Saint Petersburg State University

As manuscript

Nikolaev Aleksandr Sergeevich

Physiological basis of perception of speech of children with autism spectrum disorders

Scientific specialization 1.5.5. Human and Animal Physiology

Thesis for a Candidate degree in biological sciences

Translation from Russian

Scientific supervisor:

Doctor of biological sciences, professor

Lyakso Elena Evgenievna

Saint Petersburg

2024

TABLE OF CONTENTS

INTRODUCTION

The relevance of study

Human speech contains a variety of information; through speech, a person transmits and receives information about the world around them; a person's voice characteristics reflect their mood and state.

In studies of voice and speech, the method of auditory perceptual experiment is widely used. It has been applied when researching the possibility of determining the age and gender of the speaker (Goy et al., 2016), their height and weight (Bruckert et al., 2006). Numerous studies research the possibility of recognizing the emotional state of the speaker from speech signals (Kaya et al., 2017; Grosbras, Ross, Belin, 2018; Lyakso et al., 2021; García-Guerrero et al., 2022). The possibility of recognizing pathological conditions based on the characteristics of voice and speech is being studied (Jones et al., 2019; O'Leary et al., 2020; Verkhodanova et al., 2020). The perceptual experiment method is also used in clinical research, since the characteristics of voice and speech may serve as a diagnostic sign (Kent, 2009).

In addition to perceptual experiment, methods of instrumental analysis aimed at studying the acoustic-phonetic structure of speech are widely used. These methods were used in works studying intonation of speech messages (Ward, Hirshberg, 1985; Rockwell, 2000; Scharrer, Christmann, 2011), issues of definition and categorization of phonemes, stress location, and verbal boundaries (Sanders et al., 2002). An important area of research is the study of the acoustic characteristics of the speech of children with typical and atypical development in the process of speech formation (Kuhl, Meltzoff, 1996; Lyakso, 1998; Vorperian, Kent, 2007; Fusaroli et al., 2017).

One of the current research areas is the study of voice and speech characteristics of people with autism spectrum disorder (ASD). ASD is a group of developmental disorders that share a set of common symptoms. These symptoms include impaired or atypical social behaviour, restricted and stereotypic behaviour, and speech and language impairment (Kanner, 1943; Wing, 1993). Speech impairments in children with autism spectrum disorders may manifest at different levels of its organization: articulatory, grammatical, pragmatic. Common pathological features of the speech of children with ASD include echolalia, a poor vocabulary compared to typically developing (TD) peers, and grammatically violated phrases. A number of works consider the study of auditory perception of speech in children with ASD (Redford et al., 2017; Frolova et al., 2019; Lyakso et al., 2020). These works show the ability of listeners to determine from the speech signals of children the meaning of what was said, their gender, age, and psychoneurological state (whether it is the speech of a child with typical or atypical development).

In recent decades, the attention of researchers has been focused on studying the influence of individual characteristics of the listener, such as gender, age, or professional experience, on their perception of speech information (Smith et al., 2012; Goy et al., 2016; Lausen, Schlacht, 2018; Frolova et al. al., 2019; Frolova et al., 2020; Lyakso et al., 2020; Nikolaev, Lyakso, 2020). These studies have a practical focus for recruiting personnel that has to work with people with atypical development or create computer applications addressed to a certain circle of people. The influence of the listener's physiological and psychophysiological characteristics on the nature of their perception of speech is studied rarely (e.g., Lyakso et al., 2017), which made it possible to formulate the purpose of this study.

The purpose of the study:

To study the influence of physiological and psychophysiological characteristics of adults on the perception of information contained in the speech of children with autism spectrum disorders.

The objectives of the study:

1. To identify the features by which the lexical meaning of words of children with autism spectrum disorders and typically developing children is recognized by adults, depending on their physiological and psychophysiological characteristics, gender and age, and to establish the characteristics of children's speech that are significant for correct recognition.
2. To determine the ability to classify the psychoneurological state (either "typical development (normal)" or "autism spectrum disorder") by adults depending on their physiological and psychophysiological characteristics, gender and age, and to establish speech characteristics that are significant for correct recognition.
3. To determine the features by which the emotional state of children with autism spectrum disorders and typically developing children is recognized by adults, depending on their physiological and psychophysiological characteristics, gender and age, and to establish speech characteristics that are significant for correct recognition.

Scientific novelty

The novelty of the research is accomplished by using an integrated research approach to the information contained in the speech of typically developing children and children with autism spectrum disorders; furthermore, by obtaining original data based on Russian language material on how the auditors' individual characteristics influence their recognition of information contained in the speech of children with ASD and TD children aged 5–14 years.

For the first time within the framework of one study, there were obtained new and original data on the physiological and psychophysiological characteristics of adults who, in a perceptual experiment on the speech of children with ASD and TD children living in a Russian-speaking environment, determined information about their psychoneurological and emotional state:

1) Specific features for recognising various characteristics were identified, namely: gender and age of the auditor, level of anxiety for recognising the lexical meaning of words of children with ASD; the hearing thresholds of the auditor for recognising the meaning of words of TD children; the auditor's gender, hearing thresholds, dominant hemisphere for speech, level of anxiety for classification of the psychoneurological state of children; the auditor's gender, experience of interacting with children, hearing thresholds for recognition of the emotional state of children with ASD; the auditor's gender, age, dominant hemisphere for speech, level of anxiety for recognition of the emotional state of TD children.

2) Data on the acoustic characteristics of the speech of children with ASD and TD children influencing the recognition of the lexical meaning of words, the psychoneurological and emotional state of children with TD and ASD were obtained, namely: speech rate, duration of stressed and unstressed vowels, duration of pauses in utterances, fundamental frequency (F0) values of stressed and unstressed vowels, values of formant frequencies. Data on the acoustic characteristics of the speech of children with ASD formed the basis for identifying a set of features that can be considered as a biomarker of autism (Lyakso et al., 2021).

3) For the first time, using the material of the Russian language, dependencies have been identified between the physiological and psychophysiological characteristics of adults and their ability to determine the psychoneurological and emotional state of children with ASD and TD children, and the lexical meaning of their words. It has been shown that the recognition of information contained in the speech of children with ASD and TD children is influenced by different physiological and psychophysiological characteristics of auditors.

Theoretical and practical significance of the study

The theoretical significance of the study is in the identification of factors that influence the recognition of speech of children with autism spectrum disorder and typically developing children; in describing the physiological parameters of adults that influence the recognition of the lexical meaning of words, the psychoneurological and emotional state of children when listening to their speech signals. The acoustic characteristics of the speech of children with ASD described in the work expand the understanding of the processes of speech formation in conditions of atypical development in the period from preschool to middle school age (5–14 years).

The practical significance of the work is as follows:

1) The obtained data may be used in medical practice for determining the level of speech development of children with ASD, their emotional sphere and psychoneurological state. The identified acoustic features of the speech of children with ASD can be considered as a biomarker of autism and used as an additional diagnostic sign.

2) Data on the physiological indicators of adults (hearing thresholds, dominant hemisphere for speech, level of anxiety) can be taken into account when recruiting personnel working with children with autism spectrum disorders.

3) Data from an auditory perceptual experiment on human recognition of information contained in the speech of typically developing children and children with ASD serve as the “gold standard” for automatic recognition. The data obtained on the acoustic characteristics of the speech of children with ASD, which influence its recognition by listeners, are used in the development of methods for automatic recognition of emotional states from speech (Matveev et al., 2022; Lyakso et al., 2023), which will later be used to create systems of instant ASD diagnosis; to develop educational programs and educational games intended for children with ASD.

The results of this study are used in a number of lecture courses, namely: “Psychophysiology of speech: part 2”, “Pathology of speech behaviour”, given to graduate students of the Faculty of Biology of St. Petersburg State University; “Psychophysiology of speech activity” — for graduate students of the Faculty of Philology in the program “Artificial intelligence in modelling of speech activity”; furthermore, the obtained results are used in lectures and practical classes in the course “Fundamentals of speech technologies” for graduate students of the Faculty of Information Technologies and Programming of ITMO University.

The degree of reliability

The results of the study were obtained using certified professional equipment and software; a sufficient sample of informants and subjects (auditors) was used to obtain the necessary data. The study was conducted using standardised methods developed in the Child Speech Research Group in St. Petersburg State University. Adequate methods of statistical analysis of experimental data were used.

Approbation of the research

The thesis materials were presented at the II All-Russian Acoustic Conference (Nizhny Novgorod, 2017), at the international scientific and practical conference of young scientists “Psychology of the 21st Century” (St. Petersburg, 2017, 2019), at the international scientific conference “Problems of Ontolinguistics” (St. Petersburg, 2018 , 2019, 2021, 2023), at the VI

Congress of CIS Physiologists (Sochi, 2019), at the international conference “Video and Audio Signal Processing in the Context of Neurotechnologies” (St. Petersburg, 2022). Co-authored reports were presented at a number of international conferences: “Speech and Computer — SPECOM” (Leipzig, Germany, 2018; Istanbul, Turkey, 2019; Gurugram, India, 2022; Dharwad, India, 2023), “International Psychological Applications and Trends — InPACT” (Lisbon, Portugal, 2021), 24th International Acoustic Congress (Gyeongju, South Korea, 2021), International Conference “Language, Consciousness, Communication: Problems of the Information Society” (Moscow, 2023); 25th International Conference on Multimodal Interaction (Paris, France, 2023); XXIV Congress of the Russian Physiological Society named after Ivan Pavlov (St. Petersburg, 2023).

Publications

40 printed works have been published on the topic of the thesis, including 18 articles, of which 4 in the Higher Attestation Commission peer-reviewed scientific media, 13 in WOS and Scopus peer-reviewed scientific media.

Scope and structure of the work

The thesis comprises an introduction, a literature review, a description of the research methodology, results, their discussion, conclusions, a list of abbreviations, a list of references, and an appendix. The text of the thesis is presented on 108 pages, contains 36 figures, 42 tables. The list of references includes 156 sources, of which 116 are foreign.

This work was supported by a personal grant from the Acoustical Society of America (2019). Involvement as co-performer in grant-funded projects: RSF — No. 18-18-00063, international project RSF-DST No. 22-45-02007 (supervisor E. E. Lyakso); RFBR — No. 17-06-00503-OGN (supervisor E. E. Lyakso), 15-06-07852 (supervisor O. V. Frolova).

Main scientific results

1. For children growing up in a Russian-speaking environment, for the first time in a perceptual experiment, features of adults’ recognition of information contained in the speech of children with ASD and TD children were identified: lexical meaning (Nikolaev et al., 2018, p. 68; Lyakso et al., 2019, p. 7; Lyakso et al., 2021, pp. 1399–1400; Lyakso et al., 2022, p. 584–587), psychoneurological state (Frolova et al., 2019, p. 106–108; Lyakso et al., 2020, pp. 298–299; Lyakso et al. 2020, pp. 70, 72; Lyakso et al., 2021, pp. 1399–1401; Lyakso et al., 2022, pp. 584–587), emotional state (Lyakso et al., 2020, pp. 299; Lyakso et al., 2020, pp. 72–73; Lyakso et al., 2021, pp. 1399–1401; Lyakso et al., 2022, p. 584–587; Lyakso et al., 2023, p. 476– 477). The

author's personal participation consisted in the selection of speech material for perceptual experiments, the selection of groups of adults (auditors) listening to children's speech material, compilation of test sequences, conduct of perceptual experiments, and statistical processing of the data obtained.

2. Physiological and psychophysiological characteristics of auditors have been identified that influence their recognition of the lexical meaning of words of children with ASD and TD children, their psychoneurological and emotional state: gender, age of the auditor, everyday and professional experience of interacting with children, auditor's hearing thresholds, leading hemisphere for speech, level personal and situational anxiety (Frolova et al., 2019, p. 106–110; ; Frolova et al., 2020, p. 753; Lyakso et al., 2020, pp. 1400–1401; ; Nikolaev et al., 2020 , pp. 657; Lyakso et al., 2022, pp. 585–587). The personal participation of the author of the dissertation consisted of preparing and conducting perceptual experiments, physiological and psychophysiological testing of auditors, and statistical processing of data obtained during the experiments.

3. Data on adults' recognition of information about a child based on the characteristics of his voice were used in the development of methods for automatic classification of emotional states based on speech signals (Matveev et al., 2022, pp. 6–8; Matveev et al., 2023, p. 12–13; Lyakso et al., 2023, pp. 543–545). The author's contribution is the preparation of audio files and their annotation for automatic recognition.

4. The study describes acoustic characteristics of the speech of children with ASD and TD children: speech rate, duration of stressed and unstressed vowels, duration of pauses in utterances, values of the fundamental frequency (F0) of stressed and unstressed vowels, values of formant frequencies — which influence on the recognition of the lexical meaning of words (Nikolaev et al., 2018, pp. 69–70; Frolova et al., 2019, pp. 109–110; Lyakso et al., 2020, pp. 296–298), psychoneurological (Frolova et al., 2019, p 109; Lyakso et al., 2020, pp. 296–299) and emotional (Lyakso et al., 2020, pp. 299–300; Lyakso et al., 2023, p. 474) state of children. Linguistic features of the speech of children with ASD and TD children in dialogues with adults are described: the structure of responses, the frequency of occurrence of various parts of speech and their diversity (Nikolaev et al., 2019; pp. 95–98; Makhnytkina et al., 2021, pp. 401–404). The personal participation of the author of the thesis in obtaining the results consisted of the selection of speech material from children with ASD and TD children, instrumental analysis of speech material, and statistical data processing.

Thesis statements to be defended

1. Recognition of the lexical meaning of words of children with autism spectrum disorders is influenced by the gender, age and anxiety level of the auditor; recognition of the meaning of words of typically developing children is influenced by the auditor's hearing thresholds; fundamental frequency and vowel duration values influence the auditors' recognition of the meaning of words of children with autism spectrum disorders and typically developing children.

2. Classification of the psychoneurological state of children is influenced by the auditor’s gender, hearing thresholds, the dominant hemisphere for speech, anxiety level, as well as the features of children's speech, such as: the structure of responses in dialogues with adults, speech rate, values of fundamental frequency and duration of vowels in words.

3. Recognition of the emotional state of children with autism spectrum disorders is influenced by the auditor’s gender, experience in interacting with children, and hearing thresholds; recognition of the emotional state of typically developing children is influenced by the auditor’s gender, age, dominant hemisphere for speech, anxiety level; the correct recognition of the emotional state of children with autism spectrum disorders and typically developing children is influenced by the values of the fundamental frequency of phrases, words and stressed vowels, and the values of the variability of the fundamental frequency.

1. LITERATURE REVIEW

1.1. Method of perceptual experiment in speech research

The perceptual experiment method is widely used in studies aimed at studying speech perception (Bondarko, 1998). The method is used in linguistic, psychological, and biomedical research. Perceptual experiments are used in works concerning the study of listeners' recognition of speech signals: single phonemes, words, sentences, voice in general; their distinctions and comparisons (Liberman et al., 1967; Gelfer, 1988; Bruckert et al., 2006; Goy et al., 2016).

One of the areas of research in speech perception is the study of the acoustic-phonetic structure of speech. This includes studies of the intonation of speech messages based on material from different languages. Works that study intonation examine how listeners determine both the general type of utterance (narrative, interrogative, motivating) and its more specific characteristics, like rudeness, politeness, uncertainty, irony, etc. (Ward, Hirshberg, 1985; Rockwell, 2000; Scharrer, Christmann, 2011; Duryagin, Fokina, 2021).

The issues of isolating and describing acoustic keys for determining and categorizing vowel and consonant phonemes, stress locations, and verbal boundaries are explored. In these studies, data obtained from perceptual experiments are compared with data from spectrographic analysis of speech signals (Sanders et al., 2002; Yagunova, 2012; Krivnova et al., 2018).

Peculiarities of speech development in children with typical and atypical development are widely studied. Data on the acoustic characteristics of children's speech from birth to adolescence were obtained using material from different languages (Lyakso, 1998; Kuhl, Meltzoff, 1996; Vorperian, Kent, 2007; Ballard et al., 2012, Brehm et al., 2012; Lyakso, Grigoriev, 2013). Acoustic features of the speech of children with ASD have been described: high values of duration and F0 of vowels, F0 variability, atypical intonation, unformed or incorrect pronunciation of some consonant phonemes (Cleland et al. 2010; Boucher, 2012; Wolk, Brennan, 2016, Lyakso et al., 2017); Down syndrome: increased F0 values and vowel duration compared to these of TD children, atypical pronunciation associated with the anatomical features of the articulatory apparatus, decreased difference between the formant frequencies of cardinal vowels (/a/, /i/, / u/) (Moura et al., 2008; Kent, Vorperian, 2013; Carl et al., 2020, Frolova et al., 2019), Williams syndrome (Rossi, Giacheti, 2017; Loveall et al., 2021), hearing disorders (Frosolini et al., 2023).

Another area of speech research is the study of the possibility of determining individual characteristics of the speaker based on auditory perception. It has been shown that it is possible to determine from voice and speech the age and gender of the speaker (Goy et al., 2016), their height and weight (Bruckert et al., 2006), race and ethnicity (Purnell et al., 1999; Perrachione et al., 2010).

Auditory perception studies examine the ability of listeners to recognize speech features of people with various diagnoses, such as Down syndrome (Jones et al., 2019; O'Leary et al., 2020), autism spectrum disorder (Lyakso, Frolova, 2020), dysarthria (Schölderle, Haas, Ziegler, 2023). The ability of listeners to determine, based on auditory perception, whether speech belongs to a person with typical or atypical development is shown. One of the methods for identifying speech features characteristic of a particular diagnosis is the use of rating scales. In (Jones et al., 2019), listeners were asked to rate the speech and voice of children with Down syndrome based on a number of characteristics, such as speech loudness, monotony, rate, intelligibility, etc.

In a number of studies, the reflection (Grigorev et al., 2022) and the recognition of the emotional state of a speaker is examined. It has been shown that adults are able to determine the emotional state of a child from speech (Kaya et al., 2017); throughout childhood and adolescence, the ability to recognize the emotional state in speech increases (Grosbras, Ross, Belin, 2018; García-Guerrero et al., 2022). Listeners are able to determine the emotional state even when listening to speech in an unfamiliar language, i. e. focusing only on the characteristics of the voice, although less accurately compared to speech in the native language (Lyakso et al., 2023; Lyakso et al., 2023).

The features of recognition of the emotional state of both children and adults with typical development and children with various diagnoses are studied, the latter including ASD (Lyakso et al., 2018; Lyakso et al., 2021), Down syndrome (Lyakso et al., 2019; Lyakso et al., 2020), intellectual disabilities (Frolova et al., 2023), hearing disorders (de Jong et al., 2023). It has been shown that listeners have difficulties when recognizing the emotional state of children with atypical development compared to recognizing the state of TD children.

As a control one, the perceptual experiment method is also used in the development of systems for automatic recognition of speech information. The quality of such systems is determined by comparing with the results obtained in human speech recognition (Spille et al., 2018; Lyakso et al., 2022; Matveev et al., 2022; Matveev et al, 2023).

1.2. The influence of physiological and psychological characteristics of listeners on the determination of information contained in speech

It is being researched whether the personal characteristics of the listener influence their perception of speech material. Thus, evidence has been obtained that speech perception is influenced by the age of the listener and speaker (Smith et al., 2012; Goy et al., 2016). The work (Smith et al., 2012) showed that young listeners are better than older listeners at recognizing distorted speech signals. In (Goy et al., 2016), young and older listeners rated the pleasantness of the speaker's voice and determined their age and gender. Young listeners were more likely to rate the voices of young people as pleasant than older ones and were better at judging the age of the speaker based on the voice, while older listeners were better at identifying gender.

Cross-cultural studies have shown the influence of a listener's native language on recognizing the speaker's emotional state, and specifically when listening to pseudo-words and nonsense messages (Jiang et al., 2015; Lyakso et al., 2023).

The gender of the listener may influence the recognition of the emotional state of the speaker. In (Scherer, Scherer, 2011), listeners were offered German pseudo-sentences spoken by professional actors in five emotional states, namely joy, sadness, anger, fear, and neutral state. In (Lausen, Schlacht, 2018), the stimulus material included both semantically meaningful words and phrases, as well as pseudo-words and pseudo-sentences spoken by actors in six emotional states, namely joy, sadness, anger, fear, disgust, and neutral state. Both studies found that women were generally better than men at recognizing emotional states, but the difference in the probability of recognizing specific emotions was slight.

Listeners who have professional experience working with children with developmental disorders recognize the psychoneurological state of children from their voices better than listeners who do not have such experience (Frolova et al., 2019).

Research on the influence of hearing characteristics on speech perception has been conducted primarily in the context of hearing disorders (Carvalho et al., 2016; Hoppe et al., 2022). Studies of people with normal hearing are few. The work (Lyakso et al., 2017) shows the influence of hearing thresholds and phonemic hearing indicators on determining the meaning of what was said, determining intonation and age of children with ASD. Listeners with lower hearing thresholds are better at recognizing the meaning of children's words; listeners without phonemic hearing losses are better able to determine the intonation and age of children with ASD.

The influence of the level of anxiety on the perception of auditory and visual information is studied. In (Mattys et al., 2013), listeners with high and low level of anxiety were asked to distinguish between similarly sounding words; listeners with high anxiety performed worse on this task. High levels of anxiety may reduce success in recognizing faces in photographs (Attwood et al., 2013). A study of the influence of anxiety on emotion recognition (Kang et al., 2019) found that people with high levels of anxiety are worse at recognizing positive emotional expressions from facial expressions, but better at recognizing negative emotional manifestations. Research on the influence of anxiety on the recognition of emotions in speech is scarce and focuses mainly on patients with anxiety disorders. The data from these studies are inconsistent, both higher (Madison et al, 2021) and lower (Tseng et al., 2017) accuracy in recognizing emotional expressions in voice has been observed.

It should be noted that works considering the influence of physiological and psychological characteristics of listeners on the recognition of information contained in speech rarely use the Russian speech material, exceptions are (Lyakso et al., 2017; Frolova et al., 2019).

1.3. The process of speech production and cerebral organization of speech

Speech is a specific form of human behaviour that serves two inextricably linked aspects of their activity — communicative and cognitive (Lyakso et al., 2012).

Speech production is a strictly coordinated behavioural act, consisting of numerous gestures of the articulatory organs, through which the most important mechanism for the exchange of information between people is realized (Derkach et al., 1983). The speech signal is a sound wave and serves as a means of transmitting both verbal and nonverbal information. To create a speech signal, the speech apparatus is used, combined with the respiratory and food grinding organs (Zinder, 1979).

Human speech apparatus includes sound generator (larynx), resonator part (pharynx, nasal cavity, maxillary sinuses), energy apparatus (trachea, bronchi, lungs, diaphragm) and articulatory apparatus (oral cavity, tongue, teeth, lips, soft and hard palate) (Kodzasov, Krivnova, 2001) (Fig. 1).

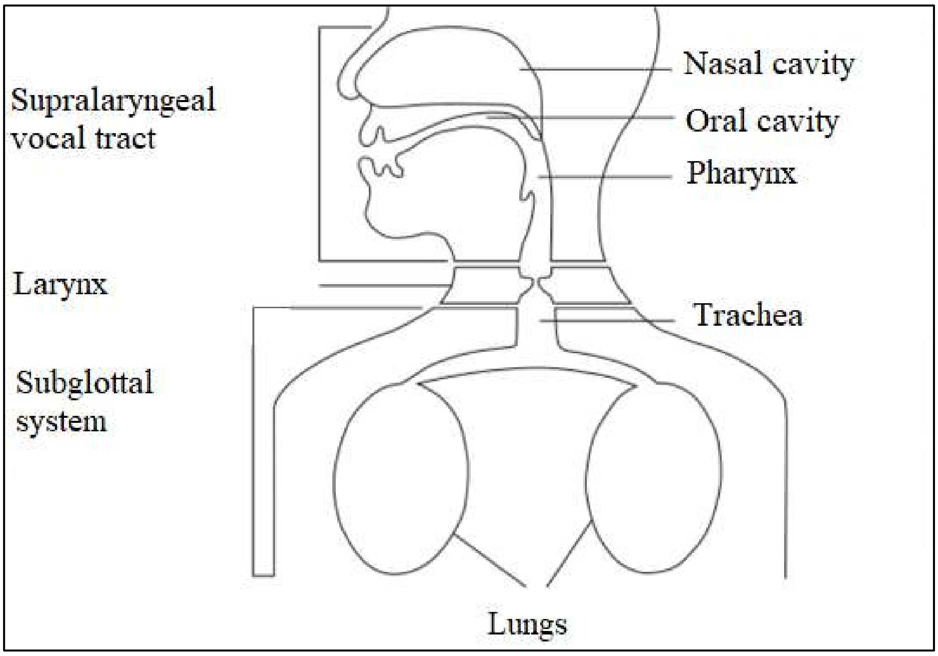


Fig. 1. Human vocal tract, schematically (Kodzasov, Krivnova, 2001)

The main energy element of speech production is air. Air flows from the lungs through the bronchi and trachea to the larynx, where sound is produced. Inside the larynx there are vocal folds, which begin to oscillate due to subglottic pressure and their own tension.

To form the voice, the vocal folds are brought together with the help of the arytenoid cartilages to which they are attached. Contraction of the vocal folds blocks the air and forms a difference between pressures above and below the folds. As a result of the pressure difference, the folds open and air begins to penetrate into the resulting gap. When the vocal folds are displaced, elastic forces arise that help the folds return to their original position. Periodic opening and closing of the glottis convert airflow into a sequence of air pressure pulses (Lyakso et al., 2012). Different configurations of the glottis create different types of vibrations of the vocal folds. As a result of such vibration, a periodic complex sound appears, consisting of a fundamental frequency (F0) and a few tens of overtons, i. e. harmonics of the fundamental frequency (Derkach et al., 1983). The resulting sound is subject to frequency filtering because of changes in the volumes of the vocal tract cavities. Changing the position of the articulatory organs affects the formation of the speech signal; each such position corresponds to a separate sound — a phoneme.

A phoneme is the minimal sound unit of language (Bondarko, 1998). All phonemes are divided into vowels and consonants. The Russian language has 6 vowels and 36 consonant phonemes. Vowels (except for whispered speech) are pronounced with the obligatory participation of the vocal folds; to pronounce consonants, the participation of the vocal folds is not necessary. To pronounce vowels, the leading role is played by the position of the lips and tongue; to pronounce consonants, the participation of the entire articulatory apparatus is necessary.

The position of the tongue when articulating vowels changes in two planes: according to its height in the oral cavity and according to its backness. For Russian language, three heights are distinguished: high, mid and low. High vowels are also called closed vowels, and low vowels are called open vowels. According to the backness, there are three rows of vowels: front, central and back. Based on the position of the lips, rounded (labialized) vowels and unrounded (non-labialized) vowels are distinguished (Bondarko, 1998) (Table 1).

Table 1. Russian vowels (Bondarko, 1998)

<table>
<tr><td rowspan="3">Height</td><td colspan="3">Backness</td></tr>
<tr><td>front</td><td>central</td><td>back</td></tr>
<tr><td colspan="2">unrounded</td><td>rounded</td></tr>
<tr><td>Close</td><td>/i/</td><td>/ ɨ/</td><td>/u/</td></tr>
<tr><td>Mid</td><td>/e/</td><td></td><td>/o/</td></tr>
<tr><td>Open</td><td></td><td>/a/</td><td></td></tr>
</table>

The most energetically expressed maxima, due to the work of the vocal tract as a resonator and filter, are called formants (Fant, 1970). The maximum number of formants depends on the length of the vocal tract. Formants characterize the distribution of signal energy along the frequency axis. Formants are designated as F1, F2, F3, etc., starting from the low-frequency part of the spectrum. The most important for distinguishing vowel sounds are low-frequency formants, primarily F1, F2 (Derkach et al., 1983), which are called acoustic keys for assigning a sound to one or another phoneme category (Chistovich et al., 1965). The frequency of the first formant is related to the height of the vowel: the frequency of the first formant is lower in high vowels and higher in low vowels. Second formant frequency is related to vowel backness: front vowels have a higher second formant frequency than back vowels.

The pronunciation of consonants is influenced by the place of articulation, the manner of articulation, the position of the vocal folds, of the tongue and of the soft palate.

Speech sounds are usually not pronounced in isolation, but in combination with each other; coarticulation occurs, which indicates the complexity of this process (Kodzasov, Krivnova, 2001).

Speech function is ensured through the work of the prefrontal, temporal, occipital areas of the cerebral cortex, as well as the work of limbic system. The activity of speech apparatus is controlled by the cerebral cortex. In speech production, the main role belongs to the left hemisphere, with the right hemisphere being actively involved (Vartanyan, 1988). In the left hemisphere there are centres associated with speech activity. These are: Broca's area (Brodmann areas 44–45) is the centre of speech production; Wernicke's area (Brodmann area 22) is the centre of speech perception; additional motor area, which is a part of the frontal lobe (Batuev, 2009). However, it is believed that the processing of speech information occurs through the interaction of two hemispheres: the left hemisphere analyzes the speech signal segment-by-segment, and the right hemisphere uses the holistic analysis, comparing the acoustic image of the signal with standards from memory (Morozov et al., 1988). The right hemisphere dominates when words with background noise are to be perceived (Galunov et al., 1986).

When perceiving speech, signals first arrive in the primary auditory cortex. Phonemic analysis and understanding of information are performed in Wernicke's area, located in the temporal lobe. When a word is pronounced, its representation is transmitted from Wernicke's area to Broca's area through the arcuate fasciculus. In Broca's area, the representation of the word is converted into an articulatory program, which then enters the facial area of the motor cortex. When reading, information first arrives in the primary visual cortex, then is sent to the angular gyrus, where the visual form of a word is connected with its sound counterpart in Wernicke's area. In modern studies it is shown that Broca's area participates in the perception of both oral and written speech (Wilson et al., 2004; Walenski et al., 2019), as well as in the processing of the syntactic structure of an utterance (Friederici, 2002).

Speech disorders, like delayed speech development, atypical speech development, regression in speech development, etc., may be caused by functional and anatomical changes or disturbances in the functioning of the peripheral and central parts of the speech and auditory systems. The most severe disorders occur when defects appear in the prenatal and/or early postnatal periods of development. Disturbances in speech development in autism spectrum disorders are associated with abnormal functioning of the speech system due to changes in prenatal ontogenesis (Lyakso et al., 2020).

1.4. Characteristics of speech development in typically developing children aged 5–14 years

The period of 5–14 years comprises preschool (5–7), primary school (8–11) and adolescent (12–14) age. During the preschool period, children master phonetics of their native language, they learn to pronounce words with different combinations of phonemes and syllables. By the age of five, the formation of the correct pronunciation of most phonemes is completed; word stress is formed, but phrasal stress is not formed yet (Lyakso, Gromova, 2005). During this period, the syntactic structure of utterances becomes more complex, response utterances acquire features of monologue speech, namely: longer duration, correctly structured text, absence of lexical and grammatical errors (Lyakso, Stolyarova, 2008). At the age of 5–6 years, children develop the ability to create text messages independently (Tseitlin, 2000). By the age of 7, children produce more remarks consisting of several phrases. At this age disruptions in children' speech breathing is observed when they pronounce complex phrases. This effect is associated with difficulties in the lexical and grammatical formatting of statements (Belyakova, Filatova, 2008). At the same age, the lexical and grammatical formation of speech is completed (Belyakova, Voloskova, 2009).

