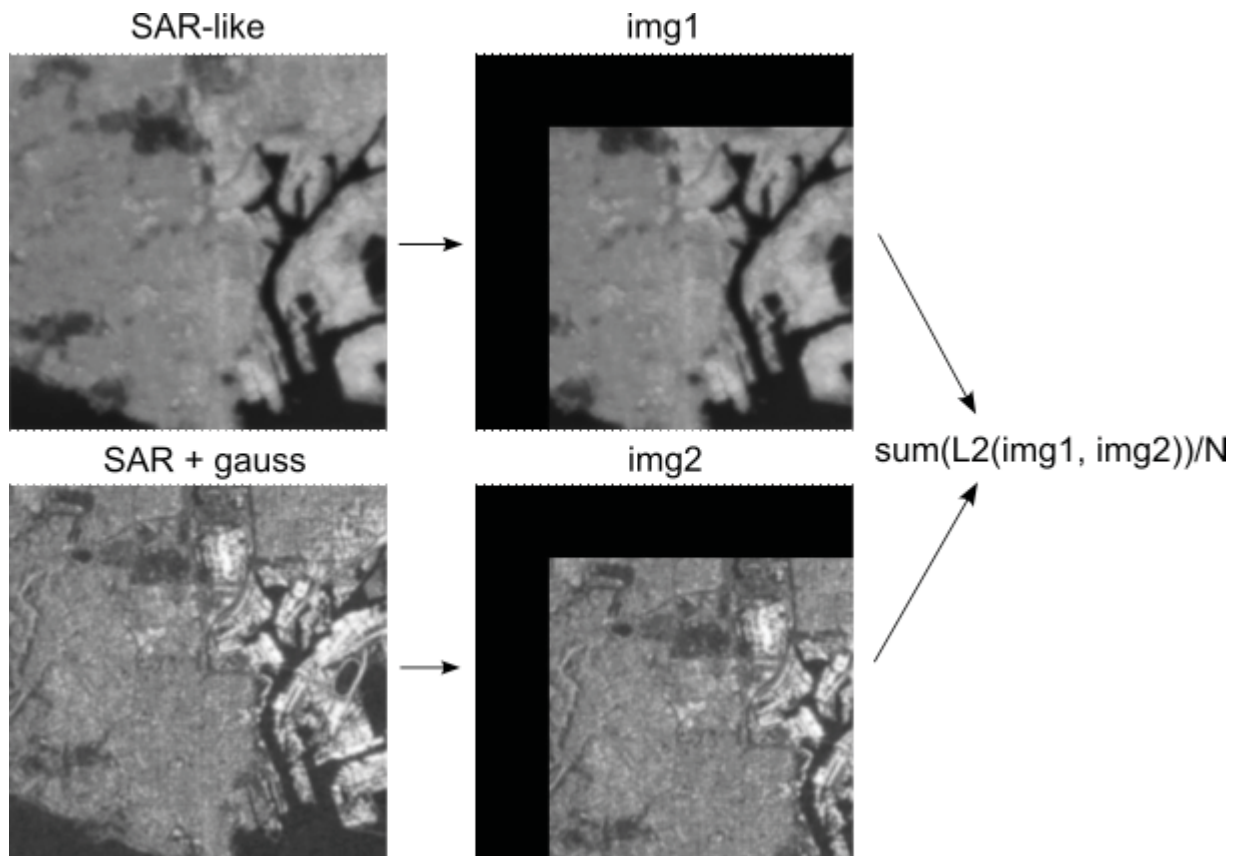
Previously in this subsection the method for obtaining a SAR-like image was described. Final part on the first step of hybrid algorithm is to find a rough alignment between SAR-like

(result of neural network) and SAR+gauss (which also was used as ground truth of neural network). To determine the translation distortion value, it is necessary to align SAR-like and SAR+gauss images. The proposed image registration method for each pairs of images calculates the similarity values of two images (in our case, SAR-like and SAR+gauss) with different translation values between images taken with a certain step (algorithm parameter). Among the all similarity values the minimum one is selected, which correspond to the best translation distortion value. The considered set of translation distortion values is limited to  $\pm 100$  pixels on each axis as mentioned earlier.

Let's pay attention at the algorithm for calculating the similarity of two images. At the input the algorithm receives two images (in our case, SAR-like and SAR+gauss) and a value which is the step between considered translation values. The algorithm looks like listed below.

Algorithm 1. Finding the best translation distortion value with the minimum similarity value.

1. A set of possible translation values taken with value and satisfying the range of permissible translation distortion values (up to  $\pm 100$  pixels for each coordinate axis, as mentioned before) is determined.
2. For each translation value perform the following:
  - a. Apply the translation value to the SAR+gauss image. Pixels outside the image are filled with zeros (Fig. 3, bottom row). At the Fig. 3 we can see that the SAR+gauss image was "moved" diagonally to down and right (img2).



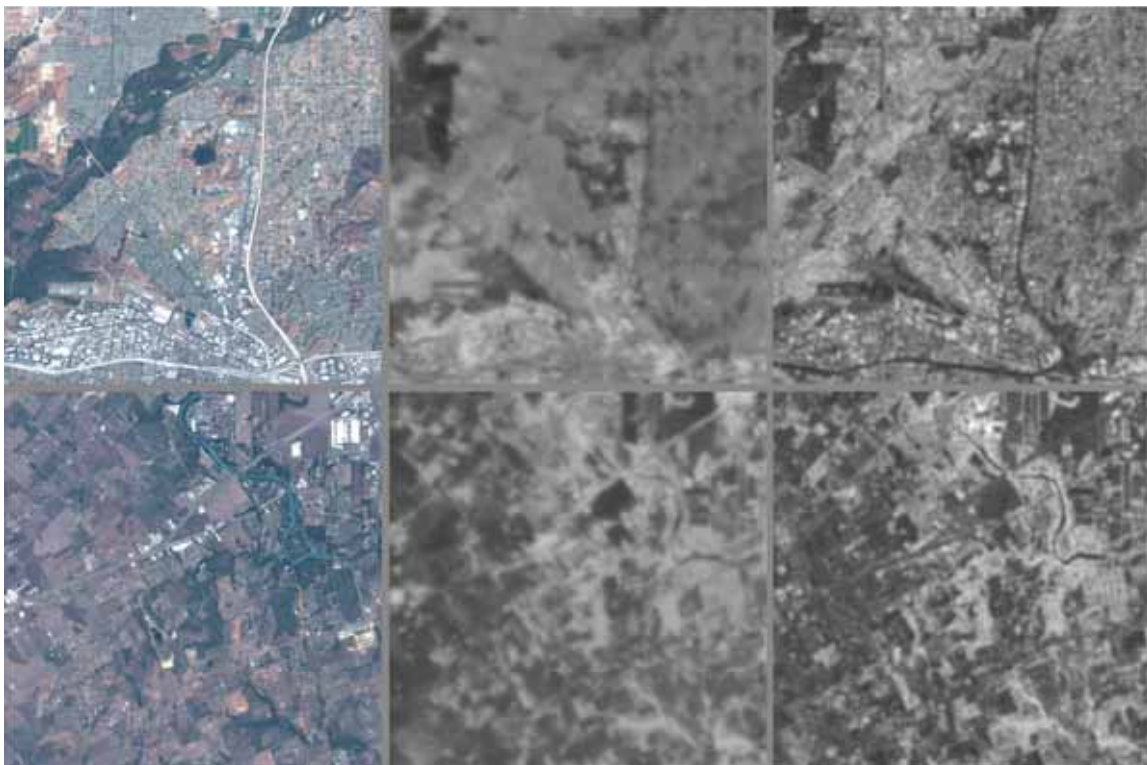
**Figure 3.** An illustration of algorithm for the calculation of the similarity value between SAR-like and SAR+gauss images for the case of translation value of 100 pixels along both coordinates (step 2 of the algorithm 1). The upper row corresponds to a SAR-like image, the lower row corresponds to a SAR+gauss. is the area of intersection of img1 and img2 images.

- b. Transform the SAR-like image by zeroing pixels outside the intersection area of the two images (taking into account the selected translation value) (Fig. 3, top row). In Fig. 3, you can see that the img1 image does not translated from SAR-like image and only the upper and left parts of the image were zeroed.
- c. Calculate similarity value as the sum of the L2 norm divided by the area of intersection of img1 and img2 images (i.e., not the zeroed part at the upper and left parts of the images).
- d. If the similarity value is less than the minimum similarity value (or this is the first iteration), then remember it and determine the current translation value as the best. Continue for all possible translation values.

The paper considered pre-alignment with pixels and pixels, which gives the accuracy of pre-alignment equal to 5 and 2.5 pixels, respectively (the window of the size  $10 \times 10$  pixels is covered by center pixel of the window  $\pm 5$  pixels). For the case of pixels we use the algorithm 1 mentioned above. For pixels, algorithm 2 was used, which is a repeated application of algorithm 1 but with lesser value. Algorithm 2 is computationally simpler than algorithm 1 with pixels. Algorithm 2 is presented below:

Algorithm 2. Modification of algorithm 1 with a decrease in the translation step.

1. Using algorithm 1 calculate the similarity value and the best translation value () with pixels and the range of considered translation values of  $[-100, 100]$  for the and coordinates.
2. Next, consider the surroundings around the founded translation value () with pixels. That is using algorithm 1 calculate the similarity value and the best translation value with the parameters: pixels and the range for coordinate:, for:.
3. An examples of a SAR-like images are shown in Fig. 4. On the left is an optical input image, on the right is a SAR+gauss image (also used as neural network ground truth), on the center is the result of the neural network (SAR-like image) obtained with a variance coefficient.



**Figure 4.** Example of successfully matched images. From left to the right: optical image, SAR-like, SAR+gauss ().

## 2.2. Second step of hybrid algorithm

This subsection describes the second stage of the hybrid algorithm, which is an algorithm for optical-to-SAR image registration based on keypoints. The general algorithm looks like:

1. Finding keypoints on two types of images.
2. Calculation of descriptors for each keypoint on two types of images.
3. Finding matching between keypoints on different types of images by comparing corresponding descriptors by some metric.
4. Finding a geometric transformation using a geometric model.

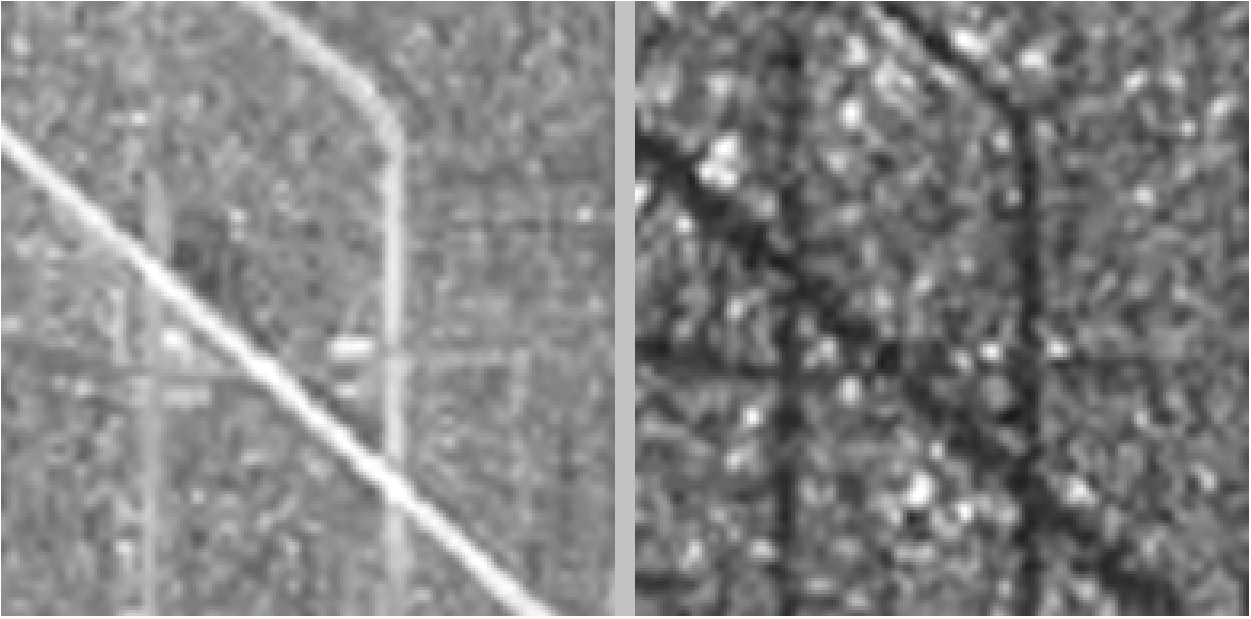
A detailed description of the algorithm used based on keypoints is presented in (Волков, 2022). Below is a brief description of the separate steps.

Algorithm use images downsampled by 2 times (to  $512 \times 512$  pixels). Optical images as RGB image and SAR image as grayscale image. To find keypoints, a SIFT detector was used with the NMS algorithm (non-max-suppression) (Yu et al., 2018) (window size 2.5 pixel) and binning (Xiong et al., 2019) with a block size of  $128 \times 128$  pixels. The NMS algorithm among any two keypoints located close enough (2.5 pixels in our case) selects the one keypoint with the highest “score” value determined by the detector. Binning is the division of an image into non-overlapping blocks. Next, the detector is applied to each block separately, after which the sets of keypoints from each block are united. This allows to get a more uniform distribution of keypoints across the image in case the same overall number of keypoints. For example, the overall number of keypoints used in this paper is 800. So for each block detector could find no more than 50 points. The method used to find keypoints is described in more detail in (Волков, 2022).

Two algorithms were studied in this paper to calculate descriptors. The first one is the SIFT descriptor (Lowe, 2004), the second one is a modification of the SIFT descriptor proposed in (Волков, 2022). The SIFT modification takes into account the inversion of gradients in some areas of the optical and SAR images. Fig. 5 shows an example of such area. On the left is a fragment of an optical image and on the right is a corresponding fragment of a SAR image. Fig. 5 shows that the road in the optical image is brighter than the road’s surrounding, instead of road on SAR image. It results in different gradients and different descriptors. SIFT modification consider gradients with opposite directions as the same direction. Detailed algorithm of SIFT descriptor modification is described in (Волков, 2022).

The descriptors were compared using the nearest neighbor method. A descriptor distance (L2 norm in our case) is calculated between each pair of keypoints on optical and SAR images. Two descriptors (and their corresponding keypoints) are considered matched if they are mutually the closest, i.e. it is impossible to find a keypoint in any image whose descriptor will be closer than chosen one. Also, the descriptor distance must be less than the specified threshold. If the mentioned conditions are not met, then the pair of keypoints are considered to be not matched.

Since only translation-type distortions are investigated in this work, the model described in (Волков, 2022) was used as a geometric model. This model is based on a histogram of translations values calculated for each pair of matched points. Gaussian filter is applied to histogram of translation. It allows to take into account not only one bin but also surrounding bins. The bin with maximum value corresponds to the best translation value.



