Russian Psychological Journal, 2022, Vol. 19, No. 4, 186–210. **doi**: 10.21702/rpj.2022.4.13

PSYCHOPHYSIOLOGY

Research article

UDC 159.91

https://doi.org/10.21702/rpj.2022.4.13

Comparison of the Mechanisms of Phonemic Awareness and Internal Pronunciation of Phonemes and Syllables: EEG and fMRI Study

Andrey O. Shevchenko[1]✉, **Aleksandr V. Vartanov**[2]
[1, 2] Lomonosov Moscow State University, Moscow, Russian Federation

✉ andreyshevchenkomsu@gmail.com

Abstract: **Introduction**. The existing speech production schemes do not fully reflect the interaction of speech perception systems and one's own internal speech. The study is aimed at comparing the connection between internal pronunciation and phonemic perception using psychophysiological methods. **Methods**. Twenty-five people took part in the EEG study. The subjects were offered auditory stimuli, and then it was necessary to internally pronounce the given phonemes or syllables with the same intonation and pronunciation as in external speech. Functional analysis of variance was used to process the results. An fMRI study was also conducted, and 30 healthy right-handed subjects took part in it. The respondents were also offered auditory stimuli, and the background, listening to the material, and listening with subsequent internal pronouncing of a given stimulus was recorded. The results were processed using a program for statistical parametric mapping and then analyzed by group statistics applying a one-sample Student t-test. **Results**. During the EEG investigation, intervals of significant differences in the structure of the evoked potential of internal pronunciation and phonemic perception were found. During the fMRI study, we obtained data that indicate both the process of phonemic perception and intentional internal pronunciation. **Discussion**. Differences in brain structures activity during internal pronunciation and perception were analyzed. Based on the data obtained by us and theoretical analysis results, a scheme of phonemic perception and internal pronunciation was proposed. This scheme represents not only the interaction of the processes of perception and speech generation but also shows the influence of articulations on the internal speech process.

Keywords: speech fMRI, speech EEG, evoked potentials, internal speech, localization, neurolinguistics, subvocalisation, phoneme, syllable, phonemic perception

Highlights:

- Phonemic perception activates conventional brain zones that are associated with speech production.
- There are differences in internal pronunciation at the level of the accompanying speech zones, in particular the cerebellum when pronouncing phonemes and syllables.

- Based on the work conducted, a scheme of phonemic perception and pronunciation was proposed.

Funding: The work was carried out with the financial support of the Russian Science Foundation, project No 20-18-00067.

For citation: Shevchenko, A.O., Vartanov, A.V. (2022). Comparison of the mechanisms of phonemic awareness and internal pronunciation of phonemes and syllables: EEG and fMRI study. *Russian psychological journal*, *19*(4), 185–210. https://doi.org/10.21702/rpj.2022.4.13

Introduction

Internal speech entails a multimodal intersection of images. One of the key elements of internal speech is internal pronunciation, which is most similar to external speech. Highlighting only the phonemic side of internal speech, we are trying to move to the level of internal pronunciation. So, according to a number of authors, the generation (production) of speech is a process with a specific hierarchy. It includes the following stages: the idea or motive of a speech utterance (motive framed in semantics), lexical selection, grammatical and syntactic construction of a phrase, and articulation of the necessary words (Gorelov & Sedov, 2001; Levelt et al., 1999; Dell, 1986; Oppenheim & Dell, 2008, 2010). Levelt's model consists of stages that have their own characteristic unit. Transitions are made from the conceptual stage to the lexical choice of a word and then to its grammatical and phonological coding. Levelt et al. (1999) claim that there is a process of collecting words through syllabification. As soon as the word is divided into syllables, the stage of syllabic phonetic coding takes place. The syllables received at the last stage turn into instructions for articulatory movements. The authors write that the already collected articulation characteristics of the most frequent syllables are stored in memory, which makes it possible not to build these syllables every time, but to extract them from memory. Levelt's model has self–control in speech production (Levelt et al., 1999), consisting of two systems: external – acoustic control and internal - control of internal representations (Levelt et al., 1999). Internal control is performed at the phonological level, not the phonetic one (Indefrey & Levelt, 2004). Another model, the Della model, includes 4 levels: semantic, syntactic, morphemic and phonological. Each level has its own rules for selecting the necessary elements for speech production (Dell, 1986). This model has a network structure and each node represents separate units of speech and interacts with other nodes at all levels. The network proposed for phonological coding consists of nodes for morphemes, syllables, syllabic components, phonemes and signs (articulatory characteristics of sounds). The network has a hierarchy of these nodes. Phonological representations in the model consist of labelled nodes denoting syllabic components. These components often are single phonemes.

Oppenheim & Dell (2010) investigated errors of phonemic similarity and lexical bias in two forms of internal speech with different degrees of articulation and concluded that articulation changes internal speech. The authors state that there is only a phonological level in internal speech, which can be influenced by articulation. Thus, the selected phonemes are able to reflect a different amount of activation beyond the phoneme level (Oppenheim & Dell, 2010). Scott et al. (2013) conducted experiments to prove the assumption that a more active participation of articulation should cause a corollary discharge of greater force. A corollary discharge is a motor command similar to a motor action command but not directly leading to action. The behavioural feature of the corollary discharge is perceptual capture – delay in perception. As a result of the study, it was proved that active involvement of the speech apparatus causes a harder corollary discharge. In addition, perceptual capture is not a simple matter of phonemic priming, and internal speech contains the information below the phoneme level. Analyzing the internal pronunciation, one can conclude that there is a clear connection between the articulatory act and internal pronunciation, while internal pronunciation is possible even without the inclusion of an explicit articulatory act because there are different levels of articulatory activation inside the internal pronunciation. Based on this, phonemic perception becomes a process related to the retention of the auditory image and its identification. Phonemic perception also involves detection – it is a reaction to an external or internal stimulus, and articulatory and phonetic accompaniment is secondary concerning the speed of recognition of the stimulus within the system of speech organization. Motor internal pronunciation has the most similar appearance to external speech. Since the tempo and intensity are preserved in the internal plan, this is represented by the internal auditory image of the spoken units. It can be assumed that the level of articulatory activation is higher. Thus, the effective construction of the BCI (brain–computer interface) is more possible precisely on articulation pronunciation (presentation of articulation) due to a bigger motor component (Sarmiento et al., 2014).