Once they start school, children develop reading and writing skills; the passive vocabulary at this age includes 7–8 thousand words and increases to 12–15 thousand by the end of primary school (Lvov, 2000). Children master abstract concepts, they form ideas about the polysemy of words, and the proportion of complex sentences in speech increases.

With age, vowel duration and frequency values lessen in children's speech. By the age of 12–15 years, they reach values characteristic of adult speech (Lee et al., 1999). Stabilization of the duration values of stressed and unstressed vowels in children may indicate that the process of speech breathing formation is completed (Hoit et al., 1990).

From 8–12 years of age, along with the processes of puberty, the values of the frequency characteristics of vowels lessen, gender differences begin to appear, the voice of boys becomes lower than that of girls (Lee et al., 1999; Lee and Iverson, 2008). Using material from different languages, the nonlinear nature of changes in F0 values in boys with sharp decreases at 6–8 and 12–15 years of age has been described, while in girls the decrease in F0 values occurs more evenly (Markova et al., 2016; Grigorev et al., 2018).

Works based on the Russian language material consider the temporal and spectral speech characteristics for ages from 5 to 16: dynamics of changes in fundamental frequency, formant frequencies of vowels, and duration of stressed and unstressed vowels is traced (Lyakso, Grigoriev, 2013; Grigoriev, Lyakso, 2014; Grigoriev, Lyakso, 2017; Lyakso et al., 2018).

1.5. Characteristics of children with autism spectrum disorders

1.5.1. Common features of autism spectrum disorders

The term "autism" was first introduced by the Swiss psychiatrist E. Bleuler in 1911 to denote immersion in internal experiences and separation from reality in schizophrenia. The first description of autism as an independent diagnosis was given by L. Kanner in 1943 in the work "Autistic Disturbances of Affective Contact", where he proposed to classify into a separate class of disorders such pathological conditions of children that come with self-isolation and start from the first years of life. These disorders got the name early childhood autism and were considered a type of schizophrenia (Kanner, 1943). In 1944, Austrian psychiatrist H. Asperger described autistic psychopathy, which was later called Asperger syndrome (Wing, 1981). In the 20th century, autistic disorders were classified as psychoses, developmental disorders, or schizophrenia.

Currently, autism spectrum disorders (ASD) are defined according to ICD-10 (International Classification of Diseases, 10th revision) as general developmental disorders and include several diagnoses, united by similar symptoms:

1) childhood autism, or Kanner syndrome (F84.0)

2) atypical autism (F84.1);

3) Rett syndrome (F84.2);

4) other childhood disintegrative disorder (F84.3);

5) overactive disorder associated with mental retardation and stereotyped movements (F84.4);

6) Asperger's syndrome (F84.5).

The complex of common symptoms observed with these diagnoses is called the autistic triad and includes disturbed or atypical social behaviour (especially interpersonal communications), limited forms of behaviour, an inclination to perform stereotyped actions, impaired language and speech. For children with ASD, the severity of autistic symptoms, the age of their onset, and the presence of a leading symptom complex are individual (Wing, 1993). ASD is detected predominantly in boys (Volkmar and Nelson, 1990), although the specific ratio remains not quite clear, which may be associated with the features of diagnostics (Harrop et al., 2024).

Childhood autism usually manifests in children between 18 and 30 months; additional symptoms of childhood autism also include auditory and tactile hypersensitivity, manifestations of aggression towards others and self-aggression, lowered pain threshold, hyperactivity, and sudden mood swings (Semyannikova, 2013).

Atypical autism is characterized either by a specific age of onset, different from childhood autism, or by absence of one of the features of the autistic triad. This diagnosis is most often found in children with mental retardation, which leads to the appearance of behaviour characteristic of autism (Neznanov et al., 2009).

Rett syndrome occurs exclusively in girls. The first developmental disorders appear between 4 months and 2.5 years. Rett syndrome is characterized by a slowdown in the rate of psychomotor development; previously acquired skills regress, children loss contact with others, breathing disorders are often observed, and convulsive seizures may occur. In future, children with Rett syndrome are characterized by severe mental retardation and movement disorders (Hagberg et al., 1983).

In childhood disintegrative disorder, the development of a child up to the age of 1.5–4 years proceeds without any anomalies, and then severe symptoms of autism appear and a significant regression of previously acquired skills occurs (Neznanov et al., 2009).

Asperger's syndrome differs from childhood autism in early speech development, retention of adaptive skills, and normal or high level of intelligence. Savantism — a condition in which a person has outstanding abilities in one or more areas, but a generally limited behaviour — is defined as a subset of Asperger's syndrome (Wing, 1993).

According to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5), ASD is defined as a spectrum of psychological characteristics describing a wide range of abnormal behaviour and difficulties in social interaction and communication, as well as severely restricted interests and highly repetitive behaviours, and includes autism (Kanner syndrome), Asperger syndrome, childhood disintegrative disorder, nonspecific pervasive developmental disorder.

Currently, standardized diagnostic questionnaires and scales are actively used to diagnose ASD. A diagnostic questionnaires is a structured conversation with parents, with focus both on the symptoms and the child’s developmental history.

The methodology of rating scales is to assign scores to certain characteristics. One of the most commonly used scales is the Childhood Autism Rating Scale (CARS). The rating on this scale is based on information received from parents. CARS contains 15 sections, each one assessing a specific area. The scale takes into account the frequency and intensity of the identified pathological signs (Schopler et al., 1980). According to this scale, scores of less than 30 indicate no autism, scores of 30–36 indicate mild or moderate autism, and scores of 37–60 indicate severe autism.

1.5.2. Speech features of children with autism spectrum disorders

One of the main impairments evident in ASD is language impairment, which manifests itself as children's underdeveloped or absent spoken language skills compared to typically developing peers (Lakso et al. 2016; Watson et al. 2011). In the early stages of ASD, speech develops normally or even rapidly, but over time the reverse process begins, manifesting itself first as a stop in speech development, and then as regression. Depending on the severity of the disorder, speech impairments can manifest themselves at different levels of its organization (articulatory, grammatical, pragmatic). The level of speech development of children with ASD can vary from well-formed speech in children with high-functioning autism (Grossman et al., 2013) to complete absence of speech (mutism). Common pathological features of speech for most children with ASD are echolalia, poor vocabulary compared to peers, and violation of the grammatical structure of phrases. Also, children with ASD have difficulties understanding other people's speech (McGregor et al., 2012); children with ASD are less likely to respond to their own name (Hatch et al., 2021). It is shown in (Kelley et al., 2006) that children with ASD are better at understanding the meaning and producing single words rather than phrasal structures.

Children with ASD exhibit "normal" and "specific" speech. "Normal speech" includes words and phrases, though violations or immaturity of different levels of its organization (articulatory, grammatical, pragmatic) are noted. "Specific" speech is represented by echolalia, repetition of syllables and the presence of a "own language" — slurred vocalizations, the meaning of which is unclear even in the context of the situation (Lyakso et al., 2016).

Children with ASD have disturbances in dialogical speech, during a conversation their speech is not addressed to the interlocutor, and there is often a lack of expression in utterances. The speech of children with ASD is represented predominantly by single words and short simple phrases (Rapin et al., 2009; Lyakso et al., 2016; Nikolaev et al., 2019; Makhnytkina et al., 2021), and nouns dominate in their vocabulary (Tek et al., 2014; Boorse et al., 2019). Children with ASD use words peculiarly, selecting words with wrong meanings, for example. In (Boucher, 2012) it is shown that children with ASD lag behind compared to TD children in mastering grammatical structures, namely: prepositions of direction, plurals, auxiliary and modal verbs, pronouns. Children with ASD differ from TD children in their use of personal, demonstrative (Terzi et al., 2019; Mazzaggio, Shield, 2020) and indefinite pronouns.

Most children with ASD have articulation disorders (Boucher, 2012), manifested in the form of incorrect or atypical pronunciation of phonemes (Wolk, Brennan, 2016; Lyakso et al., 2019), unformed affricates, incorrect pronunciation of consonant clusters, etc. (Cleland et al., 2010). A study based on the English language material showed that the speech of children with ASD at the age of three years contains a smaller set of consonants than the speech of their TD peers (Schoen et al., 2011).

In the classic work (Kanner, 1943), the speech of children with autism is described as monotonous. However, modern studies of the speech of children with ASD using material from different languages indicate high values of the fundamental frequency and its variability (Bonneh et al., 2010; Sharda et al., 2010; Filipe et al., 2014; Lyakso et al. 2017; Nikolaev et al., 2018; Lyakso et al., 2020; Patel et al., 2020; Chen et al., 2022). There are also works that show absence of significant differences between F0 values of children with ASD and TD children (Schriberg et al., 2011), which is apparently due to speech recording situations. Children with ASD have abnormal prosody and atypical word and phrase stress (Paul et al., 2005; Bonneh et al., 2010; Diehl, Paul, 2013), and a lower speech rate (Patel et al., 2020). Some studies describe the speech of children with ASD as monotonous (Redford et al., 2018), while others indicate greater intonation variability compared to typically developing peers (Diehl et al., 2009; Fusaroli et al., 2017).

According to (Lebedinskaya, Nikolskaya, 1991), speech disorders in children with ASD are divided into four types. In the first type, speech development disorders occur at 2–2.5 years, usually against the background of a disease or some other negative development. Often children lose their acquired speech skills completely, but understand speech addressed to them. The second type of disorder is characterized by delayed speech development, persistent pronunciation disorders, agrammatisms and an atypical speech rate. The third type of disorder is characterized by rapid and early speech development of children with ASD; children's vocabulary expands quickly, but their speech is monologic, contains a large number of cliches, and sometimes the speech is abrupt and scanning. With the fourth type of disorder, the early speech development of children with ASD is close in time to the development of speech in children with TD, however, at the age of 2–2.5 years, the speech activity of children decreases and speech development stops, hereon the active vocabulary of children reduces, phrasal speech almost disappears, and echolalia emerges; still, speech understanding is at a higher level than with other types of speech disorders.

The presented data indicate that the attention of researchers is mainly focused on describing different levels of speech development in children with ASD. Works aimed at studying the speech perception of informants with ASD are rare.

2. METHODS

2.1. Object of study

Participants of the study were 208 adult native Russian speakers (auditors) aged 25±4,5 years (male 100, female 106); 2 experts; 82 children aged 5–14 years: 35 children with ASD (30 boys, 5 girls) and 47 TD children (37 boys, 10 girls).

Children with ASD had CARS scores (Schopler et al., 1980) between 31 and 43 which corresponds to mild, moderate and severe autism disorder. A group of children with ASD is represented mainly by boys which corresponds to the frequency of occurrence of autism disorders (Volkmar, Nelson, 1990; Nicholas et al., 2008), therefore a group of TD children is organized in the similar way. TD children had neither diagnosed chronic diseases, nor hearing and vocal apparatus impairment.

Speech material of children was selected from "AD_Child.Ru" (Lyakso et al., 2019) database and included words and phrases cut from records of spontaneous speech. "AD_Child.Ru" database contains files of children speech recorded in standard model situations: child plays with standard set of toys; child describes a picture and answers to the experimenter's question about it; the experimenter and child are engaged in a dialogue during which experimenter asks a standard set of questions about school, friends, family, pets, hobbies etc.

The speech material of the database is recorded with professional equipment: digital recorder "Marantz PMD660" (recording frequency range is 20–16000 Hz) with remote microphone "SENNHEISER e835S" (sound recording path has a linear, uniform frequency response in the range 80–14500 Hz). File recording format is WAV. The database contains recordings of children's behaviour. Their behaviour must be taken into account when analyzing the speech material of children with atypical development because their speech can be incomprehensible without situational context. Video recordings were made with camera "SONY HDR-CX560" (maximum resolution is 1920x1080 at 50 FPS).

2.2. Design of the study

1. The study protocol is being developed.
2. Adults (auditors) for perceptual research are selected. Auditors sign an informed consent to taking part in the study.
3. Physiological and psychophysiological testing of auditors is performed: hearing thresholds, dominant hemisphere for speech, profile of functional lateral asymmetry, electrocardiogram indicators and heart rate, anxiety level are determined.

4. Speech recordings of children (children with ASD, and TD children as control group) are selected for further analysis.
5. Linguistic analysis of children's speech is conducted.
6. Perceptual experiment is carried out: test sequences for perceptual analysis are compilated, questionnaires for auditors are developed; perceptual experiment per se is performed; personal data is processed.
7. Instrumental analysis of children's speech is conducted.
8. Statistical analysis is conducted.

2.3. Research methods

2.3.1. Physiological and psychophysiological methods

All auditors had their hearing thresholds determined using pure-tone audiometry and dichotic testing (determining the coefficient of lateral preference for speech). The profile of functional lateral asymmetry was determined (PFLA, a series of tests to identify the dominant eye, ear, arm, and leg), phonemic hearing was checked, and the level of personal and situational anxiety was determined (Spielberger test).

2.3.1.1. Hearing thresholds

The hearing thresholds of auditors were determined using an AA-02 audiometer in automated mode. Frequency of presented signals was 125–8000 Hz, intensity was 0–95 dB. The pure-tone audiometry method is used to determine the hearing thresholds of pure (sinusoidal) tones. The audiometer allows to measure both air (through headphones) and bone conductivity (through a vibrator installed behind the ear, if any pathologies of the outer and middle ear are present). Hearing thresholds are measured in decibels (dB) in relation to average normal hearing thresholds. The difference between measured and normal threshold characterizes the loss of hearing. According to the classification of the World Health Organization (1997), normal hearing thresholds are in the range of 0–25 dB. During the hearing thresholds assessment, the audiometer generates pure tones in the range from 125 to 8000 Hz, which are presented to the auditor through headphones. At each frequency, a gradually increasing signal is applied in steps of 5 dB until the auditor has an auditory sensation, which is the hearing threshold. When an auditory sensation occurs, the auditor should press a button. Hearing thresholds are determined for the left and right ears separately. The results of audiometry were recorded in special audiogram forms (Fig. 2).

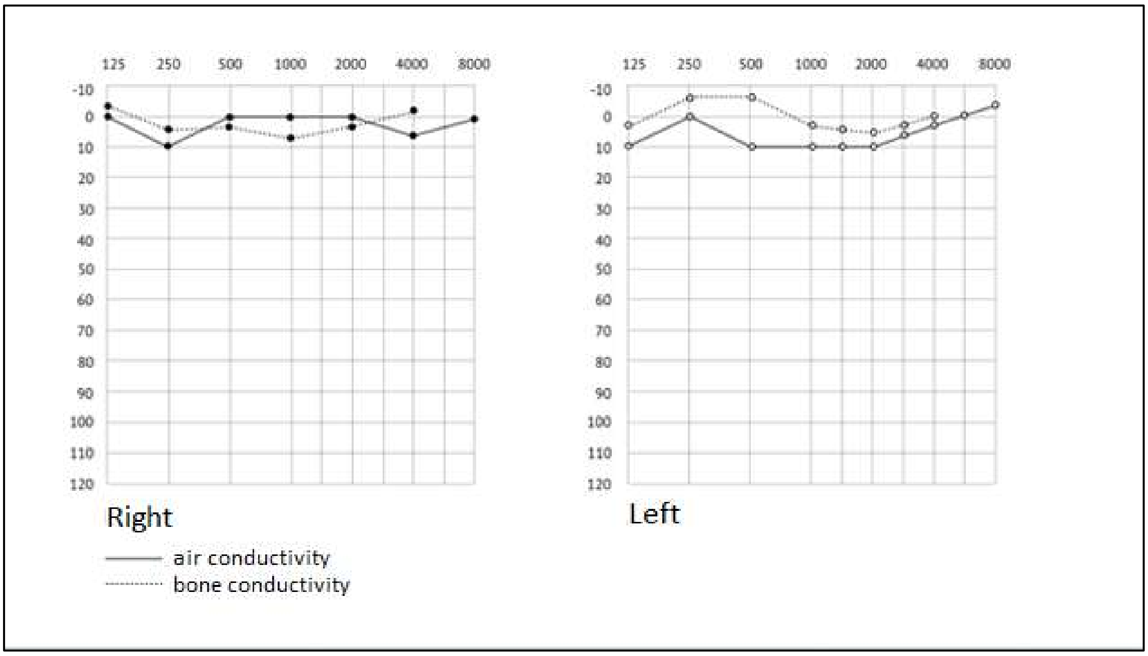


Fig. 2. An example of an audiogram with normal hearing thresholds not exceeding 25 dB at all frequencies tested.

2.3.1.2. Dichotic testing

To determine the dominant hemisphere for speech, auditors were presented through headphones a test sequence consisting of 60 pairs of words, in each pair one word was presented to the left ear, the other to the right. The auditor had to repeat the words he heard, and the experimenter noted on the form the answers from the left ear and the answers from the right ear. The lateral preference coefficient was calculated using the formula (Kimura, 1961):

$$LPC = \frac{\mathrm{Rc} - \mathrm{Lc}}{\mathrm{Rc} + \mathrm{Lc}} \times 100\%,$$

where ***Rc*** is the number of right choices, ***Lc*** is the number of left choices.

2.3.1.3. Phonemic awareness test

Auditors who were asked to determine the lexical meaning of children's words were administered a phonemic awareness test (Lyakso et al., 2012). Auditors were presented from a tablet with a two (e.g., ka-ga, ta-da) and three syllables (e.g., ba-ba-pa, fa-fa-va) test consequences selected by a speech therapist and pronounced by a professional speaker. The test included 18 sequences of syllables, 9 two-syllabic and three-syllabic, respectively. The test was presented in an open field. The auditor had to correctly repeat the syllables heard.

2.3.1.4. Determination of the profile of functional lateral asymmetry

The profile of functional lateral asymmetry was determined according to the method (Nikolaeva, 2005). The definition of PFLA included several tests to identify the leading hand, leg, eye, ear, such as opening a matchbox, catching the ball with one hand, crossing one's arms over one's chest, jumping on one leg, taking a step forward, crossing one's legs, looking through a telescope, blink one eye, put the phone to one's ear, listen with one ear to the ticking of the clock. For each of the tasks, the LPC was calculated:

$$\boldsymbol{LPC} = \frac{Rc - Lc}{Rc + Lc} \times 100\%,$$

where $\boldsymbol{Rc}$ is the number of right choices, $\boldsymbol{Lc}$ is the number of left choices.

Then the overall LPC was calculated for all tasks.

2.3.1.5. Recording an electrocardiogram and measuring heart rate

In an additional experiment, auditors listened to two tests, one of which contained emotional speech of TD children, the other that of children with ASD. While listening to the test, the auditors recorded an ECG using an "EKGK-02 Valenta" electrocardiograph in all limb leads (standard — I, II, III; and enhanced — aVL, aVR, aVF). Synchronically with ECG recording, heart rate (HR) was measured.

First, a background ECG has been recorded during one minute, then the auditors were asked to listen to an audio test, during which the ECG recording continued, and afterwards, another minute of the background ECG was recorded.

During the processing of the results, changes in heart rate and RR interval values (average, maximum and minimum) of auditors before and after listening to the tests were analyzed.

2.3.2. Psychological test for determination of the anxiety level

The Spielberger test (Spielberger, 1970; Russian-language adaptation — Khanin, 1976) for assessing anxiety consists of 20 statements related to anxiety as a state (state of anxiety, reactive or situational anxiety), and 20 statements to determine anxiety as a disposition, a personal trait (personal anxiety). In the first section of the test, the subject is asked to read each of the given statements and mark a number from 1 to 4 in the column to the right of the statement, depending on how he feels at the current moment, where 1 is "no, that's not true," and 4 is "absolutely true." In the second section of the test, the subject is asked to read and mark each of the statements in the same way, depending on how he usually feels.

The final score for each section can range from 20 to 80 points. A score of less than 30 points indicates a low level of anxiety, a score from 31 to 44 stands for a moderate level, and more than 45 points scored mean a high level of anxiety.

2.3.3. Linguistic analysis of children's speech

To assess the level of development of children's speech, a linguistic analysis of the texts of the children's dialogues with the experimenter was carried out. It included assessing of the complexity of the phrases, and calculation of the frequency of occurrence of different parts of speech based on the methodology (Lyakso et al., 2012).

When assessing the complexity of children's utterances, the phrases were divided into the following categories: 1) speech-like constructions, the meaning of which is unclear even in the context of the situation; 2) yes/no answer; 3) one word; 4) a simple phrase; 5) two simple phrases; 6) several simple phrases; 7) complex sentence.

When calculating the frequency of occurrence of different parts of speech, the following were taken into account: 1) speech-like constructions; 2) nouns; 3) verbs; 4) adjectives; 5) adverbs; 6) pronouns; 7) auxiliary parts of speech (prepositions, conjunctions, particles etc.).

To assess the diversity of the types of phrases in children's dialogues and words represented by different parts of speech, the Simpson diversity index was calculated using the formula:

$$D = \sum_{i=1}^{Z} \left[\frac{n_i(n_i-1)}{N(N-1)}\right],$$

where $\boldsymbol{n_i}$ is the number of elements in group $\boldsymbol{i}$ ($\boldsymbol{Z}$ groups in total, $\boldsymbol{\Sigma n_i} = \boldsymbol{N}$), $\boldsymbol{N}$ is the sample size. Because the value of $\boldsymbol{D}$ decreases as diversity increases, the Simpson index is often used in the form ($\boldsymbol{1 - D}$) (Rosenberg, 2007).

2.3.4. Perceptual experiment

In order to assess the possibility of adults recognizing information contained in children's speech, a perceptual experiment was conducted.

For the perceptual experiment, 14 test sequences (audio tests) were created, each containing 30 speech signals (words, phrases): 13 tests contained an equal number of signals from children with ASD and children with TD; 1 additional test contained only signals from children with ASD. Each signal in the tests was presented once, the interval between signals was 5 s.

The selection of speech material for tests containing emotional speech of children was carried out by two experts who were not involved in subsequent perceptual experiments. Basing on the recording protocol, the experts annotated the speech material into three states: “comfort — neutral state — discomfort” while watching video recordings and listening to audio files. Speech material was included in the test sequence if both experts had the same assessment (the Cohen's κ-coefficient, or consistency between experts = 1.0).

There was no preliminary training for auditors. Test sequences were presented in an open field.

Depending on the speech material contained in the test, the auditors were faced with the following tasks:

1) determining the lexical meaning of children’s words;

2) determining the psychoneurological state of children (typical or atypical development);

3) determining the emotional state of children (comfort, neutral state or discomfort) (Table 2).

Table 2. Test Sequences Information

№	Speech material	Age of children, y. o.	CARS mean score	Task — to determine:	Number of auditors
1	Words	5–7	35±4.5	1) lexical meaning 2) psychoneurological state	17
2	Words	8–9	32.8±2.6	1) lexical meaning 2) psychoneurological state	12
3	Words	10–11	30.4±0.6	1) lexical meaning 2) psychoneurological state	10
4	Words	10–11	30.4±0.6	lexical meaning	10
5	Words	12–14	31.5±1.2	1) lexical meaning 2) psychoneurological state	16
6	Phrases	5–7	38.1±3.8	psychoneurological state	17
7	Phrases	8–9	35.5±3.5	psychoneurological state	10
8	Phrases	10–11	33.6±3.8	psychoneurological state	10
9	Phrases	12–14	31.5±1.2	psychoneurological state	17
10	Phrases	5–7	33.2±4.3	emotional state	13
11	Phrases	8–9	35±2.4	emotional state	10
12	Phrases	10–11	30.8±0.9	emotional state	13
13	Phrases	12–14	31.5±1.2	emotional state	13
14	Phrases	10–14	34±4.5	emotional state	38

Before taking the tests, auditors received questionnaires in which they provided information about themselves: gender, age, experience of interacting with children (presence of younger brothers or sisters, parental status, experience of working with children).

The study engaged 201 auditors aged 25±4.5 years (96 men, 105 women; 134 with experience of interacting with children, 67 with no such experience): biology students (n = 65), IT students (department of speech technologies of ITMO University) (n = 87), linguistic students (n = 16), 1st year paediatrics students (n = 19), psychiatry residents (n = 14). A separate group consisted of doctors — child psychiatrists with at least 10 years of practice (n = 5) (Table 3).

Table 3. Information about auditors

№	Group of auditors	Number of auditors, n			Age, y. o.	Experience		
		Overall	Men	Women		domestic	professional	none
1	Biology students	65	10	55	24.2±2.7	15	31	19
2	IT students	87	67	20	23±1.9	40	12	35
3	Linguistics students	16	5	11	26.5±2.5	5	6	5
4	1st year paediatrics students	19	9	10	18.5±1.5	12	0	7
5	Psychiatry residents	14	5	9	25.1±0.9	0	14	0
6	Child psychiatrists	5	3	2	47.2±3	0	5	0

The results of the perceptual experiment were presented in the form of confusion matrices. A confusion matrix is used to describe the performance of a classification model. It is a table, where the rows correspond to the given (predicted) classes, and the columns to the actual values (real classes). The results were assessed using the following metrics:

1) Recall — completeness of the system, the proportion of samples found by the classifier that belong to a class, relative to all samples of this class in the test sample:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+TN}},$$

where ***TP*** is correctly recognized objects of class 1, ***TN*** is correctly recognized objects of class 2.

2) Precision — the accuracy of the system within a class, the proportion of samples that actually belong to this class, relative to all samples that the system assigned to this class:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FN}},$$

where ***TP*** is correctly recognized objects of class 1, ***FN*** is incorrectly recognized as objects of class 2.

3) F_1-score — F_1-measure, the harmonic mean between accuracy and recall. F_1-score tends to zero if precision or recall tends to zero:

$$\text{F}_1\text{- score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precison} + \text{Recall}}$$

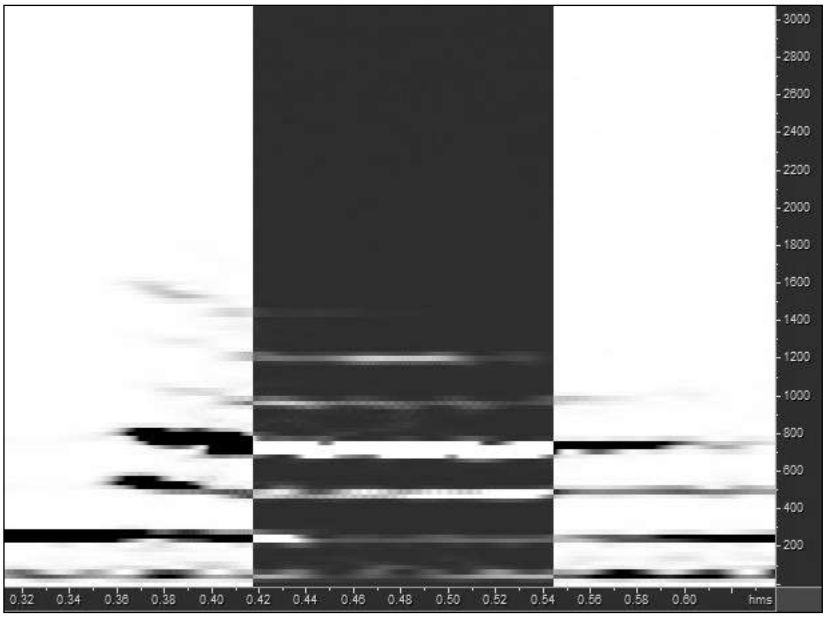
4) UAR — Unweighted Average Recall based on the mean of sensitivity and specificity, where (i) describes the number of correctly recognized elements of the ***i***-th class, $\boldsymbol{N0(i)}$ describes the total number of objects in the ***i***-th class, ***N*** describes the total number of objects, and ***k*** is the number of classes:

$$UAR = \frac{1}{k}\sum_{i=1}^{k}\frac{N_c^{(i)}}{N_0^{(i)}}.$$

2.3.5. Acoustic spectrographic analysis

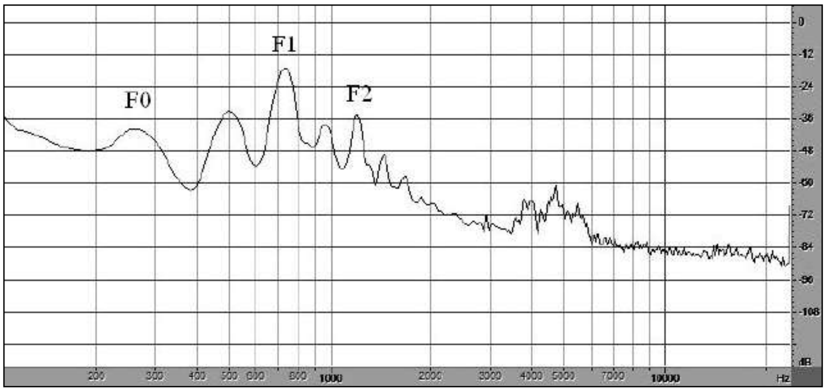
Acoustic spectrographic analysis was carried out using the Cool Edit Pro 2.0 program. The analysis included determining the duration of the word, the duration of stressed and unstressed (first pre-stressed and first post-stressed) vowels, determining the fundamental frequency (F0) and its intensity (E0) for words and vowels, the variability of F0 (the difference between the maximum (F0 max) and minimum (F0 min) the value of F0 in the word, [F0max − F0min]).

For vowels, a stationary site was determined (Fig. 3, 4). It was taken to be a section of the spectrogram, represented by a spectrum of one type, where the formant frequencies have a constant frequency and articulation changes scarcely (Bondarko, 1998). At the stationary site, the values of the first three formants (F1, F2, F3) and their intensities (E1, E2, E3) were determined. The formant intensity values were normalized with respect to the F0 intensity value (En/E0).



The horizontal axis is time, ms; vertical axis is frequency, Hz

Fig. 3. Dynamic spectrogram of a word with a highlighted stationary site of the stressed vowel /a/ in the word *vokzal* ‘railway station’



The horizontal axis is time, ms; vertical axis is intensity, dB

Fig. 4. Envelope of the spectrum of the stressed vowel /a/ in the word *vokzal* ‘railway station’

Based on the values of the first and second formants of cardinal vowels (/a/, /i/, /u/), the values of the vowel articulation index (VAI) were determined according to the formula (Roy et al, 2009; Lyakso et al., 2017):

$$VAI = (F1[a]+F2[i])/(F1[i]+F1[u]+F2[a]+F2[u]),$$

where ***F1[x]***, ***F2[x]*** are frequency values of the first and second formant of the corresponding vowel.

In phrases, the duration of the phrase, the duration of pauses, the speech rate (the number of syllables per second) were determined, and the word on which the semantic stress falls in the phrase was identified. For the selected word, the duration, the duration of stressed and unstressed vowels, the value of F0 for the word and for vowels, intensity, the values of F1–F3 and their intensities were determined.