*Figure 5. Example of inverted gradients. Optical image fragment on the left and SAR image fragment on the right.*

### 3. Experiments and results

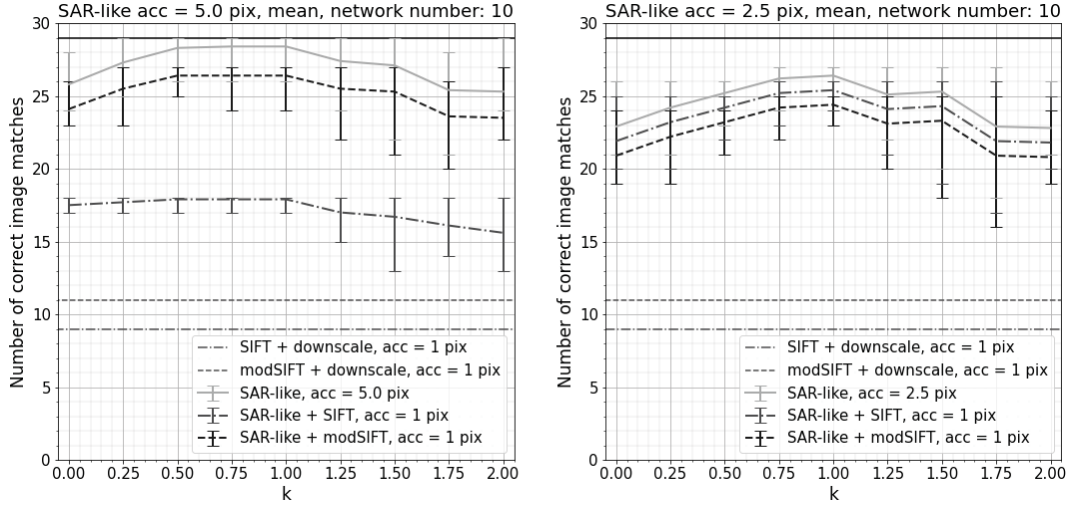
This section presents the results of the algorithms for optical-to-SAR image registration described in section 2 by the number of successfully registered pairs of images from the test part of dataset (29 pairs). Firstly, the results of the image registration of the first stage of the hybrid algorithm are estimated, i.e., the accuracy of the pre-alignment. Two cases of registration SAR-like and SAR+gauss images were considered: algorithm 1 with pixels and algorithm 2 with a final pixels, i.e., the accuracy of pre-alignment was taken 5 and 2.5 pixels respectively. In this experiment, a pair of images is considered successfully matched if the algorithm determines that the best translation value is (0, 0). In this experiment, different values of the variance multiplier in the loss function were studied. Ten neural networks were trained for each value of. Fig. 6 shows the results of the mean value for ten neural networks, and Fig. 6 shows the median value (“SAR-like, acc = 5.0 pix” for the left graph and “SAR-like, acc = 2.5 pix” for the right one).

Secondly, the results of the hybrid algorithm for both cases of pre-alignment and two algorithms based on keypoints are presented: SIFT-based algorithm (“SAR-like + SIFT”) and algorithm based on modification of SIFT (“SAR-like + modSIFT”). In this series of experiments, the final accuracy of the algorithm based on keypoints was taken to be equal to 1 pixel.

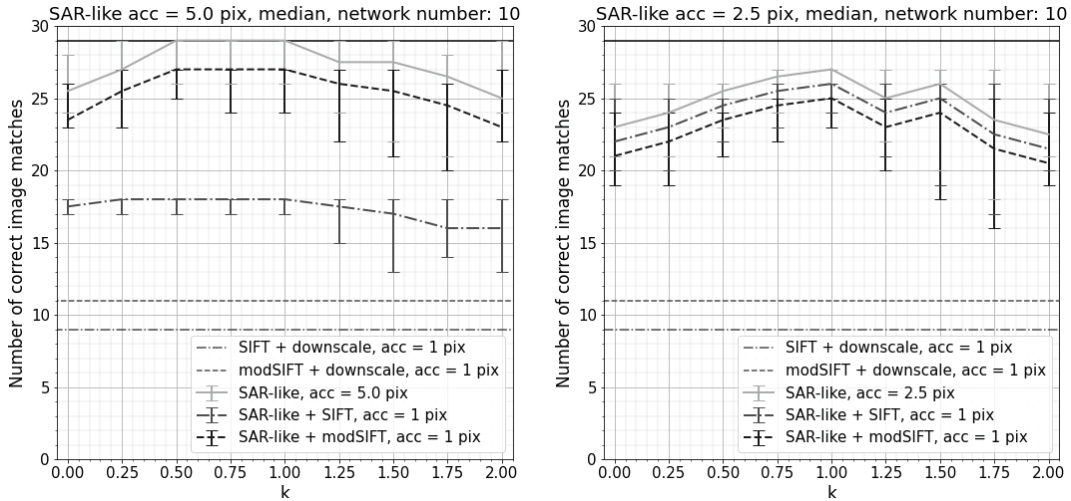
The graphs show that the highest result (26.5/29 mean value and 27/29 median value) was obtained using a hybrid algorithm with pre-alignment with an accuracy of 5 pixels and modification of SIFT with an accuracy of 1 pixel. It should be noted that with an increase in the accuracy of the pre-alignment stage to 2.5 pixels, the accuracy of the entire hybrid algorithm with modification of SIFT decreased. This is due to the fact that there are more images that were not matched at the first step of the hybrid algorithm, which, accordingly, were not matched at the second step. This shows the disadvantage of this method—if the alignment on the first step is not correct, then after the second step it will not be correct also.

Fig. 6 and Fig. 7 show that pre-alignment significantly increases the accuracy of optical-to-SAR image registration of the entire algorithm from 9/29 (without pre-alignment) to 26/29 (with

pre-alignment) in the case of SIFT and from 11/29 (without pre-alignment) to 27/29 (with pre-alignment) in the case of SIFT modification (used median values from Fig. 7).



**Figure 6.** The mean number of images successfully matched by the algorithms based on the results of ten neural networks, depending on the variance multiplier. On the left is a case with pre-alignment with an accuracy of 5 pixels, on the right—with an accuracy of 2.5 pixels.



**Figure 7.** The median value of the images successfully matched by the algorithms based on the results of ten neural networks, depending on the variance multiplier. On the left is a case with pre-alignment with an accuracy of 5 pixels, on the right—with an accuracy of 2.5 pixels.

The hybrid algorithm with the SIFT algorithm showed poor results in the case of pre-alignment with an accuracy of 5 pixels, but unlike the SIFT modification, with an increase in the accuracy of pre-alignment to 2.5 pixels, the number of matched images greatly increased to 25.5/29 mean value and 26/29 median value.

#### 4. Conclusions

This paper describes the algorithm for optical-to-SAR image registration. Image registration task has two limitations. First, algorithm consider only translation distortions without rotation and scale distortions because images supposed to be pre-aligned by georeference and in this case rotation and scale distortions have less influence on the image registration accuracy (Wang et al., 2022). Second, algorithm consider translation distortions in range  $\pm 100$  pixels by each axis because even after alignment with georeference it could be misregistration error

about 100 pixels (Ye et al., 2021). This result was obtained on Sentinel-1 and Sentinel-2 images, which used in this paper.

The paper proposes a hybrid two-step algorithm for optical-to-SAR image registration. At the first stage, a preliminary rough alignment of images takes place using a neural network based on U-Net, which predicts the SAR image (SAR-like) from optical. Next, using a pixel-by-pixel comparison of the predicted SAR-like and SAR+gauss (SAR with applied Gaussian filter) images algorithm finds an approximate alignment.

At the second stage, the image registration algorithm based on keypoints is applied using the result of the first stage as a preliminary alignment. It means that the algorithm does not consider matching of keypoints that do not correspond to the founded preliminary alignment of images taking into account the accuracy of the pre-alignment. The paper considered algorithms based on SIFT and its modification described in (Волков, 2022). The final accuracy of the overall hybrid algorithm equal to 1 pixel with pre-alignment accuracy on the first step of hybrid algorithm equal to 5 and 2.5 pixels (two cases).

It has been shown that using a neural network algorithm to find a preliminary alignment between images significantly increases the accuracy of image registration from 9/29 to 26/29 in the case of SIFT and from 11/29 to 27/29 in the case of SIFT modification.

The results show that the algorithm using SIFT has greater accuracy in the case of a higher accuracy of preliminary alignment step, instead of an algorithm with a modification of SIFT. It shows the disadvantage of the hybrid algorithm, which is sensitive to accuracy at the first step of the algorithm – if the image registration at the first step of the hybrid algorithm is not correct, then it will also be incorrect at the second step.

*This work was supported by Russian Science Foundation (Project No. 20–61–47089).*

## References

1. Волков В.В. Модифицирование метода поиска и дескрибирования устойчивых точек SIFT для сопоставления оптических и радиолокационных изображений. Сенсорные системы. 2022. 36(4): 349–365. DOI: 10.31857/S0235009222040060.
2. Волков В.В., Швец Е.А. Набор данных и метод оценки алгоритмов сопоставления оптических и радиолокационных изображений на основе устойчивых точек. Информационные технологии и вычислительные системы. 2021. 2: 44–57.
3. Сидорчук Д.С., Волков В.В. Комплексирование радиолокационных изображений и оптических снимков в видимом и тепловом диапазонах с учетом различий в восприятии яркости и цветности. Сенсорные системы. 2018. 32(1): 14–18. DOI: 10.7868/S0235009218010031.
4. Errico A., Angelino C.V., Cicala L., Persechino G., Ferrara C., Lega M., Vallario A., Parente C., Masi G., Gaetano R., Scarpa G. Detection of environmental hazards through the feature-based fusion of optical and SAR data: A case study in southern Italy. International Journal of Remote Sensing. 2015. 36(13): 3345–3367. DOI: 10.1080/01431161.2015.1054960.
5. Fan B., Huo C., Pan C., Kong Q. Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT. IEEE Geoscience and Remote Sensing Letters. 2012. 10(4): 657–661. DOI: 10.1109/LGRS.2012.2216500.
6. Gong M., Zhao S., Jiao L., Tian D., Wang S. A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. IEEE Transactions on Geoscience and Remote Sensing. 2013. 52(7): 4328–4338. DOI: 10.1109/TGRS.2013.2281391.
7. Hamdi I., Tounsi Y., Benjelloun M., Nassim A. Evaluation of the change in synthetic aperture radar imaging using transfer learning and residual network. Computer Optics. 2021. 45(4):

- 600–607. DOI: 10.18287/2412–6179-CO-814.
8. Hansson N. Investigation of Registration Methods for High Resolution SAR-EO Imagery. 2022.
9. Kunina I., Panfilova E., Gladkov A. Matching of SAR and optical images by independent referencing to vector map. In Eleventh International Conference on Machine Vision (ICMV 2018). 2019. 11041. DOI: 10.1117/12.2523132.
10. Lowe D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*. 2004. 60(2): 91–110. <https://doi.org/10.1023/B: VISI.0000029664.99615.94>
11. Ma W., Wen Z., Wu Y., Jiao L., Gong M., Zheng Y., Liu L. Remote sensing image registration with modified SIFT and enhanced feature matching. *IEEE Geoscience and Remote Sensing Letters*. 2016. 14(1): 3–7. DOI: 10.1109/LGRS.2016.2600858.
12. Paul S., Pati U. C. Automatic optical-to-SAR image registration using a structural descriptor. *IET Image Processing*. 2019. 14(1): 62–73. DOI: 10.1049/iet-ipr.2019.0389.
13. Paul S., Pati U. C. Optical-to-SAR image registration using modified distinctive order based self-similarity operator. In 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). 2018. P. 1–5. DOI: 10.1109/SCEECS.2018.8546950.
14. Plank S., Twele A., Martinis S. Landslide mapping in vegetated areas using change detection based on optical and polarimetric SAR data. *Remote Sensing*. 2016. 8(4): 307. DOI: 10.3390/rs8040307.
15. Schmitt M., Hughes L. H., Zhu X. X. The SEN1–2 dataset for deep learning in SAR-optical data fusion. *arXiv preprint arXiv:1807.01569*. 2018.
16. Shi W., Su F., Wang R., Fan J. A visual circle based image registration algorithm for optical and SAR imagery. In 2012 IEEE International Geoscience and Remote Sensing Symposium. 2012. P. 2109–2112. DOI: 10.1109/IGARSS.2012.6351089.
17. Suri S., Reinartz P. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. *IEEE Transactions on Geoscience and Remote Sensing*. 2009. 48(2): 939–949. DOI: 10.1109/TGRS.2009.2034842.
18. Wang H., Wang C., Li P., Chen Z., Cheng M., Luo L., Liu Y. Optical-to-SAR Image Registration Based On Gaussian Mixture Model. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2012. 39: 179–183.
19. Wang Z., Yu A., Zhang B., Dong Z., Chen X. A Fast Registration Method for Optical and SAR Images Based on SRAWG Feature Description. *Remote Sensing*. 2022. 14(19): 5060.
20. Wurm M., Stark T., Zhu X. X., Weigand M., Taubenböck H. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS journal of photogrammetry and remote sensing*. 2019. 150: 59–69.
21. Xiang Y., Wang F., You H. OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas. *IEEE Transactions on Geoscience and Remote Sensing*. 2018. 56(6): 3078–3090. DOI: 10.1109/TGRS.2018.2790483.
22. Xiong X., Xu Q., Jin G., Zhang H., Gao X. Rank-Based Local Self-Similarity Descriptor for Optical-to-SAR Image Matching. *IEEE Geoscience and Remote Sensing Letters*. 2019. 17(10): 1742–1746. DOI: 10.1109/LGRS.2019.2955153.
23. Ye S. P., Chen C. X., Nedzved A., Jiang J. Building detection by local region features in SAR images. *Computer Optics*. 2020. 44(6): 944–950. DOI: 10.18287/2412–6179-CO-703.
24. Ye Y., Yang C., Zhu B., Zhou L., He Y., Jia H. Improving co-registration for Sentinel-1 SAR and Sentinel-2 optical images. *Remote Sensing*. 2021. 13(5): 928.
25. Yu H., Yang W., Liu Y. Coarse-to-fine accurate registration for airborne Sar images using SAR-FAST and DSP-LATCH. *Progress In Electromagnetics Research*. 2018. 163: 89–106. DOI: 10.2528/PIER18070801.
26. Yu Q., Ni D., Jiang Y., Yan Y., An J., Sun T. Universal SAR and optical image registration via a novel SIFT framework based on nonlinear diffusion and a polar spatial-frequency descriptor. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2021. 171: 1–17.

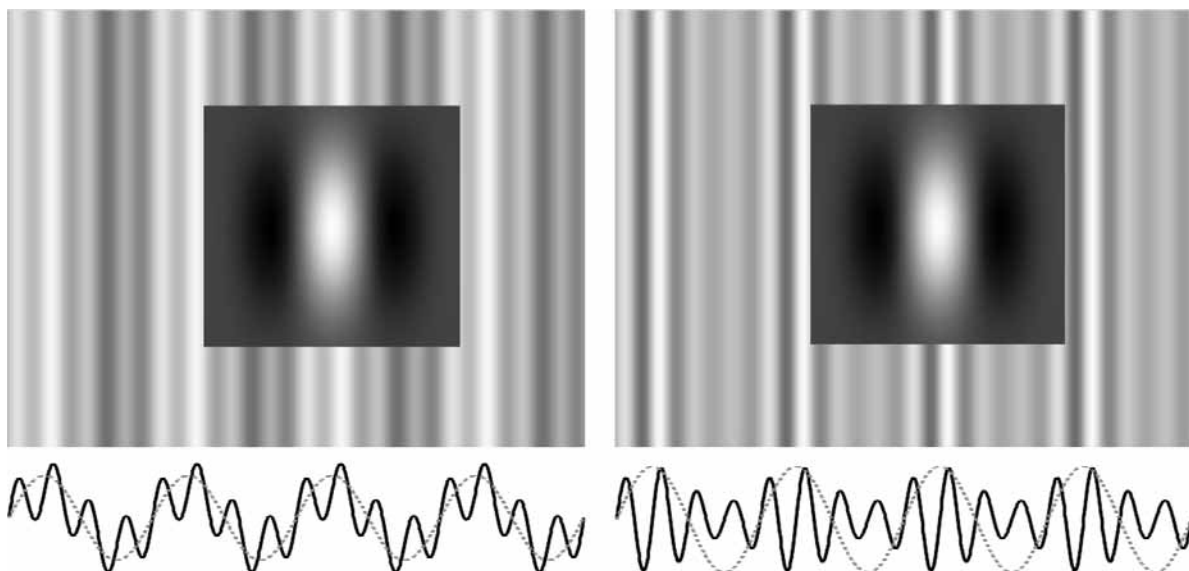
## Chapter 20. The role of spatial modulations of local visual features in a saliency map formation in humans

*Yavna D., Babenko V., Kupriyanov I.*

*Southern Federal University, Rostov-on-Don, Russia*

The second-order visual mechanisms (SOVMs) are a neuronal system that functions in the early stages of visual processing and decodes spatial changes in contrast, orientation, and spatial frequency (SF) in texture. This system provides operations such as segmentation and segregation of textures, determination of the structure, material or shape of the surface (Graham, 2011; Schofield et al., 2010). SOVMs are tuned to detect the so-called second order visual features, which are usually described in terms of spatial modulations of contrast, orientation and spatial frequency in the image. The neurophysiological substrate of these mechanisms has not yet been sufficiently studied, but it is quite clear that the decoding of the above signals of spatial inhomogeneities cannot be carried out at the level of so-called first-order visual mechanisms based on the simple striate neurons.

