Recent studies at the interface between psychiatry and neuroscience demonstrate a trend towards the investigation of single significant clinical characteristics of mental disorders in contrast to the analysis of a mental disorder as a homogeneous nosological unit. Thus, a large body of studies is focused on auditory verbal hallucinations (AVH) in schizophrenia, which are one of the core positive symptoms of the disorder and an important diagnostic criterion. Nevertheless, the neuropsychological and neurophysiological mechanisms of AVH in patients with schizophrenia remain debatable. In this paper, we will review the main neurocognitive models of AVH in schizophrenia, including models of intrusive cognitions and poor inhibitory control, a model of attentional shift to inner auditory stimuli and an inability to reallocate its resources, a model of expectation maximization, a model of working memory deficit, a model of poor source-monitoring, models of AVH within cultural-historical approach, and a model of impaired verbal self-monitoring in inner speech. The results of several neuroimaging and neurophysiological studies relevant to the models are also highlighted. We conclude that schizophrenia patients with AVH demonstrate deficits in executive functions and language, or rather a poor cross-functional interaction between them.

Keywords: auditory verbal hallucinations, schizophrenia, positive symptoms, neurocognitive models, neuropsychology, neuroimaging, neurophysiology.


Introduction

A recent trend in neurocognitive studies in psychiatry is characterized by a shift from the consideration of a mental disorder as a homogeneous nosological unit to the investigation of single clinical characteristics [26; 67]. Due to the heterogeneity of schizophrenia, separate analyses of its key clinical aspects may provide more information on the disorder than an attempt to comprehend an overall picture of its diverse characteristics.
Auditory verbal hallucinations (AVH) are a symptom common to various mental and neurological disorders (e.g., Parkinson’s disease, epilepsy, dementia, hearing impairments, bipolar disorder, and personality disorders) and are even found in nonclinical populations, but they are most prominent in the clinical picture of schizophrenia [37; 46]. AVH are one of the core positive symptoms of schizophrenia [26; 55] and an important criterion for the diagnosis of schizophrenia in the ICD-10. Kurt Schneider classified AVH as first-rank schizophrenia symptoms, including hearing thoughts spoken aloud, hearing two or more voices conversing with each other, and hearing voices commenting on a person’s behavior [55]. This symptom is found in 70–80% patients with schizophrenia [41; 46; 67; 71] and causes pronounced distress [65]; AVH can be identified when a person in the awake state hears voices not triggered by a relevant external auditory stimulus but perceived to be realistic, uncontrollable, and different from the person’s own thoughts. An unequivocal and comprehensive determination of the psychological and neurophysiological mechanisms through which the production of a patient’s own mind is perceived as something alien is still lacking; a number of models have been suggested by different authors.

This review aims to analyze basic models of the neurocognitive mechanisms of AVH in patients with schizophrenia, including the results of relevant studies applying neuroimaging and neurophysiological techniques. As far as we know, the current article is the first literature review in Russian addressing the neurocognitive mechanisms of AVH. Such a review may encourage research interest in AVH within the Russian-speaking scientific community (specialists in clinical psychology, psychophysiology, and psychiatry) and facilitate the elaboration of existing hypotheses or the generation of new ones, whereas in practical terms it may help to improve protocols of neurocognitive rehabilitation.

Notably, the first models are described relatively briefly as they are the earliest and simplest and included in more complex models as component parts.

Models of intrusive cognitions and poor inhibitory control

A model of intrusive cognitions explains the emergence of AVH in schizophrenia patients through the presence of auditory representations which do not match current external stimuli, interfere with on-going cognitive processes and disrupt them [8]; they are associated with dysfunctionally high activity in the left temporal regions [39] involved in language comprehension. The model addresses intrusive thoughts, memories, and imagery [8; 16; 51], including trauma-related intrusions [61]. Studies using questionnaires showed that intrusive thoughts were more frequently seen in schizophrenia patients with AVH, in contrast to patients without AVH and healthy individuals [51], and the severity of AVH was correlated to extra-list intrusions in memory tests [16; 17].