2.3.6. Statistical analysis

Statistical data processing was carried out in the Statistica 12 software package. To test the hypothesis about the differences between the two compared groups, the nonparametric Mann–Whitney test was used. The critical level of significance when testing statistical hypotheses (p) was taken equal to 0.05.

To assess correlations, the Spearman rank test (at a significance level of $p < 0.05$), regression and multiple regression analysis were used.

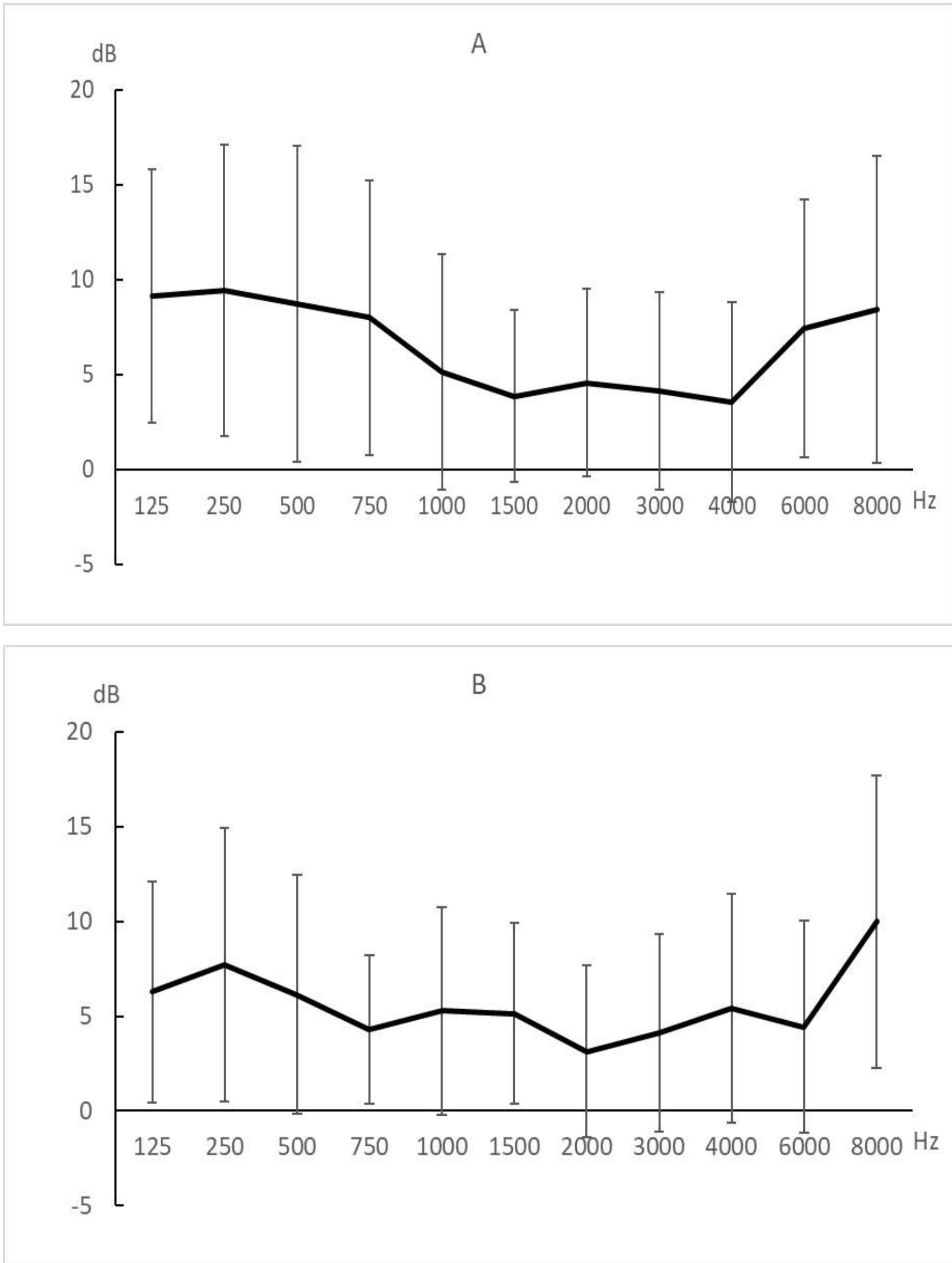
To determine the consistency between auditors when performing tasks of the perceptual experiment (determining the psychoneurological and emotional state of children), Cohen's kappa coefficient (κ) was calculated. Agreement is classified as slight when κ is in the range of 0.00–0.20, fair — 0.21–0.40, moderate — 0.41–0.60, substantial — 0.61–0.80, almost perfect — 0.81–1.00 (Landis, Koch, 1977; Juremi et al., 2017). Cohen's kappa coefficient was calculated using the *cohen_kappa_score* () function of the scikit-learn library (a machine learning library written in the Python programming language).

3. RESULTS

3.1. Physiological and psychophysiological characteristics of auditors

3.1.1. Hearing thresholds

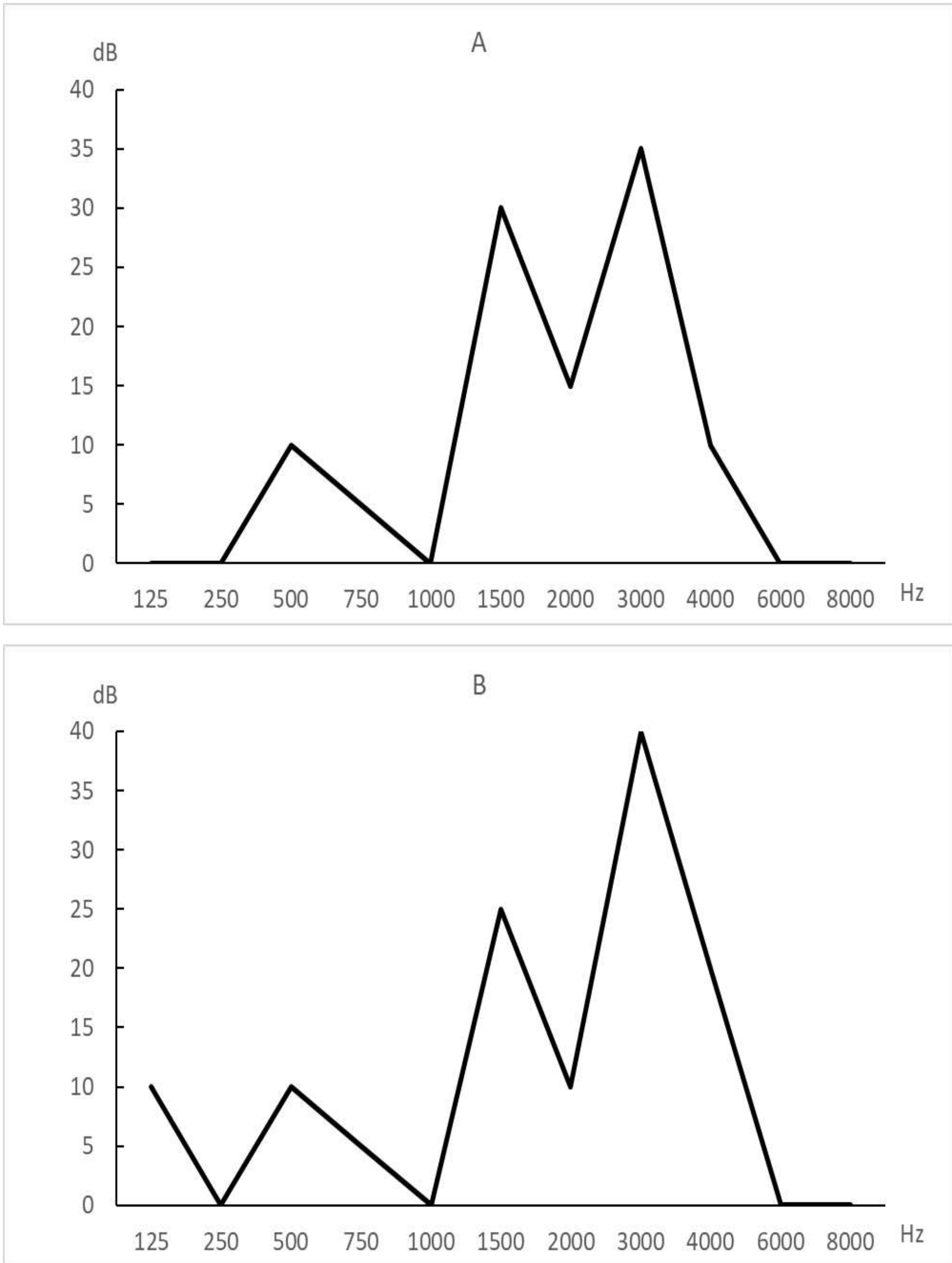
The hearing thresholds of 89% of auditors at all frequencies were within the normal range (0−25 dB): 46% of auditors had hearing thresholds in the range of 0−10 dB (hereinafter referred to as low hearing thresholds), 43% of auditors had hearing thresholds in the range of 15−25 dB (hereinafter referred to as high hearing thresholds) (Fig. 5).



The horizontal axis is frequency, Hz; the vertical axis is intensity level, dB

Fig. 5. Average audiogram of auditors for the left (A) and right (B) ear

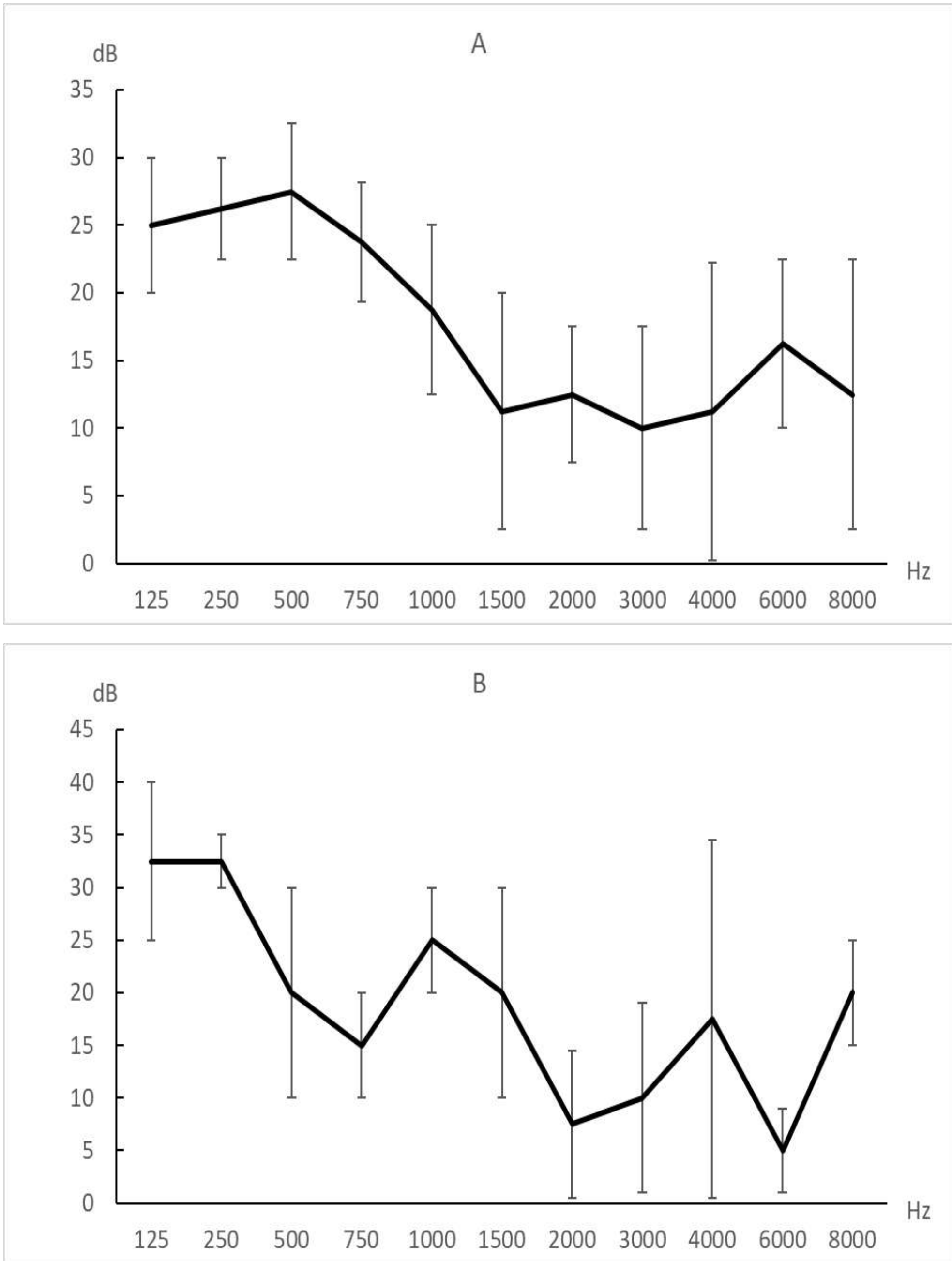
11% of auditors had hearing thresholds at different frequencies above normal. For example, 1 auditor (biology student; woman, 24 y. o.) had hearing thresholds above normal in both ears in the speech range (Fig. 6).



The horizontal axis is frequency, Hz; the vertical axis is intensity level, dB

Fig. 6. Audiogram of an auditor with elevated hearing thresholds in the speech range for the left (A) and right (B) ear

In 4 auditors (IT students: 3 men, 21, 22, 23 y. o.; biology student, woman, 29 y. o.) the hearing thresholds of the left ear were higher than normal at low frequencies and in the speech range (125−3000 Hz), in 2 auditors (IT students, 2 men, 22, 23 y. o.) those for hearing thresholds of the right ear (Fig. 7).



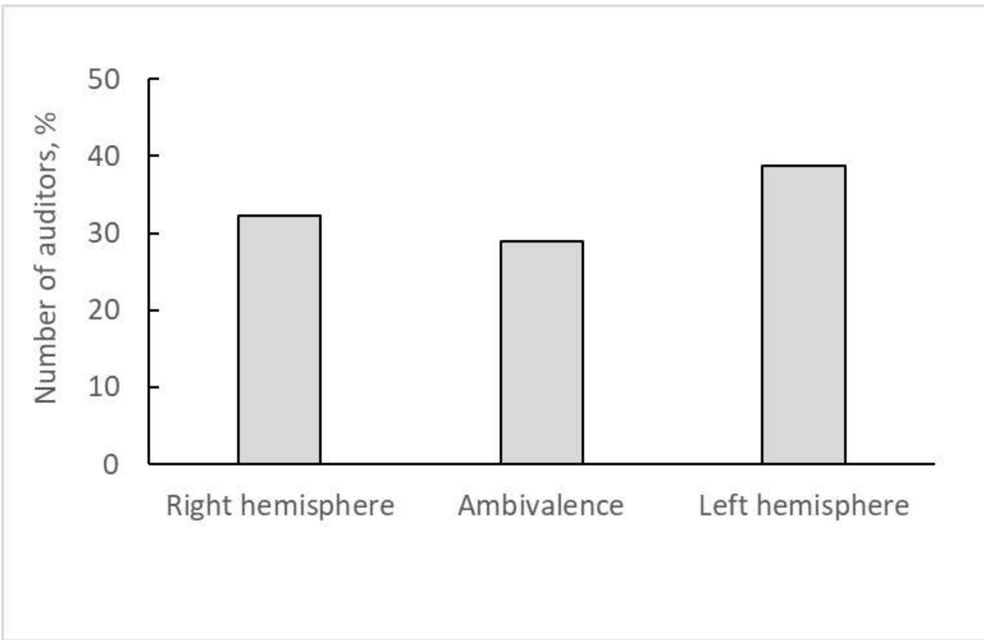
The horizontal axis is frequency, Hz; the vertical axis is intensity level, dB

Fig. 7. Average audiogram of auditors with elevated hearing thresholds at low frequencies and in the speech range for the left (A) and right (B) ear

Hearing thresholds of paediatric students, residents, and psychiatrists were not measured.

3.1.2. Dichotic testing data

According to the results of dichotic testing, 32.3% of auditors have a left-sided preference (predominance of the right hemisphere), 38.7% have a right-sided preference (predominance of the left hemisphere); 29% of auditors do not have a clear preference (ambivalence) (Fig. 8).



The vertical axis is number of auditors, %, the horizontal axis is results of dichotic testing

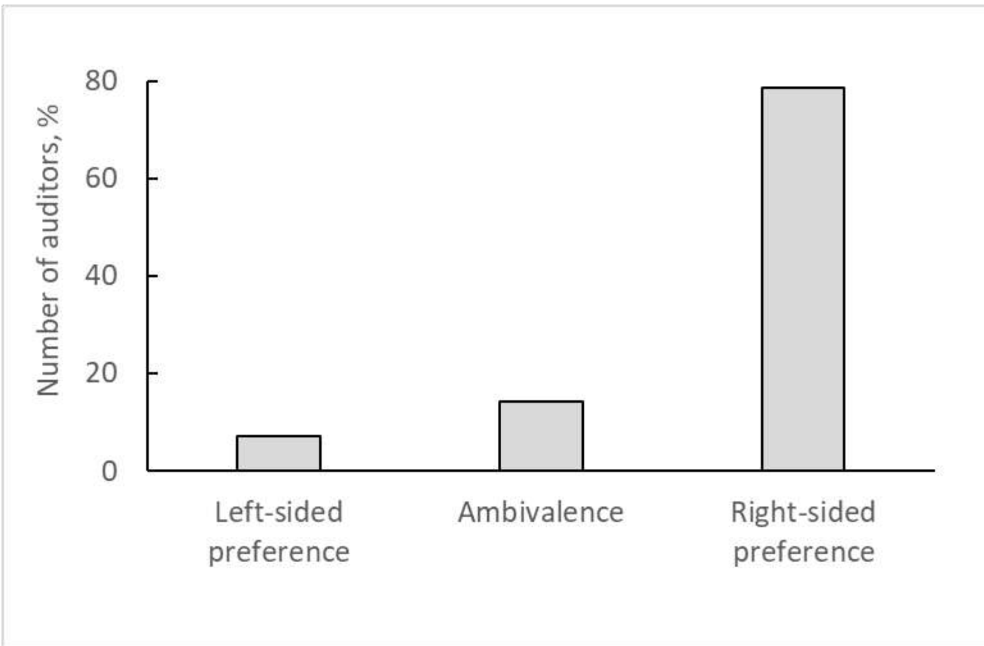
Fig. 8. Determination of the leading hemisphere for speech (dichotic testing) of auditors

3.1.3. Phonemic hearing test

During the phonemic hearing test, no phonemic perception losses were detected among the auditors.

3.1.4. Profile of functional lateral asymmetry

According to the results of PFLA determining, 7.1% of auditors have a left-sided preference (predominance of the right hemisphere), 78.6% have a right-sided preference (predominance of the left hemisphere); 14.3% of auditors have no clear preference (ambivalence) (Fig. 9).



The vertical axis is number of auditors, %, the horizontal axis is results of PFLA determination

Fig. 9. Profile of functional lateral asymmetry of auditors

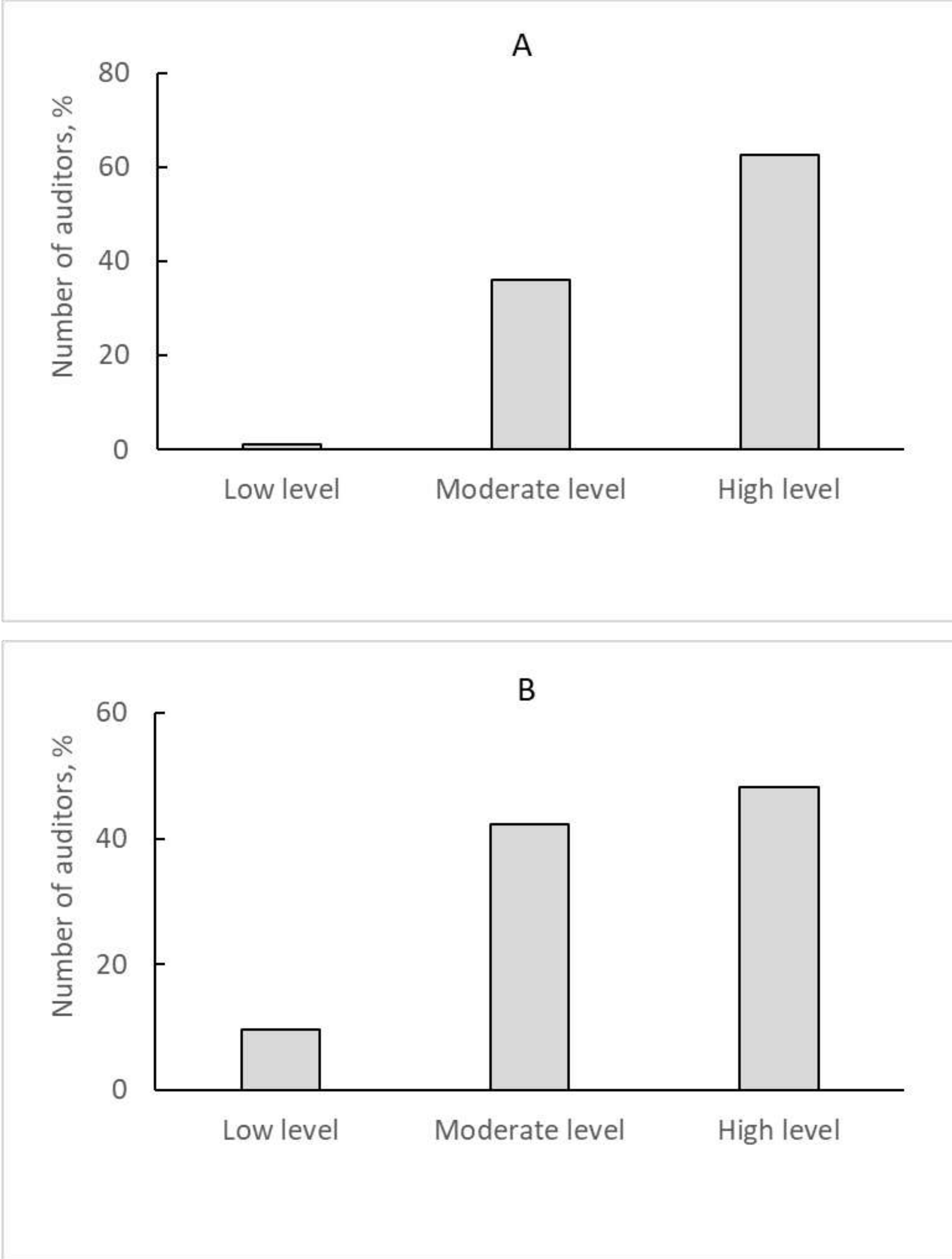
3.1.5. Characteristics of the electrocardiogram and heart rate of auditors

The average heart rate values of the auditors before listening to the test containing emotional speech of TD children were 81±11.5 beats/min; during listening — 82±10.1 beats/min; after listening — 82.9±9.8 beats/min. The average heart rate values of the auditors before listening to the test with emotional speech of children with ASD were 83±9.7 beats/min; during listening — 83.6±9.9 beats/min; after listening — 83.5±10.5 beats/min, i.e. there is a tendency for heart rate values to increase while listening to children's emotional speech, which indicates a slight change in the activation of vegetative nervous system of the listener.

The average values of RR intervals of auditors before listening to a test containing emotional speech of TD children were 0.75±0.11 s; during listening — 0.74±0.09 s; after listening — 0.74±0.09 s. The average values of RR intervals of auditors before listening to the test with emotional speech of children with ASD were 0.74±0.09 s; during listening — 0.75±0.09 s; after listening — 0.74±0.08 s. No significant changes in RR intervals were detected as a result of the experiment.

3.1.6. Anxiety level of auditors

According to the Spielberger test results, 1.2% of auditors have a low level of personal anxiety, 36.1% of auditors have a moderate level, and 62.6% of auditors have a high level of anxiety. 9.6% of auditors have a low level of situational anxiety, 42.2% of auditors have a moderate level of anxiety, 48.2% of auditors have a high level of situational anxiety (Fig. 10).



The vertical axis is number of auditors, %, the horizontal axis is anxiety level

Fig. 10. Level of personal (A) and situational (B) anxiety of auditors

3.2. Data from linguistic analysis of children's speech

Children's responses in dialogues with the experimenter were analyzed. In the speech of children with ASD, there are words, phrases, statements consisting of several phrases, complex sentences (for 12–14 years old children frequency of occurrence is 0.03) and speech-like constructions, the meaning of which is unclear even in the context of the situation (frequency of occurrence is 0.17 in 5−7 years; 0.14 — in 8−9 years; 0.23 — in 10−11 years; 0.22 — in 12−14 years). The predominant type of response in children with ASD aged 5−9 years is one word (frequency of occurrence is 0.39 in the age of 5−7 years; 0.34 — in the age of 8−9 years), in children with ASD aged 10−14 years — simple phrase (frequency of occurrence is 0.24 in 10−11 years; 0.42 — in 12−14 years). TD children use words, simple phrases, statements consisting of several phrases, and complex sentences as answers. The predominant type of response from TD children of all ages is a simple phrase (frequency of occurrence is 0.39 in 5−7 years; 0.43 — in 8−9 years; 0.49 — in 10−11 years; 0.36 — in 12−14 years) (Table 4).

Table 4. Structure of responses of children with ASD and TD children in dialogues with adults

Age, y.o,	Group	Type of phrase						
		word	simple phrase	2 simple phrases	several simple phrases	complex sentence	yes/no answer	speech-like construction
5−7	ASD	0.39	0.26	0.02	0	0	0.17	0.17
	TD	0.17	0.39	0.13	0.07	0.12	0.13	0
8−9	ASD	0.34	0.21	0.05	0.04	0	0.21	0.14
	TD	0.11	0.43	0.03	0.03	0.01	0.38	0
10−11	ASD	0.23	0.24	0.07	0.03	0	0.21	0.23
	TD	0.08	0.49	0.14	0.12	0.09	0.07	0
12−14	ASD	0.26	0.42	0.08	0	0.03	0.02	0.22
	TD	0.08	0.36	0.16	0.11	0.13	0.16	0

The diversity index (***1−D***) of response types in children with ASD in all age groups is lower than in TD children. The average value of the diversity index in children with ASD is 0.686; in TD children — 0.765 (Table 5).

Table 5. Diversity index of response types in children with ASD and TD children

Age, y. o.	Group	
	ASD	TD
5–7	0.655	0.779
8–9	0.721	0.732
10–11	0.746	0.758
12–14	0.624	0.792

An analysis of the frequency of occurrence of different parts of speech showed that in the responses of children with ASD aged 5−11 years, nouns (frequency of occurrence is 0.23 in 5−7 years; 0.24 — in 8−9 years; 0.18 — in 10−11 years old) and auxiliary parts of speech (prepositions, conjunctions, particles, etc.) (frequency of occurrence is 0.21 in 5−7 years old; 0.29 — in 8−9 years old; 0.36 — in 10−11 years old) are most often found. In the phrases of 12−14 years old children, nouns (0.3) and verbs (0.22) are most often found. In the responses of children with ASD there are incomprehensible speech-like constructions: frequency of occurrence is 0.22 in 5–7 years; 0.26 — in 8−9 years; 0.08 — in 10−11 years; 0.09 — in 12−14 years.

In the responses of TD children of all ages, nouns (frequency of occurrence is 0.29 in 5−7 years old; 0.26 — in 8−9 years old; 0.3 — in 10−11 years old; 0.18 in — 12−14 years old) and auxiliary parts of speech (frequency of occurrence is 0.22 in 5−7 years; 0.34 — in 8−9 years; 0.17 — in 10−11 years; 0.35 — in 12−14 years) predominate (Table 6).

Table 6. Frequency of occurrence of different parts of speech in children's responses in dialogues with adults

Age, y. o.	Group	Part of speech						
		noun	verb	adjective	adverb	pronoun	auxiliary parts of speech	speech-like construction
5−7	ASD	0.23	0.13	0.05	0.02	0.14	0.21	0.22
	TD	0.29	0.18	0.06	0.1	0.11	0.22	0
8−9	ASD	0.24	0.08	0.04	0.01	0.07	0.29	0.26
	TD	0.26	0.14	0.04	0.07	0.15	0.34	0
10−11	ASD	0.18	0.16	0.03	0.04	0.16	0.36	0.08
	TD	0.3	0.16	0.08	0.13	0.15	0.17	0
12−14	ASD	0.3	0.22	0.12	0.07	0.07	0.13	0.09
	TD	0.18	0.13	0.05	0.12	0.17	0.35	0

The diversity index of parts of speech in responses of children with ASD in all age groups is lower than in TD children. The average value of the diversity index in children with ASD is 0.764; in TD children — 0.795 (Table 7).

Table 7. Diversity index of parts of speech in responses of children with ASD and TD children

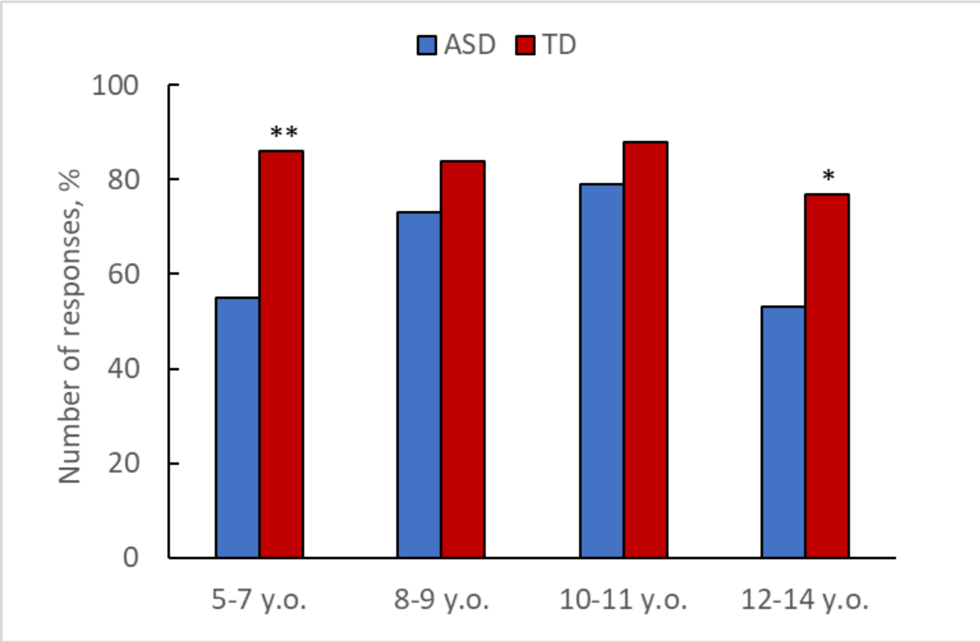
Age, y. o.	Group	
	ASD	TD
5–7	0.786	0.802
8–9	0.719	0.776
10–11	0.759	0.814
12–14	0.791	0.792

Based on correlation analysis (Spearman; p <0.05), a connection was shown between scores on the CARS scale in children with ASD and the frequency of occurrence of utterances presented in one word (r = 0.65) in children with higher scores and, accordingly, with more severe autism spectrum disorder, the frequency of occurrence of one-word utterances is higher compared to children with lower scores.