So, Figure 1 shows the receptive fields (RF) of a simple neuron superimposed on luminance-modulated (left) and contrast-modulated (right) sinusoidal gratings. Excitatory and inhibitory subfields are shown in light and dark, respectively. Taking the sum of all pixels of the RF map equal to 0, the neuron response can be numerically determined as the sum of the element-wise products of the map pixels and the signal pixel values corresponding to them. For a brightness-modulated grating, the response of a neuron will be maximum, while for a contrast-modulated grating it will be close to 0. Decoding of spatial modulations of orientation and SF also cannot be carried out by first-order visual mechanisms, at least due to too small size of the image projected onto the RF of a simple neuron and the absence of lateral interactions between cells with spatially separated RF. It would be reasonable to assume the existence of an overlying mechanism that compares and combines the responses of cells with different orientational and/or frequency tunings.



**Figure 1.** *Modulated sinusoidal gratings with superimposed RF maps of a striate neuron. The horizontal profiles of signal changes are shown below: the solid line is the modulated signal; the dotted line is the modulation function. On the left is the brightness modulation, on the right is the contrast modulation.*

To explain the mechanism of contrast modulation decoding, many researchers use the filtering–rectification–filtering (FRF) scheme, which uses the same principle as the amplitude modulation (AM) demodulator in radio communications: the signal is filtered in the carrier frequency band, then rectified, then passed through low pass filter (LP). Such a scheme can be very easily implemented on a neuronal substrate with only minimal differences from the AM demodulator. A visual cortex neuron is better suited for the role of bandpass filter than an LP filter, therefore, instead of an LP filter, the output of the mechanism can most likely be a cell with a low-frequency bandpass tuning and a sufficiently wide bandwidth. Since the late 1980s, the FRF scheme has been adopted by many researchers under various names (Landy & Graham, 2004); the first in the series of works that proposed it to explain perceptual phenomena was probably the article by Henning et al. (1975), which was somewhat in advance of its time.

The original FRF scheme is not specific to the type of the modulating feature; thus, the modulation decoder (actually AM) also passes the orientation and SF modulation signals. However, a number of psychophysical studies (Ellemberg et al., 2006; Kingdom et al., 2003; Schofield & Cruickshank, 2005) have shown that the processes that decode the modulations of different features in human interact very limitedly with each other. This indicates their possible independence; electrophysiological data also support independence (Babenko & Ermakov, 2015). However, the two-level FRF scheme of SOVMs has not lost its relevance. To solve the specificity problem, some modifications of the traditional scheme have been proposed (Kingdom et al., 2003; Бабенко & Явна, 2018), which have shown their efficiency in computational experiments. Thus, we can consider the SOVMs as a group of relatively independent mechanisms that decode different signals and have a slightly different organization.

While the initial interest in SOVMs was primarily related to their role in pattern vision (Graham, 2011), they have become increasingly important in the context of human and other mammalian visual attention studies (Frey et al., 2007; Johnson & Zarei, 2010; Khastkhodaei et al., 2016). And although there are little experimental data on the role of SOVMs in attention processes, the following assumption looks logical. Since SOVMs function at the initial, preattentive stages of visual processing, the results of their work should be used by the visual system in attention processes, including contributing to the formation of the so-called saliency map (Koch & Ullman, 1985). On the other hand, many authors of computational saliency models, which are not necessarily physiologically based, take into account spatial image inhomogeneities (Bruce & Tsotsos, 2009; Cheng et al., 2015; Valenti et al., 2009). Thus, the idea of the importance of spatial inhomogeneities for saliency is plain to see, and there are good grounds to consider SOVMs as candidates for the role of detectors of these inhomogeneities. Let us consider the problems of saliency modeling in more detail.

Following the tradition of the feature integration theory (Treisman & Gelade, 1980), Koch and Ullman (1985) used the term “saliency map” to denote a hypothetical overall conspicuity map that combines conspicuity information from various feature maps. Koch and Ullman also admitted the possibility of top-down modulating influences on the process of formation of this map. The concept of a visual saliency map was introduced as the fundamental block of visual attention control along with the mechanism of selective mapping. The work of Koch and Ullman formed the theoretical background for the computational saliency modeling; the first of the widely known models appeared in the late 1990s (Itti et al., 1998). As effective computational approaches to modeling were being developed, the possibilities of practical application of saliency models in computer vision, engineering psychology, usa-

bility, analysis of medical images, etc. began to be discussed in literature (Arun et al., 2020; Jampani et al., 2012; Medioni & Mordohai, 2005; Sun et al., 2019). Thus, the modeling of visual saliency has gained practical significance by now. Theoretical understanding of saliency mechanisms has also become deeper. Thus, a large-scale saliency network was identified by magnetic resonance imaging (Xapay30B et al., 2018).

It is possible to conditionally divide approaches to modeling saliency into traditional and based on neural networks. Thanks to use of modern neural network architectures, primarily convolutional ones, in recent years all records for the quality of model training have been broken (Borji, 2019). Finally, the success of neural network models is facilitated by the increase in the amount of data with the results of oculographic studies that are in the public domain, and the emergence of standardized and relatively easy-to-use neural network modeling tools.

In this study, we tested the hypothesis that SOVMs are directly involved in the process of forming a saliency map. The study was carried out by the method of neural network modeling. We do not draw direct analogies between artificial and natural neural networks and are aware that a convolutional network, having some fundamental similarities with the visual system (Малахова, 2017), is based on tensor mathematics and is not at all analogous to a biological system (Chollet, 2017). Although the network is sometimes capable of generating “receptive fields” that are strikingly similar to the RFs of biological neurons (Krizhevsky et al., 2017), nevertheless, our hypothesis is based on somewhat different grounds. If a neural network or other computer model, having learned to search for salient areas of a natural image, is able to detect second-order visual features, this is evidence that it has developed representations of these features as useful for solving the main problem.

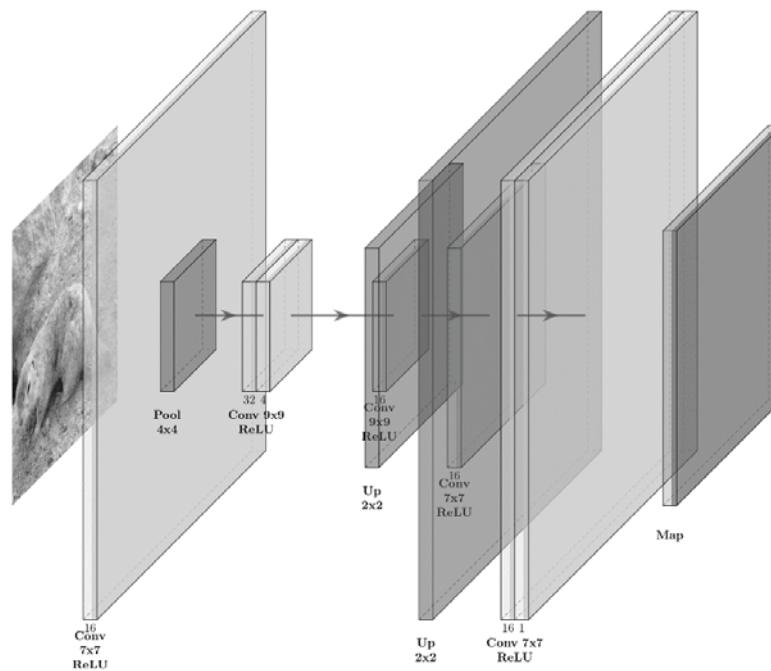
## Methods

*Model.* The model used the architecture of a simple fully convolutional autoencoder (Figure 2). The choice of this architecture and the number of trained parameters is primarily due to the speed of learning and the possibility of subsequent execution of the model on an inexpensive device with a neuroprocessor. Initially, with a similar or slightly worse result, we used similar models with a significantly larger number of parameters, and this model was the result of a consistent simplification that did not lead to worse learning outcomes. The input layer of the model receives the image in a single-channel (grayscale) representation; at the output, the model generates a probabilistic saliency map of the same size as the original raster. The Adam optimizer (Kingma & Ba, 2017) with  $\alpha = 0.00001$ , ReLU activation function, and binary crossentropy as a loss function were used. The fully convolutional architecture of the model allows processing images of any size without scaling, which is especially important when working with saliency maps. The Tensorflow library with the Keras API (Chollet, 2017) for Python was used as a development tool.

*Training data.* We have used two datasets. The first is Salicon (Saliency in Context) (Jiang et al., 2015), based on the user’s mouse movement while viewing specially preprocessed images from the Microsoft COCO (Common Objects in Context) collection (Lin et al., 2015). The second one is oculographic, compiled by the authors of this work from several of freely distributed data sets with results of eye-tracking studies.

Currently Salicon contains 10,000 training and 5,000 validating images, each of which has an empirical probabilistic ground truth salience map. Mapping data was crowdsourced on the Amazon Mechanical Turk platform by the National University of Singapore through their interactive free-viewing paradigm, which was in turn based on simulation method

of Perry and Geisler (2002). The images shown to the crowdsourcing participants were blurry, but by moving the mouse pointer, the observer could increase the detail of the area of interest to him.



**Figure 2.** The scheme of the model used. Conv is the convolutional layer, Pool is the pooling layer, Up is the upsampling layer.

To train the model on oculographic data, we formed a sample that included 11561 test and 2863 validation images with probability ground truth maps created based on the results of several eye-tracking studies with registration of screen coordinates of oculomotor fixations. The validation set contains data from all the sets used in an amount proportional to the size of these sets. The construction of probability maps was carried out either by the authors of the studies used or by the authors of this article using the Gaussian blur method for binary fixation maps (Le Meur & Baccino, 2013). If the same images were found in different data-sets, we chose those for which there were more viewing results. The oculographic datasets are described below.

CAT2000 dataset (Borji & Itti, 2015). The name of the dataset contains the abbreviation for categories, since the images are divided into 20 categories. This set includes 4000 images, but 2000 are intended for internal testing, and information about the eye movements obtained during their viewing is not provided by the authors. The authors of the collection provide both probabilistic fixation maps, as well as initial data and binary maps in mat-format.

Dalian University of Technology & OMRON Corporation (DUT-OMRON) dataset (Ruan et al., 2014). The set includes 5168 color images, being the largest set of its kind known to the authors. The images were selected from the SUN collection (Xiao et al., 2010). Fixation maps are not distributed but can be easily generated based on fixation coordinates provided as separate mat-files for each image.

MIT dataset (Judd et al., 2009). The set consists of 1003 color images of various indoor and outdoor scenes taken from Flickr.com and from the LabelMe collection (Russell et al., 2008). Both probabilistic and binary fixation maps are provided for free download.

EMOtional attention dataset (EMOd) (Fan et al., 2018). The dataset contains 698 photographs collected using the Google Image search service, as well as 321 images from the IAPS (International Affective Picture System) collection (Bradley & Lang, 2017), which can be obtained from the copyright holders upon request. We used only 698 photographs collected by the creators of EMOd. For them, as well as for images from IAPS, both probabilistic and binary fixation maps are freely available.

Fixations in Webpage Images (FiWI) dataset (Shen & Zhao, 2014). This set contains 149 full-screen images of web pages divided into 3 categories: graphic, text and mixed. The authors of the collection provide source data, binary maps, and scripts for Octave/Matlab that make it easy to build probabilistic fixation maps.

Fine-GRained Image Memorability (FIGRIM) fixation dataset (Bylinskii et al., 2015). The set contains 630 target images shown multiple times and 2157 non-target images shown only once. The images are taken from the larger (9428 images) FIGRIM dataset, created for memory research based on the SUN dataset (Xiao et al., 2010) and including 21 categories of different scenes. The authors of the set provide for free download both probability maps in jpeg and binary fixation maps in mat format, as well as the raw individual results obtained in the study and Octave/Matlab scripts for data processing.

National University of Singapore NUS3D-Saliency dataset (Lang et al., 2012). The data set includes 600 images shown to subjects in 2D and 3D modes in the free view task. We only used data related to 2D viewing. The pictures represent various scenes photographed by the authors of the set, both indoors and outdoors, using a Kinect camera, which allows, in addition to the usual raster image, to obtain a point depth map. The authors provide archives with stimuli, 2D and 3D fixation probability maps as well as original and smoothed depth maps for download.

Toronto dataset (Bruce & Tsotsos, 2005). This is one of the earliest freely distributed eye movement datasets, with only 120 images of indoor and outdoor scenes. Some scenes contain fairly salient objects, while others lack distinct areas of interest (Bruce & Tsotsos, 2009). The authors of the set provide original images, probability maps of fixations, and original fixation information.

Eye Fixations in Crowd (EyeCrowd) dataset (Jiang et al., 2014). The set contains 500 images of outdoor and indoor scenes collected using Flickr.com and Google Images. Each image contains groups of people, and their number varies greatly. The data set contains original images, fixation data in mat-format, as well as Octave/Matlab scripts that implement the model proposed by the authors and allow, among other things, to build probabilistic maps of real fixations.

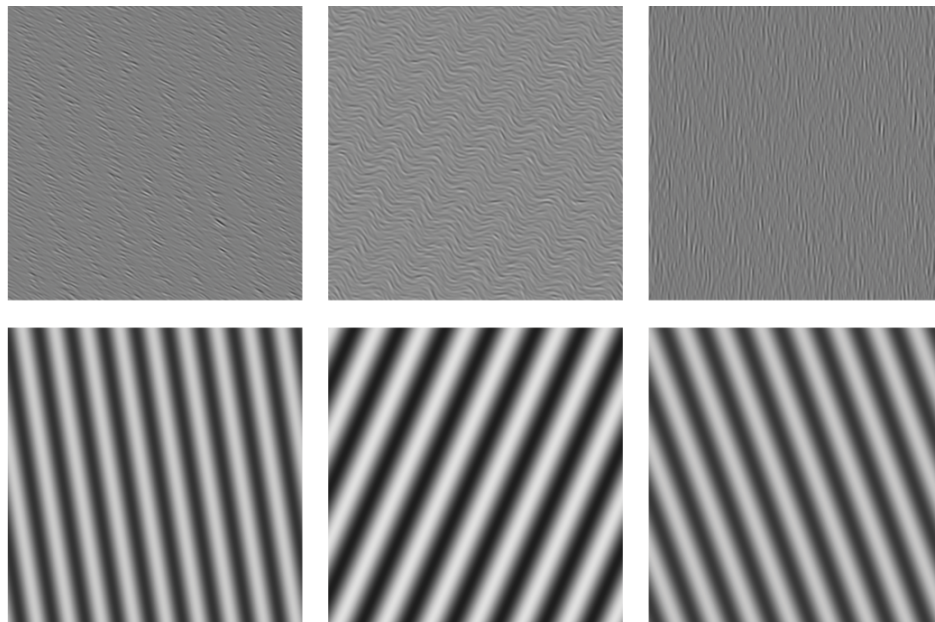
Vision and Image Understanding Lab (VIU) dataset (Eckstein, 2020; Koehler et al., 2014). The set contains 800 images of real scenes photographed by the authors, collected using search engines or taken from a previously created database (Russell et al., 2008). The authors provide free access to images and research results in text and mat formats. Probabilistic maps of fixations obtained in the free viewing task can be easily built from the data with a tabular structure contained in the FreeViewing.txt file.