The purpose of our study is to compare the relationship between internal pronunciation and phonemic perception using psychophysiological methods.

Methods

The EEG examination procedure

The study involved 25 subjects: fifteen girls and ten boys aged 17 to 28 years. All the subjects had no history of traumatic brain injuries and mental illnesses and were right-handed. The following experiment was conducted. The launch of internal pronunciation took place based on an auditory stimulus (phonemes and syllables). The following incentives were presented:

– 7 phonemes (A – [a], B –[b], F – [f], G – [g], M – [m], R – [r], U – [u]),
– 10 syllables (BA, FA, GA, MA, RA, BU, RU, MU, FU, GU).

The stimuli were presented randomly. The beginning of pronouncing was set by a special stimulus, informing the spoken phoneme. After the appearance of the setting stimulus, a pause of 500 ms was created. After it, a fixing cross and a sound signal appeared on the screen, which was the starting command of internal pronouncing. During the command to pronounce, the subjects were asked to repeat a phoneme or syllable inwardly with the same duration and intonation as in external speech.

BrainSys (BrainWin) was used as a program for recording and viewing EEG and preliminary purification of evoked potentials. The electrical activity of the brain was recorded monopolarly, using a 19-channel electroencephalograph Neuro-KM (Statokin Company, Russia). The electrodes were arranged according to the international system of 10-20% with two mostoids. The Presentation program (version 18.0 of Neurobehavioral Systems, Inc.) was used to present the stimuli.

The analysis of the obtained data was carried out using the statistical programming language R (version 3.6.3) (R Core Team, 2019) as part of the integrated *RStudio* development environment (version 1.2.5003) (RStudioTeam, 2019). As part of this work, *data.table* packages were used for efficient work with data. The primary method used in the analysis is functional dispersion. This method has the same interpretation and applies as the 'classical' analysis of variance – its main difference is that it evaluates the difference not of averages but of curves. For this reason, functional variance analysis was perfectly suited for evoked potentials (the evoked potential is a curve) and was adapted to detect significant differences in the *ERP* package (Benjamini & Hochberg, 1995).

fMRI examination procedure

The study involved 30 healthy right-handed people (9 men and 21 women aged 20 to 30 years, average age = of 24 years). All participants had no neurological or mental abnormalities, which was confirmed by a preliminary survey. All participants signed an informed consent after explaining the protocol of the experiment. Before the start of each stage, the subjects listened to standardized instruction. The strings were as follows: simple listening to the stimulus material, listening with subsequent repetition, and background strings were also recorded, during which the subject lay quietly in the tomograph. The duration of each episode was 2 minutes (the experiment took about 10 minutes in total). The continuance of each stimulus was 1500 ms, then after a pause (500 ms), an audible signal followed, which indicated the beginning of pronunciation. The stimuli were delivered in a female voice. To study the features of internal pronunciation, a series containing following:

– 7 phonemes (A – [a], B –[b], F – [f], G – [g], M – [m], R – [r], U – [u]),
– 10 syllables (BA, FA, GA, MA, RA, BU, RU, MU, FU, GU).

The presentation was made through special headphones. All stages were held in one session. Throughout the experiment, the subjects' eyes were closed.

Obtaining fMRI images

Functional MRI was performed on a Siemens MagnetomSkyra 3T MRI Machine (Siemens Medical Systems, Erlangen, Germany) to obtain a BOLD contrast (dependent on blood oxygenation). T1-weighted structural scanning was performed for each subject (TR = 2200 ms, TE = 2.48, msflipangle = 6, matrix = 256 × 256, 230 mmfieldofview, slices = 176, slice thickness = 1 mm). For functional scans, we used T2*-weighted, gradient-echo EPI sequence (TR = 3000 ms, TE = 30 ms, tilt angle = 90°, matrix = 128 × 128, field of view 260 mm, slices = 41, slice thickness = 3 mm). The movement of the head was minimized with the help of a specially selected headrest.

Image processing and data analysis were performed using the software package for statistical parametric mapping SPM12 (The Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, London, UK). Next, the group statistics were analyzed by applying the one-sample Student's t-test. Then the t-statistics were converted to the normal standard deviation (z). The voxelwise activation threshold corresponded to pFWEuncorr. < 0.001.

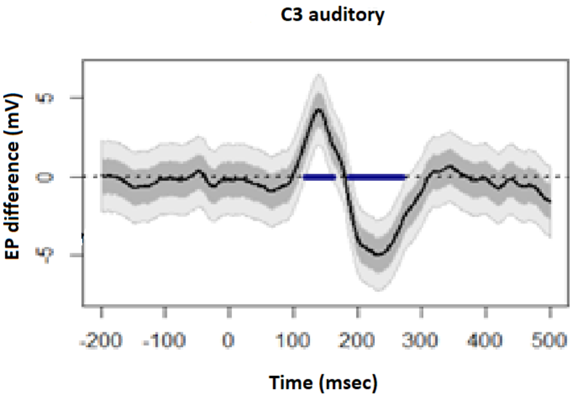
Results

The results of the EEG study, the allocation of evoked potentials for pronunciation

Checking the interaction effect of the EEG lead channel variable and the stage variable showed significant differences for phonemes (p = 0.02) and syllables (p = 0.005, respectively). As for the localization of differences, the most significant differences at the p < 0.05 level are found on channels C3 and F3, and F7. It confirms our results obtained earlier on a smaller sample (Vartanov et al., 2021). They are characterized by a positive wave in the region of 120-170 ms and a negative wave (at C3 and F3) in the interval of 200-280 ms.