3.3. Recognition of the lexical meaning of children's words by adults

3.3.1. Data from a perceptual experiment

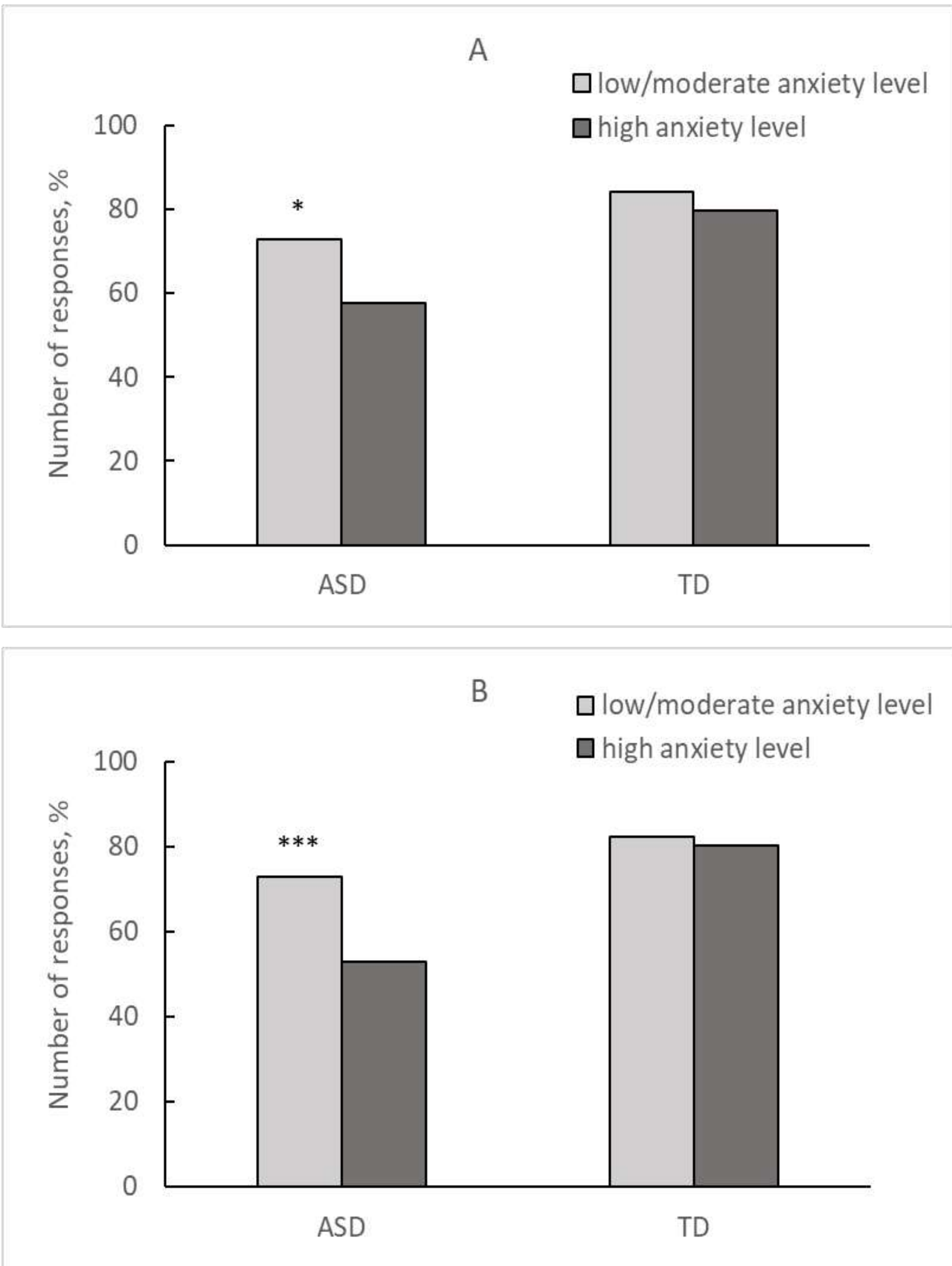
The perceptual analysis shown that adults recognize correctly the meaning of 55% of words of children with ASD aged 5–7 years, 73% of words of children aged 8–9 years, 79% of words of children aged 10–11 years and 53% of words of children at the age of 12–14 years. Adults recognize correctly the meaning of 86% of words of TD children aged 5–7 years, 84% of words of children aged 8–9 years, 88% of words of children aged 10–11 years, and 77% of words of children aged 12–14 years (Fig. 11). Auditors recognize the meaning of words of TD children aged 5–7 years and 12–14 years of age better (p <0.01; p <0.05, respectively; Mann — Whitney U-test) than the meaning of words of children with ASD of the same age.



The vertical axis is the number of correct answers from auditors, %; the horizontal axis is the age of the children, years; * — p <0.05; ** — p <0.01; Mann–Whitney test

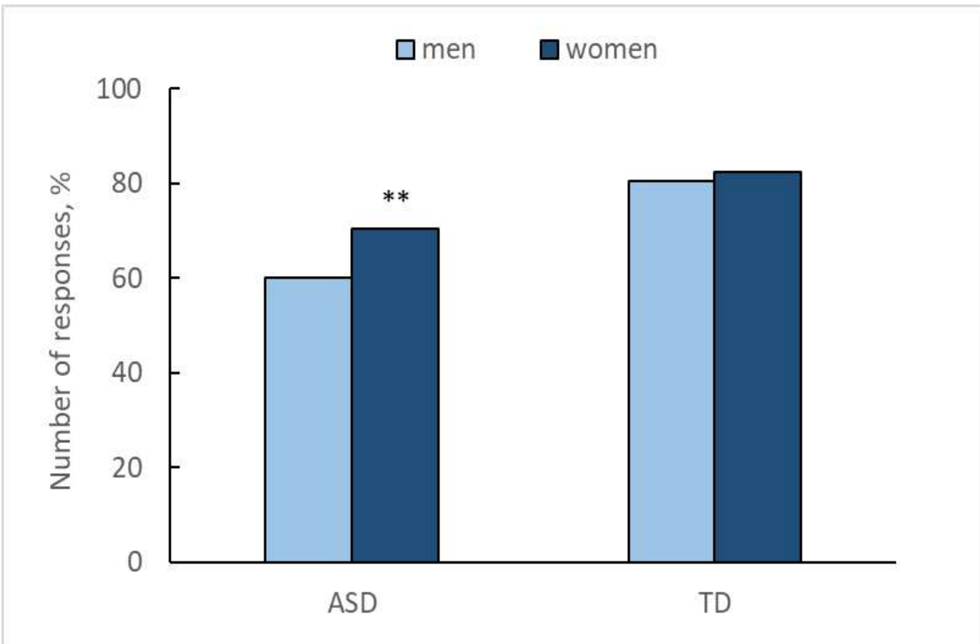
Fig. 11. Recognition of the lexical meaning of words of children with ASD and TD by adults

Female auditors recognize the meaning of words of children with ASD better than male auditors (p <0.01); no significant differences were found in recognizing the meaning of words of TD children (Fig. 13). No differences were found in recognizing the meaning of children's words between auditors who have everyday experience interacting with children and auditors who do not. Auditors with high level of personal and situational anxiety are worse at recognizing the meaning of words of children with ASD compared to auditors with moderate and low level of anxiety (p <0.05; p <0.001, respectively) (Fig. 12).



The vertical axis is the number of correct answers from auditors, %; the horizontal axis is groups of the children, years; * — $p < 0.05$; *** — $p < 0.001$; Mann–Whitney test

Fig. 12. Recognition of the meaning of children's words by auditors with different levels of personal (A) and situational (B) anxiety



The vertical axis is the number of correct answers from auditors, %; the horizontal axis is groups of the children, years; ** — p <0.01; Mann–Whitney test

Fig. 13. Recognition of the meaning of children's words by male and female auditors

Based on correlation analysis (Spearman; p <0.05), the relationship is shown:

1) between the auditor's hearing thresholds (audiometry — right ear) and recognition of the meaning of words of TD children (0.21). Auditors with low hearing thresholds are better at recognizing the meanings of TD children's words compared to auditors with high hearing thresholds;

2) between the gender of the auditor (female) and recognition of the meaning of words of children with ASD (0.29). Female auditors were better at recognizing the meaning of words of children with ASD than male auditors. The data of the correlation analysis are confirmed by the data of multiple regression analysis: $F(3,19) = 3.812$ $p < 0.028$ ($R^2 = 0.389$ $\beta = 0.502$);

3) between the age of the auditor and recognition of the meaning of words of children with ASD (0.28). Older auditors are better at recognizing the meaning of words from children with ASD than younger auditors.

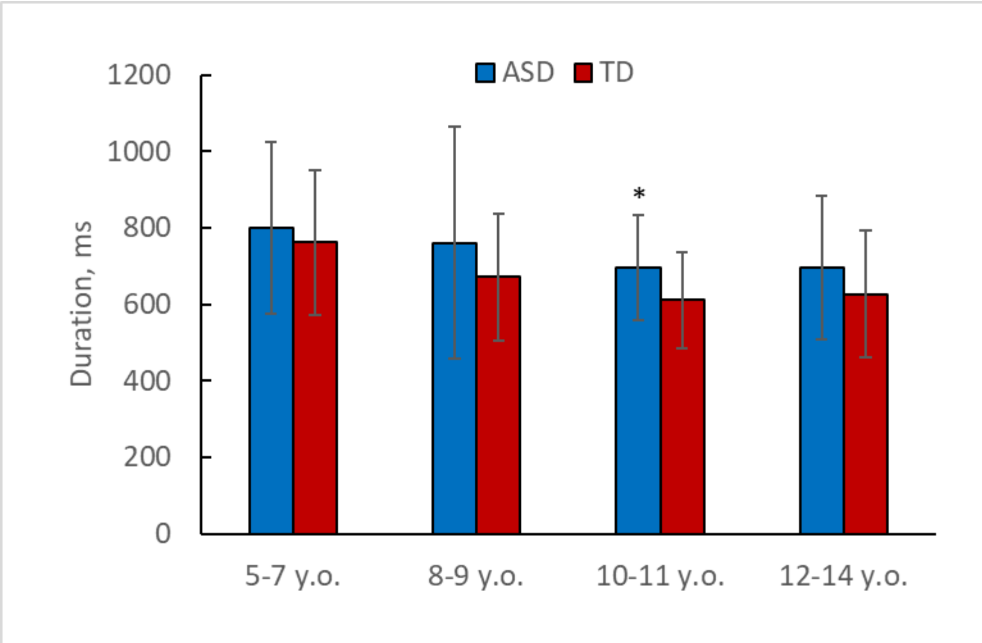
3.3.2. Acoustic characteristics of test material

3.3.2.1. Duration of words, stressed and unstressed vowels

There were no significant differences in word duration in children with ASD aged 5–9 and 12–14 years compared to TD children. In the age of 10−11 years, the duration of words in children with ASD is higher (p <0.05; Mann–Whitney test) than in TD children (Table 8, Fig. 14).

Table 8. Duration of words of children with ASD and TD children, ms

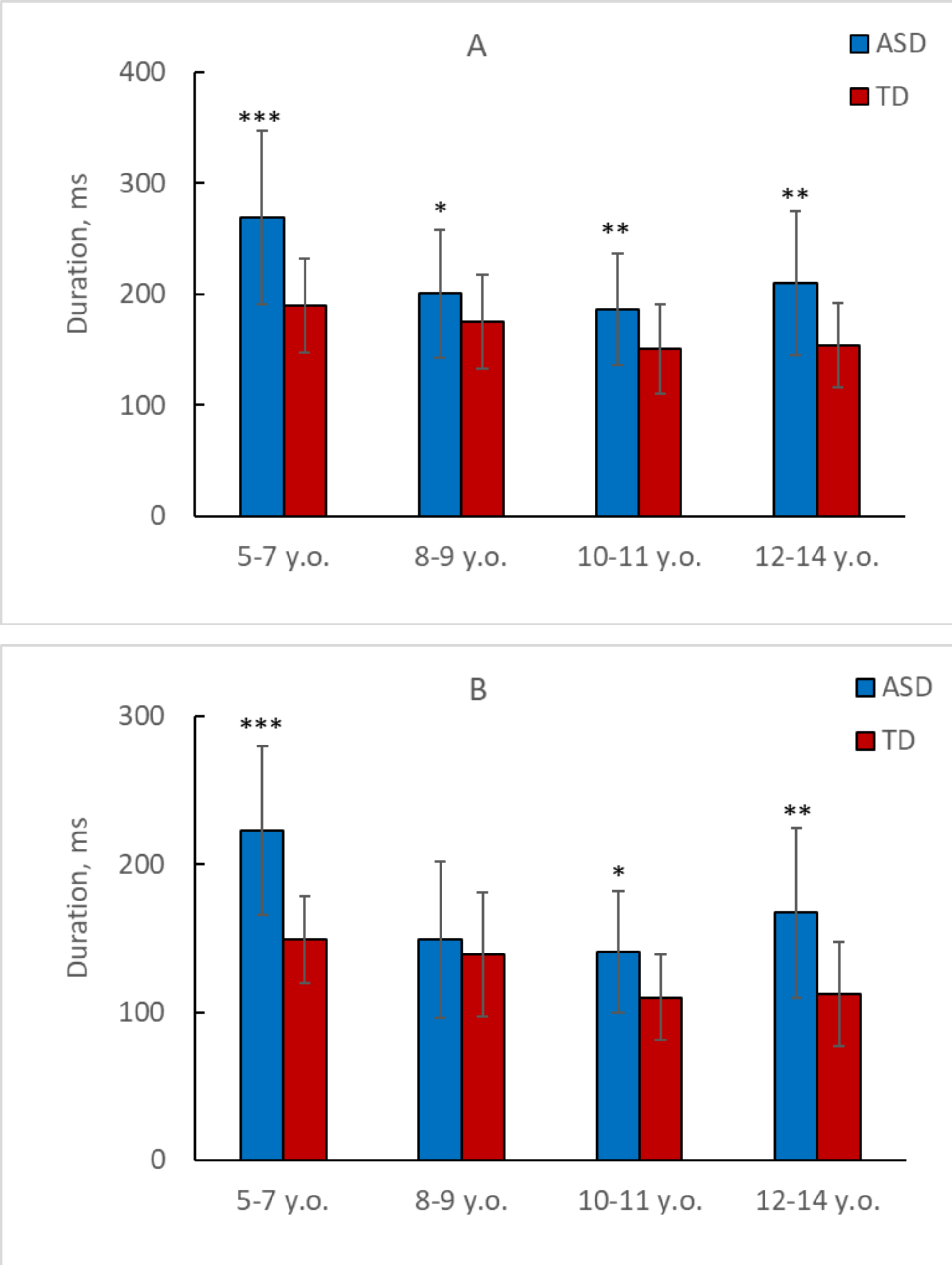
Age, y. o.	Group	
	ASD	TD
5−7	801.3±224.3 (median value 810.5)	761.9±190.5 (768)
8−9	761.1±302.9 (586)	671±167.1 (634)
10−11	696.2±138.9 (667)	611±125.9 (617)
12−14	696.5±188.3 (661.5)	627±165.7 (585)



The vertical axis is duration, ms; the horizontal axis is the age of the children, years; * — p <0.05; Mann–Whitney test

Fig. 14. Duration of words of children with ASD and TD children

The duration of stressed vowels in all age groups in children with ASD is higher (p <0.05 in 8–9 years; p <0.01 in 10–11, 12–14 years; p <0.001 in 5–7 years; Mann–Whitney test) than in TD children (Table 9, Fig. 15A). In the ages of 5–7 and 10–14 years, the duration of unstressed vowels in children with ASD is higher (p <0.05 in 10–11 years; p <0.01 in 12–14 years; p <0.001 in 5–7 years) than in TD children (Table 9, Fig. 15B).



The vertical axis is duration, ms; the horizontal axis is the age of the children, years; * — p <0.05; ** — p <0.01; *** — p <0.001; Mann–Whitney test

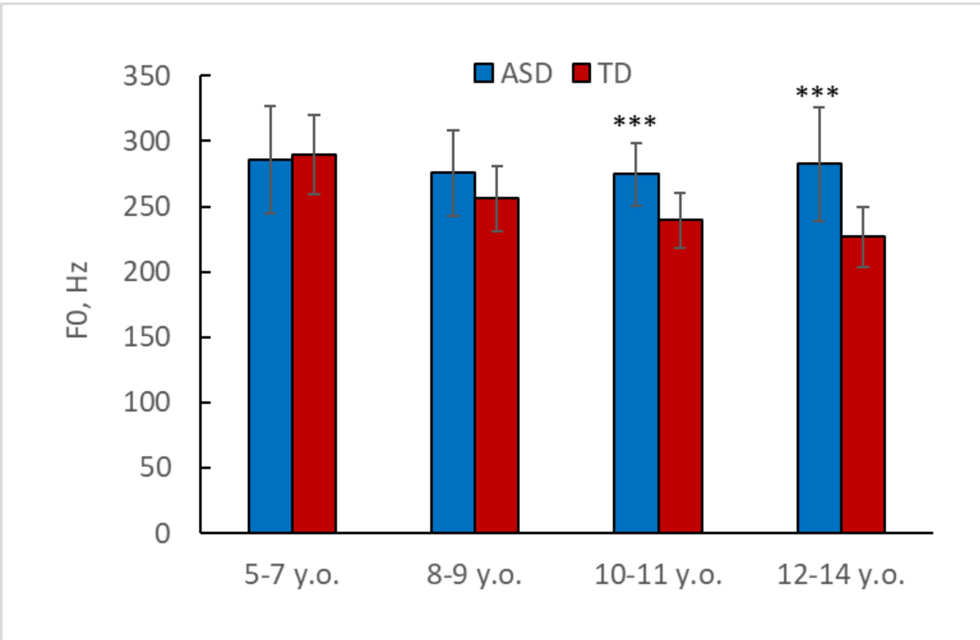
Fig. 15. Duration of stressed (A) and unstressed (B) vowels from words of children with ASD and TD children

Table 9. Duration of stressed and unstressed vowels from words of children with ASD and TD children, ms

Age, y. o.	Group			
	ASD		TD	
	stressed	unstressed	stressed	unstressed
5−7	268.8±78.6 (median value 247.5)	222.9±56.8 (200)	190±42.4 (187.5)	148.9±29.4 (144)
8−9	200.3±57.5 (191)	149.3±52.7 (125)	175.8±42.5 (158.5)	139.1±41.7 (129)
10−11	186.5±50.3 (179.5)	140.7±40.8 (137)	150.6±40.2 (148)	110.1±29 (106)
12−14	209.5±64.7 (194.5)	167.2±57.1 (159)	154.1±37.7 (148)	112.3±352 (98)

3.3.2.2. Fundamental frequency value

F0 values of words in children with ASD aged 5–9 years do not differ significantly from F0 values in TD children. In the age of 10−14 years, F0 value in children with ASD is higher (p <0.001) than in TD children (Table 10, Fig. 16).



The vertical axis is F0 value, Hz; the horizontal axis is the age of the children, years; *** — p <0.001; Mann–Whitney test

Fig. 16. Fundamental frequency values of words of children with ASD and TD children

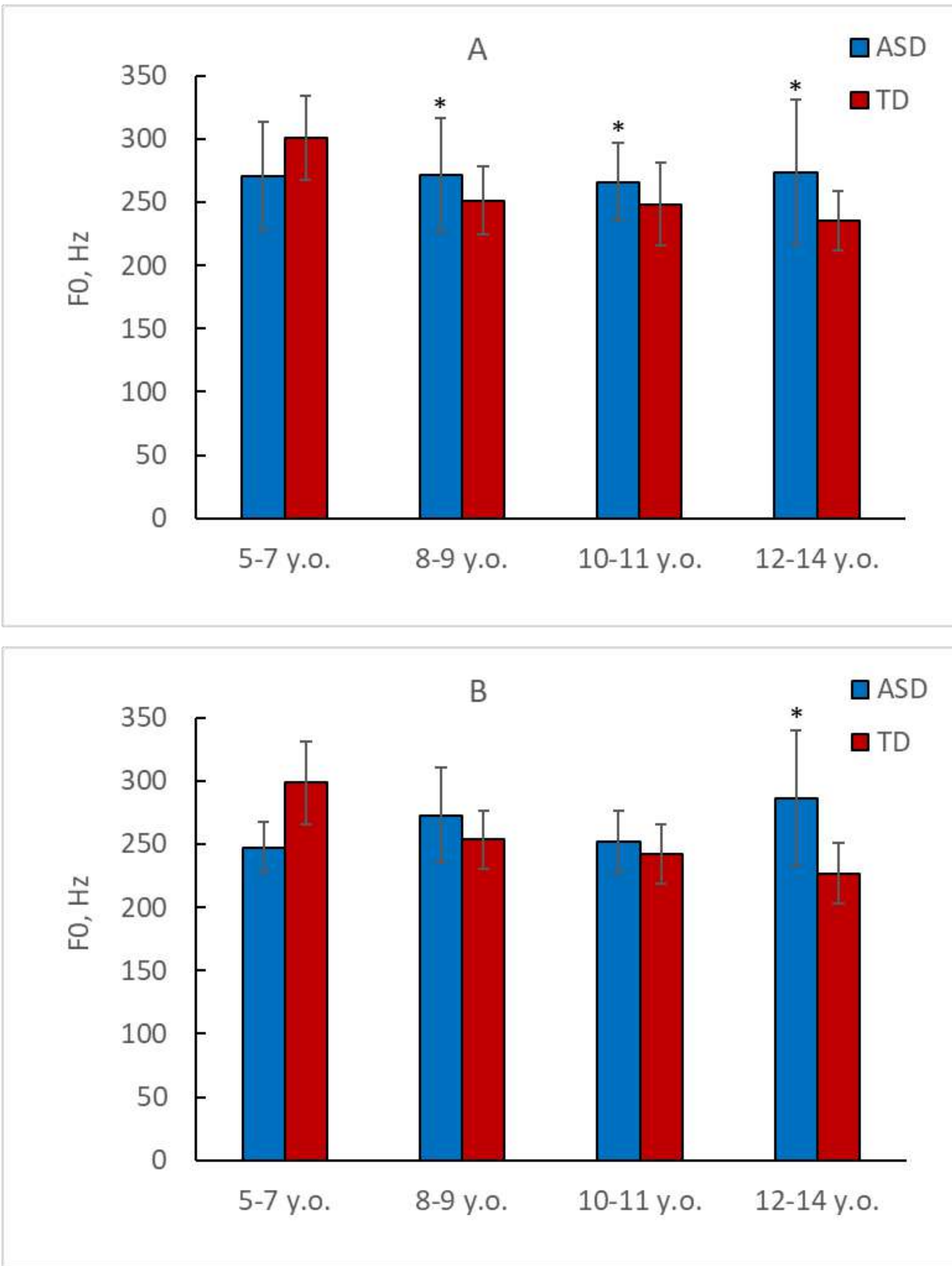
Table 10. Fundamental frequency values of words of children with ASD and TD children, Hz

Age, y. o.	Group	
	ASD	TD
5−7	285.8±40.1 (median value 281.2)	289.9±30.2 (291.2)
8−9	275.7±33 (258.3)	256.2±24.8 (234.3)
10−11	274.6±24.4 (281.2)	239.5±20.8 (234.3)
12−14	282.3±43.1 (281.2)	226.8±22.8 (234.3)

The fundamental frequency values of stressed vowels of children with ASD aged 5–7 years do not differ from the F0 values of TD children. In the age of 8−14 years, F0 values of stressed vowels in children with ASD are higher ($p < 0.05$) than in TD children (Table 11, Fig. 17A). There were no significant differences in F0 values of unstressed vowels in children with ASD and TD children aged 5–11 years. In the age of 12−14 years, F0 values of unstressed vowels in children with ASD are higher ($p < 0.05$) than in TD children (Table 11, Fig. 17B).

Table 11. Fundamental frequency values of stressed and unstressed vowels from words of children with ASD and TD children, Hz

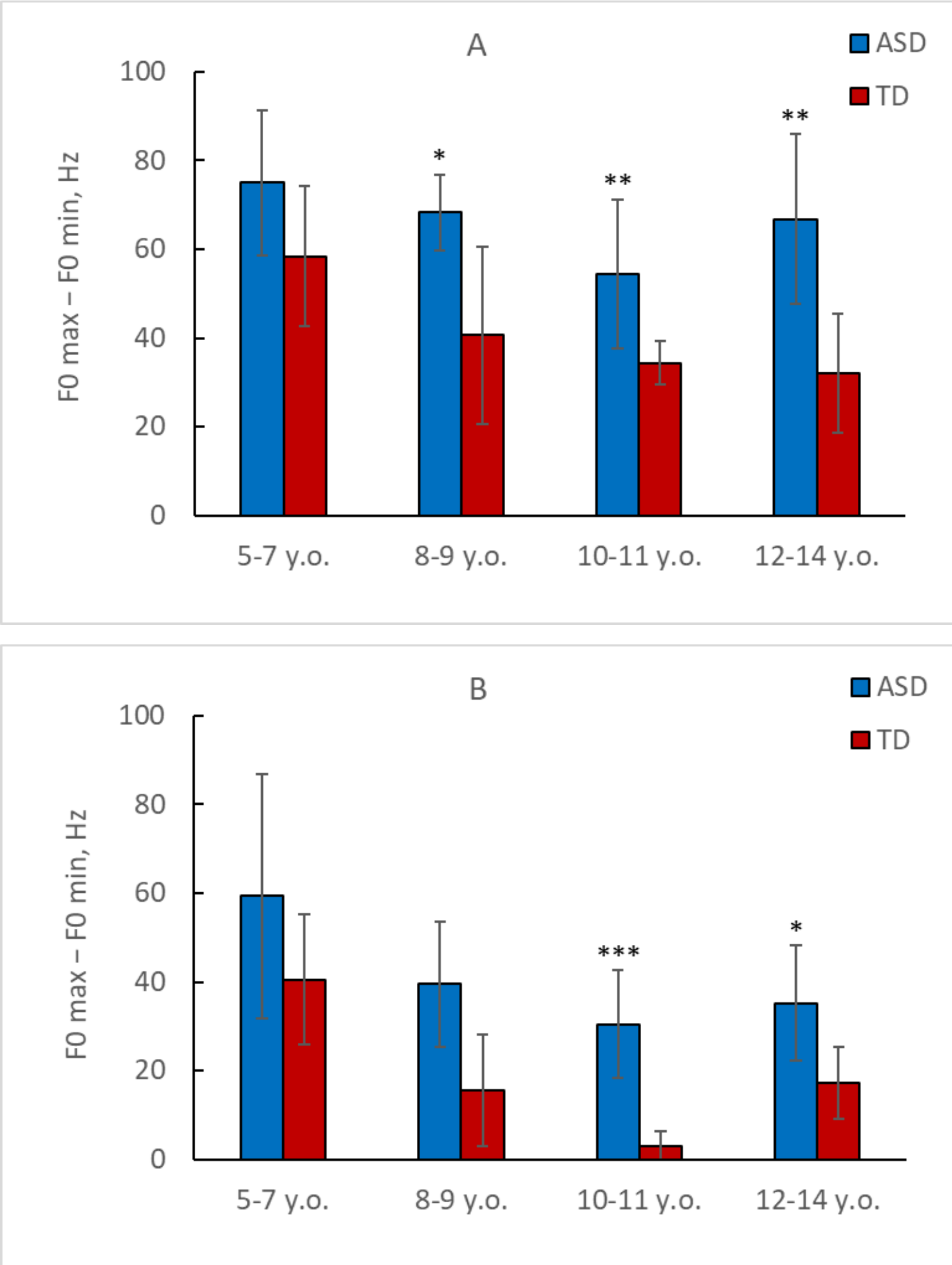
Age, y. o.	Group			
	ASD		TD	
	stressed	unstressed	stressed	unstressed
5−7	270.3±42.9 (median value is 281.2)	247.3±20.6 (234.2)	300.8±33.4 (301.4)	298.6±33.1 (281.2)
8−9	271.5±44.9 (258.3)	273.8±37.5 (269.8)	251.5±26.9 (234.3)	253.5±22.7 (234.3)
10−11	266+30.9 (281.2)	252.1+24.6 (258.3)	248.3±32.8 (234.3)	241.9±23.4 (234.3)
12−14	273.9±57 (269.8)	286.6±53.1 (269.8)	235.1±23.7 (234.3)	226.7±24.1 (234.3)



The vertical axis is F0 value, Hz; the horizontal axis is the age of the children, years; * — p <0.05; Mann–Whitney test

Fig. 17. Fundamental frequency values of stressed (A) and unstressed (B) vowels from words of children with ASD and TD children

In children with ASD aged 5−7 years, the variability of fundamental frequency values in words does not have significant differences compared to TD children. In the ages of 8−9, 10−11 and 12−14 years, the variability of F0 of words in children with ASD is higher than in TD children (p <0.05; p <0.01; p <0.01, respectively) (Table 12, Fig. 18A). The variability of F0 of stressed vowels does not differ significantly in 5–9 years of age in children with ASD and TD children. In the ages of 10−11 and 12−14 years, the variability of F0 of stressed vowels in children with ASD is higher than in TD children (p <0.001; p <0.05, respectively) (Table 12, Fig. 18B)



The vertical axis is F0 variability value, Hz; the horizontal axis is the age of the children, years; * — p <0.05; ** — p <0.01; *** — p <0.001; Mann–Whitney test

Fig. 18. Fundamental frequency variability of words (A) and stressed vowels (B) from the words of children with ASD and TD children

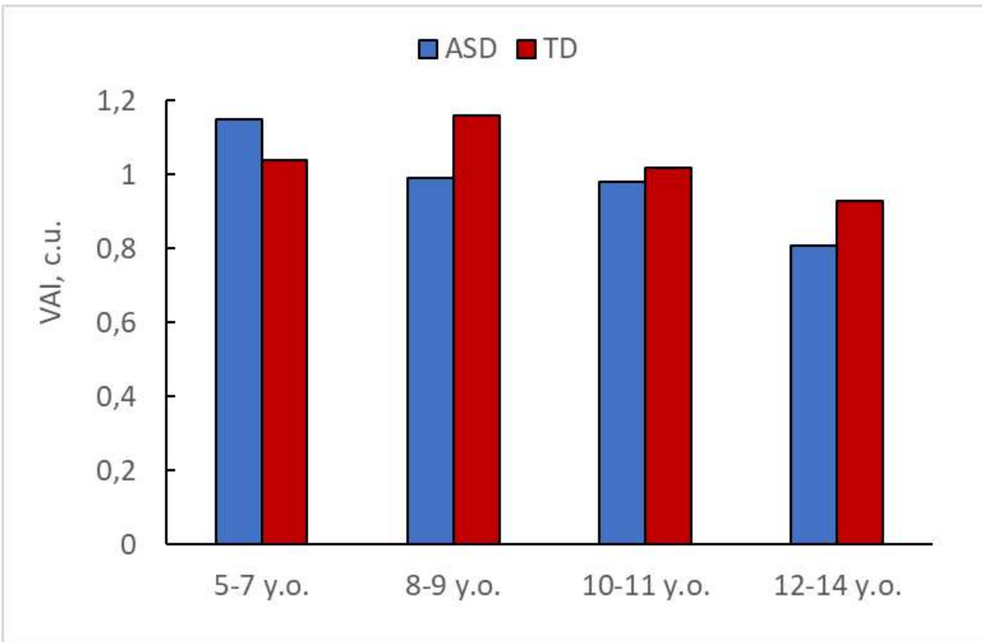
Table 12. Fundamental frequency variability of words and stressed vowels from the words of children with ASD and TD children, Hz

Age, y. o.	Group			
	ASD		TD	
	word	stressed vowel	word	stressed vowel
5−7	75±16.3 (median value 78.1)	59.4±27.5 (46.9)	58.4±15.8 (54.4)	40.5±14.7 (46.9)
8−9	68.3±8.5 (62.5)	39.5±14.1 (31.2)	40.6±19.9 (46.8)	15.6±12.5 (15.6)
10−11	54.4±16.9 (46.9)	30.5±12.2 (31.3)	34.4±4.9 (32.3)	3.1±4.9 (0)
12−14	66.8±19.2 (58.7)	35.2±13 (30)	32±13.4 (31.3)	17.2±8.1 (15.6)

3.3.2.3. Vowel articulation index values

The average values of the articulation index (VAI) of stressed vowels in the age of 5−7 years are higher in children with ASD compared to TD children. In the age of 8−14 years, VAI values are higher in TD children.