Object and Semantic Images and Eye-tracking (OSIE) dataset (Xu et al., 2014). The dataset contains 700 images of various everyday scenes taken outdoors and indoors and containing several dominant objects of different semantic categories. The images were either created by the authors or obtained using Flickr and Google Images. Authors freely distribute source

images, experimental data in mat-format and software for working with them in the form of Matlab and C source files, as well as compiled mex files. Unfortunately, there are specific requirements for the software environment to run the code. The authors of this article used the fixation coordinates contained in the mat-file and wrote their own script for building probabilistic maps.

*Test data.* The test sample was generated from modulated textures containing explicitly specified second order visual features. The textures were created by summing randomly positioned raster micropatterns based on 2D Gabor function values (Grigorescu et al., 2003) using the method described by Prins and Kingdom (2003). The Gabor micropattern, which perfectly matches the RF configuration of a simple visual cortex neuron (Jones & Palmer, 1987), is often used as a standard stimulus in psychophysics. Contrast, orientation, and spatial frequency were modulated, respectively, by multiplying the values of the Gabor function by the contrast gain or by varying the values of the Gabor function arguments  $\theta$  or  $\lambda$ , which determine the orientation and spatial frequency of the micropattern (Grigorescu et al., 2003). The contrast gain and the magnitude of changes in  $\lambda$  and  $\theta$ , depending on the spatial position, were set by the value of a two-dimensional periodic function, calculated as a sinusoid with a phase shift, at a point corresponding to the position of the micropattern.

To make sure that the model is invariant to various combinations of texture properties, in addition to the modulated local feature (contrast, orientation, or spatial frequency), the slope of the carrier and envelope axes, the carrier and envelope wavelengths, the phase shift of the envelope, the average brightness, and the overall contrast were randomly varied in the synthesized textures. For each texture, a pixel map of modulating function values was created. The pixel size of all images and maps was  $640 \times 640$ . The examples of the modulated textures and maps are shown in Figure 3.



**Figure 3.** An example of pairs of images from a test sample. Above is a modulated texture, below is a map of the modulating function values. From left to right: modulations of contrast, orientation, and SF.

*Evaluation metrics.* The following distribution-based metrics (Bylinskii et al., 2017) were used to evaluate the models:

1. Similarity metric or histogram intersection (*SIM*). The metric compares the result predicted by the model and the empirically derived probabilistic map as histograms, normalizing the values of each pixel so that the sum of all values of each raster is equal to 1. Then both normalized histograms are compared pointwise, and for each corresponding pair of pixels, the smaller of two values is determined; the results obtained are pooled. The final value of 1 indicates a complete match of the model with empirical data, while 0 indicates a complete lack of similarity.

2. Kullback-Leibler divergence (*KL*). This metric evaluates differences between probability distributions. There are several varieties of this metric; here we consider the variant described by Bylinskii et al. (2017) as KL-Judd and implemented in Python by Zanca et al. (2018):

where  $\epsilon = \textit{eps}$  is a regularization constant introduced to avoid the appearance of zero values. The smaller the value of *KL*, the closer the generated map *P* to ground truth map *Q*.

3. Pearson's correlation coefficient (*CC*). This metric is widely used in statistics to assess the closeness of the probabilistic relationship of two variables to a linear one. When the variables are linearly related, the coefficient is |1|, in the absence of a relation it is 0. A coefficient value close to 1 indicates the coincidence of the simulation result with the ground truth map. The shift of maps relative to each other can greatly affect the *CC* value and sometimes even the sign of the coefficient.

4. Earth mover's distance (*EMD*). This metric, also called Wasserstein distance (Добрушин, 1970), takes into account the spatial distance between the simulated and empirical maps, measuring the distance between two probability density distributions. Intuitively speaking, this metric considers the probability density as the amount of conditional "ground" in the columns of discrete distributions and measures the minimum "cost" of transforming the model distribution into an empirical one; the cost is directly proportional to the probability density and the distance by which this density must be "drag". We adopted a type of metric used in MIT/Tübingen Saliency Benchmark (Bylinskii et al., 2017):

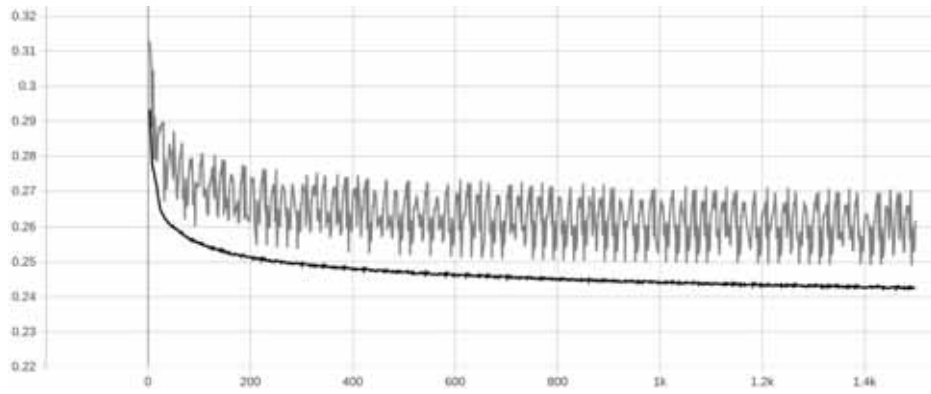
where the following conditions should be met:

Here *i* and *j* are not indexes of columns and lines in raster, but indexes of bins in discrete distributions. Each *f* value is the value of the density transferred from *i*-th to *j*-th bin, and the *d* is the distance between these bins.

A fairly quick EMD calculation algorithm was implemented by Pele and Werman (2009) in C++. Based on this code, they created the package for Python. The mex-compiled code is also used by Z. Bylinskii in her Matlab implementation of the EMD metric. To call the function of O. Pele and M. Werman, it is necessary to form a matrix of distances between the distributions bins; to do this, we rewrote Z. Bylinskii's Matlab code in Python.

## Model training

The training was aimed at the fitting the weights that ensure the maximum coincidence of the generated map with a ground truth one. The training process was carried out for 1,500 epochs with preservation of the weights which provided the best result on validation data. The training was conducted on the NVIDIA Tesla V100 16GB card, since it was necessary to provide a fairly large batch size: with a decrease the size, the dynamics of the loss function on validation sample became very unstable. Dynamics of training the model on Salicon dataset with a batch size of 256 images is shown in the Figure 4.



**Figure 4.** The dynamics of loss function during training the model on Salicon dataset. The black curve is for training data, and the gray curve is for validation one.

## Results and discussion

After training, we conducted preliminary testing to evaluate the quality of our models. Tests were carried out using the center bias correction, calculated by Kroner et al. (2020) on MIT data. The model trained on oculographic data was tested on Salicon dataset with the following results:  $SIM = 0.562$ ,  $KL = 0.659$ ,  $CC = 0.377$ ,  $EMD = 1.063$ . The model trained on Salicon demonstrated the following results on MIT data:  $SIM = 0.345$ ,  $KL = 1.429$ ,  $CC = 0.286$ ,  $EMD = 2.412$  and these results on CAT data:  $SIM = 0.448$ ,  $KL = 1.103$ ,  $CC = 0.310$ ,  $EMD = 3.103$ . The best results were achieved on VIU data:  $SIM = 0.519$ ,  $KL = 0.798$ ,  $CC = 0.316$ ,  $EMD = 0.790$ . Based on reference values of MIT/Tuebingen Saliency Benchmark, with consideration of simplicity of the model, we can admit that the training was successful. An example of how the model works is shown in the Figure 5.



**Figure 5.** Examples demonstrating the work of the model trained on Salicon. From top to bottom: the original image; ground truth map; saliency map created by the model. The original images and maps are taken respectively from the Microsoft COCO and Salicon collections.

The trained models were tested on modulated textures (Table 1). Both models were able to respond well to all spatial modulations. Regarding to the modulations of contrast and spatial frequency, they actually act as demodulators reproducing even the signal phase. To a greater extent, this relates to contrast modulations. High-contrast areas are usually more salient than low-contrast ones. To a somewhat lesser extent, this also relates to spatial frequency modulations: low-frequency regions are usually more salient, but a human eye is quite often drawn to high-frequency regions as well. Indeed, in an eye-tracking study, one can observe gaze movements between the extrema of the function modulating the spatial frequency (Yavna et al., 2018).

**Table 1.** The results of testing the models on modulated textures ( $n = 960$ ). 1 – the model was trained on eye movements data, 2 – on Salicon.  $X$  is an average,  $s$  is a standard deviation.

	Contrast modulation				Orientation modulation				SF modulation			
	1		2		1		2		1		2	
	$X$	$s$	$X$	$s$	$X$	$s$	$X$	$s$	$X$	$s$	$X$	$s$
SIM	0.788	0.057	0.752	0.069	0.629	0.085	0.614	0.136	0.681	0.055	0.712	0.070
KL	0.099	0.067	0.159	0.102	0.220	0.144	0.277	0.254	0.219	0.098	0.188	0.102
CC	0.662	0.220	0.576	0.268	-0.011	0.390	-0.030	0.550	0.321	0.236	0.411	0.318
EMD	0.277	0.119	0.429	0.200	0.445	0.185	0.490	0.213	0.462	0.153	0.469	0.205

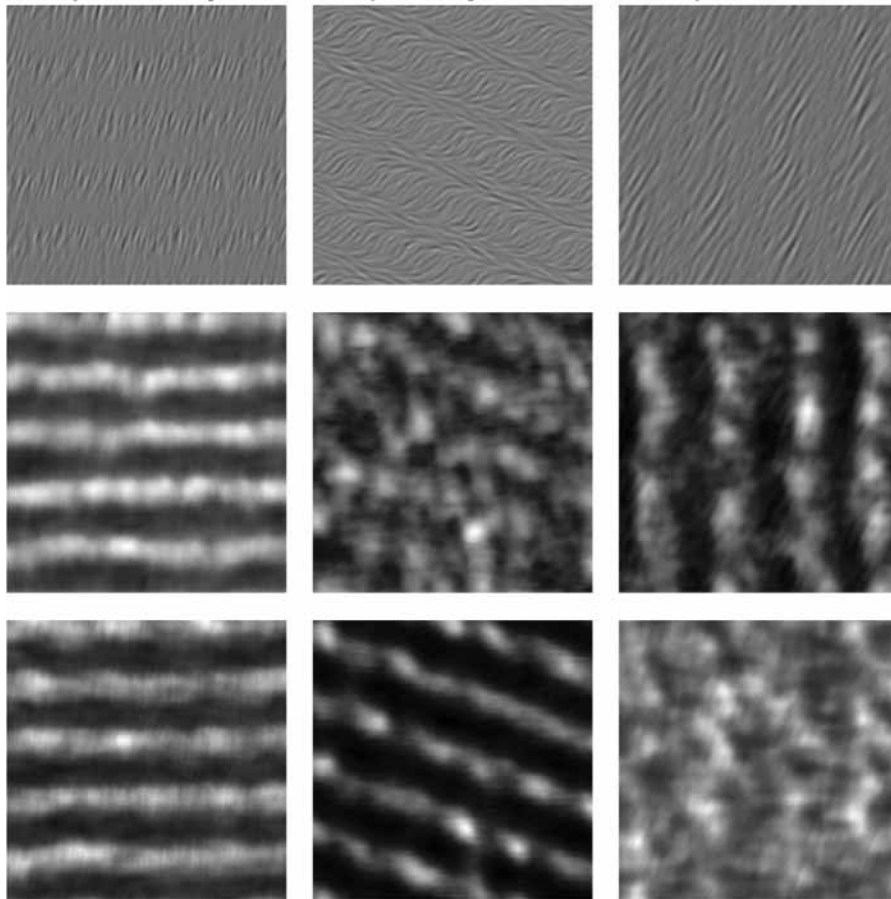
Zero averaged correlation with a very large deviation of coefficient values in tests performed on orientationally modulated textures, while the values of other similarity metrics are quite large, reflects the complex nature of the “phase shift” of saliency relative to the extrema of the orientation modulation function. Apparently, this offset depends on the orientation of the carrier and envelope. According to our preliminary eye-tracking data, areas where the orientations of the carrier and envelope coincide are more preferable for attention; this statement is supported statistically (Yavna et al., 2018), but a detailed direct empirical study of this issue is desirable.

Thus, the model trained on a relatively large and rather heterogeneous set of oculographic data has developed a mechanism for detecting second-order features. The same applies to the model trained on Salicon, the result of which is somewhat inferior to the first, but nevertheless remains high.

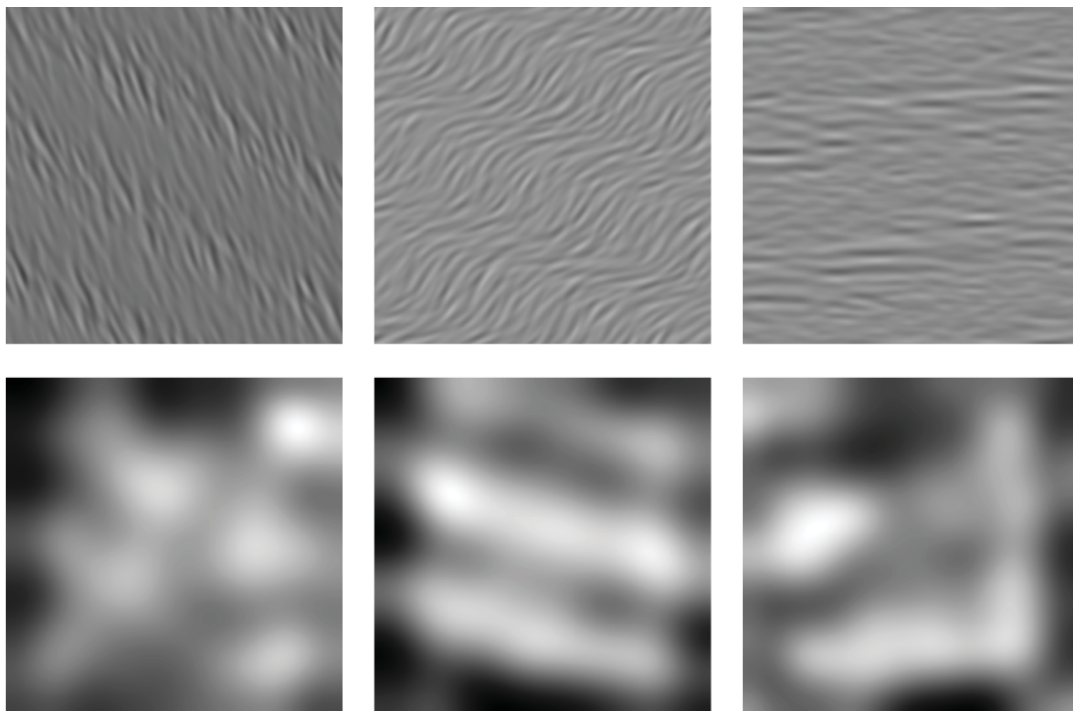
It was not a surprise for us that not only our models can find the second order features in an image. For example, the Deepgaze II model (Kummerer et al., 2017) usually prefers high contrast (surrounded by low contrast) and low frequency (surrounded by high frequency) areas of images. With certain ratios of orientation of the carrier and envelope, the model detects orientation modulations quite well (Figure 7). To a slightly lesser extent, but nevertheless, the more modern Unisal (Droste et al., 2020) model has such properties too.

## Conclusion

It was shown that the computer model, having learned to search for the salient areas of the image, is able to detect the second order visual features. This is true regardless of nature of the training data. It can be both probabilistic oculomotor fixation maps built using modern eye tracking technologies, and saliency maps generated in the specially organized procedure for collecting coordinates of computer manipulator movements (Salicon). The result indicates the importance of second-order visual features for saliency modeling and confirms our hypothesis. Thus, models of visual mechanisms of the second order can act as modules in the construction of complex models of saliency.