According to a model considered as an independent one by some authors [8], schizophrenia patients with AVH demonstrate a deficit of an inhibitory component of executive functions, namely the inhibition of the above-mentioned intrusive cognitions [9; 60; 69]. This may be related to the aberrant functioning of the frontoparietal executive network [41]. Waters et al. [69], using the Hayling Sentence Completion Test and Inhibition of Currently Irrelevant Memories task, revealed that an inhibition deficit correlated to AVH frequency but was not associated with other schizophrenia symptoms. Badcock et al. [9]
also used the latter task and showed that patients with AVH made more inhibition errors than patients without AVH and healthy controls, which might point to a deficit of selectivity of memory in this patient group. Soriano et al. [60] replicated and extended these findings by applying the task measuring the ability to intentionally forget recently learned information. In the study by Toh et al. [64], only current voice-hearers, but not past or never voice-hearers with schizophrenia, were characterized by inhibitory impairment. Taken together, the first and the second models allow us to explain the sense of lacking voluntary control over voices and their intrusive nature.

Model of an attentional shift to inner auditory stimuli and an inability to reallocate its resources, and a model of expectation maximization

The third group of models is based on the idea that schizophrenia patients during AVH, partly due to attention switching difficulties, spend most of their attentional resources (and expectations) on listening to voices, which results in an aberrant processing of external stimuli [67]. According to Friston’s Expectation-Maximization algorithm [29] based on a Bayesian framework (for details see [42]), AVH may arise because of the greater weight of prior expectations compared to sensory input, while the level of uncertainty is underestimated. Consequently, the incoming prediction error (i.e., the difference between the bottom-up signal and top-down prediction) causes the patient to be unable to correct expectations, and a false inference is made that the voices are real. Benrimoh et al. [12] continued developing Friston’s ideas and focused on the activity of a subject who may listen to voices, try to ignore them or answer them. Using computational simulation, the authors found that the weight of prior beliefs (i.e., the level of confidence in them) depends on beliefs about the reliability of incoming sensory data as well as the monitoring of a subject’s own actions (‘beliefs about policies’). Thus, the evaluation of incoming sensory information as imprecise does not allow the person to correct the false-positive hypothesis about the presence of voices; however, high confidence that the person is listening at the moment is also necessary for the emergence of AVH [12]. Importantly, the contribution of the subject’s activity (i.e., listening) to the emergence of AVH was emphasized as early as 1970 in the works by the Russian pathopsychologist Susanna Rubinshtein [5]. Horga and Abi-Dargham [37] develop Friston’s ideas and suggest that an important contribution to the reevaluation of prior expectations in AVH may be made by aberrant functioning of the dopaminergic system and networks including the striatum, which leads to a permanent sense of perceptual uncertainty and does not allow a patient to dynamically adjust to its changing level; therefore, he or she predominantly relies on prior expectations but not on incoming sensory data.

Studies using the methodology of Signal Detection Theory (SDT) rely on the idea that perception always takes place under some uncertainty, and the detection of a stimulus depends both on perceptual sensitivity (i.e., an ability to detect a signal if it is really present) and perceptual bias. The experimental procedure implies a presentation of auditory stimuli masked by white noise and requires a participant to indicate the moment of voice detection. In such studies, patients with AVH demonstrate a false-positive response bias but not impaired perceptual sensitivity [14; 18; 24].

According to the results of neuroimaging studies, schizophrenia patients with AVH miss more target auditory stimuli and demonstrate less activation in the primary auditory
cortex of the left hemisphere in response to a target stimulus compared to patients without AVH [28]. A meta-analysis [45] revealed that increased activation of the left primary auditory cortex was found in schizophrenia patients with AVH in the absence of external stimulation, while decreased activation was seen when they listened to speech. Dichotic listening studies identified the absence of a right-ear advantage in speech perception, which is common for healthy individuals, in patients with AVH [33; 40]. These results may seem to contradict the above data; in some cases, schizophrenia patients give increased attention to external stimuli (i.e., false-positive responses), while in other cases the patients miss them. The level of noise may be an important factor, as it additionally loads the auditory system and prompts patients to listen, which caused AVH in the studies by Rubinshtein and led to false-positive detections in SDT studies.

**Model of working memory deficit**

Working memory is one of the main components of executive functions that allows individuals to maintain and manipulate information that is necessary for the current activity, online. Working memory involves language subcomponents, namely rehearsal in the phonological loop. According to a number of studies, a working memory deficit plays an important role in the AVH mechanism [19; 32; 44; 63].

Bruder et al. [19] divided schizophrenia patients into two groups, those with and those without a core impairment in auditory information processing, based on tone discrimination test performance. The severity of AVH was associated with a verbal working memory deficit in patients with intact auditory information processing. The same patient group performed a verbal working memory task worse than both patients without AVH with intact auditory information processing and healthy individuals. Jenkins et al. [44], using hierarchical binary logistic regression, revealed that working memory (assessed with the MATRICS battery) predicted the presence of AVH in schizophrenia patients. Similar results were obtained in other studies [32; 63] regarding verbal working memory.