Figure 1

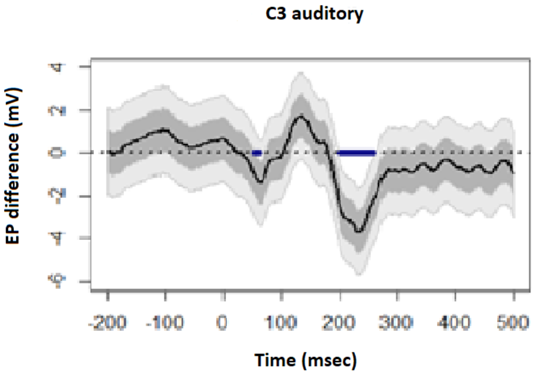
Differences between evoked potentials (pronouncing phonemes minus phonemic perception)



The moments of significant differences are marked in blue.

Figure 2

Differences between evoked potentials (pronouncing syllables minus phonemic perception)



The moments of significant differences are marked in blue.

When comparing the phonemic perception of stimuli with their pronunciation, there are significant differences at the $p < 0.001$ level, while, if you look at the syllables on average, the channels C3 and F3 are again the most pronounced. If we analyze the components of the EP, then the negative wave with a latency of 200 ms is more pronounced.

Functional MRI examination of phonemic perception and internal pronunciation

As a result of the analysis, structures significantly related to phonemic perception and intentional pronunciation were identified. The coordinates of the detected voxels are presented in the MNI format. The voxelwise activation threshold corresponded to pFWEuncorr. < 0.001, the most significant structures with $p < 0.05$ are highlighted in grey. Thus, in comparison of motor internal pronunciation with phonemic perception, the cerebellum plays a significant role in series with motor pronouncing syllables inwardly (Table 1). At the same time, the T- and Z-values are markers of the contribution to the description of the model, relative to which Table 1 is filtered. BA is the number of the cytoarchitectonic field according to Brodman.

Table 1

Comparison of the mechanisms of pronouncing phonemes and syllables with their phonemic perception

L / R	Structure	BA	Cluster	T	Z	P (FWE)	X {mm}	Y {mm}	Z {mm}
				Internal pronunciation of syllables compared to phonemic perception					
L	Cerebellum	–	1	6.362920	4.993979	0.00000030	–24	–52	–46
L	Cerebellum exterior	–	1	6.337997	4.980898	0.00000032	–14	–56	–46
L	Cerebellum	–	575	6.362920	4.993979	0.00000030	–24	–52	–46
L	Subcallosal area	BA25	44	5.823698	4.701207	0.00000129	–2	14	–4
R	MP Temporal Pole	BA38	6	4.443623	3.848925	0.00005932	44	12	–40
R	SecVisual	BA18	19	4.339899	3.778314	0.00007895	20	–70	–1
R	Subcallosal area	BA25	4	4.289875	3.743913	0.00009059	2	14	–4
L	SecVisual	BA18	6	4.204751	3.684852	0.00011442	–18	–80	–4
R	Inferior occipital gyrus (SecVisual)	BA18	11	4.197034	3.679465	0.00011686	30	–92	2
R	Cerebellum exterior	–	3	4.156387	3.651	0.00013061	16	–82	–46
R	Superior frontal gyrus	BA6	1	3.92514	3.486165	0.000245	18	–6	77
R	Inferior occipital gyrus	BA18	3	3.918942	3.481679	0.00024914	34	–92	–1

L	Fusiform gyrus	BA37	10	3.862814	3.44089	0.00028990	–36	–58	–7
R	Cerebellum exterior	–	22	3.861920	3.440238	0.00029060	22	–82	–46
R	Fusiform gyrus	BA37	1	3.858072	3.43743	0.00029363	54	–68	–1
R	Medial frontal cortex (orbFrontal)	BA11	4	3.851278	3.43247	0.00029906	12	34	–13
L	Precentral gyrus	BA6	2	3.790266	3.387731	0.00035237	–36	–14	68
L	Medial frontal cortex (orbFrontal)	BA11	6	3.77112	3.37362	0.00037093	–8	30	–16
R	Cerebellum exterior	–	1	3.714886	3.331975	0.00043116	22	–60	–49
R	Inferioir occipital gyrus (VisualAssoc)	BA19	3	3.711242	3.329267	0.00043538	52	–66	5
R	Inferior occipital gyrus	BA18	1	3.679365	3.305516	0.00047401	32	–84	–7
L	Orbital part of the inferior frontal gyrus	BA47	2	3.663915	3.293972	0.00049391	–42	30	–4
R	Cerebellum exterior	–	4	3.634835	3.27218	0.00053361	4	–74	–43
R	Lingual gyrus	BA19	1	3.604819	3.249604	0.00057783	16	–42	–10
R	Cerebellum exterior	–	1	3.572326	3.225069	0.00062971	18	–70	–46
L	Brain Stem (pyramid area)	–	1	3.545394	3.204658	0.00067612	–10	–40	–43

R	Cerebellum exterior	–	2	3.542027	3.202101	0.00068215	24	–64	–49
R	Medial frontal cortex	BA11	1	3.541181	3.201458	0.00068367	12	38	–13
R	Fusiform gyrus	BA37	1	3.539396	3.200102	0.0006869	42	–56	–13
L	Lateral ventricle (Arg-Retrolimb)	BA30	2	3.536704	3.198057	0.00069178	–26	–56	11
R	Cerebellum exterior	–	1	3.530372	3.193243	0.00070342	42	–72	–34
R	Cerebellum white matter	–	3	3.511992	3.179248	0.00073829	14	–42	–37
R	Occipital Pole	BA18	2	3.498797	3.169181	0.00076435	14	–100	11
L	Cerebral white matter (y lateral ventricle)	–	1	3.493938	3.16547	0.00077416	–24	–38	26
R	Superior occipital gyrus	BA18	2	3.492151	3.164105	0.00077780	22	–82	14
R	Cerebellum exterioir	–	1	3.451802	3.133195	0.00086457	16	–68	–43
L	Lingual gyrus (Visual-Associated)	BA19	1	3.444823	3.127833	0.00088050	–22	–62	–7
L	Cerebellum exterior	–	1	3.444173	3.127333	0.0008826	–16	–74	–43
L	Cerebellum exterior	–	1	3.422555	3.110694	0.00093324	–12	–68	–22
Internal pronunciation of phonemes compared to phonemic perception									
L	Supramarginal gyrus	BA40	15	4.748875	3.870373	0.00005433	–52	–34	32
L	Precuneus (Dorsal)	BA31	3	4.15238	3.508109	0.00022565	–18	–54	38

L	Fusiform gyrus	BA37	7	4.126642	3.491703	0.00023996	–26	–54	–10
L	Superior parietal lobule (VisMotor)	BA7	3	3.973716	3.392849	0.00034585	–32	–48	65
L	Cuneus (visual associative)	BA19	2	3.948858	3.376556	0.000367	–6	–86	35

Based on the obtained coordinates, the images were made with color-highlighted zones of activation of intentional internal pronunciation in comparison with phonemic awareness (Fig. 3, Fig. 4). The gradient of changing dynamics is reflected on the right side of the images.