VAI of stressed vowels in 5−7 years of age is 1.15 in children with ASD, 1.04 in TD children; in 8−9 years old: 0.99 in children with ASD, 1.16 in TD children; in 10−11 years old: 0.98 in children with ASD, 1.02 in TD children; in 12−14 years old: 0.81 in children with ASD; 0.93 in TD children (Fig. 19).



The vertical axis is vowel articulation index values, c.u.; the horizontal axis is the age of the children, years

Fig. 19. Vowel articulation index values of stressed vowels of children with ASD and TD children

Thus, the words of children with ASD are characterized by larger values of the duration of stressed and unstressed vowels, larger values of F0 of stressed vowels, and a wider variability of F0 compared to the words of TD children. In the age of 8−14 years, the values of the vowel articulation index in children with ASD are lower than in TD children, however, in the age of 5−7 years, the vowel articulation index in children with ASD are higher compared to TD children.

3.3.2.4. Connections between auditors' recognition of the meaning of children's words and the acoustic characteristics of test material

Based on correlation analysis (Spearman, $p < 0.05$), connections are shown between auditors' recognition of the lexical meaning of words of children with ASD and TD children and the F0 values of words ($r = -0.18$), stressed vowels (-0.21), F0 variability of words (-0.21). The relationship between the recognition of the lexical meaning of words by children with ASD and TD children and the duration of the stressed vowel (-0.18) was shown — the data of the correlation analysis were confirmed by the data of multiple regression analysis: $F (5,144) = 3.935$ $p < 0.002$ ($R^2 = 0.089$ $\beta = -0.234$).

Auditors are worse at recognizing the meaning of words with high F0 values of word, stressed vowel, high values F0 variability of words, and high values of duration of a stressed vowel compared to words with lower values.

Based on the data of multiple regression analysis, a relationship is shown between auditors' recognition of the meaning of children's words and the maximum F0 values of stressed vowels: $F (5,144) = 3.935$ $p < 0.002$ ($R2 = 0.089$ $\beta = -0.779$) — auditors recognize worse the meaning of words with high maximum F0 values of stressed vowels.

3.4. Recognition of the psychoneurological state of children by adults

3.4.1. Data from a perceptual experiment

When performing a task to determine the psychoneurological state of children, auditors listened to tests containing children's words and tests containing children's phrases.

When listening to tests containing words of children aged 5–7 years, auditors classified 54% of the signals of children with ASD into the "atypical development" category; auditors classified 78% of the signals of TD children into the "typical development" category. The average recall of recognition (UAR) was 0.66 (Table 13).

Table 13. Confusion matrix for recognition of the psychoneurological state of children 5−7 years old (words), % of correct answers by auditors

Group	ASD	TD
ASD	**54**	46
TD	22	**78**
Recall	0.54	0,78
Precision	0.71	0,63
F_1-score	0.61	0,7
UAR	0.66	

In tests containing words of children aged 8−9 years, auditors classified 37% of signals of children with ASD as “atypical development”; auditors classified 87% of signals of TD children as “typical development”. The average recall of recognition was 0.62 (Table 14).

Table 14. Confusion matrix for recognition of the psychoneurological state of children 8−9 years old (words), % of correct answers by auditors

Group	ASD	TD
ASD	**37**	63
TD	13	**87**
Recall	0.37	0,87
Precision	0.74	0,58
F_1-score	0.49	0,7
UAR	0.62	

In tests containing words of children aged 10−11 years, auditors classified 40% of signals of children with ASD as “atypical development”; auditors classified 86% of signals of TD children as “typical development”. The average recall of recognition was 0.63 (Table 15).

Table 15. Confusion matrix for recognition of the psychoneurological state of children 10−11 years old (words), % of correct answers by auditors

Group	ASD	TD
ASD	**40**	60
TD	14	**86**
Recall	0.4	0,86
Precision	0.74	0,59
F_1-score	0.51	0,7
UAR	0.63	

In tests containing words of children aged 12−14 years, auditors classified 56% of signals of children with ASD as “atypical development”; auditors classified 73% of signals of TD children as “typical development”. The average recall of recognition was 0.65 (Table 16).

Table 16. Confusion matrix for recognition of the psychoneurological state of children 12−14 years old (words), % of correct answers by auditors

Group	ASD	TD
ASD	**56**	44
TD	27	**73**
Recall	0.56	0,73
Precision	0.67	0,62
F_1-score	0.61	0,67
UAR	0.65	

When listening to tests containing phrases of children aged 5−7 years, auditors classified 73% of the signals of children with ASD as “atypical development”; auditors classified 90% of the signals of TD children as “typical development”. The average recall of recognition was 0.82 (Table 17).

Table 17. Confusion matrix for recognition of the psychoneurological state of children 5−7 years old (phrases), % of correct answers by auditors

Group	ASD	TD
ASD	**73**	27
TD	10	**90**
Recall	0.73	0,9
Precision	0.88	0,77
F_1-score	0.8	0,83
UAR	0.82	

In tests containing phrases of children aged 8–9 years, auditors classified 52% of signals of children with ASD as “atypical development”; auditors classified 91% of signals of TD children as “typical development”. The average recall of recognition was 0.72 (Table 18).

Table 18. Confusion matrix for recognition of the psychoneurological state of children 8−9 years old (phrases), % of correct answers by auditors

Group	ASD	TD
ASD	**52**	48
TD	9	**91**
Recall	0.52	0,91
Precision	0.85	0,65
F_1-score	0.65	0,76
UAR	0.72	

In tests containing phrases of children aged 10–11 years, auditors classified 63% of signals of children with ASD as “atypical development”; auditors classified 97% of signals of TD children as “typical development”. The average recall of recognition was 0.8 (Table 19).

Table 19. Confusion matrix for recognition of the psychoneurological state of children 10−11 years old (phrases), % of correct answers by auditors

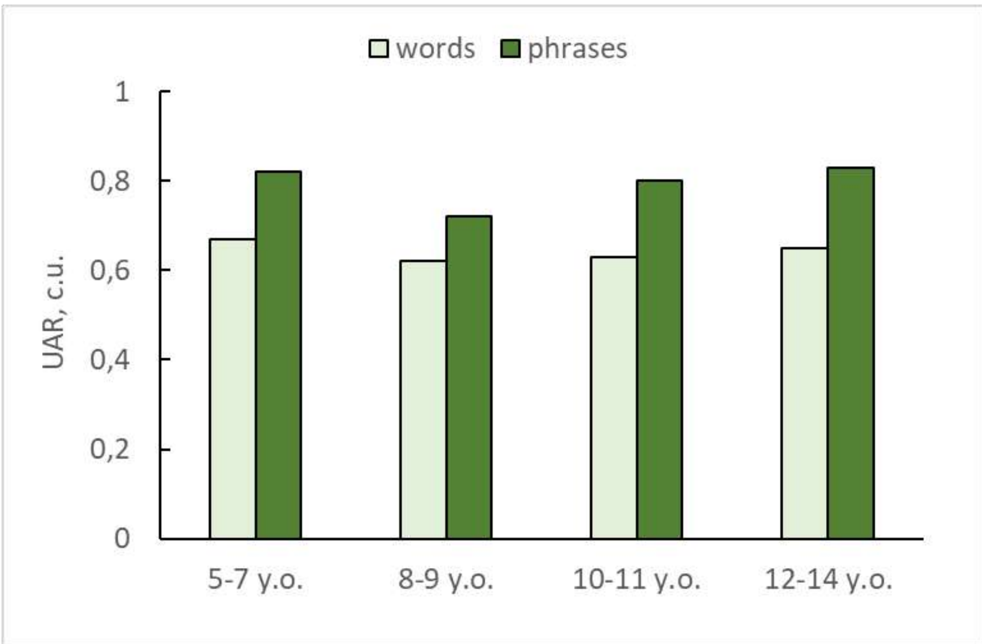
Group	ASD	TD
ASD	**63**	27
TD	3	**97**
Recall	0.63	0,97
Precision	0.95	0,78
F_1-score	0.76	0,87
UAR	0.8	

In tests containing phrases of children aged 12–14 years, auditors classified 73% of signals of children with ASD as “atypical development”; auditors classified 93% of signals of TD children as “typical development”. The average recall of recognition was 0.83 (Table 20).

Table 20. Confusion matrix for recognition of the psychoneurological state of children 12−14 years old (phrases), % of correct answers by auditors

Group	ASD	TD
ASD	**73**	27
TD	7	**93**
Recall	0.73	0,93
Precision	0.91	0,78
F_1-score	0.81	0,85
UAR	0.83	

Auditors are better at recognizing the psychoneurological state of children in tests containing phrases than in tests containing children's words: the average recall for tests containing children's words is 0.64; the average recall for tests containing children's phrases was 0.79. Average recall values were maximum in tests where the state of children aged 5−7 (0.67 — words; 0.82 — phrases) and 12−14 years (0.65 — words; 0.83 — phrases) was determined. The minimum values of average recall were in tests where the state of children 8−9 years old was determined: 0.62 — words; 0.72 — phrases (Fig. 20).



The vertical axis is unweighted average recall (UAR) values, c. u.; the horizontal axis is the age of the children, years

Fig. 20. Average recall of recognition of the psychoneurological state of children by auditors

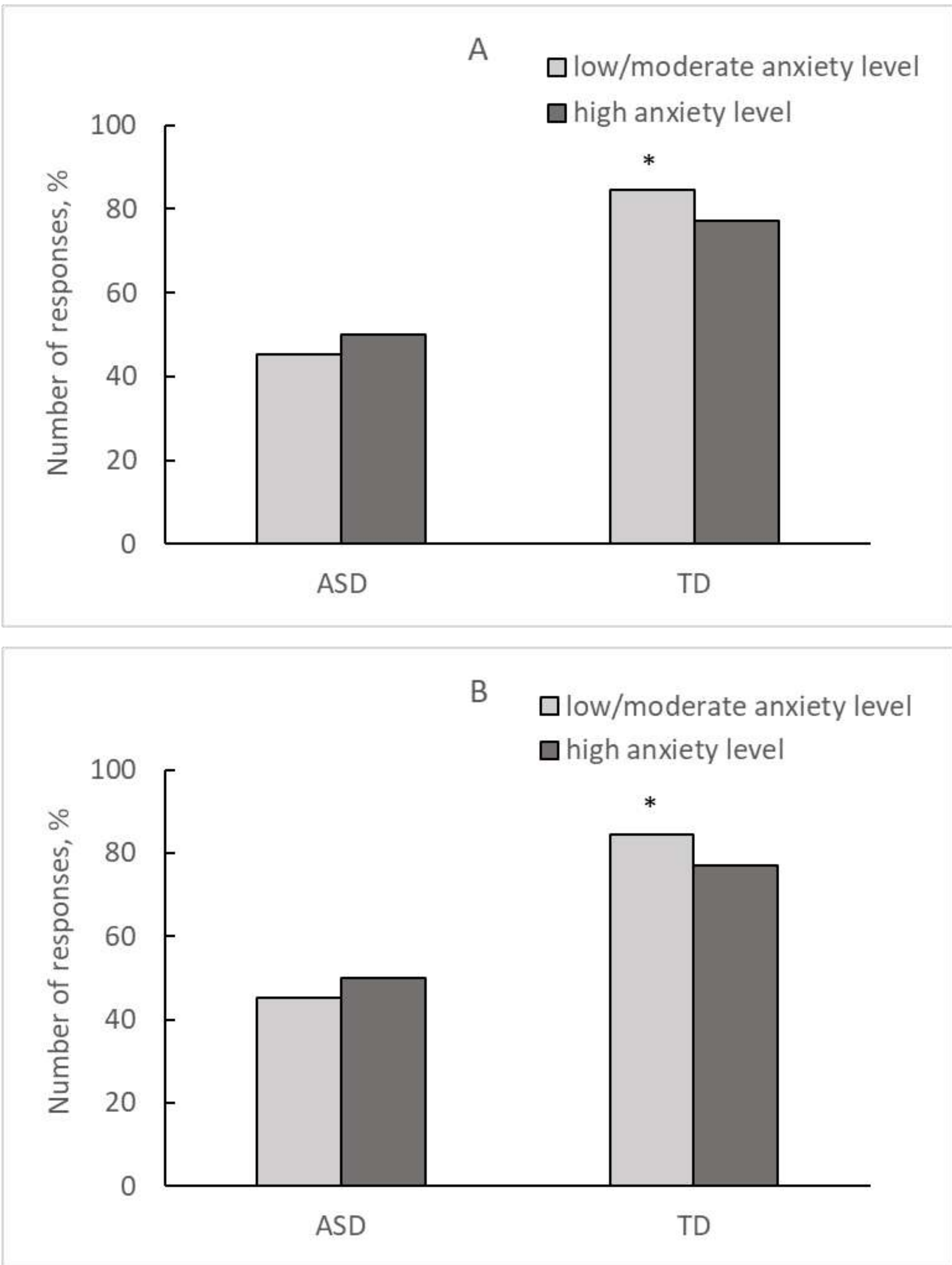
When recognizing the psychoneurological state in tests containing words, auditors demonstrated fair agreement (average κ-value is 0.358). In tests containing children's phrases, auditors' agreement was moderate (average κ-value is 0.511) (Table 21).

Table 21. Agreement of auditors in recognizing the psychoneurological state of children

Test material	Children's age, y. o.	κ
Words	5–7	0.340
	8–9	0.419
	10–11	0.347
	12–14	0.327
Phrases	5–7	0.426
	8–9	0.562
	10–11	0.538
	12–14	0.519

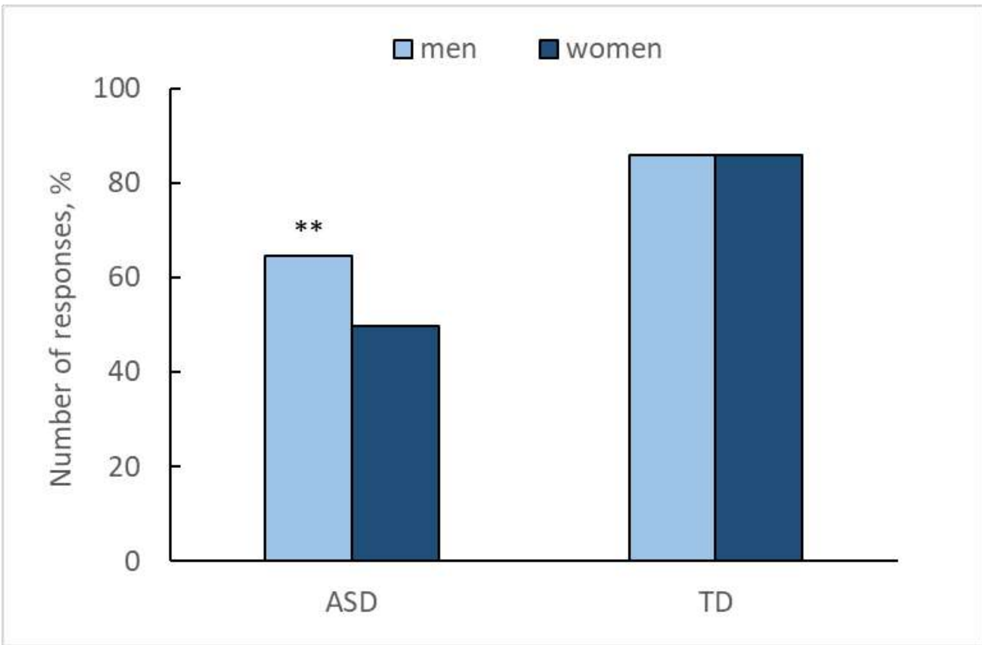
Male auditors are better in recognizing the state of children with ASD than female auditors ($p < 0.01$); no significant differences were found in recognizing the state of TD children (Fig. 22).

Auditors with high level of personal and situational anxiety recognize the state of TD children worse than auditors with low and moderate level of anxiety ($p < 0.05$) (Fig. 21).



The vertical axis is the number of correct answers from auditors, %; the horizontal axis is the groups of children, years; * — $p < 0.05$; Mann–Whitney test

Fig. 21. Recognition of the psychoneurological state of children by auditors with different levels of personal (A) and situational (B) anxiety



The vertical axis is the number of correct answers from auditors, %; the horizontal axis is the groups of children, years; ** — $p < 0.01$; Mann–Whitney test

Fig. 22. Recognition of the psychoneurological state of children by male and female auditors

Based on correlation analysis (Spearman, $p < 0.05$), a connection is shown between recognizing the state of children with ASD and:

1) results of dichotic testing (left hemisphere) (0.38): auditors with the dominant left hemisphere recognize the state of children with ASD better than auditors with the dominant right hemisphere. The data of the correlation analysis are confirmed by the data of multiple regression analysis: $F (5, 83) = 5.242$ $p < 0.0003$ ($R^2 = 0.239$ $\beta = 0.39$);

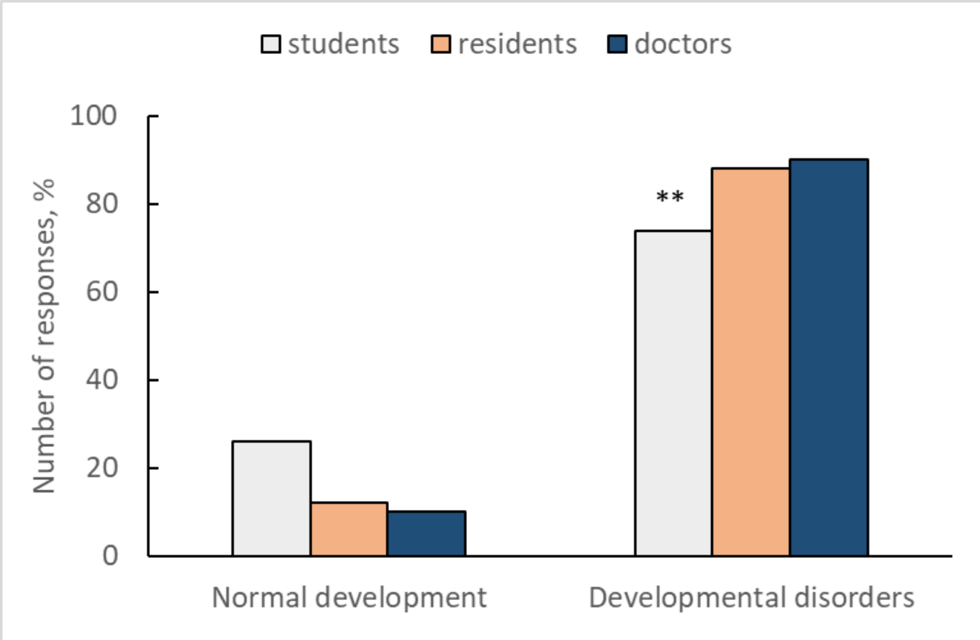
2) gender of the auditor (female) (-0.33): female auditors recognize the state of children with ASD worse than male auditors. The data of the correlation analysis are confirmed by the data of multiple regression analysis: $F (5, 83) = 7.031$ $p < 0.0002$ ($R^2 = 0.298$ $\beta = -0.312$);

3) level of personal anxiety (-0.26): auditors with a high level of anxiety recognize the state of children with ASD worse compared to auditors with low and moderate level of anxiety.

Based on data from multiple regression analysis, a relationship is shown between recognition of the state of children with ASD and hearing thresholds (left ear): $F (5, 83) = 5.2416$ $p < 0.0003$ ($R^2 = 0.239$ $\beta = 0.274$) — auditors with low hearing thresholds recognize the state of children with ASD better than auditors with high hearing thresholds.

A correlation was found between recognition of the child’s psychoneurological state and scores on the CARS scale (0.37).

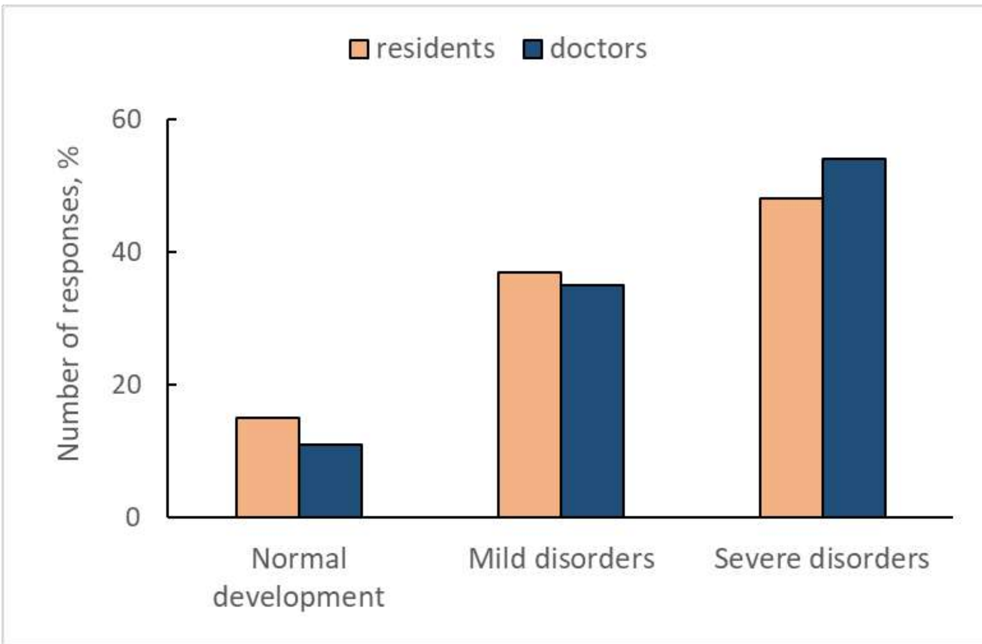
In an additional experiment, first-year students of St. Petersburg State Medical University (n = 19), psychiatry residents (n = 14) and psychiatrists (n = 5) took part in a test containing 30 phrases of children with ASD aged 10−14 years. While listening to the test samples, the auditors were asked to determine the psychoneurological state of children by category: normal development or developmental disorders. Students classified a smaller ($p < 0.05$) number of signals from children with ASD into the "atypical development" category than residents and doctors (74%, 88%, 90% of answers, respectively) (Fig. 23).



The vertical axis is the number of correct answers from auditors, %; the horizontal axis is the answer options; ** — $p < 0.01$; Mann–Whitney test

Fig. 23. Determination of the psychoneurological state of children with ASD by students, residents and doctors: normal development — developmental disorders

The task for residents and doctors included classifying the state of children in three categories: normal development — mild disorders — severe disorders. Residents classified 37% of signals of children with ASD into the "mild disorders" category, and 48% of signals into the "severe disorders" category. Doctors classified 35% of the signals of children with ASD into the "mild disorders" category, and 54% of the signals into the "severe disorders" category (Fig. 24). There were no differences in determining the state of children with ASD between residents and doctors.



The vertical axis is the number of correct answers from auditors, %; the horizontal axis is answer options

Fig. 24. Determination of the psychoneurological state of children with ASD by residents and doctors: normal development — mild disorders — severe disorders

3.4.2. Acoustic characteristics of test material (words)

3.4.2.1. Duration of words, stressed and unstressed vowels

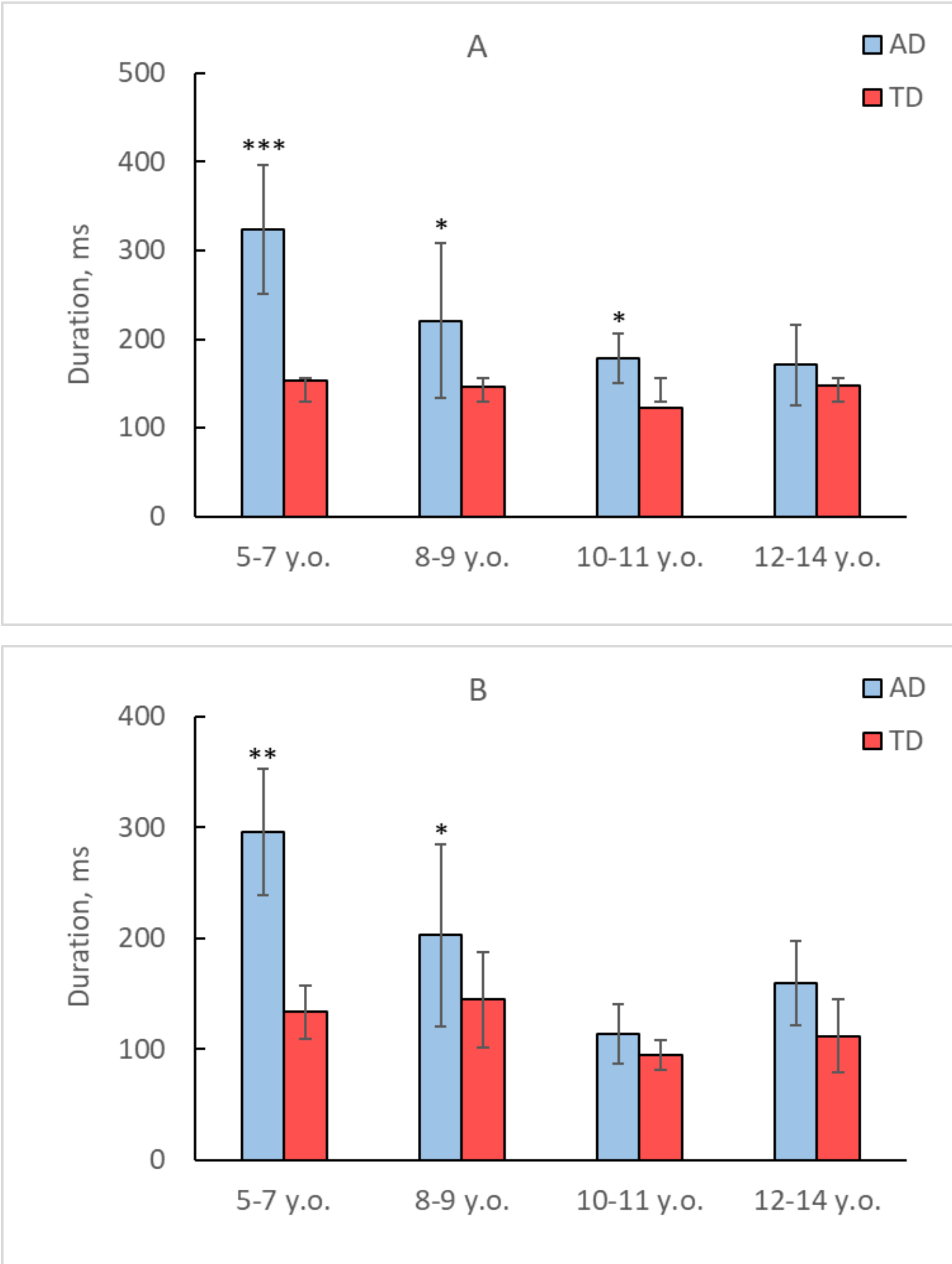
Words classified by auditors into the categories "typical development" and "atypical development" (recognition probability — 0.75−1.0) do not differ in duration, which is apparently due to the high variability of duration within groups (Table 22).

Table 22. Duration of words classified by auditors as "atypical development" and words classified by auditors as "typical development", ms

Age, y. o.	Group	
	atypical development	typical development
5−7	937.9±222.5 (median — 952)	926.3±169.9 (898)
8−9	1322.5±999.8 (725.5)	667.8±185.1 (650)
10−11	769.8±237 (591)	628.5±135.8 (630.5)
12−14	599.7±113 (575)	774.8±116 (781)

Stressed vowels from words of children 5−7, 8−9, and 10−11 years old, classified as "atypical development", have a longer duration ($p < 0.001$; $p < 0.05$; $p < 0.05$, respectively) compared with vowels from words classified as "typical development" (Table 23, Fig. 25A).

Stressed vowels from the words of children aged 12–14 years do not differ significantly. Unstressed vowels from words of children 5−7 and 8−9 years old classified as “atypical development” have a longer duration ($p < 0.01$; $p < 0.05$, respectively) compared to vowels from words classified as “typical development” (Table 23, Fig. 25B). Unstressed vowels from the words of children aged 10–14 years do not have significant differences between the groups.