**Figure 6.** The demodulation of spatial modulation signals by our saliency models (from top to bottom): original modulated textures; signals demodulated by the model trained on eye movements data; signals demodulated by the model trained on Salicon.



**Figure 7.** The demodulation of spatial modulation signals by the DeepGaze II saliency model (from left to right): original modulated textures; signals demodulated by the model.

Modern neural networks, while providing relatively high performance in saliency modeling, nevertheless do not yet reach human performance in the corresponding tasks and are not universal, demonstrating good results often only on certain sets of saliency data. At the same time, the application of knowledge about the mechanisms of biological vision in combination with modern deep learning technologies would contribute to the emergence of more advanced and, importantly, less resource-intensive saliency models and machine vision systems based on them. With our work, we hope to make a certain contribution to the development of such systems.

From a psychophysiological point of view, our results can be considered as another confirmation that second-order visual mechanisms play an important role in the processes of deploying the visual attention of a human observer.

## References

1. Бабенко, В. В., & Явна, Д. В. (2018). Конкуренция за внимание пространственных модуляций градиентов яркости. *Российский Психологический Журнал*, 15(3), 160–189. <https://doi.org/10.21702/rpj.2018.3.8>
2. Добрушин, Р. Л. (1970). Задание системы случайных величин при помощи условных распределений. *Теория Вероятностей и Её Применения*, 15(3), 469–497.
3. Малахова, Е. Ю. (2017). Представление информации в нейронных сетях при распознавании семантики изображений. *Материалы XXIII Съезда Физиологического Общества Имени И. П. Павлова*, 1642–1644.
4. Хараузов, А. К., Васильев, П. П., Шелепин, Ю. Е., Соколов, А. В., & Фокин, В. А. (2018). Взаимодействие крупномасштабных нейронных сетей при распознавании текстур. In *Нейротехнологии* (pp. 244–256). ООО “Издательство BBM.” <https://www.elibrary.ru/item.asp?id=41144088>
5. Arun, N. T., Gaw, N., Singh, P., Chang, K., Hoebel, K. V., Patel, J., Gidwani, M., & Kalpathy-Cramer, J. (2020). Assessing the validity of saliency maps for abnormality localization in medical imaging. *ArXiv:2006.00063 [Cs]*. <http://arxiv.org/abs/2006.00063>
6. Babenko, V. V., & Ermakov, P. N. (2015). Specificity of brain reactions to second-order visual stimuli. *Visual Neuroscience*, 32, E011. <https://doi.org/10.1017/S0952523815000085>
7. Borji, A. (2019). Saliency Prediction in the Deep Learning Era: Successes, Limitations, and Future Challenges. *ArXiv:1810.03716 [Cs]*. <http://arxiv.org/abs/1810.03716>
8. Borji, A., & Itti, L. (2015). CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research. *ArXiv:1505.03581 [Cs]*. <http://arxiv.org/abs/1505.03581>
9. Bradley, M. M., & Lang, P. J. (2017). International Affective Picture System. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–4). Springer International Publishing. [https://doi.org/10.1007/978-3-319-28099-8\\_42-1](https://doi.org/10.1007/978-3-319-28099-8_42-1)
10. Bruce, N. D. B., & Tsotsos, J. K. (2005). Saliency Based on Information Maximization. *NIPS*.
11. Bruce, N. D. B., & 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>
12. Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, 116, 165–178. <https://doi.org/10.1016/j.visres.2015.03.005>
13. Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2017). What do different evaluation metrics tell us about saliency models? *ArXiv:1604.03605 [Cs]*. <http://arxiv.org/abs/1604.03605>
14. Cheng, M.–M., Mitra, N. J., Huang, X., Torr, P. H. S., & Hu, S.–M. (2015). 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>
15. Chollet, F. (2017). *Deep Learning with Python* (1st ed.). Manning Publications Co.
  16. Droste, R., Jiao, J., & Noble, J. A. (2020). Unified Image and Video Saliency Modeling. *ArXiv:2003.05477 [Cs]*, 12350, 419–435. [https://doi.org/10.1007/978-3-030-58558-7\\_25](https://doi.org/10.1007/978-3-030-58558-7_25)
  17. Eckstein, M. (2020). “What do saliency models predict?” Koehler, Guo, Zhang, Eckstein, *Journal of Vision*, 2014 [Data set]. Mendeley. <https://doi.org/10.17632/8RJ98PP6KM.1>
  18. 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>
  19. Fan, S., Shen, Z., Jiang, M., Koenig, B. L., Xu, J., Kankanhalli, M. S., & Zhao, Q. (2018). Emotional Attention: A Study of Image Sentiment and Visual Attention. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7521–7531. <https://doi.org/10.1109/CVPR.2018.00785>
  20. 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>
  21. 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 Res.*, 51(13), 1397–1430. <https://doi.org/10.1016/j.visres.2011.02.007>
  22. Grigorescu, C., Petkov, N., & Westenberg, M. A. (2003). Contour detection based on nonclassical receptive field inhibition. *IEEE Trans. Image Processing*, 12(7), 729–739.
  23. Henning, G. B., Hertz, B. G., & Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyses spatial patterns in independent bands of spatial frequency. *Vision Res.*, 15(8–9), 887–897.
  24. Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259. <https://doi.org/10.1109/34.730558>
  25. Jampani, V., Ujjwal, Sivaswamy, J., & Vaidya, V. (2012). Assessment of computational visual attention models on medical images. *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing–ICVGIP ’12*, 1–8. <https://doi.org/10.1145/2425333.2425413>
  26. Jiang, M., Huang, S., Duan, J., & Zhao, Q. (2015). SALICON: Saliency in Context. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1072–1080. <https://doi.org/10.1109/CVPR.2015.7298710>
  27. Jiang, M., Xu, J., & Zhao, Q. (2014). Saliency in Crowd. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision–ECCV 2014* (pp. 17–32). Springer International Publishing. [https://doi.org/10.1007/978-3-319-10584-0\\_2](https://doi.org/10.1007/978-3-319-10584-0_2)
  28. Johnson, A., & Zarei, A. (2010). Second-order saliency predicts observer eye movements when viewing natural images. *Journal of Vision*, 10(7), 526–526. <https://doi.org/10.1167/10.7.526>
  29. Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6), 1233–1258.
  30. Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. *2009 IEEE 12th International Conference on Computer Vision*, 2106–2113. <https://doi.org/10.1109/ICCV.2009.5459462>
  31. Khastkhodaei, Z., Jurjut, O., Katzner, S., & Busse, L. (2016). Mice can use second-order, contrast-modulated stimuli to guide visual perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 36(16), 4457–4469. <https://doi.org/10.1523/JNEUROSCI.4595-15.2016>
  32. Kingdom, F. A. A., Prins, N., & Hayes, A. (2003). Mechanism independence for texture-modulation detection is consistent with a filter-rectify-filter mechanism. *Vis. Neurosci.*,

- 20(1), 65–76. <https://doi.org/10.1017/S0952523803201073>
33. Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *ArXiv: 1412.6980 [Cs]*. <http://arxiv.org/abs/1412.6980>
  34. Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4(4), 219–227.
  35. Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of Vision*, 14(3), 14–14. <https://doi.org/10.1167/14.3.14>
  36. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
  37. Kroner, A., Senden, M., Driessens, K., & Goebel, R. (2020). Contextual encoder–decoder network for visual saliency prediction. *Neural Networks*, 129, 261–270. <https://doi.org/10.1016/j.neunet.2020.05.004>
  38. Kummerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017). Understanding low- and high-level contributions to fixation prediction. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 4789–4798.
  39. Landy, M., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences* (pp. 1106–1118). MIT Press.
  40. Lang, C., Nguyen, T. V., Katti, H., Yadati, K., Kankanhalli, M., & Yan, S. (2012). Depth Matters: Influence of Depth Cues on Visual Saliency. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Eds.), *Computer Vision—ECCV 2012* (Vol. 7573, pp. 101–115). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-33709-3\\_8](https://doi.org/10.1007/978-3-642-33709-3_8)
  41. Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behavior Research Methods*, 45(1), 251–266. <https://doi.org/10.3758/s13428-012-0226-9>
  42. Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). Microsoft COCO: Common Objects in Context. *ArXiv:1405.0312 [Cs]*. <http://arxiv.org/abs/1405.0312>
  43. Medioni, G., & Mordohai, P. (2005). Saliency in Computer Vision. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 583–585). Academic Press. <https://doi.org/10.1016/B978-012375731-9/50099-9>
  44. Pele, O., & Werman, M. (2009). Fast and robust earth mover’s distances. *2009 IEEE 12th International Conference on Computer Vision*, 460–467. <https://doi.org/10.1109/ICCV.2009.5459199>
  45. Perry, J. S., & Geisler, W. S. (2002). *Gaze-contingent real-time simulation of arbitrary visual fields* (B. E. Rogowitz & T. N. Pappas, Eds.; pp. 57–69). <https://doi.org/10.1117/12.469554>
  46. Prins, N., & Kingdom, F. A. A. (2003). Detection and discrimination of texture modulations defined by orientation, spatial frequency, and contrast. *J. Opt. Soc. Am. A*, 20(3), 401–410.
  47. Ruan, X., Tong, N., & Lu, H. (2014). How far we away from a perfect visual saliency detection-DUT-OMRON: a new benchmark dataset. *FCV2014*. 20th Korea-Japan Joint Workshop on Frontiers of Computer Vision. <http://saliencydetection.net/dut-omron/download/FCV2014.pdf>
  48. Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision*, 77(1), 157–173. <https://doi.org/10.1007/s11263-007-0090-8>
  49. Schofield, A. J., & Cruickshank, A. (2005). Transfer of tilt after-effects between second-order cues. *Spatial Vision*, 18(4), 379–397. <https://doi.org/10.1163/1568568054389624>
  50. Schofield, A. J., Rock, P. B., Sun, P., Jiang, X., & Georgeson, M. A. (2010). What is second-order vision for? Discriminating illumination versus material changes. *Journal of Vision*, 10(9), 2–2. <https://doi.org/10.1167/10.9.2>

51. Shen, C., & Zhao, Q. (2014). Webpage Saliency. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision—ECCV 2014* (Vol. 8695, pp. 33–46). Springer International Publishing. [https://doi.org/10.1007/978-3-319-10584-0\\_3](https://doi.org/10.1007/978-3-319-10584-0_3)
52. Sun, X., Houssin, R., Renaud, J., & Gardoni, M. (2019). A review of methodologies for integrating human factors and ergonomics in engineering design. *International Journal of Production Research*, 57(15–16), 4961–4976. <https://doi.org/10.1080/00207543.2018.1492161>
53. Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
54. Valenti, R., Sebe, N., & Gevers, T. (2009). Image saliency by isocentric curvedness and color. *2009 IEEE 12th International Conference on Computer Vision*, 2185–2192. <https://doi.org/10.1109/ICCV.2009.5459240>
55. Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492. <https://doi.org/10.1109/CVPR.2010.5539970>
56. Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision*, 14(1), 28–28. <https://doi.org/10.1167/14.1.28>
57. Yavna, D. V., Alekseeva, D. S., & Babenko, V. V. (2018). Gaze fixations on spatially modulated textures. *10th Conference of the Lithuanian Neuroscience Association / LNA and 2nd International Symposium on Visual Physiology, Environment and Perception / VISPEP 2018*, 77.
58. Zanca, D., Serchi, V., Piu, P., Rosini, F., & Rufa, A. (2018). FixaTons: A collection of Human Fixations Datasets and Metrics for Scanpath Similarity. *ArXiv:1802.02534 [Cs]*. <http://arxiv.org/abs/1802.02534>

## Chapter 21. Visual Rehabilitation Technologies Based on Neuroplasticity

**Zueva M. V., Neroeva N. V., Kotelin V.I., Zhuravleva A. N., Tsapenko I. V., Fadeev D. V.**

*Helmholtz National Medical Research Center of Eye Diseases, Moscow, Russia*

### Introduction

Retinal degenerative diseases related to the age include age-related macular degeneration (AMD), primary open-angle glaucoma (POAG), and diabetic retinopathy (DR)—the leading causes of vision loss and blindness. The lack of effective therapy requires identifying of new approaches for therapeutic strategies (Choudhary M., Malek G., 2016). In neurodegenerative retinal diseases, non-medicinal methods of neuroprotection and neurorehabilitation are widely used in order to protect and restore the structure and function of nervous tissue. Various visual rehabilitation methods based on the neuroplasticity are currently developed aimed to restore damaged neural networks and compensate functional deficits (Serruya M.D., Kahana M.J., 2008; Baroncelli L. et al., 2010; Rosa A.M. et al., 2013; Alwis D.S., Rajan R., 2014; Baroncelli L., Lunghi C., 2021). The methods of neuroprotection and neurorehabilitation slow down the clinical course of degenerative disease by suppressing primary events (neuroprotection) or by enhancing regenerative mechanisms (neurorehabilitation) (Francardo V. et al., 2017; Gidday J.M., 2018). Neuroprotection applied at the early stage of the disease activates adaptive plasticity slowing down the disease progression. In more advanced stages changes in the retina and brain reflect the non-adaptive plasticity (Francardo V. et al., 2017; Gidday J.M., 2018).

Structural plasticity of the retina (remodeling) is well-described in various retinal diseases, including POAG, AMD, DR, and others. Neurodegenerative diseases lead to the lesions of dendrites and axons of retinal neurons, first of all, of bipolar cells (BC) and ganglion cells (RGCs). These cells lose synaptic connections and become apoptotic. It is the more important because inter-neuronal connections in the retina are very plastic (O'Brien J., Bloomfield S.A., 2018) and play a key role in visual processing.

Visual impairment (due to glaucoma, AMD and corneal opacities) occurs in more than 80% of people aged 60 and over (Bourne R.R. et al., 2013). Methods of visual rehabilitation include a wide range of therapeutic measures aimed at improving the quality of life of patients. Approaches of visual rehabilitation include visual prosthetics, cell and gene therapy, optogenetics, etc. (Menon A., Vijayavenkataraman S., 2022; Cehajic-Kapetanovic J. et al., 2022; Cuenca N. et al., 2014; Zhang X. et al., 2020).

### Visual Restoration Therapy

In different computer technologies of vision perception training, stimuli are presented in the blind zone and at the border of the blind and intact zones, while the observer performs the task of identifying the useful signal or differences between the signals and tracking moving stimuli. Compensatory therapy includes long-term (over several weeks) patient education, which makes it possible to effectively analyze visual information (Trauzettel-Klosinski S. et al., 2011; Jaquin-Courtois S. et al., 2013; Sahraye A. et al., 2020). Computer training programs for visual rehabilitation use methods of augmented (AR) and virtual (VR) reality. It has been shown (Dehn L.B. et al., 2020) that VR cognitive training improved spatial visual memory, visual fields and symptoms of depression. These results confirm the promise of using visual rehabilitation methods in AR and VR conditions. However, further study of the mechanisms of therapeutic effects and optimal conditions and training regimes are required.

In patients with homonymous visual field defects, visual functions can be improved with the help of intensive visual training in the zone of residual vision, i.e., in the area of relative scotoma (Zihl J., von Cramon D., 1979; Kasten E., Sabel B.A., 1995; Sabel B.A. et al. 2011). It has been shown that such training (VRT method—vision restoration therapy) expands residual vision in patients with postchiasmatic visual defects, improves contrast sensitivity and visual fields. Sabel, Kasten and

colleagues have proposed various computer-assisted treatments for patients with cerebral blindness, demonstrating improvement in visual function and visual field defects after severe traumatic brain injury (TBI) and cerebral stroke (Kasten E., Sabel B.A., 1995; Kasten E. et al., 2000; Sabel B.A., Castaing E., 2000; Ajina S., Kennard C., 2012). In patients with glaucoma, ART training led to an increase in photosensitivity, but only when using a high-resolution perimeter rather than standard perimeters (Gudlin J. et al., 2008; Sabel B.A., 2014).