Figure 3

Internal pronunciation of syllables compared with phonemic awareness

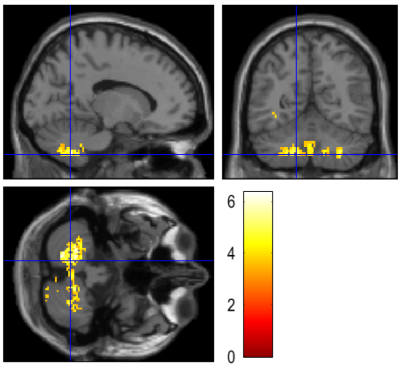
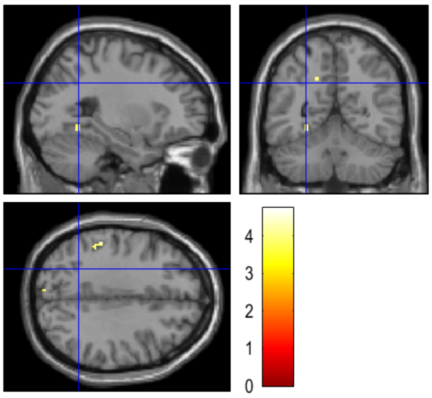


Figure 4

Internal pronunciation of phonemes compared with phonemic awareness



Phonemic awareness was also compared with the background. The most expressed structures were the BA22 zone in the left and right hemispheres (Table 2).

Table 2

Comparison of phonemic awareness mechanisms with the background

L / R	Structure	BA	Clus-ter	T	Z	P (FWE)	X {mm}	y {mm}	z {mm}
				Phonemic awareness compared with the background					
R	STG (Superior temporal gyrus)	BA22	8	7.058488	5.34254	0.00000005	50	–38	11
L	STG (Superior temporal gyrus)	BA22	4	6.715026	5.17432	0.00000011	–36	–38	8
R	Cerebellum	–	1587	6.015718	4.807855	0.00000076	18	–36	–34
L	Posterior insula	BA13	61	5.648526	4.601521	0.0000021	–28	–22	17
R	Thalamus	–	44	5.30348	4.39823	0.00000546	12	–22	5
R	Agranular retrolimbic area	BA30	115	5.283346	4.386076	0.00000577	16	–40	2
L	Calcarine cortex	BA17	20	5.253253	4.367849	0.00000627	–18	–74	5
R	Insula	BA13	62	5.206018	4.339092	0.00000715	30	–12	20
L	Cerebellum	–	21	5.178558	4.32229	0.00000772	–2	–58	–10
L	Calcarine cortex	BA17	56	5.108249	4.278987	0.00000939	–12	–86	14
R	Calcarine cortex	BA1	55	5.085045	4.264607	0.00001001	22	–36	44

L / R	Structure	BA	Cluster	T	Z	P (FWE)	X {mm}	y {mm}	z {mm}
R	Lateral ventricle	–	22	5.026877	4.228361	0.00001177	16	–34	17
L	Lateral ventricle		14	5.013901	4.220236	0.00001220	–22	–36	17
R	Inferior temporal gyrus	BA20	46	4.986055	4.202754	0.00001318	38	–10	–28
R	Superior temporal gyrus	BA41	9	4.965448	4.189775	0.00001396	44	–24	–1
R	Lateral orbital gyrus	BA47	46	4.905086	4.151549	0.00001651	40	36	–4
R	Posteriorc ingulate gyrus	BA23	10	4.887183	4.140153	0.00001735	10	–14	29
L	Thalamus	–	78	4.730111	4.038985	0.00002684	–26	–24	2
R	Hippocampus	–	13	4.700709	4.019812	0.00002912	30	–24	–16
R	Posterior cingulate gyrus	BA31	13	4.651894	3.987812	0.00003334	4	–42	41
R	Calcarine cortex	BA17	25	4.587883	3.945534	0.00003981	4	–76	11

L / R	Structure	BA	Clus-ter	T	Z	P (FWE)	X {mm}	y {mm}	z {mm}
R	Fusiform gyrus	BA37	17	4.556784	3.924864	0.00004339	28	–32	–22
R	Lingual gyrus	BA18	25	4.501969	3.888222	0.00005049	14	–68	2
L	Amygdala	–	44	4.478426	3.872402	0.00005388	–28	–4	–19
L	Cuneus	BA18	5	4.359511	3.79174	0.0000748	–4	–76	32
L	Superior temporal gyrus	BA22	2	4.357996	3.790704	0.00007511	–50	6	–13
R	Middle frontal gyrus	BA9	14	4.347326	3.783403	0.00007735	38	22	26
L	Lateral ventricle	–	7	4.336168	3.775757	0.00007976	–8	–24	23
L	Visual assoc	BA19	9	4.335631	3.775389	0.00007988	–30	–70	–4
R	Visual assoc	BA19	6	4.317257	3.762772	0.00008402	34	–70	–7

L / R	Structure	BA	Cluster	T	Z	P (FWE)	X {mm}	y {mm}	z {mm}
R	Thalamus	–	5	4.312673	3.75962	0.00008509	2	–22	8
R	Precentral gyrus (PrimVisual)	BA4	3	4.269074	3.729542	0.00009591	30	–20	38
R	Anterior orbital gyrus (Orbital Frontal)	BA11	4	4.255989	3.720481	0.00009942	24	40	–10
R	Postcentral gyrus (PrimSensory)	BA1	10	4.244459	3.712484	0.00010262	38	–28	41
L	Precentral gyrus medial segment	BA24	6	4.238666	3.708462	0.00010426	–20	–20	41
R	Precentral Gyrus (PreMot+-SuppMot)	BA6	13	4.211751	3.689733	0.00011224	28	–20	59
R	Precentral gyrus medial segment	BA24	8	4.18588	3.671668	0.00012049	18	–14	38