The vertical axis is duration, ms; the horizontal axis is the age of the children, years; * — $p < 0.05$; ** — $p < 0.01$; *** — $p < 0.001$; Mann–Whitney test

Fig. 25. Duration of stressed (A) and unstressed (B) vowels from words classified by auditors into the categories “atypical development” (AD) and “typical development” (TD)

Table 23. Duration of stressed and unstressed vowels from words classified by auditors as "atypical development" and words classified by auditors as "typical development", ms

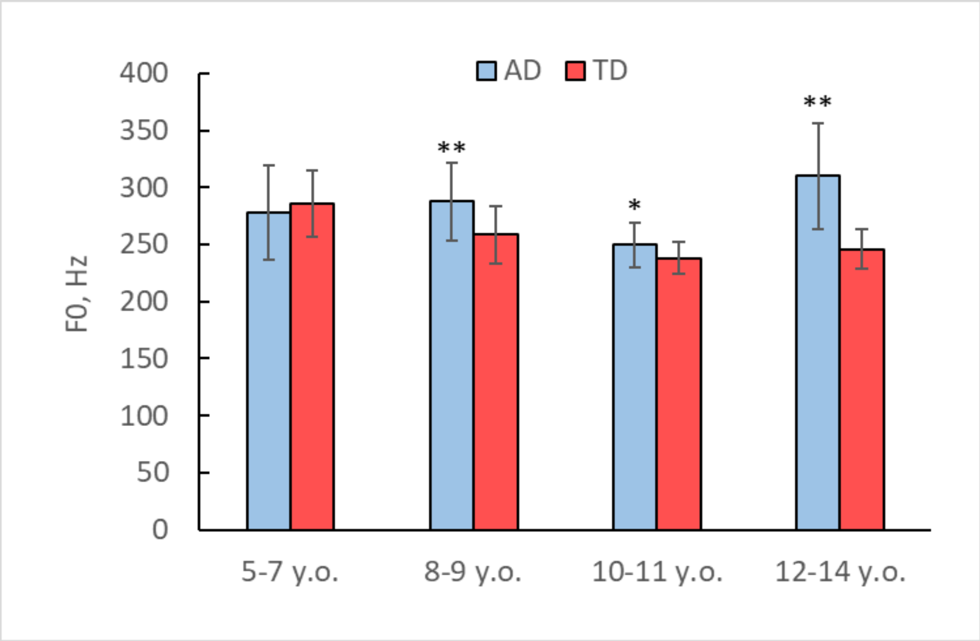
Age, y. o.	Group			
	atypical development		typical development	
	stressed	unstressed	stressed	unstressed
5−7	323.7±72.5 (median — 302)	295.7±58.9 (255)	153.8±36.2 (176)	133.5±24.1 (132.5)
8−9	220.8±87.8 (198)	202.7±82.5 (202.5)	146.1±20.9 (146)	144.5±43.1 (128.5)
10−11	178.6±27.9 (167)	113.4±26.9 (112)	122.7±30 (120.5)	95±13.3 (97.5)
12−14	170.8±45.1 (189)	159.8±38.2 (151)	148.3±23.8 (149)	111.6±33.1 (111)

3.4.2.2. Fundamental frequency values of words, stressed and unstressed vowels

In children aged 5−7 years, no significant differences were found in F0 values of words. Words of children aged 8−9, 10−11 and 12−14 years old classified as "atypical development" have higher F0 values ($p < 0.01$; $p < 0.05$; $p < 0.01$ respectively) as compared with words classified as "typical development" (Table 24, Fig. 26).

Table 24. Fundamental frequency values of words classified by auditors as "atypical development" and words classified by auditors as "typical development", Hz

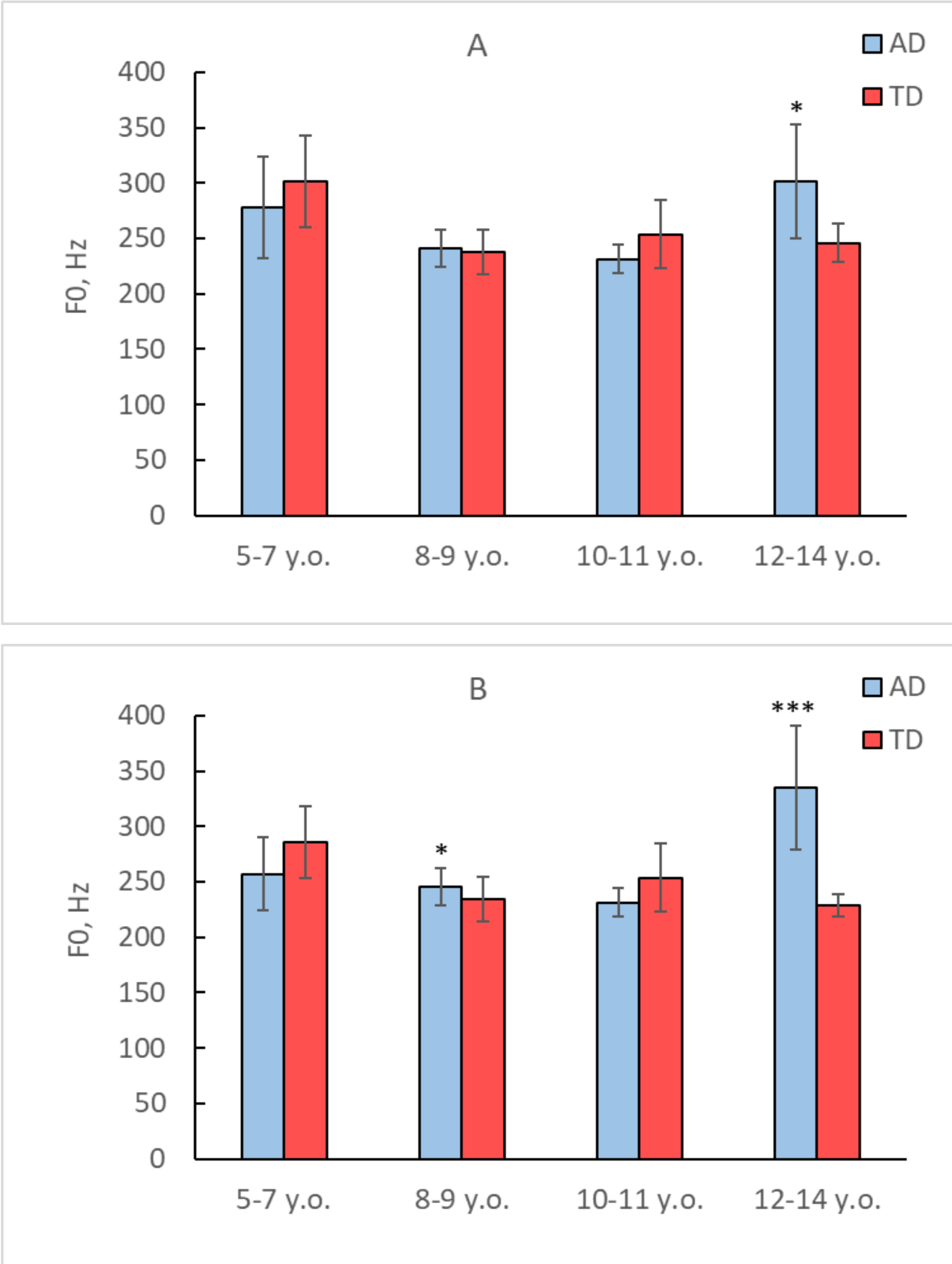
Age, y. o.	Group	
	atypical development	typical development
5−7	277.8±41.2 (median — 281.2)	285±29.1 (281.2)
8−9	287.5±34.3 (281.2)	258.7±25.4 (234.3)
10−11	249.5±19.7 (258.3)	238.2±14.3 (234.3)
12−14	309.9±46.6 (301. 4)	246±17.6 (234.3)



The vertical axis is F0 value, Hz; the horizontal axis is the age of the children, years; * — p <0.05; * — p <0.01; Mann–Whitney test

Fig. 26. Fundamental frequency values of words classified by auditors into the categories "atypical development" (AD) and "typical development" (TD)

The words of children aged 5−11 years old, classified into the categories of "typical" and "atypical development", do not differ in F0 values of stressed vowels. In words classified as "atypical development", in children aged 12-14 years, F0 values of stressed vowels are higher (p <0.05) compared to words classified as "typical development" (Table 25, Fig. 27A). Unstressed vowels from the words of children 5−7 and 10−11 years old do not have significant differences in F0 values between the groups. In words of children aged 8−9 and 12−14 years classified as "atypical development" F0 values of unstressed vowels are higher (p <0.05; p <0.001, respectively) compared to words classified as "typical development" (Table 25, Fig. 27B).



The vertical axis is F0 value, Hz; the horizontal axis is the age of the children, years; * — p <0.05; *** — p <0.001; Mann–Whitney test

Fig. 27. Fundamental frequency values of stressed (A) and unstressed (B) vowels from words classified by auditors into the categories “atypical development” (AD) and “typical development” (TD)

Table 25. Fundamental frequency values of stressed and unstressed vowels from words classified by auditors as "atypical development" and words classified by auditors as "typical development", Hz

Age, y. o.	Group			
	atypical development		typical development	
	stressed	unstressed	stressed	unstressed
5−7	277.8±46.1 (median is 281.2)	257.1±33.2 (234.2)	301.4±41.8 (291.3)	285.5±32.5 (281.2)
8−9	241.6±16.8 (246.3)	241.6±16.8 (246.3)	237.7±20 (234.3)	237.9±20 (234.3)
10−11	231.5±30.6 (234.3)	231.5±13 (234.3)	253.2±30.6 (234.3)	253.8±30.6 (234.3)
12−14	301.3±51.8 (301. 4)	335.2±56.1 (301. 4)	246±17.6 (234.3)	228.5±10.2 (234.3)

Thus, the words of children classified as "atypical development" are characterized by larger values of duration and fundamental frequency of vowels compared to words classified as "typical development".

3.4.3. Acoustic characteristics of test material (phrases)

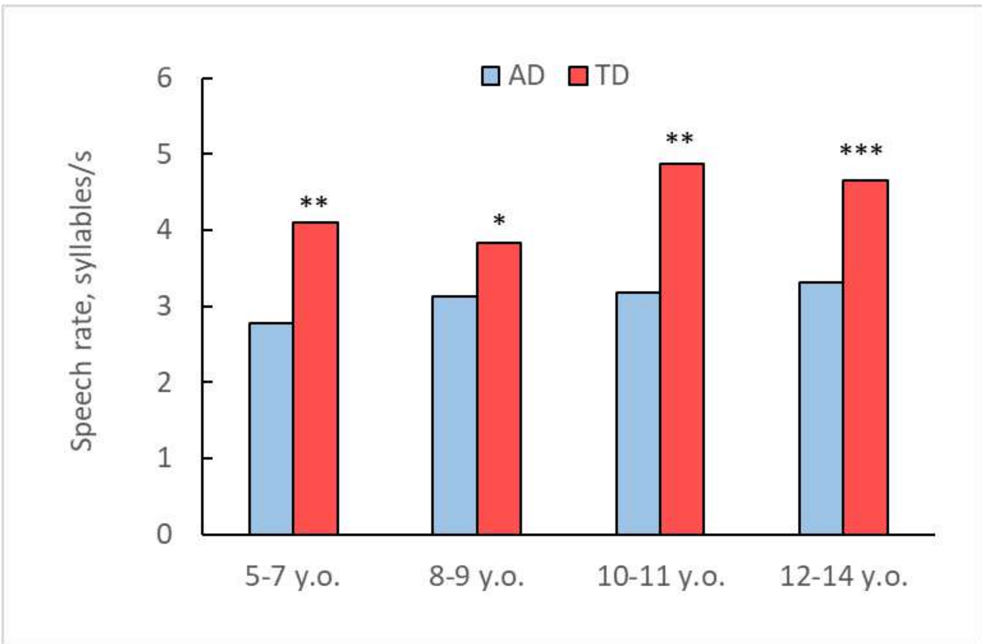
3.4.3.1. Number of words in phrases and speech rate

Phrases of children classified by auditors as "atypical development" are characterized by a smaller number of words compared to phrases classified by auditors as "typical development" (5−7 years — $p < 0.001$; 8−9, 10−11 years — $p < 0.01$; 12−14 years — $p < 0.05$) (Table 26).

Table 26. Number of words in phrases classified by auditors into the categories of "atypical development" and "typical development"

Age, y. o.	Atypical development	Typical development
5−7	1.6±0.6	3.7±1.3
8−9	1.7±0.6	3.2±0.8
10−11	2.5±0.8	4.7±1.3
12−14	3.5±1.9	4.5±1.6

Phrases classified as "typical development" are distinguished by higher speed of speech compared to phrases classified as "atypical development" ($p < 0.01$ for 5−7 years; $p < 0.05$ for 8−9 years; $p < 0.01$ for 10−11 years; $p < 0.001$ for 12−14 years) (Fig. 28).



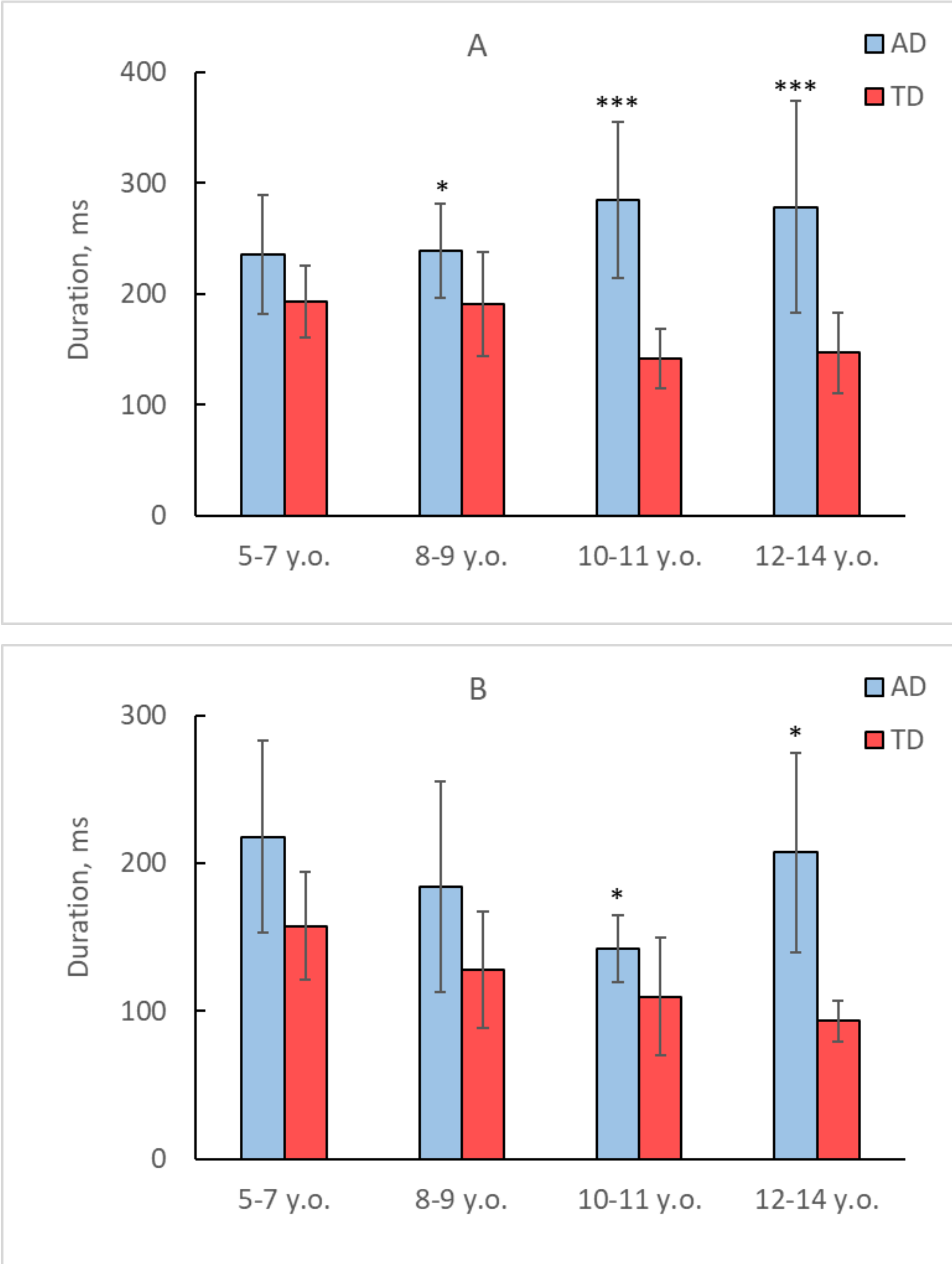
The vertical axis is speech rate, syllables/s; the horizontal axis is the age of the children, y. o.; * — $p < 0.05$; ** — $p < 0.01$; *** — $p < 0.001$; Mann–Whitney test

Fig. 28. Speech rate of phrases classified by auditors into the categories "atypical development" (AD) and "typical development" (TD)

3.4.3.2. Duration of phrases, words, stressed and unstressed vowels

Speech material classified into the categories "atypical development" and "typical development" does not differ in the duration of phrases and pauses.

Stressed vowels of 5–7 years old children do not differ significantly in duration. Stressed vowels of children aged 8−9, 10−11 and 12−14 years from phrases classified as "atypical development" have a longer duration ($p < 0.05$; $p < 0.001$; $p < 0.001$ — respectively) compared to vowels from phrases classified as "typical development" (Table 27, Fig. 29A). There were no significant differences in the duration of unstressed vowels in children aged 5−9 years. Unstressed vowels of children aged 10−11 and 12−14 years from phrases classified as "atypical development" have a longer duration ($p < 0.05$) compared to vowels from phrases classified as "typical development" (Table 27, Fig. 29B).



The vertical axis is duration, ms; the horizontal axis is the age of the children, years; * — p <0.05; *** — p <0.001; Mann–Whitney test

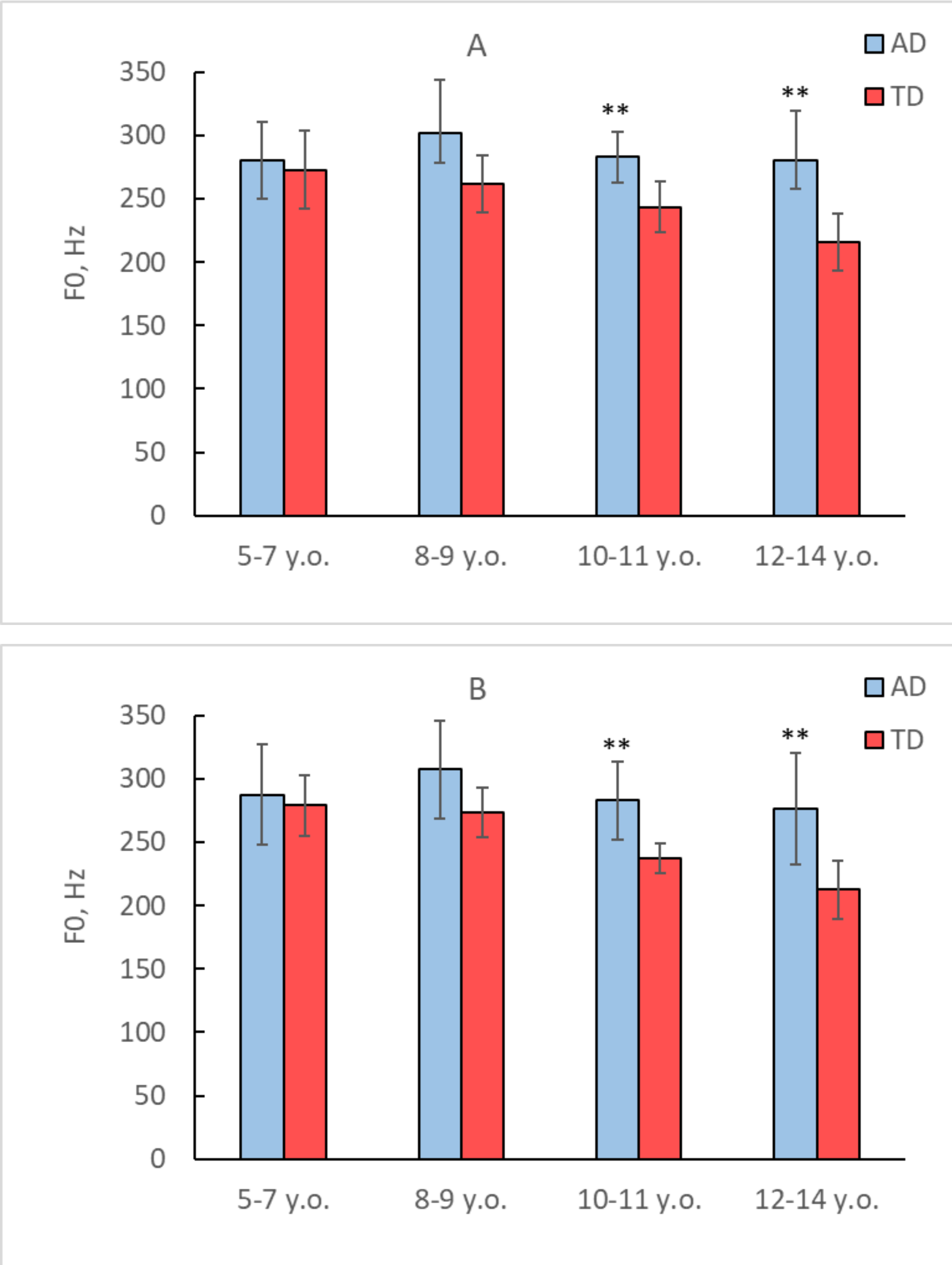
Fig. 29. Duration of stressed (A) and unstressed (B) vowels from phrases classified by auditors as “atypical development” (AD) and “typical development” (TD)

Table 27. Duration of stressed and unstressed vowels from words in phrases classified by auditors as "atypical development" and words classified by auditors as "typical development", ms

Age, y. o.	Group			
	atypical development		typical development	
	stressed	unstressed	stressed	unstressed
5−7	235.6±53.8 (median is 221.5)	217.8±64.9 (191.5)	193.1±32.1 (187)	157.5±36.4 (148)
8−9	238.7±42.1 (251)	184.3±71.4 (147.5)	190.1±47.1 (163)	128±39.4 (121.5)
10−11	284.8±70.7 (267.5)	142.4±22.5 (137)	141.5±26.8 (151)	209.9±40.1 (95.5)
12−14	278.5±95.1 (237)	207.3±67.6 (194.5)	146.6±36.3 (137)	93.3±14.1 (95)

3.4.3.3. Fundamental frequency values of phrases, words, stressed and unstressed words

Fundamental frequency values of phrases and word do not differ significantly in children aged 5−9 years. In children aged 10−11 and 12−14 years, F0 values for phrases and words in phrases classified as "atypical development" are higher ($p < 0.01$) than F0 values in phrases classified as "typical development" (Fig. 30). In children aged 8−9 and 12−14 years, the maximum F0 values of phrases and words in phrases classified as "atypical development" are higher ($p < 0.01$) compared to the corresponding speech material classified as "typical development."

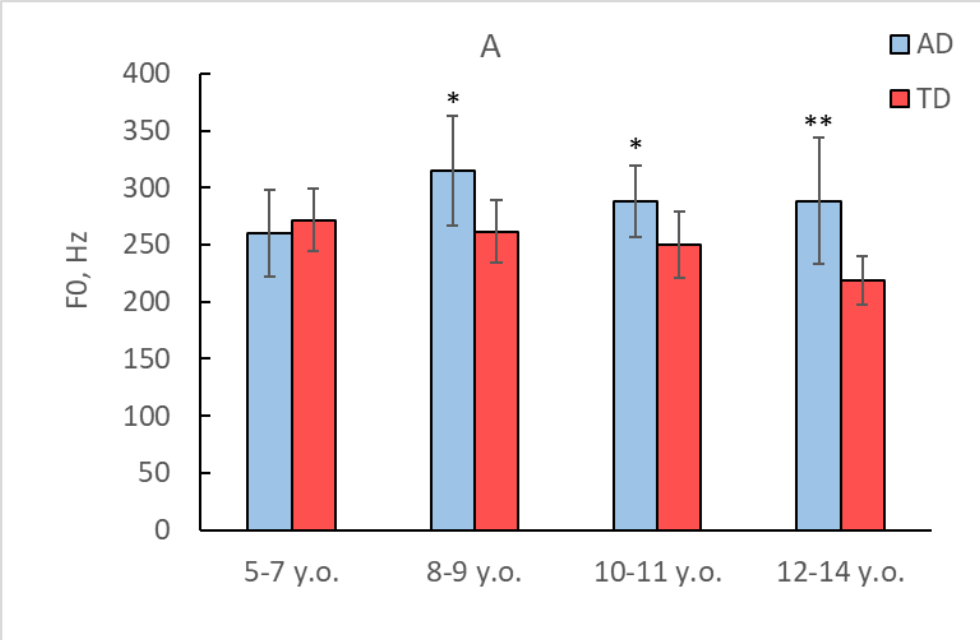


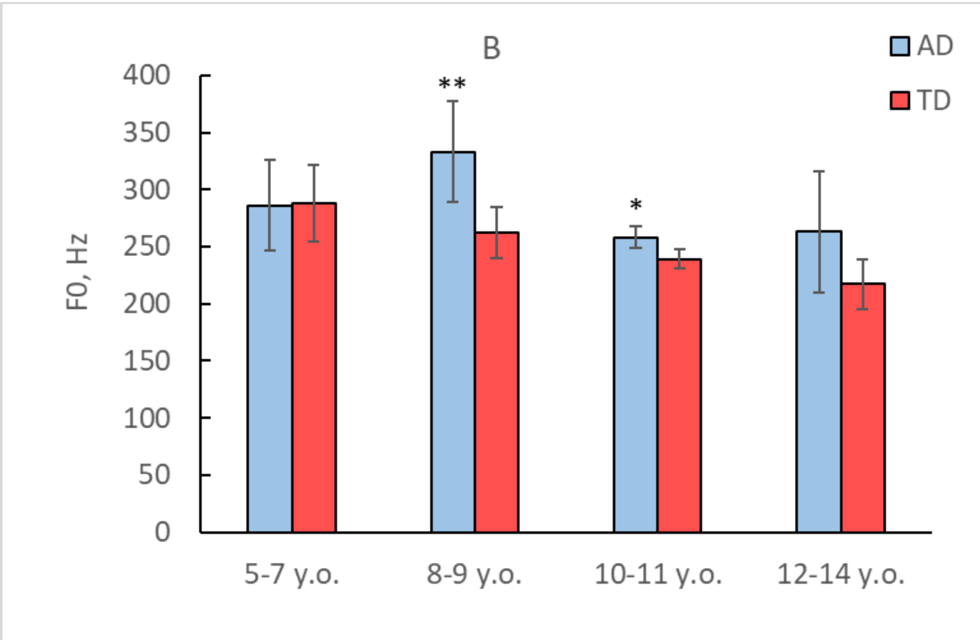
The vertical axis is fundamental frequency values, Hz; the horizontal axis is the age of the children, years; ** — p <0.01; Mann–Whitney test

Fig. 30. Fundamental frequency values of phrases (A) and words (B) from phrases classified by auditors as "atypical development" (AD) and "typical development" (TD)

In the age of 5−7 years, F0 values of stressed vowels do not differ. F0 values of stressed vowels in children aged 8−9, 10−11 and 12−14 years are higher ($p < 0.05$; $p < 0.05$; $p < 0.001$, respectively) in phrases classified as "atypical development" than in phrases classified as "typical development" (Table 28, Fig. 31A).

There are no significant differences in F0 values of unstressed vowels in children 5−7 and 12−14 years old. F0 values of unstressed vowels in children aged 8−9 and 10−11 years are higher (p <0.05; p <0.01, respectively) in phrases classified as “atypical development” than in phrases classified as “typical development” (Table 28, Fig. 31B).





The vertical axis is fundamental frequency values, Hz; the horizontal axis is the age of the children, years; * — p <0.05; ** — p <0.01; Mann–Whitney test

Fig. 31. Fundamental frequency values of stressed (A) and unstressed (B) vowels of words from phrases classified by auditors into the categories “atypical development” (AD) and “typical development” (TD)

Table 28. Fundamental frequency values of stressed and unstressed vowels from words classified by auditors into the category “atypical development” and words classified by auditors into the category “typical development”, Hz

Age, y. o.	Group			
	atypical development		typical development	
	stressed	unstressed	stressed	unstressed
5−7	259.8±38.3 (median is 234.3)	287.5±39.7 (281.2)	271.7±27.2 (281.2)	287.9±33.9 (281.2)
8−9	314.3±48.1 (281.2)	333.1±44.3 (314.8)	361.7±27.4 (257.8)	262.4±22.5 (281.2)
10−11	288.5±31.3 (281.2)	258.3±9.5 (258.3)	249.9±29.2 (234.3)	239±8.4 (234.3)
12−14	288.4±55.2 (301.4)	262.9±53.4 (246.3)	218.7±20.8 (234.3)	217.3±21.7 (234.3)

The speech material of children classified by auditors as “atypical development” is characterized by a smaller number of words, by a lower speech rate compared to the corresponding material classified by auditors as “typical development”, by longer values of the duration of stressed and unstressed vowels, and by greater F0 values of phrases, words, stressed and unstressed vowels.

3.4.4. Connections between auditors’ recognition of the psychoneurological state of children and the acoustic characteristics of the test material

Based on correlation analysis (Spearman, $p < 0.05$), connections are shown between classifying children’s speech signals as “atypical development” and:

1) the number of words in a phrase (-0.47): auditors classify phrases with fewer words as “atypical development”;

2) speech rate (-0.5): auditors classify phrases with a lower speech rate as “atypical development”; the data of the correlation analysis are confirmed by the data of multiple regression analysis: $F (6, 113) = 8.305$ $p < 0.0000$ ($R^2 = 0.269$ $\beta = -0.552$);

3) the type of utterance in the test (-0.47): auditors are less likely to classify children’s speech signals presented as a phrase or several phrases as “atypical development” compared to signals presented as a word or speech-like constructions;

4) phrase duration (-0.35): auditors classify phrases with a shorter duration as “atypical development”;

5) duration of the stressed vowel (0.47); unstressed vowel (0.24): the data of the correlation analysis are confirmed by the data of multiple regression analysis: F (6, 79) = 8.781 p <0.0000 (R^2 = 0.355 β = 0.267) — in signals classified as "atypical development" there is more duration of stressed and unstressed vowels than in signals classified as "typical development";

6) duration of pauses (0.18): auditors classify signals with longer pauses as "atypical development";

7) F0 values of phrases (0.39), words (0.36), stressed vowels (0.3): auditors classify signals with higher F0 values as "atypical development";

8) maximum F0 values of phrases (0.46) — the data of the correlation analysis are confirmed by the data of multiple regression analysis: F (7, 112) = 7.701 p <0.0000 (R^2 = 0.283 β = 0.548); words (0.45), stressed vowels (0.3), F0 variability in phrases (0.42) — in signals classified by auditors as "atypical development" there are higher maximum F0 values of phrases, words and stressed vowels, higher F0 variability in phrases.

Based on multiple regression analysis data, a connection is shown between classifying speech signals as "atypical development" and:

1) minimum F0 values of phrases F (7, 112) = 7.701 p <0.0000 (R^2 = 0.83 β = -0.472) — in signals classified by auditors as "atypical development" there are higher minimum F0 values of phrases;

2) F0 variability values of stressed vowels F (6, 79) = 8.781 p < 0.0000 (R^2 = 0.355 β = 0.477) — in signals classified by auditors as "atypical development" there are higher F0 variability values in phrases;

A correlation is shown between scores on the CARS scale and the acoustic characteristics of speech: maximum F0 values of words (0.56), F0 values of unstressed vowels (0.7), F0 intensity values of unstressed vowels (0.36).