### **Biofeedback Microperimetry**

For visual rehabilitation of patients with visual impairment due to AMD, a positive effect of visual training and other methods based on neuroplasticity on the reorganization of neural connections in the primary visual cortex has been shown (Tarita-Nistor L. et al., 2009; Plank T. et al., 2014; Maniglia M. et al., 2018). These trainings resulted in some improved visual acuity and fixation stability in patients with central vision loss, and these changes correlated with a moderate change in the BOLD fMRI response in the primary visual cortex (Plank T. et al., 2014).

In the macular zone pathologies, other methods of visual rehabilitation based on the brain and visual system plasticity include strategies of biofeedback rehabilitation with using a microperimeter with or without analysis of VEP in real time.

Patients with low visual acuity and unstable fixation use extrafoveal retinal regions (preferred retinal loci – PRL) to compensate for the loss of central vision. The use of microperimeter biofeedback systems (Nidek MP-1) in AMD patients with eccentric fixation improved their fixation stability and reading ability (Vingolo E.M. et al., 2007; Vingolo E.M., Salvatore S., Cavarretta S., 2009). More recent publications verified a high potential of this biofeedback method in patients with macular pathology, confirming an increase in macular sensitivity, visual acuity and the PRL relocation (Meleth A.D. et al., 2011; Vingolo E.M., Salvatore S., Limoli P.G., 2013; Morales M.U. et al., 2015; Sborgia G. et al., 2020; Qian T. et al., 2022). Recently, a training of fixation with the microperimeter biofeedback (MP-3) in AMD (7 patients) and myopic maculopathy (10 patients) (Qian T. et al., 2022) was showed a significant improvement in best corrected visual acuity (BCVA), central sensitivity, reading speed, and fixation stability.

The acoustic biofeedback method based on VEP also promoted the PRL relocation from the affected macular zone to the intact adjacent area of the retina, and it was associated with the primary visual cortex re-modeling (Verdina T. et al., 2020). The efficiency of acoustic biofeedback training system based on real time VEP was found for visual functions in patients with amblyopia, AMD, and TBI (Esposito Veneruso P. et al., 2014; Chiari M. et al., 2014; Verdina T. et al., 2020). The results showed that VEP biofeedback visual rehabilitation in the AMD patients can improve visual acuity, and contrast sensitivity, and retinal fixation, improving reading ability and quality of life.

### **Stimulation Therapy for Visual Rehabilitation**

Other approach for restoring vision in retinal pathology and cortical blindness involves the use of different ways of non-invasive stimulation therapy. They include transcranial electrical stimulation (TcES), transcorneal (TES) or transdermal (TdES) electrical and magnetic stimulation and light therapy.

Transcranial direct current stimulation (tDCS) combined with the VRT training was demonstrated to reduce defects in the visual field and improve physical activity in patients with chronic cerebral stroke (Plow E.B. et al., 2012). TdES with biphasic pulses (10 ms, 20 Hz, 1.0 mA) after 2-weeks course improved the visual acuity and MD-index in patients with retinitis pigmentosa (Miura G. et al., 2019). However, other studies did not mark a statistically significant increase in visual functions and in visual field after prolonged course of transcorneal electrical stimulation in patients with advanced retinitis pigmentosa (Schatz A. et al., 2017; Wagner S.K. et al., 2017). Transcranial random noise stimulation (tRNS) and anodic transcranial direct current stimulation (a-tDCS) have also been used for visual rehabilitation (Herpich F. et al., 2019). It has been shown that 10 days of visual training in combination with tRNS may be more effective than either method alone (Herpich F. et al., 2019).

Thus, studies indicate a high potential for combined approaches that include various types of electrical stimulation and visual training. Such combinations can significantly reduce the training course (from several months to weeks). In addition, research results indicate a high ability of the visual cortex of an adult to change upon stimulation. However, in most of these works there is no convincing evidence of the positive effects of the studied method of improving vision, there is no sufficiently good evidence base. Van Nispen and coauthors (van Nispen R. M. A. et al., 2020) studied the evidence base of 44 clinical trials and made a conclusion that using different methods of visual rehabilitation in patients with low vision do not have a significant impact to health-related quality of life. However, they found evidence that psychological therapy and some methods of improving vision can increase vision-related quality of life, compared with traditional treatment.

Physical training promotes neuronal survival, optimizes physiological responses to social and physical stress factors, and stimulates synaptic plasticity, neurogenesis and angiogenesis (Mahalakshmi B. et al., 2020). Learning new languages, playing musical instruments, and playing computer games have been shown to slow age-related cognitive decline (Green C. S., Bavelier D., 2007). However, the efficiency of the training programs is not good enough and was not verified with the evidence-based medicine (Dockx K. et al., 2016). Thus, the search for new strategies of visual rehabilitation based on reactivation of the visual system plasticity is relevant now. The reactivation (maximization) of retina and brain plasticity is needed to provide their maximal response to therapeutic impact and homeostatic reorganization of neural networks and visual functions in patients with neurodegenerative retinal pathology.

The high brain sensitivity to external rhythmic force and capacity to synchronize its dominant activity with the external rhythms are well-known (Barlow J. S., 1960; Oster G., 1973; Huang T. L., Charyton C., 2008; Tang H.-Y. et al., 2014, 2016; Zhang Y. S. et al., 2016). For neurorehabilitation, in patients with personality and sleep disorders, gait disturbance and cognitive decline, brain injuries sensory (audio, visual) rhythmic stimulation is widely applied. The mechanism of such influences is related with the “brainwave entrainment” (BWE) (Tang H.-Y. et al., 2014, 2016; Zhang Y. S. et al., 2016; Srinivasan A. et al., 2020), and the methods of sensory stimulation consider that various pathological conditions are characterized by various dominating fluctuations in the ongoing brain activity (Shin Y. W. et al., 2011; Uhlhaas P. J. et al., 2008; Westlake K. P. et al., 2012; Laaksonen K. et al., 2013; Tang H.-Y. et al., 2015; Crowell A. L. et al., 2012; Heinrichs-Graham E. et al., 2014; Herz D. M. et al., 2014; Andreou C. et al., 2014;).

Light stimulation of low intensity under certain parameters can influence brain activity and improve episodic memory in adults. Visual flickering stimulation in the frequency range of 9.5–11.0 Hz changed spontaneous cortical activity on the EEG, improving episodic memory in young and old people (Williams J. H., 2001; Williams J. et al., 2006). This effect was sensitive to the stimulation regimen. In cognitively healthy elderly, flicker stimulation improved word recall in the long term after memorization, but only when 10.2 Hz stimuli were applied. At flicker frequencies below 9.0 Hz and above 11.0 Hz, no positive effect was observed. (Williams J. et al., 2006).

It should be noted that the rhythms of sound or light signals of a constant frequency in stimulation therapy can improve electrical activity in the local EEG range, but cannot restore the fractal dynamics of brain activity characteristic of a healthy person.

We believe that the sign of the effect can differ for rhythm with strongly determined frequency and for nonlinear fluctuations of the stimuli parameters, such as deterministic-chaotic (named "fractal") behavior.

### **Impact of Natural and Artificial Fractals**

Fractal geometry and natural noise affect the adaptation of living systems to environmental conditions (Halley J. M., Inchausti P., 2004; Dey S. et al., 2016). Fractal noise (pink noise or flickering noise) has been shown to describe river flow dynamics, fluctuations in climate data, traffic, human speech and music, and more (Halley J. M., 1996; Storch D. et al., 2002; Vasseur D. A., Yodzis P., 2004).

There is an extensive evidence base for the fractal structure of natural objects, geometry and

physiological processes in a healthy person, characterized by the properties of scale invariance in space and time (Mandelbrot B., 1983; Goldberger A.L., Rigney D.R., West B.J., 1990; Goldberger A.L. et al., 2002; Ivanov P. Ch. et al., 1998; Hausdorff J.M. et al., 1995; Taylor R.P. et al., 2005, 2011; Sejdić E., Lipsitz L.A., 2013; Zueva M.V. 2013, 2015; Di Ieva A. et al., 2014). Dynamics of healthy physiological processes including activity of a healthy brain are characterized by the fractal complexity (Teich M.C. et al., 1997; Lehnertz K., 1999; Rabinovich M.I., Abarbanel H.D.I., 1998; Goldberger A.L. et al., 2002; Tan C.O. et al., 2009; Manor B., Lipsitz L.A., 2013; Di Ieva A. et al., 2015). It is fundamentally important that the fractal dynamics of physiological processes is lost in diseases (Peng C.K. et al., 2002; Goldberger A.L., 1997; Hausdorff J.M. et al., 1995; 2001) and is replaced by weakly correlated or strongly periodic (deterministic) fluctuations with a constant frequency (Manor B., Lipsitz L.A., 2013).

The fractal dimension (D) (Mandelbrot B., 1983) estimates the complexity of fractal structures and noise. Natural fractals have D in the range of 1.3–1.9. Most natural fractals, such as tree branches, river beds, seashores, mountain ranges, etc., have typical average D values from 1.3 to 1.5 (Taylor R.P. et al., 2011). The fractal environment largely determines human behavior visual processing and perception that depend on visual complexity of the objects (Field D.J., 1987; Field D.J., Brady N. 1997; Spehar B. et al., 2003; Taylor R.P. et al., 2011).

Animal studies describe a positive for visual activity and structure effects of a multimodal training included an environment enriched with complex sensory stimuli, cognitive tasks, and motor activity (Alwis D.S., Rajan R., 2014; van Praag H. et al., 2000; Mora F. et al., 2007; Maya-Vetencourt J.F. et al., 2011; Baroncelli L. et al., 2010; Baroncelli L. et al., 2012). It has been shown that such “environmental enrichment” (EE) can activate neuroplasticity (Baroncelli L. et al., 2012). The therapeutic effects of EE can be determined to some extent by the same factors as the effect of natural environmental patterns on a person.

Research has shown that the aesthetic appeal of natural and artificial fractals depends on the visual complexity of those fractals. Most people prefer fractals of medium complexity 1.3–1.4 (Aks D., Sprott J., 1996; Wise J.A., Rosenberg E., 1986; Taylor R.P. et al., 2005; Taylor R.P., 2006; Taylor R.P., Sprott J.C., 2008; Taylor R.P. et al., 2011; Salingeros N.A., 2012).

Note also that the most pleasant music has a fractal dimension near to 1.4 (Hazard C. et al., 1998–1999). For most musical compositions, fractal dynamics corresponds to the middle range D, as in natural fractals (Bigerelle M., Iost A., 2000). This may explain the influence of music on the cortical and subcortical areas involved in the processing of auditory, cognitive, and emotional signals (Särkämö T., et al., 2014), as well as the influence of music on the EEG, mental activity, and human behavior (Petsche H., 1996; Yuan Q. et al., 2000; Staricoff R.L., 2004; Jausovec N., et al., 2006; Särkämö T. et al., 2008; Särkämö T. et al., 2014; Nozaradan S., 2014; Benoit C.–E., et al., 2014; Braun Janzen T. et al., 2022). Evidence that music education and regular practice activates neuroplasticity and improves brain structure and function supports the suggestion that music may be a useful approach for neurorehabilitation (Yuan Q. et al., 2000; Sink M. et al., 2011; Anderson S., 2013; Särkämö T. et al., 2014). A particularly significant effect was found in people listening to Mozart's sonata 448 (the Mozart effect) (Allen K., Blascovich J., 1994; Bigerelle M., Iost A., 2000; Hsü K.J., Hsü A., 1991; Jenkins J.S., 2001; Jausovec N. et al., 2006; Roth E.A., Smith K.H., 2008; Balshem H. et al., 2011).

It is noteworthy that many works of art are fractals and are invariant on a temporal or spatial scale (Ulrich R.S., 1984; Mandelbrot B.B., 1993; Taylor R.P. et al., 1999; Taylor R.P., 2006; Abbott A., 2006; Salingeros N.A., 2012; Hazard C. et al., 1998–1999; Zueva M.V., 2015). However, the influence of fractal patterns of the nature, works of art, and playing musical instruments or listening to music may be not effective enough in improving patient's condition in many cases due to very low potential of neuroplasticity in these patients (Gilbert C.D., Li W., 2012; Zueva M.V., 2018a, b). The activation of adaptive neuroplastic changes in the retina (and brain) may require the use of artificial geometric or dynamic fractals. Dynamic visual fractals, such as fractal flicker, can reactivate neuroplasticity and positively affect the visual system when used in visual rehabilitation strategies (Zueva M.V., 2013, 2015, 2018a, b; Cheng W. et al., 2014). Taking into account the published evidence, fractal optical stimuli with D values between 1.3 and 1.5 can be considered the

most hopeful for visual stimulation (Zueva M. V., 2018a, b, Zueva M. V. и соавт., 2019; Neroev V. V. et al., 2022a, b).

Because the complexity of the retinal and brain activity is lost in patients with neurodegenerative diseases (Sejdić E., Lipsitz L.A., 2013; Jeong J., 2004; Dauwels J. et al., 2011; Al-Nuaimi A.H. et al., 2018; Goldberger A.L. et al., 2002; Manor B., Lipsitz L.A., 2013), light stimulation at a constant frequency will not correspond to the desired healthy dynamics (Zueva, 2015). Please note that in degenerative diseases, there is a parallel decrease in the potential of neuroplasticity and the loss of the fractal dynamics of physiological rhythms in patients. We can assume the relationship of these processes and their mutual influence on the progression of the disease and the expected effectiveness of visual rehabilitation approaches.

### **Fractal Phototherapy: Recent Experience and Perspectives**

Earlier, we create a LED device generating fractal fluctuations in the brightness on the homogenous screen of photostimulator based on the Weierstrass-Mandelbrot function (Zueva M. V., Karankevich A. I., 2018; Zueva M. V., Zuev T. A., Karankevich A. I., 2019; Neroev V. V. et al., 2022) and tested it in the restricted clinical research in 98 patients with POAG (Zueva M. V. et al., 2018). The most significant influences of fractal photostimulation ( $D=1.4$ , 10–12 lux, 10 sessions) on the perimetric index MD we found for initial glaucoma, however, it was documented also for advanced stage of POAG (Zueva M. V. et al., 2018; Zueva M. V. et al., 2019). These results confirmed that fractally time-invariant visual cues with a  $D$  value close to natural fractals could potentially be a promising method of visual rehabilitation.

To study the mechanisms of the effect of fractal photostimulation on the retina and the optimal mode of stimulation therapy, experimental studies on animals is necessary with objective assessment the retinal activity and morphology of the retina. In 12 healthy rabbits, we perform the recording of electroretinogram (ERG) and optical coherence tomography (OCT) of the retina before and in different time after the beginning of prolonged course of fractal photostimulation (up to 90 days) (Neroev V. V. et al., 2022c). The photostimulation was provided daily with 20-minute sessions,  $D = 1.4$ . OCT did not show changes in the retina thickness throughout all the follow-up period. Any negative effect to retinal activity was also absent. A positive impact of fractal photostimulation was found consisted of a significant increase in the b-wave amplitude of the photopic and scotopic ERGs and decrease in the latency of b-wave peaks. The most significant increase in the ERG amplitude was shown for ERG responses on flickers 8.3–12 Hz. These data indicate the predominant influence of fractal light stimulation on the cones activity. The obtained data proved the feasibility of the application of a four-week course of photic stimulation in further studies of the pathology on animal models and in clinical studies. However, considering a significant increase in the 8.3 Hz-flicker ERG amplitude already after a one-week course of stimulation, 1–2-week fractal therapy should be seen as an alternative to a 30-day course in clinical studies. The data also suggested that even long-term exposure to fractal optical signals is safe for the retinal structure and can be used in clinical studies without harm or risk of overdose.