L / R	Structure	BA	Cluster	T	Z	P (FWE)	X {mm}	y {mm}	z {mm}
R	Calcarine cortex (Sec-Visual)	BA18	17	4.176546	3.665136	0.00012360	18	–84	17
R	Thalamus	–	2	4.148016	3.645119	0.00013363	6	–26	–13
R	Parahipp	BA36	10	4.147383	3.644674	0.00013387	16	–38	–10
R	Precuneus (VisMotor)	BA7	11	4.143725	3.642102	0.00013521	12	–68	47
R	Thalamus	–	8	4.028056	3.560136	0.00018533	4	–18	5
L	Postcentral gyurs (PrimSensory)	BA1	5	4.022578	3.556223	0.00018811	–62	–16	20
L	Cerebellum	–	13	3.98835	3.531713	0.00020644	–34	–68	–31
R	Cerebellum	–	4	3.944522	3.500171	0.00023248	24	–62	–49

L / R	Structure	BA	Cluster	T	Z	P (FWE)	X {mm}	y {mm}	z {mm}
L	Occipital fusiform gyrus (VisualAssoc)	BA19	2	3.944196	3.499936	0.00023268	–26	–76	–4
L	Anterior cingulate gyrus	BA24	1	3.940487	3.497258	0.00023503	–14	–4	41
L	Agranular retrolimbic area	BA30	2	3.93505	3.493331	0.00023852	–22	–52	11
R	Cerebellum	–	2	3.918813	3.481586	0.00024923	32	–58	–43
R	Middle frontal gyrus (Front-EyeFields)	BA8	19	3.913408	3.47767	0.0002529	42	8	53
R	Brain Stem	–	3	3.909098	3.474547	0.00025586	6	–26	–7
R	Angular Gyrus	BA39	1	3.892334	3.462379	0.00026771	46	–64	23
R	Temporal pole	BA38	5	3.874402	3.449335	0.00028098	54	2	–31
R	Precentral gyrus (PreMot+-SuppMot)	BA6	7	3.866411	3.443513	0.00028710	30	4	41
L	Cerebellum	–	4	3.849235	3.430977	0.0003007	–28	–76	–25

L / R	Structure	BA	Cluster	T	Z	P (FWE)	X {mm}	y {mm}	z {mm}
R	Anterior frontal gyrus (Orbital Frontal)	BA11	1	3.835686	3.42107	0.00031188	20	46	–7
L	Insula	BA13	2	3.827527	3.415095	0.00031880	–34	–4	2
R	Anterior prefrontal cortex	BA10	3	3.817274	3.407578	0.00032771	26	50	–1
R	Globus pallidum	–	5	3.809014	3.401515	0.00033507	24	–6	–7
L	Supra-marginal gyrus	BA40	1	3.801684	3.39613	0.00034173	–30	–32	32
R	Posterior cingulate cortex	BA31	7	3.792591	3.389443	0.00035017	0	–36	41
L	Middle occipital gyrus (Visual-Assoc)	BA19	2	3.772275	3.374472	0.00036979	–36	–68	11
R	Inferior temporal gyrus (Prim-Sensory)	BA1	1	3.764789	3.368946	0.00037728	46	–16	–25
L	Cerebellum	–	5	3.759167	3.364793	0.00038301	–44	–60	–40

L / R	Structure	BA	Cluster	T	Z	P (FWE)	X {mm}	y {mm}	z {mm}
R	Cerebellum	–	4	3.758329	3.364174	0.00038387	50	–56	38
L	Calcarine cortex	BA17	6	3.74434	3.353825	0.00039851	–2	–74	11
L	Frontal operculum (Broca-Operc)	BA44	1	3.739974	3.350591	0.00040320	–42	12	14

Discussion

As follows from the study results, considering the analysis of evoked potentials on temporal electrodes (Fig. 1, Fig. 2) as a basis, it can be assumed that the main difference between internal pronunciation and perception is the earlier components of the evoked potential (100–200 ms). This may indicate both the process of launching the command to pronounce (which should not have occurred with passive perception) and the effect of waiting for the starting stimulus (this is best seen in the series with syllables). The absence of identified differences in late components is actually an indicator of the similarity of the response process in both phonemic awareness and internal pronunciation. That is, apparently, a similar internal auditory representation arises when perceiving a phonemic stimulus.

Analyzing the obtained fMRI data, it is possible to identify a number of general patterns in phonemic listening and internal pronunciation. Thus, the Wernicke's area (BA22L, and its homologous BA22R) is activated during both phonemic awareness and internal pronunciation. Several authors claim the existence of two different systems for the perception and production of speech. For example, when the Broca's area is damaged, speech production may be affected, but speech perception and understanding remain intact. In case of injury in the Wernicke's area, the opposite may occur, which tells us about the difference in systems for these two processes (Morais & Kolinsky, 1994). However, we assume that these systems are more than likely interconnected.

In a series of comparisons of the mechanisms of pronunciation of phonemes and syllables with their phonemic awareness (Table 1), activation of the BA37L zone was found. It is known that this zone is associated with pronunciation to a visual stimulus, and there is an assumption that this zone is responsible for phonemic awareness (Devlin et al., 2003). This may be relevant for studies with visual stimulus display (Ardila et al., 2015; Flowers et al., 2004). There is also a hypothesis about the role of this zone in the rhythmic construction of speech process (Booth et al., 2002a, 2002b), which confirms the different role of tempo in motor pronunciation in comparison with phonemic awareness.