During working memory task performance, schizophrenia patients with AVH demonstrated decreased activation in the left temporoparietal regions involved in speech comprehension and verbal working memory compared to a clinical control group, and the decrease was negatively correlated to AVH severity [71].

Hoffman and McGlashan [36] developed a computer simulation of working memory during recognition of a single word in a sentence. Two conditions were used for the presentation of a word, normal and degraded phonetic input, with the latter simulating a reliance on working memory. When connections between different network layers were disrupted, the system produced spontaneous percepts of words in the absence of phonetic input and recognized words poorly under degraded phonetic input. Schizophrenia patients with AVH made more word recognition errors compared to control groups, which was best explained by an overpruned model with changed activation of the network elements [36]. A detailed psychological interpretation of the contribution of verbal working memory deficits to the AVH mechanism is lacking; some authors [32; 71] suggest that AVH interfere with external auditory stimuli, exploiting working memory resources (this hardly differs from interpretations of the previous model).
Model of poor source-monitoring of speech

The next model prioritizes an impairment of source-monitoring, a metacognitive function which allows an individual to attribute his or her mental experience or actions, including language production, to an external or internal source [68]. In typical tasks, participants are asked to determine whether a particular mental action (e.g., utterance, movement) was performed by themselves or by someone else. For instance, participants produce an association to each word in a series of sequentially presented words and then identify the words generated by themselves as well as the presented and non-presented words. A meta-analysis of studies with different tasks [68] revealed a deficit of source-monitoring in schizophrenia patients with AVH, in contrast to patients without AVH, which took place at early stages of information processing (perception but not memory). In a study by Mechelli et al. [49], schizophrenia patients with AVH misidentified their own speech as being that of somebody else more often than clinical and nonclinical control groups. In healthy individuals and patients without AVH, the effective connectivity of the left superior temporal gyrus with the anterior cingulate gyrus was higher during listening to alien speech versus self-generated speech, while patients with AVH demonstrated the reverse picture [49]. Simons et al. [58] found that differences between listening to speech and language production in the activation of left superior temporal gyrus were less pronounced in schizophrenia patients with AVH than in controls. Therefore, the perception of self-generated speech in schizophrenia patients with AVH may rely on brain mechanisms underlying the perception of another person’s speech in healthy individuals.

According to some authors, misattribution in patients with AVH may be related to dysfunction of the right hemisphere [21; 59], which is involved in online monitoring of incoming and transmitted information, maintaining of the integrity of a mental model of a situation as well as the correspondence of language and thinking to reality, knowledge of the world, and life experience [1].

However, the reasons for source-monitoring deficits in AVH remain unclear in this model. In our opinion, they are highlighted in a model of poor verbal self-monitoring in inner speech.

Models of AVH within cultural-historical approach

A model of inner speech impairment in schizophrenia patients with AVH by Charles Fernyhough [26] is rooted in Lev Vygotsky's ideas of the social genesis of higher mental functions as well as the dialogical nature, abbreviation, and predicativity of inner speech. Fernyhough describes four levels of inner speech development, namely external dialogue, private speech, expanded inner speech, and condensed inner speech. Two possible mechanisms of AVH are suggested: a disrupted internalization of inner speech (i.e., a developmental impairment) and a compensatory re-expansion of inner speech under conditions of stress and cognitive challenge. However, the explanation of inner speech misattribution remains unclear in this model.

A model describing the psychological mechanisms of pathological alienation was proposed by Ignaty Zhuravlev [2]. According to this model, AVH are perceived by a patient as resulting from the influence of an alien will because their main aspect, the sense of
uncontrollability and being made by someone else, is also common to the perception of an external objective world. Pathological alienation, in the author's opinion, is related to an impaired development of subjectivity. In normal conditions, subjectivity is organized through the distinction between a subject and an object, with the possibility of boundary shifting, but nevertheless, a stable range of its localization. If the boundary shifts outside this range, a thought may develop into AVH [2]. Subjectivity is considered a higher mental function; therefore, in its development, it acquires the possibility to be under voluntary control. This implies a polarization between an individual's own and an alien possession, with a further interiorization of this distinction. Zhuravlev showed that schizophrenia patients in psychosis produced utterances not addressed to an interlocutor with increased frequency, with their further objectification as AVH often dialogical in content [2].