Thus, the classification of children's speech signals into the category of "atypical development" is influenced by a smaller number of words in a phrase, low speech rate, type of response (asingle word or a speech-like construction), shorter duration of a phrase, longer duration of vowels and pauses, high F0 values of phrases, words and vowels, high values of F0 variability.

3.5. Recognition of the emotional state of children by adults

3.5.1. Data from a perceptual experiment

When recognizing the emotional state of children with ASD aged 5−7 years, auditors were better at identifying the state of discomfort (63% of correct answers) and neutral (60% of correct answers) than the state of comfort (58% of correct answers). The average recall of recognition (UAR) was 0.6 (Table 29).

Table 29. Confusion matrix for recognition of the emotional state of children with ASD, 5−7 years old: comfort — neutral state — discomfort, % of correct answers by auditors

State	Comfort	Neutral	Discomfort
Comfort	**58**	28	14
Neutral	8	**60**	32
Discomfort	15	22	**63**
Recall	0.58	0.6	0.63
Precision	0.72	0.55	0.58
F_1-score	0.64	0.57	0.6
UAR	0.6		

When recognizing the emotional state of children with ASD aged 8−9 years, auditors were better at identifying the state of comfort (78% of correct answers) and discomfort (52% of correct answers) than the neutral state (34% of correct answers). The average recall of recognition was 0.55 (Table 30).

Table 30. Confusion matrix for recognition of the emotional state of children with ASD, 8−9 years old: comfort — neutral state — discomfort, % of correct answers by auditors

State	Comfort	Neutral	Discomfort
Comfort	**78**	16	6
Neutral	52	**34**	14
Discomfort	36	12	**52**
Recall	0.78	0.34	0.52
Precision	0.47	0.55	0.72
F_1-score	0.59	0.42	0.6
UAR	0.55		

When recognizing the emotional state of children with ASD aged 10−11 years, auditors were better at identifying the state of discomfort (78% of correct answers) and neutral (74% of correct answers) than the state of comfort (43% of correct answers). The average recall of recognition was 0.65 (Table 31).

Table 31. Confusion matrix for recognition of the emotional state of children with ASD, 10−11 years old: comfort — neutral state — discomfort, % of correct answers by auditors

State	Comfort	Neutral	Discomfort
Comfort	**43**	46	11
Neutral	11	**74**	15
Discomfort	12	9	**78**
Recall	0.43	0.74	0.78
Precision	0.65	0.57	0.75
F_1-score	0.52	0.64	0.76
UAR	0.65		

When recognizing the emotional state of children with ASD aged 12−14 years, auditors were better at identifying the state of comfort (60% of correct answers) and discomfort (54% of correct answers) than the neutral state (53% of correct answers). The average recall of recognition was 0.56 (Table 32).

Table 32. Confusion matrix for recognition of the emotional state of children with ASD, 12−14 years old: comfort — neutral state — discomfort, % of correct answers by auditors

State	Comfort	Neutral	Discomfort
Comfort	**60**	23	17
Neutral	18	**53**	29
Discomfort	28	18	**54**
Recall	0.6	0.53	0.54
Precision	0.57	0.56	0.54
F_1-score	0.58	0.54	0.54
UAR	0.56		

When recognizing the emotional state of TD children aged 5−7 years, auditors were better at identifying the state of comfort (74% of correct answers) and neutral (69% of correct answers) than the state of discomfort (46% of correct answers). The average recall of recognition was 0.63 (Table 33).

Table 33. Confusion matrix for recognition of the emotional state of TD children, 5−7 years old: comfort — neutral state — discomfort, % of correct answers by auditors

State	Comfort	Neutral	Discomfort
Comfort	**74**	23	3
Neutral	26	**69**	5
Discomfort	35	18	**46**
Recall	0.74	0.69	0.46
Precision	0.55	0.63	0.85
F_1-score	0.63	0.66	0.6
UAR	0.63		

When recognizing the emotional state of TD children aged 8−9 years, auditors were better at identifying the state of comfort (44% of correct answers) and neutral (52% of correct answers) than the state of discomfort (24% of correct answers). The average recall of recognition was 0.4 (Table 34).

Table 34. Confusion matrix for recognition of the emotional state of TD children, 8−9 years old: comfort — neutral state — discomfort, % of correct answers by auditors

State	Comfort	Neutral	Discomfort
Comfort	**44**	38	18
Neutral	28	**52**	20
Discomfort	16	60	**24**
Recall	0.44	0.52	0.24
Precision	0.5	0.35	0.39
F_1-score	0.47	0.42	0.3
UAR	0.4		

When recognizing the emotional state of TD children aged 10−11 years, auditors were better at identifying the state of comfort (71% of correct answers) and neutral (83% of correct answers) than the state of discomfort (46% of correct answers). The average recall of recognition was 0.63 (Table 35).

Table 35. Confusion matrix for recognition of the emotional state of TD children, 10−11 years old: comfort — neutral state — discomfort, % of correct answers by auditors

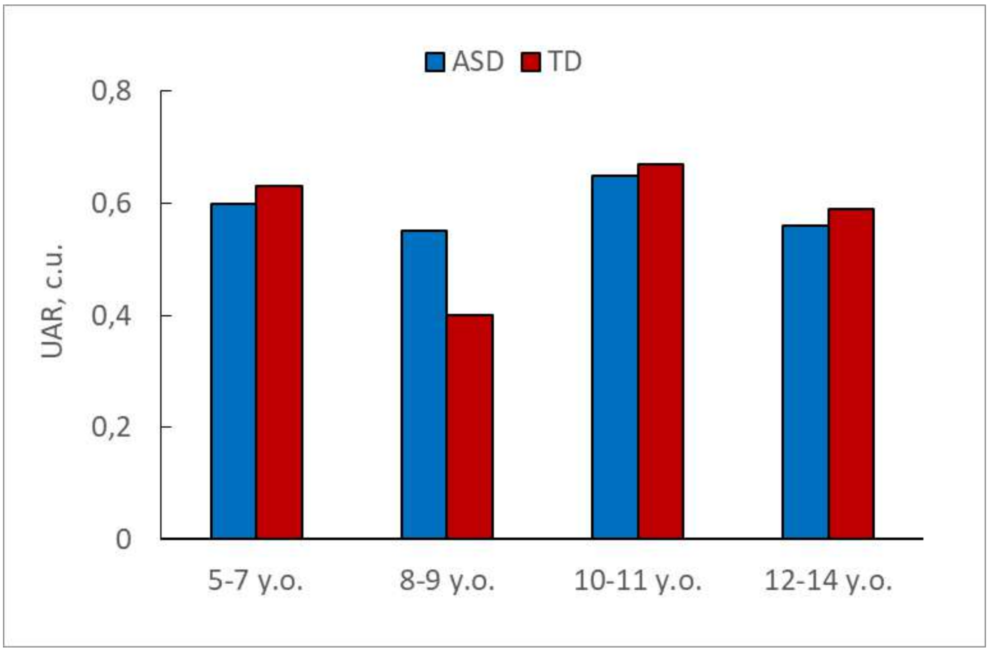
State	Comfort	Neutral	Discomfort
Comfort	**71**	29	0
Neutral	12	**83**	5
Discomfort	5	49	**46**
Recall	0.71	0.83	0.46
Precision	0.81	0.52	0.9
F_1-score	0.76	0.64	0.61
UAR	0.67		

When recognizing the emotional state of TD children aged 12−14 years, auditors were better at identifying the state of comfort (63% of correct answers) and neutral (67% of correct answers) than the state of discomfort (46% of correct answers). The average recall of recognition was 0.59 (Table 36).

Table 36. Confusion matrix for recognition of the emotional state of TD children, 12−14 years old: comfort — neutral state — discomfort, % of correct answers by auditors

State	Comfort	Neutral	Discomfort
Comfort	**63**	27	10
Neutral	22	**67**	11
Discomfort	10	44	**46**
Recall	0.63	0.67	0.46
Precision	0.66	0.49	0.69
F_1-score	0.64	0.57	0.55
UAR	0.59		

Auditors determine the emotional state of TD children aged 5−7, 10−11 and 12−14 years better than the emotional state of children with ASD. In the age of 8−9 years, the emotional state is better defined in children with ASD than in TD children. The values of average recall were maximum in tests for determining the emotional state of 10−11 years old children (0.65 for ASD; 0.67 for TD) and 5−7 years old children (0.6 for ASD; 0.63 for TD). The minimum values of average recall were in tests to determine the state of children 8−9 years old: 0.55 for ASD; 0.4 for TD (Fig. 32).



The vertical axis is unweighted average recall (UAR) values, c. u.; the horizontal axis is the age of the children, years

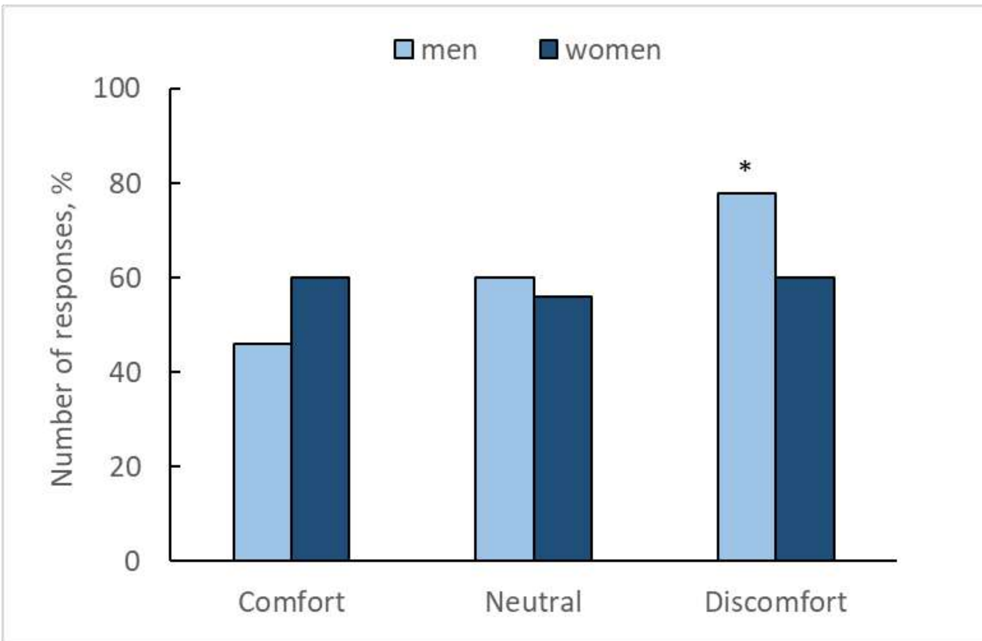
Fig. 32. Average recall of recognition of the emotional state of children by auditors

When recognizing the emotional state of children 5–7 years old, auditors demonstrated moderate agreement (κ = 0.408); for other age groups, agreement was fair (0.223 for 8–9 years old; 0.378 for 10–11 years old; 0.292 for 12–14 years old). In all age groups, auditor agreement is higher when recognizing states of comfort and discomfort compared to the neutral state (Table 37).

Male auditors are better at recognizing the state of discomfort in children with ASD than female auditors ($p < 0.05$); no significant differences were found in recognizing state of comfort and neutral state (Fig. 33).

Table 37. Agreement of auditors in recognizing the emotional state of children

Children's age, y. o.	Emotional state	κ
5–7	comfort	0.425
	neutral	0.315
	discomfort	0.489
	all states	0.408
8–9	comfort	0.295
	neutral	0.152
	discomfort	0.204
	all states	0.224
10–11	comfort	0.419
	neutral	0.308
	discomfort	0.428
	all states	0.378
12–14	comfort	0.317
	neutral	0.267
	discomfort	0.293
	all states	0.292



The vertical axis is the number of correct answers from auditors, %; the horizontal axis is groups of the children, years; * — $p < 0.05$; Mann–Whitney test

Fig. 33. Recognition of the emotional state of children with ASD by male and female auditors

Based on correlation analysis (Spearman, $p < 0.05$), the following connections are shown:

1) between the age of the auditor and the recognition of the neutral state in TD children ($r = -0.68$): older auditors are worse at recognizing the neutral state of TD children;

2) between the gender of the auditor (female) and recognition of the discomfort state in children with ASD (-0.73): female auditors are worse at recognizing the state of discomfort in children with ASD;

3) between the results of dichotic testing (left hemisphere) and recognition of the discomfort state of TD children (0.48): auditors with the dominant left hemisphere are better at recognizing the state of discomfort in TD children;

4) the auditor's experience is connected (0.82) with recognizing the neutral state in children with ASD: auditors with experience of interacting with children are better at recognizing the neutral state in children with ASD;

5) the level of personal (-0.91) and situational (-0.66) anxiety is connected with recognition of the neutral state of TD children: auditors with a high level of anxiety are worse at recognizing the neutral state in TD children.

Based on multiple regression analysis, connections are shown between the auditor's professional experience in interacting with children, the auditor's age, hearing thresholds, the level of personal and situational anxiety, and the recognition of states of comfort and discomfort in children with ASD and TD children (Table 38).

Table 38. Connections between the individual characteristics of auditors and recognition of the emotional state of children with ASD and TD children. Regression analysis data

R^2	F	Independent variables	β	SE β	B	SE B	t	p
ASD								
Dependent variable: comfort state							t (15)	
0.315	(1,15) 6.911	Hearing thresholds (audiometry for left ear)	-0.562	0.214	-29.33	11.155	-2.629	0.019
TD								
Dependent variable: comfort state							t (5)	
0.916	(6,5) 9.089	Personal anxiety	-0.59	0.223	27.193	8.819	-2.651	0.045
		Situational anxiety	-0.906	0.168	-1.985	0.367	-5.407	0.003
		Professional experience	-0.934	0.195	-32.503	6.795	-4.783	0.005
Dependent variable: discomfort state							t (24)	
0.943	(5, 6) 19.911	Personal anxiety	0.689	0.151	30.523	6.721	4.541	0.001
		Auditor's age	-0.692	0.113	-4.736	0.771	-6.142	0.001

R^2 is the square of the correlation coefficient (R); SE is standard error; B is regression coefficient

Auditors with low hearing thresholds are better in recognizing the state of comfort in children with ASD. Auditors with a high level of anxiety are worse at recognizin the state of comfort in TD children, but better at recognizing the state of discomfort; auditors with professional experience of interacting with children are better in recognizing the state of comfort in TD children. Older auditors are less able to identify the state of discomfort in TD children.

3.5.2. Acoustic characteristics of test material

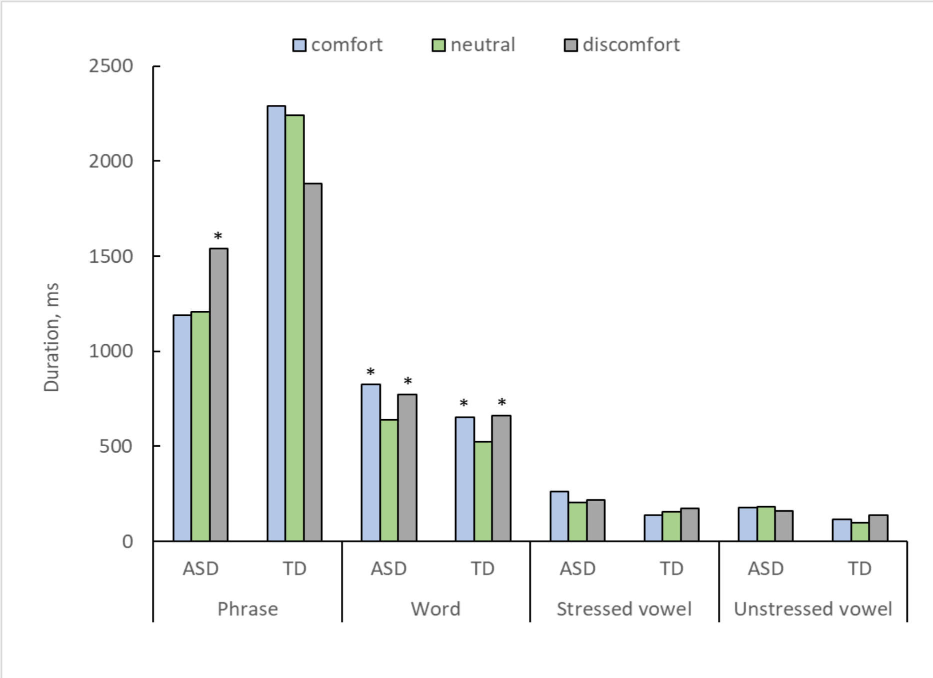
In signals reflecting a state of comfort, children with ASD had a higher duration of words compared to the neutral state and a lower duration of pauses ($p < 0.05$). In signals reflecting a state of discomfort, the duration of phrases and words is higher than in the neutral state and the duration of pauses is lower ($p < 0.05$). TD children in the comfort and discomfort states had longer word duration compared to the neutral state ($p < 0.05$). The duration of stressed and unstressed vowels in all three emotional states does not differ significantly in both groups of children (Table 39, Fig. 34). The fundamental frequency values of phrases, words and stressed vowels in children with

ASD are higher in the states of comfort (p <0.01) and discomfort (p <0.01; p <0.001; p <0.001 for phrases, words, stressed vowels, respectively) than in the neutral state. F0 values of unstressed vowels in children with ASD and TD children are higher in the discomfort state compared to the neutral state (p <0.01) (Table 39, Fig. 35). In the states of comfort and discomfort, children with ASD had a higher F0 variability of phrases (p <0.01), words (p <0.01), stressed vowels (p <0.05) than in the neutral state. TD children had a higher F0 variability of phrases (p < 0.05) and words (p < 0.05) in the comfort and discomfort conditions than in the neutral state. F0 variability of stressed vowels in TD children is higher in the discomfort state than in the neutral state (p <0.05) (Table 39, Fig. 36).

Table 39. Differences between the acoustic characteristics of children's speech signals, reflecting different emotional states

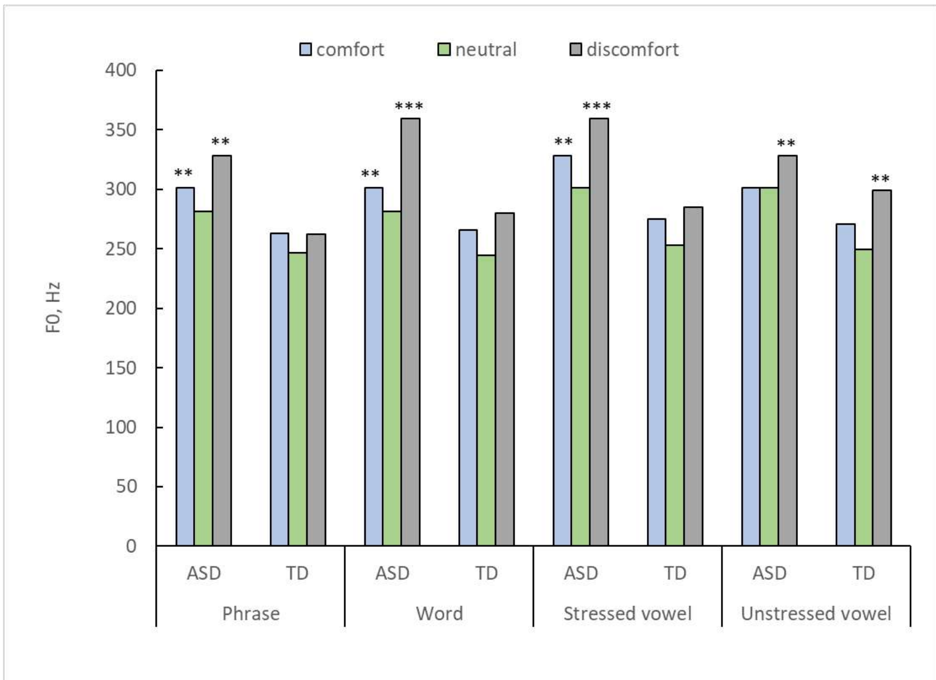
Characteristics	Emotional state					
	comfort (c)		neutral (n)		discomfort (d)	
	ASD	TD	ASD	TD	ASD	TD
Phrase duration					D> N*	
Word duration	C> N*	C> N*			D> N*	D> N*
Pause duration			N> C* N> D*			
F0 of phrase	C> N**				D> N**	
F0 of word	C> N**				D> N***	
F0 of stressed vowel	C> N**				D> N***	
F0 of unstressed vowel					D> N**	D> N**
F0 variability of phrase	C> N**	C> N*			D> N**	D> N*
F0 variability of word	C> N**	C> N*			D> N**	D> N*
F0 variability of stressed vowel	C> N*				D> N*	
F3 /i/					D> C> N*	

* — p <0.05; ** — p <0.01; *** — p <0.001; Mann–Whitney test



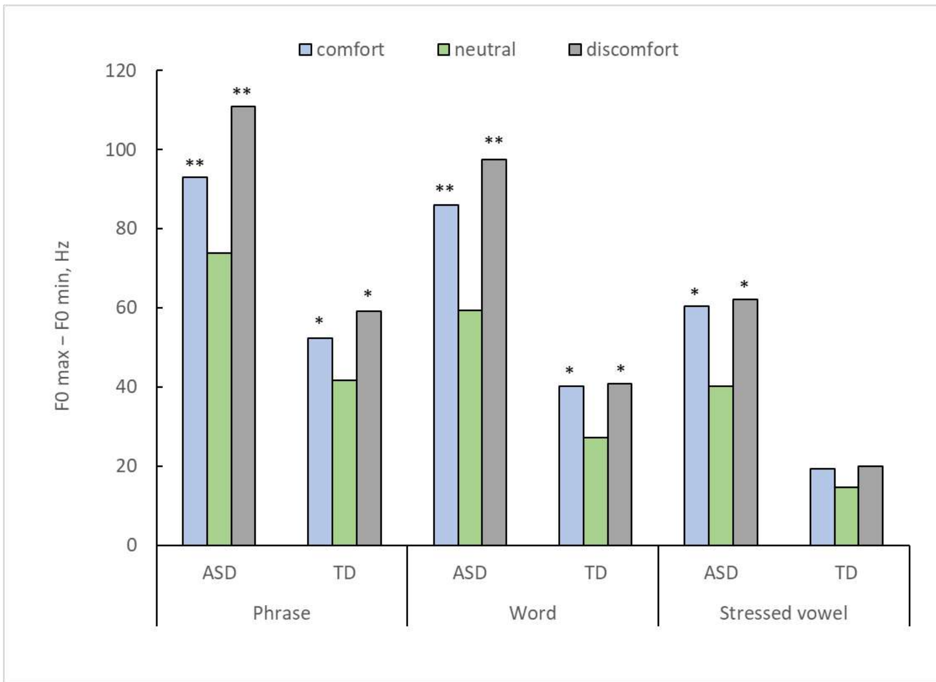
The vertical axis is duration, ms; the horizontal axis is characteristics and groups of children; * — p <0.05; Mann–Whitney test

Fig. 34. Duration of phrases, words and vowels of children with ASD and TD children in speech signals reflecting different emotional states



The vertical axis is fundamental frequency value, Hz; the horizontal axis is characteristics and groups of children; * — p <0.05; ** — p <0.01; *** — p <0.001; Mann–Whitney test

Fig. 35. Fundamental frequency values of phrases, words and stressed vowels of children with ASD and TD children in speech signals reflecting different emotional states



The vertical axis is F0 variability value, Hz; the horizontal axis is characteristics and groups of children; * — p <0.05; ** — p <0.01; Mann–Whitney test

Fig. 36. Fundamental frequency variability of phrases, words and stressed vowels of children with ASD and TD children in speech signals reflecting different emotional states

The values of the third formant of the stressed vowel /i/ in children with ASD in the comfort state are higher (p <0.05) than in the neutral state; for the state of discomfort the named values are higher than for the state of comfort (p <0.05) (Table 39).

Differences in the acoustic characteristics of emotional speech signals of children with ASD correctly classified by auditors (recognition probability is 0.75–1.0) compared with signals classified incorrectly (recognition probability is 0.0–0.25) were shown. Correctly recognized signals reflecting a state of comfort are characterized by larger minimum F0 values of phrases (p <0.05; Mann–Whitney test), larger maximum F0 values of stressed vowels (p <0.05); signals reflecting a neutral state are characterized by smaller maximum and minimum F0 values of phrases (p <0.05; p <0.001 respectively); of words (p <0.05), smaller maximum F0 values of stressed vowels (p <0.05). For signals reflecting a state of discomfort, a smaller intensity is characteristic for F0 values of stressed vowels (p <0.05) (Table 40).

The TD children's signals which were classified by auditors correctly and reflected a state of comfort are characterized by larger F0 values of phrases and words, larger maximum F0 values of phrases, larger F0 variability values of phrases, and by larger F0 intensity of stressed vowels (p <0.05); signals which reflected a state of discomfort are characterized by greater F0 intensity of stressed vowels (p <0.05) (Table 41). No differences were found between the signals reflecting the neutral state.

Table 40. Differences in the acoustic characteristics of emotional speech signals of children with ASD, recognized by auditors with high and low probability

Characteristics	Emotional state					
	Comfort		Neutral		Discomfort	
	recognition probability		recognition probability		recognition probability	
	0.75–1.0	0.0–0.25	0.75–1.0	0.0–0.25	0.75–1.0	0.0–0.25
max. F0 of phrase, Hz			303.9±35.7	339.6±29.3*		
min. F0 of phrase, Hz	283.9±31.9*	259.5±20.2	219±14.3	281±29.4***		
max. F0 of word, Hz			300.5±38	339.6±29.3*		
min. F0 of word, Hz			238.9±31.7	295.3±43.7*		
min. F0 of stressed vowel, Hz	383.4±57.7*	317.9±45.6	293.9±41.8	339.7±29.4*		
F0 intensity of stressed vowel, dB					-32.9 ±9.2	-24.3 ±9.3*

max. — maximum value; min. — minimum value; * — p <0.05; *** — p <0.001; Mann–Whitney test

Table 41. Differences in the acoustic characteristics of emotional speech signals of TD children, recognized by auditors with high and low probability

Characteristics	Emotional state			
	Comfort		Discomfort	
	recognition probability		recognition probability	
	0.75–1.0	0.0–0.25	0.75–1.0	0.0–0.25
F0 of phrase, Hz	286±31.7*	239.1±7.6		
F0 of word, Hz	290.6±48.6*	239.1±7.7		
max. F0 of phrase, Hz	319.8±32.9*	266.5±25.7		
F0 variability of phrase, Hz	69.2±21.7*	27.38±21.9		
F0 intensity of stressed vowel, dB	-29.6±7.34*	-43.2±13.3	-29.2±5.8*	-36.2±2.8

max. — maximum value; * — p <0.05

3.5.3. Connections between auditors' recognition of children's emotional state and the acoustic characteristics of test material

Based on correlation analysis (Spearman, p <0.05), the following connections are shown:

1) between auditors' recognition of the neutral state in children with ASD and F0 values of: phrases (-0.39), words (-0.44), stressed vowels (-0.48), unstressed vowels (-0. 32); maximum and minimum F0 values of: phrases (-0.49 — F0 max; -0.5 — F0 min), words (-0.47; -0.43), stressed vowels (-0.5; - 0.41) — signals assigned to the neutral state have lower F0 values compared to signals assigned to the states of comfort and discomfort;

2) between auditors' recognition of the neutral state in children with ASD and F0 variability of phrases (-0.27): signals assigned to the neutral state have lower values of F0 variability;

3) between auditors' recognition of the state of discomfort in children with ASD and the F0 values of: phrases (0.41), words (0.42), stressed vowels (0.32), unstressed vowels (0.36); maximum and minimum F0 values of: phrases (0.32 — F0 max; 0.31 — F0 min), words (0.3; 0.29), stressed vowels (0.29; 0.31) — signals related to a state of discomfort have higher F0 values;

4) between auditors' recognition of the state of comfort in TD children and F0 values of: phrases (0.34), words (0.34), stressed vowels (0.3), unstressed vowels (0.31); maximum and minimum F0 values of: phrases (0.38 — F0 max; 0.37 — F0 min), words (0.26; 0.37), stressed vowels (0.29; 0.3) — signals related to the state of comfort have higher F0 values;

5) between auditors' recognition of the neutral state in TD children and F0 values of: phrases (-0.32), words (-0.36), stressed vowels (-0.27), unstressed vowels (-0.36); maximum and minimum F0 values of: phrases (-0.34 — F0 max; -0.32 — F0 min), words (-0.27; -0.36); minimum F0 values of stressed vowels (-0.3) — signals classified as a neutral state have lower F0 values;

6) between auditors' recognition of the neutral state in TD children and the F0 variability values of phrases (-0.26): signals assigned to the neutral state have lower values of F0 variability.

Based on multiple regression analysis, connections are shown between children's speech rate, duration, F0 values of phrases and vowels, F3 values and recognition of their emotional state by auditors (Table 42).

Table 42. Connections between auditors' recognition of children's emotional state and acoustic characteristics of test material. Regression analysis data

R^2	F	Independent variables	β	SE β	B	SE B	t	p
ASD								
Dependent variable: comfort state							t (21)	
0.811	(17,21) 5.312	Speech rate	-0.669	0.188	-0.261	0.073	-0.247	0.0002
Dependent variable: neutral state							t (24)	
0.697	(14, 24) 3.935	F0 of phrase	-0.431	0.191	-0.002	0.001	-2.264	0.0016
		F0 of unstressed vowel	-0.942	0.443	-0.003	0.001	-2.125	0.0016
Dependent variable: discomfort state							t (24)	
0.679	(14, 24) 3.618	Phrase duration	-0.641	0.276	-0.001	0.001	-2.321	0.0028
		Minimum F0 of phrase	-0.569	0.273	-0.004	0.002	-2.083	0.0028
TD								
Dependent variable: comfort state							t (13)	
0.927	(26, 13) 6.346	F3 of stressed vowel	0.575	0.244	0.001	0.001	2.356	0.0006
		F0 of phrase	-0.815	0.336	-0.005	0.002	-2.429	0.0006
Dependent variable: discomfort state							t (16)	
0.909	(23, 13) 6.97	Phrase duration	-0.692	0.169	-0.001	0.001	-4.089	0.0001
		Stressed vowel duration	0.486	0.141	0.001	0.001	3.461	0.0001

R^2 is the square of the correlation coefficient (R); SE is standard error; B is regression coefficient

Auditors classify signals of children with ASD with a slower speech rate as a state of comfort. Auditors classify the signals of children with ASD with lower F0 values of phrases and unstressed vowels as a neutral state. Auditors classify signals from children with ASD with a shorter duration and low minimum F0 values of phrase as a state of discomfort.

TD children's signals, classified by auditors as a state of comfort, are characterized by low F0 values of phrases and high F3 values of stressed vowels. Auditors classify signals of children with a shorter phrase duration and a longer duration of stressed vowels as a state of discomfort.

The correlation is shown between CARS scores and:

1) F0 intensity of unstressed vowels (0.64) in signals reflecting the state of comfort;

2) maximum F0 values of phrases (0.58), words (0.53), stressed vowels (0.5); minimum F0 values of phrases (0.52) and words (0.52) in signals reflecting the neutral state;

3) duration of unstressed vowels in signals reflecting the state of discomfort (0.65).