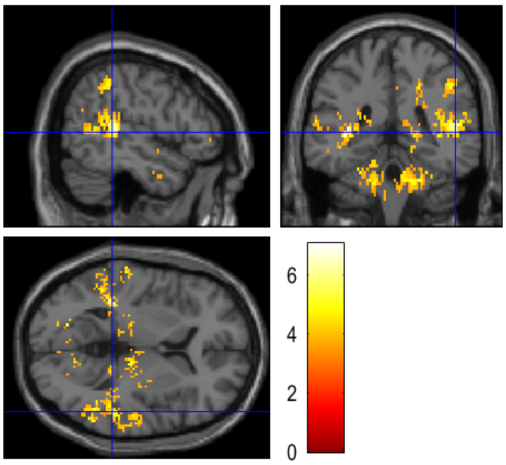
The pronunciation of syllables to oneself includes the BA6L zone, associated with motor speech planning and phonological planning (Shuster & Lemieux, 2005; Fox et al., 2000), which is also clearly seen in our study. Furthermore, in the study by Shuster & Lemieux (2005), activation of this zone was presented during the pronunciation of monosyllabic words. Another activated zone during pronunciation of syllables is BA47L; it is associated with semantic coding, as well as with the processes of recalling and memory retrieval (Robinson et al., 2015; Demb et al., 1995; De Carli et al., 2007). This suggests that the activation of this zone during the syllable pronunciation is associated with the extraction of sound images in the syllable from memory. In the series with the pronunciation of syllables, the activity of the BA30L zone was revealed. This zone (Posterior cingulated BA30L) is associated with semantic coding (Patel et al., 2006). However, the data obtained in the course of research bring this into question, since the experimental model of the study did not assume the presence of semantic coding. Perhaps this zone can also be associated with phonological coding and with memory processes that are launched to search for a word with the same syllable and the similar meaning.

One of the most interesting results obtained in this study was the discovery of a difference in cerebellar activation at the moment of pronunciation of phonemes and syllables. It appears that the cerebellum contains a model of "reverse dynamics" (according to Ziegler, 2016). In support of this version, we can note that the cerebellum is involved in learning new patterns, evaluation, and compensatory response to auditory and somatosensory errors. For this reason, some researchers consider the cerebellum an auxiliary component in speech production models. It is known that after cerebellum injury, motor ataxia is possible, leading to impaired sound pronunciation (Ziegler, 2016). Indefrey & Levelt (2004) reported that areas involved in articulatory planning are more likely to be activated with external pronunciation than with internal pronunciation. At the same time, the areas involved in the syllable formation process should be activated regardless of external or internal pronunciation. Indefrey & Levelt (2004) mention that the bilateral motor and sensory areas, the right dorsal motor area, the right supplementary motor area (SMA), the left and right medial cerebellum, the bilateral thalamus, and the midbrain are most likely involved in the planning and execution of articulatory movements. Furthermore, the left ventral precentral gyrus, the bilateral middle anterior temporal gyri, the left fusiform gyrus, the left thalamus, and the right medial cerebellum were observed more frequently in external articulation tasks than in internal pronunciation tasks. However, in our study, activation of the right and left areas of the cerebellum was recorded in the pronunciation of syllables and phonemes during internal speech. Moreover, the cerebellum was activated in listening tasks compared to the background. That is, we can conclude that the auxiliary role of the cerebellum manifests itself not only at the level of external speech, but also at the level of internal speech. The role is differentiated at the level of phonemic units (phonemes, syllables, etc.): the more complex the unit, the more the role of the cerebellum is expressed.

In the pronunciation of phonemes, in comparison with listening, the BA40L zone was active, associated with phonological operations and semantic processing (McDermott et al., 2003; Chou et al., 2006). It is of great interest that this zone was generally active both during motor pronunciation of phonemes and during phonemic awareness (although there was a difference in the spatial localization of the maximum of differences). Therefore, we can assume that in this zone there are subzones responsible for various aspects of phonological operations. In addition, the demonstrated fact may testify that the conscious and unconscious representation of phonemes relies on different, but close areas of the brain (Morais & Kolinsky, 1994). Phonemic awareness is presumably related to the process of detection and is directly related to the production of speech. Taking this into consideration, one can state that there is a close relationship between the structures responsible for phonemic listening and internal pronunciation, but the present differences should not be overlooked. Traditionally, the zones BA44, BA22, BA17, and BA39 are associated with perception, speech production, and reading. These zones were also presented in our study in phonemic awareness compared with the background (Fig. 5).