**Model of poor verbal self-monitoring in inner speech**

We assume that the model of poor verbal self-monitoring in inner speech [13; 25; 30] includes a majority of the previously described models as component parts and that it is the most integrative and elaborated model. According to this model, AVH in schizophrenia patients arise due to impaired self-monitoring in inner speech production and insufficient attenuation of its sensory consequences (disruption of corollary discharge). As the perception of inner speech is not coupled with signals indicating that an individual initiated this process himself or herself, a mental event (i.e., inner speech production) does not match its expected sensory consequences. Therefore, an individual is not ready to categorize his or her experience as inner speech and misattributes it to another source.

In terms of brain mechanisms, this deficit may rely on an increased activation of language areas due to aberrant functional connectivity with regions involved in executive functions [6]. Thus, increased bottom-up signals from the secondary auditory cortex involved in speech perception may be coupled with decreased top-down signals from the dorsolateral prefrontal, anterior cingulate, and supplementary motor cortices [6; 15]. Hugdahl [38] suggests additionally including the parietal regions contributing to an attentional shift to inner stimuli in this model. Waters et al. [67] underline an important role of negative emotions in the triggering and chronification of AVH in the model.

Expectations, imagery, and memories individualizing the content of AVH, a lack of insight, delusional interpretations of AVH along with a false-positive perceptual bias and an inhibitory deficit are considered top-down processes [67]. The model is supported by data on the associations between the predominant language of AVH in bilinguals and an earlier age of language acquisition, more frequent language use, and subjectively higher language proficiency [34]. The involvement of an anticipatory corollary discharge mechanism in language processes in healthy individuals was demonstrated in several studies [56; 57]. Neurophysiological studies revealed that in healthy individuals but not in schizophrenia patients with AVH, amplitude characteristics of the N1 component of auditory event-related potentials reflected a dampening of the auditory cortex during language production [27]. The coherence of the theta rhythm between the frontal and temporal regions of the left hemisphere was higher in language production than in listening to speech in healthy individuals but not in patients with AVH [27]. Further studies found a delay in auditory cortex suppression, associated with a decreased fractional anisotropy of the arcuate fasciculus, in schizophrenia patients with AVH [70].
As inner speech underlies executive functions (according to the ideas originating from Lev Vygotsky and supported by many contemporary authors; see paragraphs 1.2 and 1.3 in [4]), its impairments may lead to poor performance in a range of tasks loading different components of executive functions, including inhibition, switching, and working memory [53], in schizophrenia patients with AVH. These components were mentioned in previous models. At the same time, an executive deficit per se may contribute to inner speech impairments [53].

In the first group of fMRI studies relevant to this model, AVH are considered to be associated with a stable aberration of functional brain architecture (trait studies), and schizophrenia patients with AVH are compared to clinical and nonclinical control groups. During imagining sentences versus listening to them, healthy individuals had decreased activation of the left superior temporal gyrus, and this decrease was less pronounced in schizophrenia patients with AVH [58]. Activation of the cingulate gyrus in healthy participants was higher in imagining sentences than listening to them, while patients demonstrated the reverse trend [58]. The severity of AVH in schizophrenia patients was associated with decreased functional connectivity between the left dorsolateral prefrontal cortex and left temporal regions in sentence completion [47]. Studies using resting-state fMRI also revealed aberrant functional connectivity between the anterior cingulate cortex and left temporal regions in patients with AVH [20; 66]. At the same time, according to our study [4], which included schizophrenia patients with a history of AVH, patients without a history of AVH, and healthy individuals, decreased functional connectivity between the anterior cingulate cortex and the superior temporal gyrus bilaterally was not a specific trait of patients with AVH but was a common characteristic of all schizophrenia patients. A specific trait of patients with AVH was decreased functional connectivity between the left inferior frontal gyrus, which is involved in language production, and the anterior cingulate cortex.

The second group of fMRI studies is based on the idea that AVH are related to a temporary change in brain functioning (state studies), including the increased activation of a network involved in language production and perception [7; 15; 31; 43]. The subjective reality of voices in patients with schizophrenia spectrum disorders is correlated with the functional connectivity of the inferior frontal gyrus with temporal regions and the anterior cingulate cortex [54]. According to some authors, brain activation during AVH is not lateralized to any hemisphere [7], while other data suggest that the symptom is related to predominant right hemisphere activation [15; 23; 59]. Sommer and Diederen [59] propose that the key mechanism of AVH is the insufficient inhibition of right-hemisphere language areas by the anterior cingulate cortex. The authors suggest that negative content, intrusive aspect, and a lack of voluntary control are similar for AVH in schizophrenia and “automatic speech” in aphasia due to left hemisphere injury and right hemisphere disinhibition. An additional argument provided by the authors concerns the data indicating that language processes in healthy individuals rely on the inhibition of right-hemisphere homologues of language areas [59]. At the same time, a recent meta-analysis by Barber et al. [11] did not replicate the activation of inferior frontal and superior temporal gyri in any hemisphere during AVH in schizophrenia patients (perhaps due to an application of conservative statistical thresholds); however, they revealed a cluster of activation in the left insula which, on the one hand, is involved in language, and on the other hand, is a component of the salience network.
Data on therapy of AVH with noninvasive brain stimulation also support this model of AVH [10; 48; 50]. For instance, transcranial direct current stimulation with the cathode (inhibitory effect) placed over the left temporoparietal junction and the anode (excitatory effect) placed over the left prefrontal cortex reduced treatment-resistant AVH in schizophrenia patients [50].