In a recent study, retinal pigment epithelium atrophy in one eye was modeled in six rabbits by subretinal administration of bevacizumab as described in (Neroeva N. V. et al., 2020). After simulation of atrophy, 3 rabbits received daily fractal phototherapy and 3 control rabbits received a 20-minute exposure to a constant LED lamp (15–20 lux) (data not published). Our data revealed a more significant positive dynamics of ERG recovery after a course of fractal phototherapy compared with a course of simple 20-minute light exposure.

### **Conclusion**

Reactivation of neuroplasticity is necessary to provide the maximal visual system response to visual rehabilitation technologies in patients with degenerative pathology to compensate the functional deficit. Methods of visual rehabilitation include a wide range of therapeutic measures aimed at improving the vision-related quality of life of patients. A special role is played by methods and technologies of visual rehabilitation based on the neuroplasticity of the visual system.

Among them, considerable attention today is drawn to such methods as visual perception training (Visual Restoration Therapy), feedback microperimetry, various types of stimulation therapy, environmental enrichment with multimodal stimulation and therapy by natural fractals, music and art therapy and fractal photostimulation. The application of natural and artificial fractal signals in the stimulation therapy can help reactivating the adaptive (homeostatic) neuroplasticity of the visual system and the brain and improve the efficiency of visual rehabilitation.

We draw attention to the appropriateness of using in visual rehabilitation new technologies and devices based on fractal fluctuation in the visual signals. Clinical study and animal research show significant positive effect of fractal light therapy on light sensitivity and the ERG and verify a prospect of new approaches in neurorehabilitation.

## References

1. Abbott A. Fractals and art: in the hands of a master. *Nature*. 2006;439(7077):648–50. doi: 10.1038/439648a.
2. Ajina S., Kennard C. Rehabilitation of damage to the visual brain. *Rev Neurol (Paris)*. 2012;168(10):754–61.
3. Aks D., Sprott J. Quantifying aesthetic preference for chaotic patterns. *Empir Stud Arts*. 1996; 14(1): 1–16.
4. Allen K., Blascovich J. Effects of music on cardiovascular reactivity among surgeons. *JAMA*. 1994;272:882–884.
5. Al-Nuaimi A.H., Blūma M., Al-Juboori S.S., Eke C. S., Jammeh E., Sun L., Ifeakor E. Robust EEG Based Biomarkers to Detect Alzheimer's Disease. *Brain Sciences*. 2021;11(8):1026.
6. Alwis D. S., Rajan R. Environmental enrichment and the sensory brain: the role of enrichment in remediating brain injury. *Front Syst Neurosci*. 2014;8:156.
7. Anderson S. Auditory training: evidence for neural plasticity in older adults. *Perspect. Hear Hear Disord Res Res Diagn*. 2013;17:37–57.
8. Andreou C., Nolte G., Leicht G., Polomac N., Hanganu-Opatz I.L., Lambert M. et al. Increased resting-state gamma-band connectivity in first-episode schizophrenia. *Schizophr Bull*. 2014;41:930–9.
9. Balshem H., Helfand M., Schünemann H. J., Oxmand A. D., Kunz R., Brozek J. et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401–406.
10. Barlow J.S. Rhythmic activity induced by photic stimulation in relation to intrinsic alpha activity of the brain in man. *Electroencephalogr Clin Neurophysiol*. 1960;12(2):317–26.
11. Baroncelli L., Bonaccorsi J., Milanese M., Bonifacino T., Giribaldi F., Manno I. et al. Enriched experience and recovery from amblyopia in adult rats: impact of motor, social and sensory components. *Neuropharmacology*. 2012; 62: 2388–97.
12. Baroncelli L., Braschi C., Spolidoro M., Begenisic T., Sale A., Maffei L. Nurturing brain plasticity: impact of environmental enrichment. *Cell Death Differ*. 2010;17:1092–103.
13. Baroncelli L., Lunghi C. Neuroplasticity of the visual cortex: in sickness and in health. *Exp Neurol*. 2021;335:113515.
14. Benoit C.–E., Dalla Bella S., Farrugia N., Obrig H., Mainka S., Kotz S.A. Musically cued gait-training improves both perceptual and motor timing in Parkinson's disease. *Front Neurosci*. 2014;8, Article 494, 11 pages.
15. Bigerelle M., Iost A. Fractal dimension and classification of music. *Chaos Solitons Fractals*. 2000;11:2179–92.
16. Bourne R.R., Stevens G.A., White R.A., Smith J.L., Flaxman S.R., Price H. et al. Vision Loss Expert Group. Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Glob Health*. 2013;1(6): e339–49.
17. Braun Janzen T., Koshimori Y., Richard N.M., Thaut M.H. Rhythm and Music-Based Interventions in Motor Rehabilitation: Current Evidence and Future Perspectives. *Front Hum Neurosci*. 2022;15:789467.

18. Cehajic-Kapetanovic J., Singh M. S., Zrenner E., MacLaren R. E. Bioengineering strategies for restoring vision. *Nat Biomed Eng.* 2022.
19. Cheng W., Law P. K., Kwan H. C., Cheng R. S. Stimulation Therapies and the Relevance of Fractal Dynamics to the Treatment of Diseases *Open J Regenerative Medicine.* 2014;3(4):73–94.
20. Chiari M., Savi C., Battagliola E., Saviola D., De Tanti A. Visual rehabilitation with Retimax vision trainer in patients with severe Acquired brain injury: Report of two cases. *Neuropsychol. Trends.* 2014;15:35–40.
21. Choudhary M., Malek G. Rethinking Nuclear Receptors as Potential Therapeutic Targets for Retinal Diseases. *J Biomol Screen.* 2016;21(10):1007–18.
22. Crowell A. L., Ryapolova-Webb E. S., Ostrem J. L., Galifianakis N. B., Shimamoto S., Lim D. A., Starr Ph. A. Oscillations in sensorimotor cortex in movement disorders: an electrocorticography study. *Brain.* 2012;135(Pt.2):615–30.
23. Cuenca N., Fernández-Sánchez L., Campello L., Maneu V., De la Villa P., Lax P., Pinilla I. Cellular responses following retinal injuries and therapeutic approaches for neurodegenerative diseases. *Prog Retin Eye Res.* 2014;43:17–75.
24. Dauwels J., Srinivasan K., Reddy M. R., Musha T. Slowing and Loss of Complexity in Alzheimer's EEG: Two Sides of the Same Coin? *Int J Alzheim Dis.* 2011;2011, Art ID539621, 10 pages.
25. Dehn L. B., Piefke M., Toepper M., Kohsik A., Rogalewski A., Dyck E., Botsch M., Schäbitz W. R. Cognitive training in an everyday-like virtual reality enhances visual-spatial memory capacities in stroke survivors with visual field defects. *Top Stroke Rehabil.* 2020;27(6):442–52.
26. Dey S., Proulx S. R., Teotónio H. Adaptation to temporally fluctuating environments by the evolution of maternal effects. *PLoS Biol.* 2016;14(2): e1002388. doi: 10.1371/journal.pbio.1002388
27. Di Ieva A., Esteban F. J., Grizzi F., Klonowski W., Martín-Landrove M. Fractals in the neurosciences, part II: clinical applications and future perspectives. *Neuroscientist.* 2015; 21(1):30–43.
28. Di Ieva A., Grizzi F., Jelinek H., Pellionisz A. J., Losa G. A. Fractals in the neurosciences, part I: general principles and basic neurosciences. *Neuroscientist.* 2014;20(4):403–17.
29. Dockx K., Bekkers E. M. J., Van den Bergh V., Ginis P., Rochester L., Hausdorff J. M., Mirelman A., Nieuwboer A. Virtual reality for rehabilitation in Parkinson's disease. *Cochrane Database of Systematic Reviews.* 2016; Issue 12. Art. No: CD010760.
30. Esposito Veneruso P., Ziccardi L., Magli G., Falsini B., Magli A. Short-term effects of vision trainer rehabilitation in patients affected by anisometropic amblyopia: electrofunctional evaluation. *Doc Ophthalmol.* 2014;129(3):177–89.
31. Field D. J. Relationships between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am.* 1987;4(12): 2379–94.
32. Field D. J., Brady N. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vis Res.* 1997;37:3367–83.
33. Francardo V., Schmitz Y., Sulzer D., Cenci M. A. Neuroprotection and neurorestoration as experimental therapeutics for Parkinson's disease. *Exp Neurol.* 2017;298:137–47.
34. Giddy J. M. Adaptive plasticity in the retina: Protection against acute injury and neurodegenerative disease by conditioning stimuli. *Cond Med.* 2018;1(2):85–97.
35. Gilbert C. D., Li W. Adult visual cortical plasticity. *Neuron.* 2012;75(2):250–64.
36. Goldberger A. L. Fractal variability versus pathologic periodicity: complexity loss and stereotypy in disease. *Perspect Biol Med.* 1997;40:543–61.
37. Goldberger A. L., Amaral L. A. N., Hausdorff J. M., Ivanov P. C., Peng C. K., Stanley H. E. Fractal dynamics in physiology: alterations with disease and aging. *Proc Natl Acad Sci USA.* 2002;99: 2466–72.
38. Goldberger A. L., Rigney D. R., West B. J. Chaos and fractals in human physiology. *Sci Am.* 1990;262:42–49.
39. Green C. S., Bavelier D. Action-video-game experience alters the spatial resolution of vision.

- Psychol. Sci. 2007;18(1): 88–94.
40. Gudlin J., Mueller I., Thanos S., Sabel B.A. Computer based vision restoration therapy in glaucoma patients: a small open pilot study. *Restor Neurol Neurosci.* 2008;26(4–5):403–12.
  41. Halley J.M. Ecology, evolution and 1/f-noise. *Trends Ecol Evol.* 1996;11:33–37
  42. Halley J.M., Inchausti P. The increasing importance of 1/f noise as models of ecological variability. *Fluctuation and Noise Letters.* 2004;4(2): R1–26.
  43. Hausdorff J.M., Ashkenazy Y., Peng C.K., Ivanov P.C., Stanley H.E., Goldberger A.L. When human walking becomes random walking: fractal analysis and modeling of gait rhythm fluctuations. *Physica A.* 2001;302:138–47.
  44. Hausdorff J.M., Peng C.–K., Ladin Z., Wei J.Y., Goldberger A.L. Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. *J Appl Physiol* 1995;78: 349–58.
  45. Hazard C., Kimport C., Johnson D. Fractal Music. Research Project (1998–1999). Available on January 2016 at: <http://www.tursiops.cc/fm/>
  46. Heinrichs-Graham E., Wilson T.W., Santamaria P.M., Heithoff S.K., Torres-Russotto D., Hutter-Saunders J.A.L. et al. Neuromagnetic evidence of abnormal movement-related beta desynchronization in Parkinson's disease. *Cereb Cortex.* 2014;24:2669–78.
  47. Herpich F., Melnick M.D., Agosta S., Huxlin K.R., Tadin D., Battelli L. Boosting learning efficacy with noninvasive brain stimulation in intact and brain-damaged humans. *J Neurosci.* 2019;39(28):5551–61.
  48. Herz D.M., Florin E., Christensen M.S., Reck C., Barbe M.T., Tscheuschler M.K. et al. Dopamine replacement modulates oscillatory coupling between premotor and motor cortical areas in Parkinson's disease. *Cereb Cortex.* 2014; 24(11):2873.
  49. Hsü K.J., Hsü A. Self-similarity of the 1/f Noise called music. *Proc Natl Acad Sci USA.* 1991;88:3507–9.
  50. Huang T.L., Charyton C. A comprehensive review of the psychological effects of brainwave entrainment. *Altern Ther Health Med.* 2008;14(5):38–50.
  51. Ivanov P. Ch., Amaral L.A., Goldberger A.L., Stanley H.E. Stochastic feedback and the regulation of biological rhythms. *Europhys Lett.* 1998;43:363–8.
  52. Jacquin-Courtois S., Bays P.M., Salemm R., Leff A.P., Husain M. Rapid compensation of visual search strategy in patients with chronic visual field defects. *Cortex.* 2013;49(4):994–1000.
  53. Jausovec N., Jausovec K., Gerlic I. The influence of Mozart's music on brain activity in the process of learning. *Clin Neurophysiol.* 2006;117(12):2703–14.
  54. Jenkins J.S. The Mozart effect. *J R Soc Med.* 2001;94(4):170–2.
  55. Jeong J. EEG dynamics in patients with Alzheimer's disease. *Clin Neurophysiol.* 2004;115(7): 1490–505.
  56. Kasten E., Poggel D.A., Sabel B.A. Computer-based training of stimulus detection improves color and simple pattern recognition in the defective field of hemianopic subjects. *J Cogn Neurosci.* 2000;12(6):1001–12.
  57. Kasten E., Sabel B.A. Visual field enlargement after computer training in brain-damaged patients with homonymous deficits: an open pilot trial. *Restor Neurol Neurosci.* 1995;8(3):113–27.
  58. Laaksonen K., Helle L., Parkkonen L., Kirveskari E., Mäkelä J.P., Mustanoja S., Tatlisu-mak T., Kaste M., Forss N. Alterations in spontaneous brain oscillations during stroke recovery. *PLoS One.* 2013;8: e61146.
  59. Lehnertz K. Non-linear time series analysis of intracranial EEG recordings in patients with epilepsy—an overview. *Int J Psychophysiol.* 1999;34:45–52.
  60. Mahalakshmi B., Maurya N., Lee S.D., Bharath Kumar V. Possible Neuroprotective Mechanisms of Physical Exercise in Neurodegeneration. *Int J Mol Sci.* 2020;21(16):5895.
  61. Mandelbrot B. The fractal geometry of nature. Macmillan, 1983.
  62. Mandelbrot B.B. Fractals and an Art for the Sake of Science. In: Emmer M, ed., *The Visual Mind*, Cambridge, MA: MIT Press, 1993, pp. 11–14.