Figure 5

Phonemic awareness compared with the background

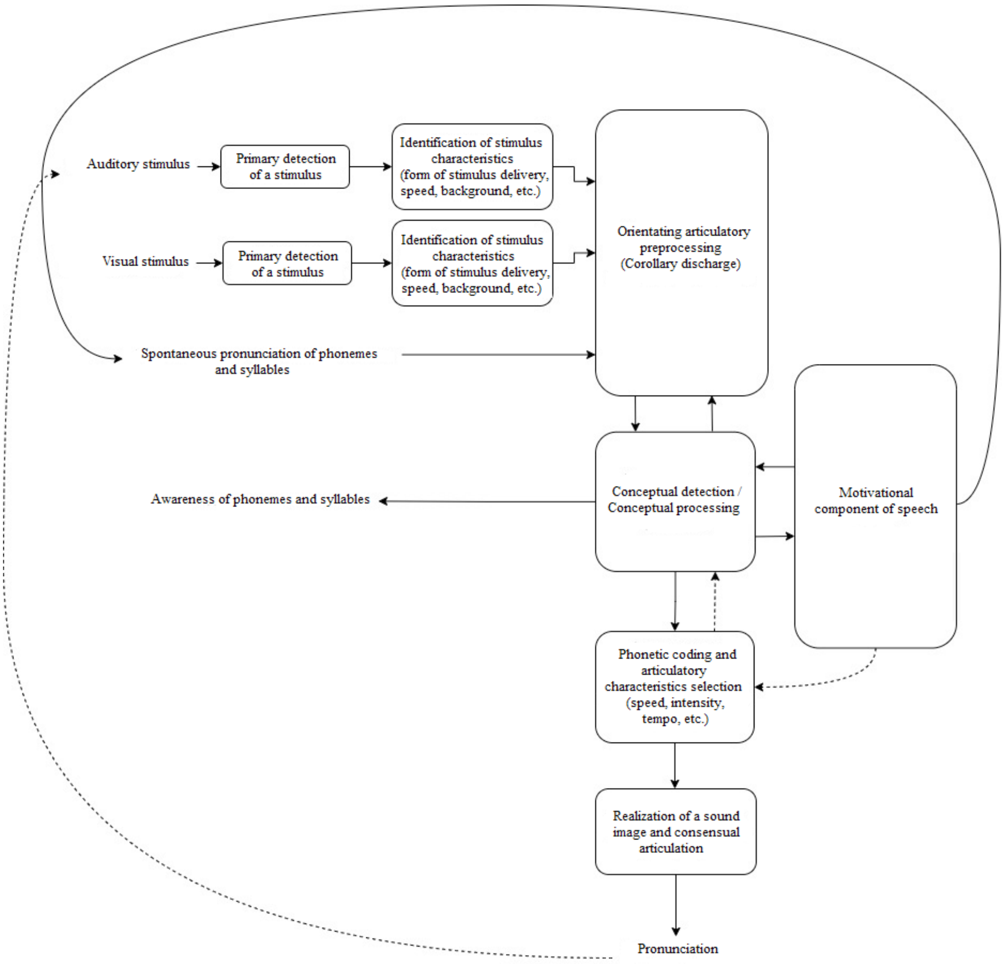


In addition, the BA24L zone, which was active during phonemic awareness, should be viewed a little closer. Researchers associate this zone with the naming of objects process (Garn et al., 2009; Kiyosawa et al., 1996), and with semantic and phonological fluency (Whitney et al., 2009). It is necessary to pay attention to the BA38R zone, which showed activation in a series of phonemic awareness. In a study by Nakamura et al. (2001), this zone was active in the course of recognition of familiar voices. There is a possibility that the activation of this zone in our study was caused by the display of auditory stimuli. Phonemic listening involves the internal representation of an auditory stimulus through the activation of conventional areas associated with speech. Listening to speech, even phonemic, stimuli is associated with the processes of detection and repetition, but the rhythmic component may differ. Also, in the process of communication, the systems of production and perception of speech constantly interact. These two systems not only work in harmony, but also interact in a special way at different levels of processing. These facts should be considered when searching for common in neural networks of perception and speech production (Indefrey & Levelt, 2004).

Based on the results obtained and the theoretical analysis performed, it is possible to present the following scheme of phonemic awareness and pronunciation, inspired by the Levelt model (Fig. 6).

Figure 6

The scheme of phonemic awareness and pronunciation



In contrast to the Levelt scheme (Indefrey & Levelt, 2004), we consider the plan only at the level of phonemic units. In this case, we believe that the motivational component of speech is an important condition even for a phonemic plan. The reason is that at this level, the assessment and planning of a potential result take place. The detection of phonemes and syllables perceived based on visual or auditory input is encoded differently in perception, as shown in our other studies (Vartanov et al., 2021), which is also consistent with the Levelt model. The stage of orientating articulatory preprocessing carries out the process of disintegrating a phoneme into components (subphonemic quanta) or commands for motor action (kinakemes) and launches a corollary discharge to certain areas of the brain. It is what distinguishes our model from the Levelt model. At the stage of conceptual detection, a compilation of separate perceptual features is carried out in a single image of a phonemic unit. It also includes what V. Levelt described as a "conceptualizer" (Levelt et al., 1999). This image is influenced by a form of stimulus delivery, its tonal characteristics, and the motivation a person has at the time of receiving the stimulus. At the stage of phonetic coding, a program is formed for articulatory actions related to the future auditory image that should result from pronunciation. There characteristics for the pronunciation of phonemes or syllables are laid, with consideration of the influence of the motivational component and the image formed at the level of conceptual detection. For example, a speaker's voice can set intonation in a speaking task based on an auditory stimulus. In addition, at this stage, the received program is compared with the image that was formed in the conceptual detection buffer to check accuracy. In case of an error, the program of motor actions is corrected. When the command to pronounce the received image is delivered to the block of primary detection of a stimulus in order to correct the integral image. During mental pronunciation (when external speech is inhibited), the signal from the level of phonetic coding goes back to the conceptual detection unit. It ensures the process of the emergence of a sound image of a phoneme or syllable in the absence of an external stimulus.

Conclusion

1. The triggering of the pronunciation command in comparison with phonemic listening has been confirmed to be reflected in the early components of the evoked potential. However, no differences were found in the late components (from 300 to 500 ms) of the amplitude, which can be explained in terms of the corollary discharge.
2. Phonemic awareness is associated with internal pronunciation. With phonemic awareness, the activation of conventional zones occurs, which are also associated with the production of speech.
3. There are differences in motor pronunciation at the level of concomitant speech zones, in particular, the cerebellum, when pronouncing phonemes and syllables. This is also found in awareness in comparison with the background. One of the possible explanations is the emulation of existing functional cortical connections by this zone. These results complement the Levelt model. We believe that the expression of articulation (muscle) commands is also reflected at the level of concomitant speech zones.

4. Activation of the BA38R zone Indicates the influence of intonational characteristics (in particular, the parameters of the speaker's voice) on phonemic pronunciation and awareness. This largely confirms the idea of the influence of the starting stimulus on the speech process.
5. Based on the results obtained and the theoretical review performed, a scheme (Fig. 6) of phonemic awareness and pronunciation was proposed, inspired by the Levelt model.