**Conclusions**

We have considered the main models of the neurocognitive mechanisms of AVH in patients with schizophrenia. Although the reviewed models focus on different aspects of AVH, the models overlap and complement each other. The first model prioritizes the intrusive cognitions (thoughts, memories, or imagery) interfering with on-going cognitive processes and disrupting them. The next models address the reasons for the inability to avoid these intrusions, namely a deficit of executive functions such as inhibition, switching, and working memory, which explains the sense of lacking voluntary control over the voices heard by a patient. Specifically, a group of models emphasizes an attentional shift to inner stimuli and an inability to switch to external stimuli. Thus, according to a Bayesian model of expectation maximization, AVH may arise due to the dominance of prior expectations over novel incoming information and an underestimation of the uncertainty level. A number of authors [5; 12] underline the importance of an individual’s activity (i.e., listening) in the psychological mechanism of AVH. Another significant model is based on the idea of poor source-monitoring of the mental experience, in particular language production, i.e., its misattribution to external or internal sources. This misattribution error is explained by the dialogical nature of inner speech and its development through interiorization in the models within a cultural-historical approach. Finally, in our opinion, the most integrative and elaborated model is that of poor verbal self-monitoring in inner speech. According to this model, AVH may arise due to an impaired anticipation of the sensory consequences of inner speech; since such prediction is absent, a patient is unable to categorize the ongoing experience as his or her own inner speech and misattributes it to another source. Methodological limitations of existing fMRI studies of the brain mechanisms of AVH in schizophrenia patients are briefly discussed elsewhere [3].

Based on our review of neurocognitive models of AVH in schizophrenia patients, we can conclude that this clinical group demonstrates deficits in executive functions and language, or rather a deficit in the cross-functional interaction between them that is particularly evident in the last model. This conclusion coincides with the data suggesting that deficits in executive functions [52; 62] and language [21; 22; 35; 72] are the key cognitive disturbances in schizophrenia patients. In this context, it is possible that schizophrenia patients with AVH have specific traits which differentiate them from other schizophrenia patients, along with neuropsychological impairments somehow common for all patients with schizophrenia.

Limitations of the current literature review are associated with the fact that it is not a systematic review. These limitations are due to a large number of publications that cannot be analyzed within the scope of one article as well as difficulties in the formulation of precise criteria for the literature search within the defined issue. Nevertheless, it is possible to conduct a systematic review of articles within each of the described models.
References


Информация об авторах

Паникратова Яна Романовна, кандидат психологических наук, младший научный сотрудник лаборатории нейровизуализации и мультимодального анализа, ФГБНУ «Научный центр психического здоровья», г. Москва, Российская Федерация, ORCID: https://orcid.org/0000-0002-5698-4251, e-mail: panikratova@mail.ru

Лебедева Ирина Сергеевна, доктор биологических наук, заведующая лабораторией нейровизуализации и мультимодального анализа, ФГБНУ «Научный центр психического здоровья», г. Москва, Российская Федерация, ORCID: https://orcid.org/0000-0002-0649-6663, e-mail: irina.lebedeva@ncpz.ru

Information about the authors

Yana R. Panikratova, PhD in Psychology, Junior Research Scientist at the Laboratory of Neuroimaging and Multimodal Analysis, Mental Health Research Center, Moscow, Russia, ORCID: https://orcid.org/0000-0002-5698-4251, e-mail: panikratova@mail.ru

Irina S. Lebedeva, Doctor of Sciences in Biology, Head of the Laboratory of Neuroimaging and Multimodal Analysis, Mental Health Research Center, Moscow, Russia, ORCID: https://orcid.org/0000-0002-0649-6663, e-mail: irina.lebedeva@ncpz.ru