63. Maniglia M., Soler V., Cottureau B., Trotter Y. Spontaneous and training-induced cortical plasticity in MD patients: Hints from lateral masking. *Sci. Report.* 2018;8:90.
64. Manor B., Lipsitz L.A. Physiologic complexity and aging: implications for physical function and rehabilitation. *Prog Neuropsychopharmacol Biol Psychiatry.* 2013;45:287–93.
65. Maya-Vetencourt J.F., Tiraboschi E., Spolidoro M., Castrén E., Maffei L. Serotonin triggers a transient epigenetic mechanism that reinstates adult visual cortex plasticity in rats. *Eur. J. Neurosci.* 2011;33:49–57.
66. Meleth A.D., Mettu P., Agrón E., Chew E.Y., Sadda S.R., Ferris F.L., Wong W.T. Changes in retinal sensitivity in geographic atrophy progression as measured by microperimetry. *Invest Ophthalmol Vis Sci.* 2011;52(2):1119–26.
67. Menon A., Vijayavenkataraman S. Novel vision restoration techniques: 3D bioprinting, gene and stem cell therapy, optogenetics, and the bionic eye. *Artif Organs.* 2022;46(8):1463–1474.
68. Miura G., Sugawara T., Kawasaki Y., Tatsumi T., Nizawa T., Baba T., Hanaoka H., Yamamoto S. Clinical Trial to Evaluate Safety and Efficacy of Transdermal Electrical Stimulation on Visual Functions of Patients with Retinitis Pigmentosa. *Sci Rep.* 2019;9(1):11668.
69. Mora F., Segovia G., del Arco A. Aging, plasticity and environmental enrichment: structural changes and neurotransmitter dynamics in several areas of the brain. *Brain Res Rev.* 2007;55:78–88.
70. Morales M.U., Saker S., Amoaku W.M. Bilateral eccentric vision training on pseudovitelliform dystrophy with microperimetry biofeedback. *BMJ Case Rep.* 2015;2015: bcr2014207969.
71. Neroev V.V., Zueva M.V., Manakhov P.A., Neroeva N.V., Shan A.V., Chujkin N.K., Fadeev D.V. [Method for improving the functional activity of the visual system using fractal phototherapy using a stereoscopic display]. RU2773684C1, Priority 2021–10–04, 2022a.
72. Neroev V.V., Zueva M.V., Neroeva N.V., Fadeev D.V., Kotelin V.I., Sumin S.L., Buryi E.V. [Stimulator with complex-structured optical signals and method for operation thereof]. RU211969U1, Priority 2022–03–31, 2022b.
73. Neroev V.V., Zueva M.V., Neroeva N.V., Fadeev D.V., Tsapenko I.V., Okhotsimskaya T.D., Kotelin V.I., Pavlenko T.A., Chesnokova N.B. [Impact of fractal visual stimulation on healthy rabbit retina: functional, morphometric and biochemical studies]. *Russian Ophthalmological Journal.* 2022c;15(3):99–111. (In Russ.)
74. Neroeva N.V., Neroev V.V., Ilyukhin P.A., Karmokova A.G., Losanova O.A., Ryabina M.V. et al. [Modeling the atrophy of retinal pigment epithelium]. *Russian Ophthalmological Journal.* 2020;13(4):58–63. (In Russ.)
75. Nozaradan S. Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Phil Trans R Soc B.* 2014;369:20130393.
76. O'Brien J., Bloomfield S.A. Plasticity of Retinal Gap Junctions: Roles in Synaptic Physiology and Disease. *Ann. Rev. Vis. Sci.* 2018; 4:79–100.
77. Oster G. Auditory beats in the brain. *Sci Amer.* 1973;229(4): 94–102.
78. Peng C.K., Mietus J.E., Liu Y., Lee C., Hausdorff J.M., Stanley H.E., Goldberger A.L., Lipsitz L.A. Quantifying fractal dynamics of human respiration: age and gender effects. *Ann Biomed Eng.* 2002;30:683–92.
79. Petsche H. Approaches to verbal, visual and musical creativity by EEG coherence analysis. *Int J Psychophysiol.* 1996;24(1–2):145–59.
80. Plank T., Rosengarth K., Schmalhofer C., Goldhacker M., Brandl-Rühle S., Greenlee M.W. Perceptual learning in patients with macular degeneration. *Front Psychol.* 2014;5:1189.
81. Plow E.B., Obretenova S.N., Fregni F., Pascual-Leone A., Merabet L.B. Comparison of visual field training for hemianopia with active versus sham transcranial direct cortical stimulation. *Neurorehabil Neural Repair.* 2012;26(6):616–26.
82. Qian T., Xu X., Liu X., Yen M., Zhou H., Mao M. et al. Efficacy of MP-3 microperimeter biofeedback fixation training for low vision rehabilitation in patients with maculopathy. *BMC*

- Ophthalmol. 2022;22:197.
83. Rabinovich M.I., Abarbanel H.D.I. The role of chaos in neural systems. *Neuroscience*. 1998;87:5–14
  84. Rosa A.M., Silva M.F., Ferreira S., Murta J., Castelo-Branco M. Plasticity in the human visual cortex: an ophthalmology-based perspective. *Biomed Res Int*. 2013;2013:568354.
  85. Roth E.A., Smith K.H. The Mozart effect: evidence for the arousal hypothesis. *Percept Mot Skill*. 2008;107:396–402.
  86. Sabel B.A., Gudlin J. Vision restoration training for glaucoma: a randomized clinical trial. *Jama Ophthalmology*. 2014;132:381–9.
  87. Sabel B.A., Henrich-Noack P., Fedorov A., Gall C. Vision restoration after brain and retina damage: the "residual vision activation theory". *Prog Brain Res*. 2011; 192: 199–262.
  88. Sabel B.A., Kasten E. Restoration of vision in patients with brain injury. *Brain Injury Source*. 2000;1:12–39.
  89. Sahraie A., Treveltham C.T., MacLeod M.J., Murray A.D., Olson J.A., Weiskrantz L. Increased sensitivity after repeated stimulation of residual spatial channels in blind-sight. *Proc Natl Acad Sci U S A*. 2006;103(40):14971–6.
  90. Salingaros N.A. Fractal art and architecture reduce physiological stress. *J Biourbanizm*. 2012;2:11.
  91. Särkämö T., Ripollés P., Vepsäläinen H., Autti T., Silvennoinen H.M., Salli E. et al. Structural changes induced by daily music listening in the recovering brain after middle cerebral artery stroke: a voxel-based morphometry study. *Front Hum Neurosci*. 2014;8:245.
  92. Särkämö T., Tervaniemi M., Laitinen S., Forsblom A., Soinila S., Mikkonen M. et al. Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*. 2008;131:866–76.
  93. Sborgia G., Niro A., Tritto T., Albano V., Sborgia L., Sborgia A. et al. Microperimetric bio-feedback training after successful inverted flap technique for large macular hole. *J Clin Med*. 2020;9(2):556.
  94. Schatz A., Pach J., Gosheva M., Naycheva L., Willmann G., Wilhelm B., Peters T., Bartz-Schmidt K.U., Zrenner E., Messias A., Gekeler F. Transcorneal electrical stimulation for patients with retinitis pigmentosa: A prospective, randomized, sham-controlled follow-up study over 1 year. *Invest Ophthalmol Vis Sci*. 2017;58:257–69.
  95. Sejdić E., Lipsitz L.A. Necessity of noise in physiology and medicine. *Comput Methods Programs Biomed*. 2013;111(2):459–470.
  96. Serruya M.D., Kahana M.J. Techniques and devices to restore cognition. *Behav Brain Res*. 2008;192:149–65.
  97. Shin Y.W., O'Donnell B.F., Youn S., Kwon J.S. Gamma oscillation in schizophrenia. *Psychiatry Investig*. 2011;8(4):288–96.
  98. Sink M., Hossain M., Kato T. Non-Linear Analysis of psychophysiological effects of auditory stimuli using fractal mining. 2011.
  99. Spehar B., Clifford C., Newell B., Taylor R.P. Universal aesthetic of fractals. *Copmuters Graphics*. 2003;27(5):813–20.
  100. Srinivasan A., Karuppathal E., Venkatesan K.R., Kalpana R. Brainwave Entrainment through External Sensory Stimulus: A Therapy for Insomnia (1784). *Neurology*. 2020;94(15 Supplement).
  101. Staricoff R.L. Arts in health: a review of the medical literature. Research report 36. London: Arts Council England. 2004.
  102. Storch D., Gaston K.J., Cepák J. Pink landscapes: 1/f spectra of spatial environmental variability and bird community composition. *Proc Biol Sci*. 2002;269(1502):1791–6.
  103. Tan C.O., Cohen M.A., Eckberg D.L., Taylor J.A. Fractal properties of human heart period variability: physiological and methodological implications. *J Physiol*. 2009;587(Pt 15):3929–41.
  104. Tang H.–Y., Riegel B., McCurry S.M., Vitiello M.V. Open-Loop Audio-Visual Stimula-

- tion (AVS): A Useful Tool for Management of Insomnia? *Appl Psychophysiol Biofeedback*. 2016;41(1):39–46.
105. Tang H.-Y., Riegel B., McCurry S.M., Vitiello M. V. Open-Loop Audio-Visual Stimulation: A pilot study of its potential for sleep induction in older adults. *Appl Psychophysiol Biofeedback*. 2015; 40:183–88.
  106. Tang H.-Y., Vitiello M. V., Perlis M., Mao J.J., Riegel B.A. Pilot Study of Audio-Visual Stimulation As a self-Care Treatment or Insomnia in Adults with Insomnia and Chronic pain. *Appl Psychophysiol Biofeedback*. 2014;39(39):219–25.
  107. Tarita-Nistor L., González E. G., Markowitz S. N., Steinbach M. J. Plasticity of fixation in patients with central vision loss. *Vis Neurosci*. 2009;26(5–6):487–94.
  108. Taylor R. P., Micolich A., Jonas D. Fractal analysis of Pollock's drip paintings. *Nature*. 1999;399: 422.
  109. Taylor R. P., Micolich A. P., Jonas D. Fractal analysis: revisiting Pollock's drip paintings (Reply). *Nature*. 2006;444: E10–E11.
  110. Taylor R. P., Spehar B., Donkelaar P. V., Hagerhall C. M. Perceptual and physiological responses to Jackson Pollock's fractals. *Front Hum Neurosci*. 2011;5:60.
  111. Taylor R. P., Spehar B., Wise J. A., Clifford C. W., Newell B. R., Hagerhall C. M., Purcell T., Martin T. P. Perceptual and physiological responses to the visual complexity of fractal patterns. *Nonlinear Dynamics Psychol Life Sci*. 2005; 9:89–114.
  112. Taylor R. P., Sprott J. C. Biophilic fractals and the visual journey of organic Screen-savers. *J Non-Linear Dyn Psychol Life Sci*. 2008;12:117–29.
  113. Teich M. C., Heneghan C., Lowen S. B., Ozaki T., Kaplan E. Fractal character of the neural spike train in the visual system of the cat. *J Opt Soc Am A*. 1997;14:529–46.
  114. Trauzettel-Klosinski S. Current methods of visual rehabilitation. *Dtsch Arztebl Int*. 2011;108:871–8.
  115. Uhlhaas P. J., Singer W. Abnormal neural oscillations and synchrony in schizophrenia. *Nat Rev Neurosci*. 2010;11(2):100–13.
  116. Ulrich R. S. View through a window may influence recovery from surgery. *Science*. 1984;224(4647):420–1.
  117. van Nispen R. M. A., Virgili G., Hoeben M., Langelaan M., Klevering J., Keunen J. E. E. et al. Low vision rehabilitation for better quality of life in visually impaired adults (Review). *Cochrane Database of Systematic Reviews* 2020, Issue 1. Art. No.: CD006543.
  118. van Praag H., Kempermann G., Gage F. H. Neural consequences of environmental enrichment. *Nat Rev Neurosci*. 2000;1:191–198.
  119. Vasseur D. A., Yodzis P. The color of environmental noise. *Ecology*. 2004;85(4):1146–52.
  120. Verdina T., Piaggi S., Ferraro V., Russolillo V., Peschiera R., Chester J. et al. Efficacy of biofeedback rehabilitation based on visual evoked potentials analysis in patients with advanced age-related macular degeneration. *Sci Rep*. 2020;10(1):20886.
  121. Vingolo E. M., Cavarretta S., Domanico D., Parisi F., Malagola R. Microperimetric biofeedback in AMD patients. *Appl Psychophysiol Biofeedback*. 2007;32, 185–89.
  122. Vingolo E. M., Salvatore S., Cavarretta S. Low-vision rehabilitation by means of MP-1 biofeedback examination in patients with different macular diseases: A pilot study. *Appl Psychophysiol Biofeedback*. 2009;34(2):127–33.
  123. Vingolo E. M., Salvatore S., Limoli P. G. MP-1 biofeedback: luminous pattern stimulus versus acoustic biofeedback in age related macular degeneration (AMD). *Appl Psychophysiol Biofeedback*. 2013;38(1):11–6.
  124. Wagner S. K., Jolly J. K., Pefkianaki M., Gekeler F., Webster A. R., Downes S. M., Maclaran R. E. Transcorneal electrical stimulation for the treatment of retinitis pigmentosa: results from the TESOLAUK trial. *BMJ Open Ophthalmol*. 2017;2(1): e000096.
  125. Westlake K. P., Hinkley L. B., Bucci M., Guggisberg A. G., Byl N., Findlay A. M., Henry R. G., Nagarajan S. S. Resting state alpha-band functional connectivity and recovery after stroke. *Exp Neurol*. 2012;237:160–69.

126. Williams J., Ramaswamy D., Oulhaj A. 10 Hz flicker improves recognition memory in older people. *BMC Neurosci.* 2006;7(5): Article number 21.
127. Williams J.H. Frequency-specific effects of flicker on recognition memory. *Neuroscience.* 2001;104:283–6.
128. Wise J.A., Rosenberg E. The Effects of Interior Treatments on Performance Stress in Three Types of Mental Tasks. Technical Report, Space Human Factors Office, NASA-ARC, Sunnyvale, CA. 1986.
129. Yuan Q., Liu X.H., Li D.C., Wang H.L., Liu Y.S. [Effects of noise and music on EEG power spectrum]. *Space Med Med Eng (Beijing).* 2000;13:401–6. (In Chinese)
130. Zhang X., Li S., Tang Y., Guo Y., Gao S. Intractable ocular diseases and treatment progress. *AAPS PharmSciTech.* 2020;21(6):236. doi: 10.1208/s12249-020-01774-1
131. Zhang X., Tenerelli K., Wu S., Xia X., Yokota S., Sun C. et al. Cell transplantation of retinal ganglion cells derived from hESCs. *Restor. Neurol. Neurosci.* 2020;38:131–140. doi: 10.3233/RNN-190941
132. Zhang Y.S., Guo D., Xu P., Zhang Y., Yao D. Robust frequency recognition for SSVEP-based BCI with temporally local multivariate synchronization index. *Cogn. Neurodyn.* 2016;10:505–11.
133. Zihl J., von Cramon D. Restitution of visual function in patients with cerebral blindness. *J Neurol Neurosurg Psychiatry.* 1979;42(4):312–322.
134. Zueva M. V. [The prospect of the use of nonlinear stimulation therapy in the treatment of traumatic brain injury and maintaining cognitive functions in elderly]. V.M. Bekhterev Review of Psychiatry and Medical Psychology. 2018a; 2:36–43. (In Russ.).
135. Zueva M. V. Dynamic fractal flickering as a tool in research of non- linear dynamics of the evoked activity of a visual system and the possible basis for new diagnostics and treatment of neurodegenerative diseases of the retina and brain. *World Appl Sci J.* 2013;27(4):462–8.
136. Zueva M. V. Fractality of sensations and the brain health: the theory linking neurodegenerative disorder with distortion of spatial and temporal scale-invariance and fractal complexity of the visible world. *Front Aging Neurosci.* 2015;7:135.
137. Zueva M. V. Technologies of Nonlinear Stimulation: Role in the Treatment of Diseases of the Brain and the Potential Applications in Healthy Individuals. *Human Physiology.* 2018b;44(3):289–299. <https://doi.org/10.1134/S0362119718030180>
138. Zueva M. V., Karankevich A. I. [Method of brain training]. RU2680185C1, Priority 2017–10–23, 2018. (In Russ)
139. Zueva M. V., Kovalevskaya M.A., Donkareva O.V., Karankevich A.I., Tsapenko I.V., Taranov A.A., Antonyan V.B. [Fractal Phototherapy in Neuroprotection of Glaucoma]. *Ophthalmology in Russia.* 2019;16:317–28. (In Russ.)
140. Zueva M. V., Kovalevskaya M.A., Donkareva O.V., Starikova M.A., Karankevitch A.I., Taranov A.A., Antonyan V.B. [The impact of complex-structured optical signals on color perception and light sensitivity in patients with suspicion of glaucoma and primary open-angle glaucoma]. *J Clin Exp Ophthalmol.* 2018; 9.
141. Zueva M. V., Zuev T.A., Karankevich A. I. [Stimulator with complex-structured optical signals and method for operation thereof]. RU2671199C1. Priority 2017–10–23, 2019.

Научное издание

# NEUROTECHNOLOGY AND SENSORY COMMUNICATION

Коллективная монография

Под редакцией Ю. Е. Шелепина

Компьютерная верстка: *Мецгерин В. В.*

Подписано в печать 12. 12. 2022. Формат 60 × 84 1/.  
Бумага офсетная. Гарнитура Times. Печать цифровая.  
Усл. печ. л. 36,94. Тираж 500 экз. Заказ № 2069.

Отпечатано в Издательстве ВВМ.  
198095, Санкт-Петербург, ул. Швецова, 41.

# NEUROTECHNOLOGY AND SENSORY COMMUNICATION 2022