References

Ardila, A., Bernal, B., & Rosselli, M. (2015). Language and visual perception associations: Meta-analytic connectivity modeling of Brodmann area 37. *Behavioural Neurology*, *2015*. https://doi.org/10.1155/2015/565871

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological*), *57*(1), 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2002a). Modality independence of word comprehension. *Human Brain Mapping*, *16*(4), 251–261. https://doi.org/10.1002/hbm.10054

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2002b). Functional anatomy of intra- and cross-modal lexical tasks. *NeuroImage*, *16*(1), 7–22. https://doi.org/10.1006/nimg.2002.1081

Chou, T.-L., Booth, J. R., Bitan, T., Burman, D. D., Bigio, J. D., Cone, N. E., Lu, D. & Cao, F. (2006). Developmental and skill effects on the neural correlates of semantic processing to visually presented words. *Human Brain Mapping*, *27*(11), 915–924. https://doi.org/10.1002/hbm.20231

De Carli, D., Garreffa, G., Colonnese, C., Giulietti, G., Labruna, L., Briselli, E., Ken, S., Macrì, M. A., & Maraviglia, B. (2007). Identification of activated regions during a language task. *Magnetic Resonance Imaging*, *25*(6), 933–938. https://doi.org/10.1016/j.mri.2007.03.031

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *The Journal of Neuroscience*, *15*(9), 5870–5878. https://doi.org/10.1523/JNEUROSCI.15-09-05870.1995

Devlin, J. T., Matthews, P. M., & Rushworth, M. F. S. (2003). Semantic processing in the left inferior prefrontal cortex: A combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, *15*(1), 71–84. https://doi.org/10.1162/089892903321107837

Flowers, D. L., Jones, K., Noble, K., VanMeter, J., Zeffiro, T. A., Wood, F. B., & Eden, G. F. (2004). Attention to single letters activates left extrastriate cortex. *NeuroImage*, *21*(3), 829–839. https://doi.org/10.1016/j.neuroimage.2003.10.002

Fox, P. T., Ingham, R. J., Ingham, J. C., Zamarripa, F., Xiong, J.-H., & Lancaster, J. L. (2000). Brain correlates of stuttering and syllable production: A PET performance-correlation analysis. *Brain*, *123*(10), 1985–2004. https://doi.org/10.1093/brain/123.10.1985

Garn, C. L., Allen, M. D., & Larsen, J. D. (2009). An fMRI study of sex differences in brain activation during object naming. *Cortex*, *45*(5), 610–618. https://doi.org/10.1016/j.cortex.2008.02.004

Gorelov, I. N., Sedov, K. F. (2001). *Fundamentals of psycholinguistics. Textbook* (3rd ed., rev. & add.). Labyrint. (in Russ).

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Kiyosawa, M., Inoue, C., Kawasaki, T., Tokoro, T., Ishii, K., Ohyama, M., Senda, M., & Soma, Y. (1996). Functional neuroanatomy of visual object naming: A PET study. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *234*, 110–115. https://doi.org/10.1007/BF00695250

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

McDermott, K. B., Petersen, S. E., Watson, J. M., & Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia*, *41*(3), 293–303. https://doi.org/10.1016/s0028-3932(02)00162-8

Morais, J., & Kolinsky, R. (1994). Perception and awareness in phonological processing: The case of the phoneme. *Cognition*, *50*(1–3), 287–297. https://doi.org/10.1016/0010-0277(94)90032-9

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., & Kojima, S. (2001). Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia*, *39*(10), 1047–1054. https://doi.org/10.1016/s0028-3932(01)00037-9

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537. https://doi.org/10.1016/j.cognition.2007.02.006

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, *38*, 1147–1160. https://doi.org/10.3758/MC.38.8.1147

Patel, R. S., Bowman, F. D., & Rilling, J. K. (2006). Determining hierarchical functional networks from auditory stimuli fMRI. *Human Brain Mapping*, *27*(5), 462–470. https://doi.org/10.1002/hbm.20245

Plotkin, V. Ya. (1993). *Phonological quanta*. Siberian publishing company VO «Nauka». (in Russ).

Robinson, J. L., Barron, D. S., Kirby, L. A. J., Bottenhorn, K. L., Hill, A. C., Murphy, J. E., Katz, J. S., Salibi, N., Eickhoff, S. B., & Fox, P. T. (2015). Neurofunctional topography of the human hippocampus. *Human Brain Mapping*, *36*(12), 5018–5037. https://doi.org/10.1002/hbm.22987

Sarmiento, L. C., Lorenzana, P., Cortes, C. J., Arcos, W. J., Bacca, J. A., & Tovar, A. (2014). Brain computer interface (BCI) with EEG signals for automatic vowel recognition based on articulation mode. In *5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC*) (pp. 1–4). https://doi.org/10.1109/brc.2014.6880997

Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America*, *133*(4). https://doi.org/10.1121/1.4794932

Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, *93*(1), 20–31. https://doi.org/10.1016/j.bandl.2004.07.007

Vartanov, A. V., Suyuncheva, A. R., Shevchevko, A. O. (2021). Mechanisms of internal pronunciation and perception in different types of external initiation. *Intelligent systems. Theory and applications*, *25*(4), 302–306. (in Russ.).

Whitney, C., Weis, S., Krings, T., Huber, W., Grossman, M., & Kircher, T. (2009). Task-dependent modulations of prefrontal and hippocampal activity during intrinsic word production. *Journal of Cognitive Neuroscience*, *21*(4), 697–712. https://doi.org/10.1162/jocn.2009.21056

Ziegler, W. (2016). Chapter 1 – The Phonetic Cerebellum: Cerebellar Involvement in Speech Sound Production. *The Linguistic Cerebellum*, 1–32. https://doi.org/10.1016/B978-0-12-801608-4.00001-3

Received: November 03, 2022
Revision received: December 12, 2022
Accepted: December 20, 2022

Author Contributions

A. O. Shevchenko designed the study, processed the results, wrote the text of the article.

A. V. Vartanov designed the study, processed the results, wrote the text of the article.

Author Details

Andrey Olegovych Shevchenko – Graduate student, Lomonosov Moscow State University, Moscow, Russian Federation; Researcher ID: GLQ-7645-2022, SPIN-code: 9970-3878, ORCID ID: https://orcid.org/0000-0002-9118-2617, e-mail: andreyshevchenkomsu@gmail.com

Alexander Valentinovich Vartanov – Cand. Sci. (Psychology), Associate Professor, Senior Research Fellow, Lomonosov Moscow State University, Moscow, Russian Federation; Scopus Author ID: 6603018783, Researcher ID: D-9907-2012, ORCID ID: https://orcid.org/0000-0001-8844-9643, e-mail: a_v_vartanov@mail.ru

Conflict of Interest Information

The authors have no conflicts of interest to declare.