4. DISCUSSION

The study revealed the influence of physiological and psychophysiological characteristics of auditors on the features of their recognition of information contained in children's speech signals.

It has been shown that adults are able to recognize the lexical meaning of words of children with ASD and TD children correctly; still, adults determine the lexical meaning of words of children with ASD worse compared to the words of TD children. The age of children with ASD and TD children does not influence the probability of recognition the lexical meaning of words by auditors. The obtained data clarify the results of previous studies conducted on speech material of children with ASD aged 5−14 years (Lyakso et. al, 2017) and 4−16 years (Lyakso et al., 2021). However, these works show a lower number of correct answers from auditors who listened to the tests, which may be due to different organization of test sequences, to the complexity of the test material and to varying severity of autistic disorders in children whose speech material was used for perceptual experiments, because in children with ASD the features of the manifestation of speech disorders are extremely individual.

Recognition of the lexical meaning of words in children with ASD is influenced by the gender of the auditor: female auditors are better at identifying the meaning of children’s words than male auditors, this is probably due to the fact that women have more experience communicating with children than men and therefore understand better their speech. The work shows that older auditors are better at recognizing the lexical meaning of words in children with ASD, which may also be due to more experience communicating with children.

Auditor hearing thresholds influence the recognition of the lexical meaning of words: auditors with low hearing thresholds are better able to recognize the meaning of words in TD children. The results obtained complement the data described in (Lyakso et al., 2017), which showed that auditors with low hearing thresholds are better at recognizing the lexical meaning of words in children with ASD.

High level of anxiety influences the word recognition in children with ASD negatively. In a study examining the effects of anxiety on word perception (Mattys et al., 2013), auditors with high and low levels of anxiety were asked to distinguish similar-sounding words, and auditors with high levels of anxiety also performed worse on this task. It is likely that auditors with high levels of anxiety may have reduced attention span (Pacheco-Unguetti et al., 2010; Najmi et al., 2011), and this leads to them being less able to recognize the lexical meaning of children's words during listening tests, where a response is required immediately after the signal is presented.

Adults are able to recognize the psychoneurological state of children from speech signals correctly. Adults recognize the state of TD children better than the state of children with ASD. The age of children does not affect the recognition of their psychoneurological state by adults. Recognition of the psychoneurological state of children is influenced by the organization of the test material: auditors recognize the state of children better by phrases than by individual words. The obtained data are consistent with the results of studies conducted on speech material of children with ASD aged 11−12 years (Frolova et al., 2019) and 4−16 years (Lyakso et al., 2021; Lyakso et al., 2021; Lyakso et al., 2022).

Speech signals of children with ASD, recognized by auditors as belonging to children with atypical development, are characterized by high F0 values and its variability in phrases, words, stressed vowels, and low speech rate. These acoustic characteristics are distinctive features of the speech of children with ASD (Bonneh et al., 2010; Sharda et al., 2010; Lyakso et al. 2017; Redford et al., 2017; Patel et al., 2020). Auditors also classified speech signals of children with ASD with a small number of words in a phrase as “atypical development.” This may indicate that auditors, when recognizing a child’s psychoneurological state, rely not only on the characteristics of his voice, but also on the grammatical structure of the utterance.

Male auditors are better at recognizing the psychoneurological state of children with ASD. It is likely that men have less experience communicating with children and are less familiar with the peculiarities of children’s speech development and, therefore, have increased demands on children’s speech, classifying any speech signals with impaired sound pronunciation as “atypical development.” Auditors with domestic and professional experience of interacting with children recognize the psychoneurological state of children better. The obtained data expand the results of the work (Frolova et al., 2019), in which, using the speech material of children with ASD, Down syndrome and TD children aged 11–12 years, it was shown that auditors with experience of working with children recognize their psychoneurological state better compared to auditors without experience.

Auditors with low hearing thresholds (0−10 dB) recognize the psychoneurological state of children with ASD better. This complements the data of the work (Grigoriev, Gorodny, 2020), where it was shown that hearing thresholds influence the correct recognition of the state of children with ASD, Down syndrome and TD children aged 5−7 years. Auditors with a dominant left hemisphere recognize the state of children with ASD in speech better; this may be due to the fact that the left hemisphere is believed to carry out segment-by-segment analysis of the speech signal, isolating and identifying speech sounds (Galunov et al., 1986). It can be assumed that auditors rely on the articulatory characteristics of children, when recognizing the speech material of children with ASD as belonging to children with atypical development.

Auditors with high level of anxiety are worse at recognizing the psychoneurological state of children with ASD and TD children compared to auditors with low and moderate levels of anxiety, which may also be associated with a decrease in concentration during test tasks.

It has been shown that adults are able to recognize the emotional state of children with ASD and TD children correctly. In children with ASD, adults are better at recognizing states of comfort and discomfort than neutral states; in TD children, adults recognize states of comfort and neutral better than states of discomfort. Adults are better at determining the emotional state of children with TD than children with ASD. This is not entirely consistent with the data obtained in (Lyakso et al., 2016). According to this work, adults better recognize the emotional state of children with ASD compared to TD children, but the work used speech material from TD children aged 4–7 years. Probably, when recognizing the emotional state of older children with TD, auditors focus more on the meaning of the statement than on the prosodic characteristics of speech, while the meaning of the statements of children with ASD is not always clear.

Male auditors are better at recognizing the discomfort state of children with ASD than female auditors; when recognizing other emotional states of children with ASD and the emotional state of TD children, no differences related to the gender of the auditor were identified. In works that study the influence of the auditor’s gender on the recognition of emotions, the data are contradictory. Studies (Scherer, Scherer, 2011; Lausen, Schlacht, 2018) show that women are better at recognizing emotional states in general from speech signals, but there are no differences when recognizing specific emotional states. Another study (Paulmann et al., 2008) showed no gender differences in emotion recognition. However, in these works, professional actor's speech was used as test material, and the task for auditors included assigning speech signals to one of seven categories: joy, sadness, fear, anger, disgust, surprise, neutral state.

Auditors who have domestic and professional experience of interacting with children recognize the emotional state of children with ASD and TD children better, which confirms the data obtained in the study (Lyakso et al., 2021). Younger auditors are better at recognizing the emotional state of children with disabilities; no differences were found in recognizing the emotional state of children with ASD. The study (Paulmann et al., 2008) showed using adult actors' speech that young auditors were better at recognizing emotional states compared to middle-aged auditors.

The influence of the leading hemisphere in speech on the recognition of the emotional state of children is revealed: auditors with the leading left hemisphere are better at recognizing the emotional state of TD children, but not at that of children with ASD. It is likely that, when determining the emotional state of TD children, auditors rely primarily on the semantics of the utterance, while the meaning of the phrases of children with ASD may not always be clear outside the context of the situation.

Auditors with a high level of anxiety are worse at recognizing the state of comfort and neutral in TD children, but better at recognizing the state of discomfort. Other studies of the influence of anxiety on emotion recognition have obtained similar data: people with high levels of anxiety are worse at recognizing positive emotional manifestations, and better at recognizing negative ones, both in facial expression (Kang et al., 2019) and in speech signals (Koizumi et al. , 2011; Heffer et al., 2022), however, there is evidence that high level of anxiety negatively affects the accuracy of emotional state recognition from speech in general (Tseng et al., 2017).

Thus, the study shows the influence of the auditor's age, gender, domestic and professional experience on their recognition of the lexical meaning of children's words, the psychoneurological and emotional state of children with ASD and TD children. It is shown that the hearing thresholds and the dominant hemisphere in auditor's speech influence the recognition of the lexical meaning of words in TD children and the psychoneurological state of children with ASD. Auditors with a high level of anxiety are worse at recognizing the lexical meaning of words and the psychoneurological state of children, but better at recognizing the state of discomfort in TD children. The study was the first to analyze the relationship between the level of anxiety, the age of auditors and their ability to recognize information contained in children's speech signals.

CONCLUSION

1. It has been shown that adults recognize correctly the lexical meaning of 65% of words of children with ASD and 84% of words of TD children. The meaning of words of children with ASD is better recognized by female auditors than by male auditors. Auditors in the older age group recognize the lexical meaning of words in children with ASD better. Auditors with low hearing thresholds recognize the lexical meaning of TD children's words better. The probability of recognizing the lexical meaning of children's words is associated with acoustic characteristics: the duration of the stressed vowel and the values of fundamental frequency of words and vowels.

2. It has been shown that the accuracy of adults' recognition of the psychoneurological state of children is higher for phrases than for individual words. Adults recognize the state of TD children better than the state of children with ASD. Auditors with domestic and professional experience interacting with children are better able to recognize the psychoneurological state of children with ASD and TD children. Auditors with high level of anxiety recognize the psychoneurological state of children worse. The psychoneurological state of children with ASD is better determined by male auditors, by auditors with the left dominant hemisphere for speech, and with low hearing thresholds. Speech signals classified by auditors as belonging to children with atypical development are represented by single words or simple phrases and are characterized by high F0 values and its variability, high values of vowel duration, and low speech rate.

3. It has been shown that the accuracy of recognition by adults of the emotional state of children with ASD is maximum for states of comfort and discomfort, and for TD children — for states of comfort and neutral. The accuracy of recognizing the emotional state of children with ASD is lower than that of TD children. Female auditors and auditors with experience of interacting with children recognize the emotional state of children with ASD better. Auditors with low hearing thresholds recognize the state of comfort in children with ASD better. Older auditors are worse at identifying the state of discomfort in TD children. Auditors with a dominant left hemisphere for speech recognize the emotional state of TD children better. Auditors with a high level of anxiety are worse at recognizing the state of comfort and neutral TD of children, and better at recognizing the state of discomfort. Speech signals of children, reflecting a state of comfort and discomfort, differ from signals reflecting a neutral state in the duration of phrases, words and vowels, the values and variability of fundamental frequency of phrases, words and vowels.

Thus, the work shows the ability of adults to recognize the lexical meaning of words in TD children and children with ASD, as well as the psychoneurological and emotional state of children.

It was found that female auditors recognize the lexical meaning of words of children with ASD better, male auditors recognize the psychoneurological state of children with ASD and the state of discomfort of children with ASD better. Older auditors are better at recognizing the lexical meaning of words in children with ASD and worse at recognizing the emotional state of children with TD. Auditors with domestic and professional experience of interacting with children recognize the psychoneurological and emotional state of children better. Auditors with low hearing thresholds recognize the lexical meaning of words of TD children and the psychoneurological state of children with ASD better. Auditors with a dominant left hemisphere determine the psychoneurological state of children with ASD by speech better. Auditors with a high level of anxiety recognize the lexical meaning of words of children with ASD, the psychoneurological state of children with ASD and children with TD worse; they are worse at recognizing the state of comfort and neutral state TR of children worse, but better when it is the state of discomfort.

LIST OF ABBREVIATIONS

LPC — lateral preference coefficient

PFLA — profile of functional lateral asymmetry

ASD — autism spectrum disorders

TD — typically developing

F0 — fundamental frequency

HR — hearth rate

ECG — electrocardiogram

VAI — vowel articulation index

UAR — unweighted average recall

REFERENCES

1. Batuev A.S. Fiziologiya vysshei nervnoi deyatel'nosti i sensornykh sistem: uchebnik dlya vuzov / A.S. Batuev. — 3-e izd. — SPb: Piter, 2009. — 317 s.
2. Belyakova L.I. Logopediya. Dizartriya / L.I. Belyakova, N.N. Voloskova. — M.: Gumanitarnyi izdatel'skii tsentr VLADOS, 2009. — 287 s.
3. Belyakova L.I. Mekhanizm rechedvigatel'nogo akta v svete logopedicheskogo analiza / L.I. Belyakova, Yu.O. Filatova // Rech' rebenka: problemy i resheniya / pod red. T.N. Ushakovoi. — M.: Institut psikhologii RAN, 2008. — S. 40–54.
4. Bondarko L.V. Fonetika sovremennogo russkogo yazyka / L.V. Bondarko. — SPb: Izdatel'stvo SPbGU, 1998. — 276 s.
5. Vartanyan I.A. Neirofiziologicheskie osnovy rechevoi deyatel'nosti / I.A. Vartanyan // Mekhanizmy deyatel'nosti mozga cheloveka. Chast' 1. Neirofiziologiya cheloveka / Pod red. N.P. Bekhterevoi. — L.: Nauka, 1988. — S. 608–667.
6. Galunov V.I. Vzaimodeistvie dvukh polusharii v protsesse obrabotki rechevoi informatsii / V.I. Galunov, I.V. Koroleva, G.G. Shurgaya // Akustika rechi i slukha / Pod red. L.A. Chistovich. — L.: Nauka, 1986. — S. 127–143.
7. Grigor'ev A.S., Gorodnyi V.A. Vliyanie psikhofiziologicheskikh pokazatelei ispytuemykh na uspeshnost' raspoznavaniya rechevogo materiala v pertseptivnom eksperimente // Trudy Vserossiiskoi akusticheskoi konferentsii: Materialy III Vserossiiskoi konferentsii, Sankt-Peterburg, 21–25 sentyabrya 2020 goda. 2020. — S. 708–712.
8. Grigor'ev A.S., Lyakso E.E. Akusticheskie kharakteristiki glasnykh iz slov detei 14–16 let // Uchenye zapiski fizicheskogo fakul'teta Moskovskogo Universiteta. 2017. — № 5.
9. Grigor'ev A.S., Lyakso E.E. Slukhovoe vospriyatie slov detei 5–8 let // Sensornye sistemy. 2014. — № 3. — S. 28–35.
10. Derkach M.F. Dinamicheskie spektry rechevykh signalov / M.F. Derkach, R.Ya. Gumetskii, B.M. Guba, M.E. Chaban. — L'vov: Vysshaya shkola, 1983 — 168 s.
11. Duryagin P.V., Fokina M.V. Vospriyatie prosodii bezlichnykh fraz so znacheniem otkaza nositelyami russkogo yazyka: lingvisticheskii i paralingvisticheskii aspekty // Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya. 2021. — №69. — S. 94–121.
12. Zinder L.R. Obshchaya fonetika / L.R. Zinder. — M.: Vysshaya shkola, 1979. — 312 s.
13. Kodzasov S.V., Krivnova O.F. Obshchaya fonetika / S.V. Kodzasov, O.F. Krivnova. — M.: Izd-vo Rossiiskogo gosudarstvennogo gumanitarnogo universiteta, 2001. — 592 s.
14. Krivnova O.F., Knyazev S.V., Smirnova O.S. Intonatsionnoe chlenenie i segmentiruyushaya sila slovorazdelov v zvuchashchem tekste (dannye pertseptivnogo

eksperimenta) // Trudy instituta russkogo yazyka im. V.V. Vinogradova. 2018. — № 17. — S. 128–140.

15. Lebedinskaya K.S., Nikol'skaya O.S. Diagnostika rannego detskogo autizma: nachal'nye proyavleniya / K.S. Lebedinskaya, O.S. Nikol'skaya. — M.: Prosveshchenie, 1991. — 53 s.
16. L'vov M.R. Osnovy teorii rechi / M.R. L'vov. — M.: Izdatel'skii tsentr «Akademiya», 2002 — 248 s.
17. Lyakso E.E. Neverbal'nye vokalizatsii mladentsev na rannikh etapakh "dorechevogo" perioda razvitiya // Novosti otolaringologii i logopatologii. 1998. — №4 (16). — C. 23–31.
18. Lyakso E.E., Frolova O.V., Smirnov A.G., Kurazhova A.V., Gaikova Yu.S., Bednaya E.D., Grigor'ev A.S. Uroven' rechevogo razvitiya detei na etape formirovaniya navyka chteniya // Psikhologicheskii zhurnal. 2012. — T. 33. № 1. — S. 73–87.
19. Lyakso E.E. Golosovoi portret rebenka s tipichnym i atipichnym razvitiem / E.E. Lyakso, O.V. Frolova, S.V. Grechanyi, Yu.N. Matveev, O.V. Verkholyak, A.A. Karpov; pod red. E. E. Lyakso, O. V. Frolovoi. — SPb: Izdatel'sko-poligraficheskaya assotsiatsiya vysshikh uchebnykh zavedenii, 2020. — 204 s.
20. Lyakso E.E., Grigor'ev A.S. Dinamika dlitel'nosti i chastotnykh kharakteristik glasnykh na protyazhenii pervykh semi let zhizni detei // Rossiiskii fiziologicheskii zhurnal im. I.M. Sechenova. 2013. — №9 (99). — S. 1097–1110.
21. Lyakso E.E., Ogorodnikova E.A., Alekseev N.P. Psikhofiziologiya slukhovogo vospriyatiya: uchebnoe posobie / E.E. Lyakso, E.A. Ogorodnikova, N.P. Alekseev. — SPb.: SPbGIPSR, 2013. — 111 s.
22. Lyakso E.E., Ogorodnikova E.A., Alekseev N.P. Fiziologiya slukha i rechi: uchebno-metodicheskoe posobie / E.E. Lyakso, E.A. Ogorodnikova, N.P. Alekseev. — SPb: Rech', 2012. — 168 s.
23. Lyakso E.E., Stolyarova E.I. Spetsifika realizatsii rechevykh navykov 4–5-letnikh detei v dialoge // Psikhologicheskii zhurnal. 2008. — T. 29. № 3. — S. 48–57.
24. Lyakso E.E., Frolova O.V., Grigor'ev A.S., Gorodnyi V.A. Osobennosti rechi, fiziologicheskie i psikhofiziologicheskie kharakteristiki detei s rasstroistvami autisticheskogo spektra // Rossiiskii fiziologicheskii zhurnal im. I.M. Sechenova. 2018. — № 5 (104). — S. 600–615.
25. Lyakso E.E., Frolova O.V., Grigor'ev A.S., Sokolova V.D., Yarotskaya K.A. Raspoznavanie vzroslymi emotsional'nogo sostoyaniya tipichno razvivayushchikhsya

detei i detei s rasstroistvami autisticheskogo spektra // Rossiiskii fiziologicheskii zhurnal im. I.M. Sechenova. 2016. — № 6 (102). — S. 729–741.

26. Lyakso E.E., Frolova O.V., Nikolaev A.S., Grigor'ev A.S. Pertseptivnyi analiz vzroslymi rechi detei s rasstroistvami autisticheskogo spektra, sindrom Dauna, umstvennoi otstalost'yu // Rossiiskii fiziologicheskii zhurnal im. I.M. Sechenova. 2021. — №11 (107). — S. 1395–1407.
27. Lyakso E.E., Frolova O.V., Nikolaev A.S, Gorodnyi V.A., Grigor'ev A.S., Kurazhova A.V. Pol, vozrast, sostoyanie rebenka: pertseptivnoe issledovanie rechi detei 8–9 let tipichno razvivayushchikhsya, s rasstroistvami autisticheskogo spektra i sindromom Dauna // Teoreticheskaya i prikladnaya lingvistika. 2020. — T. 6. № 3. — S. 61–76.
28. Morozov V.P. Voprosy funktsional'noi asimmetrii mozga: monografiya / V.P. Morozov, I.A. Vartanyan, V.I. Galunov; pod. red. V.P. Morozova. — L.: Nauka, 1988. — 135 s.
29. Neznanov N.G. Psikhiatriya / N.G. Neznanov, Yu.A. Aleksandrovskii, L.M. Bardenshtein, V.D. Vid, V.N. Krasnova, Yu.V. Popov. — M.: GEOTAR-Media, 2009. — 512 s.
30. Nikolaev A.S., Lyakso E.E. Raspoznavanie vzroslymi rechi detei 10-14 let s rasstroistvami autisticheskogo spektra // Zhurnal evolyutsionnoi biokhimii i fiziologii. 2020. Tom 56. № 7. S. 657.
31. Nikolaev A.S., Frolova O.V., Gorodnyi V.A., Lyakso E.E. Kharakteristika otvetnykh replik detei 5–11 let s rasstroistvami autisticheskogo spektra v dialogakh so vzroslymi // Voprosy psikholingvistiki. 2019. — № 4 (42). — S. 92–105.
32. Nikolaev A.S., Frolova O.V., Balyakova A.A., Lyakso E.E. Raspoznavanie vzroslymi znacheniya slov detei s rasstroistvami autisticheskogo spektra // Voprosy psikhicheskogo zdorov'ya detei i podrostkov. 2018. — T. 18. № 4. — S. 64–74.
33. Nikolaeva E.I. Levorukii rebenok: diagnostika, obuchenie, korrektsiya / E.I. Nikolaeva. — SPb: DETSTVO-PRESS, 2005 — 128 s.
34. Rozenberg G.S. Neskol'ko slov ob indekse raznoobraziya Simpsona // Samarskaya Luka: problemy regional'noi i global'noi ekologii. 2007. — T. 16. №3 (21). — S. 581–584.
35. Semyannikova A.A. Rasstroistva autisticheskogo spektra: klassifikatsii, opredelenie ponyatii, simptomy // Psikhologiya i pedagogika: metodika i problemy prakticheskogo primeneniya. 2013. — №32. — S. 67–71.
36. Fant G. Analiz i sintez rechi / G. Fant; per. s angl. V.S. Lozovskogo i N.V. Bakhmutovoi pod red. N.G. Zagoruiko. — Novosibirsk: Nauka, 1970 — 167 s.
37. Khanin Yu.L. Kratkoe rukovodstvo k primeneniyu shkaly reaktivnoi i lichnostnoi trevozhnosti Ch.D. Spilbergera / Yu.L. Khanin. — L.: LNIITEK, 1976. — 40 s.

38. Tseitlin S.N. Yazyk i rebenok. Lingvistika detskoi rechi: uchebnoe posobie / S.N. Tseitlin. — M.: VLADOS, 2000. — 240 s.
39. Chistovich L.A., Kozhevnikov V.A. Rech'. Artikulyatsiya i vospriyatie / L.A. Chistovich, V.A. Kozhevnikov. — M., L.: Nauka, 1965. — 241 s.
40. Yagunova E.V. Issledovanie pertseptivnoi ustoichivosti fonem kak elementov rechevoi tsepi // Chelovek govoryashchii: issledovaniya XXI veka / Pod red. L.A. Verbitskoi, N.K. Ivanovoi. — Ivanovo, 2012. — S. 67–75.
41. Attwood A.S., Penton-Voak I.S., Burton A.M., Munafò M.R. Acute Anxiety Impairs Accuracy in Identifying Photographed Faces // Psychological Science. 2013. — N. 24 (8). — P. 1591−1594.
42. Ballard K.J., Djaja D., Arciuli J., James D.J.H., van Doorn J. Developmental trajectory for production of prosody: lexical stress contrastivity in children 3 to 7 years and adults // Journal of Speech, Language, and Hearing Research. 2012. — V. 55. — P. 1716–1735.
43. Bonneh Y.S., Levanov Y., Dean-Pardo O., Lossos L., Adini Y. Abnormal speech spectrum and increased pitch variability in young autistic children // Frontiers in Human Neuroscience. 2011. — N. 4 (237). — P. 1–7.
44. Boorse J., Cola M., Plate S., Yankowitz L., Pandey J., Schultz R.T., Parish-Morris J. Linguistic markers of autism in girls: evidence of a “blended phenotype” during storytelling // Molecular Autism. 2019. — V. 10:14.
45. Boucher J. Research review: structural language in autistic spectrum disorder – characteristics and causes // Journal of Child Psychology and Psychiatry. 2012. — N. 53 (3). — P. 219–233.
46. Brehm S.B., Weinrich B.D., Sprouse D.S., May S.K., Hughes M.R. An examination of elicitation method on fundamental frequency and repeatability of average airflow measures in children age 4:0 – 5:11 years // Journal of Voice. 2012. — V. 26 (6). — P. 721–725.
47. Bruckert L., Liénard J.S., Lacroix A., Kreutzer M., Leboucher G. Women use voice parameters to assess men's characteristics // Proceedings of the Royal Society B: Biological Sciences. 2006. — N. 273. — P. 83–89.
48. Carl M., Kent R.D., Levy E.S., Whalen D.H. Vowel acoustics and speech intelligibility in young adults with Down syndrome // Journal of Speech, Language, and Hearing Research. 2020. — N. 63(3). — P. 674–687.
49. Carvalho G.M., Santos M.D., Pauna H.F., Guimarães A.C., Curi S.B., Jeronymo D., Porto P.R., Bianchini W.A., Castilho A.M., Sartorato E.L. Evaluation of Audiometric Thresholds and Speech Perception Sentence Test in Adults and Elderly After Cochlear Implantation // Current Aging Science. 2016. — N. 9 (3). — P. 224–228.

50. Chen F., Cheung C.Ch., Peng G. Linguistic tone and non-linguistic pitch imitation in children with autism spectrum disorders: A cross-linguistic investigation // Journal of Autism and Developmental Disorders. 2022. — N. 52 (5). — P. 2325–2343.
51. Cleland J., Gibbon F.E., Peppé S.J., O'Hare A., Rutherford M. Phonetic and phonological errors in children with high functioning autism and Asperger syndrome // International Journal of Speech-Language Pathology. 2010. — N. 12 (1). — P. 69–76.
52. de Jong T.J., Hakkesteegt M.M., van der Schroeff M.P., Vroegop J.L. Communicating Emotion: Vocal Expression of Linguistic and Emotional Prosody in Children With Mild to Profound Hearing Loss Compared With That of Normal Hearing Peers // Ear and Hearing. 2023. — V. 10.1097.
53. Diehl J., Paul R. Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems by children with autism spectrum disorders // Applied Psycholinguistics. 2013. — N. 34 (1). — P. 135–161.
54. Diehl J., Watson D., Bennetto L., McDonough J., Gunlogson C. An acoustic analysis of prosody in high-functioning autism // Applied Psycholinguistics. 2009. — N. 30. — P. 385–404.
55. Filipe M.G., Frota S., Castro S.L., Vicente S.G. Atypical prosody in Asperger syndrome: perceptual and acoustic measurements // Journal of Autism and Developmental Disorders. 2014. — N. 44 (8). — P. 1972–1981.
56. Frenkel T.I., Bar-Haim Y. Neural activation during the processing of ambiguous fearful facial expressions: an ERP study in anxious and nonanxious individuals // Biological Psychology. 2011. — N. 88 (2–3). — P. 188–195.
57. Friederici A.D. Towards a neural basis of auditory sentence processing // Trends in Cognitive Sciences. 2002. — N. 6 (2). — P. 78–84.
58. Frolova O.V., Grechanyi S.V., Kaliev A., Grigorev A.S., Gorodnyi V.A., Nikolaev A.S., Lyakso E.E. Perceptual features of speech in children with atypical development // Zhurnal evolyutsionnoi biokhimii i fiziologii. 2020. — Tom 56. № 7. — C. 753.
59. Frolova O., Gorodnyi V., Nikolaev A., Grigorev A., Grechanyi S., Lyakso E. Developmental disorders manifestation in the characteristics of the child’s Voice and speech: Perceptual and acoustic study // Lecture Notes in Computer Science. 2019. — V. 11658. — P. 103–112.
60. Frolova O., Nikolaev A., Grave P., Lyakso E. Speech Features of Children with Mild Intellectual Disabilities // Companion Publication of the 25th International Conference on Multimodal Interaction (ICMI '23 Companion). 2023. — P. 406–413.

61. Frosolini A., Fantin F., Tundo I., Pessot N., Badin G., Bartolotta P., Vedovelli L., Marioni G., de Filippis C. Voice Parameters in Children With Cochlear Implants: A Systematic Review and Meta-Analysis // Journal of voice. 2023. — S0892-1997(23)00021-8.
62. Fusaroli R., Lambrechts A., Bang D., Bowler D.M., Gaigg S.B. Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis // Autism Research. 2017. — N. 10 (3). — P. 384–407.
63. García-Guerrero M.A., Peña J., Zubiaurre-Elorza L., Benítez D., Hernández A.M., Oribe A., Ojeda, N. Voice, Body Cues and Facial Expression in Emotion Recognition of Spanish Children and Adolescents: The Validation of Bell-Lysaker Emotion Recognition Test // The Spanish journal of psychology. 2022. — V. 25: e7.
64. Gelfer M.P. Perceptual attributes of voice: Development and use of rating scales // Journal of Voice. 1988. — N. 2. — P. 320–326.
65. Goy H., Pichora-Fuller M.K., van Lieshout P. Effects of age on speech and voice quality ratings // Journal of the Acoustical Society of America. 2016. — N. 139 (4). — P. 1648.
66. Grigorev A., Frolova O., Lyakso E. Acoustic Features of Speech of Typically Developing Children Aged 5–16 Years // Artificial Intelligence and Natural Language. AINL 2018. Communications in Computer and Information Science. 2018. — V. 930.
67. Grigorev A., Kurazhova A., Kleshnev E., Nikolaev A., Frolova O., Lyakso E. An Electroglottographic Method for Assessing the Emotional State of the Speaker // Lecture Notes in Computer Science. 2022. — V. 13721. — P. 214–225.
68. Grosbras M.H., Ross P.D., Belin, P. Categorical emotion recognition from voice improves during childhood and adolescence // Scientific reports. 2018. — V. 8(1): 14791.
69. Grossman R.B., Edelson L.R., Tager-Flusberg H. Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism // Journal of Speech, Language, and Hearing Research. 2013. — V. 56 (3). — P. 1035–1044.
70. Hagberg B., Aicardi J., Dias K., Ramos O. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. // Annals of neurology. 1983. — V. 14. N. 4. — P. 471–479.
71. Hatch B., Iosif A.M., Chuang A., de la Paz L., Ozonoff S., Miller M. Longitudinal Differences in Response to Name Among Infants Developing ASD and Risk for ADHD // Journal of Autism and Developmental Disorders. 2021. — N. 51 (3). — P. 827–836.
72. Harrop C., Tomaszewski B., Putnam O., Klein C., Lamarche E., Klinger L. Are the diagnostic rates of autistic females increasing? An examination of state-wide trends // Journal of Child Psychology and Psychiatry — Published online January 5, 2024. doi:10.1111/jcpp.13939

73. Heffer N., Gradidge M., Karl A., Ashwin C., Petrini K. High trait anxiety enhances optimal integration of auditory and visual threat cues // Journal of Behavior Therapy and Experimental Psychiatry. 2022. — V. 74:101693.
74. Hoit J. Hixon T., Watson P., Morgan W. Speech breathing in children and adolescents // Journal of Speech, Language, and Hearing Research. 1990. — N. 33. — P. 51–69.
75. Hoppe U., Hast A., Hocke T. Speech Perception in Bilateral Hearing Aid Users with Different Grades of Asymmetric Hearing Loss // Frontiers in Neuroscience. 2022. — V. 15:715660.
76. Jiang X., Paulmann S., Robin J., Pell M.D. More than accuracy: Nonverbal dialects modulate the time course of vocal emotion recognition across cultures // Journal of experimental psychology. Human perception and performance. 2015. — N. 41(3). — P. 597–612.
77. Jones H.N., Crisp K.D., Kuchibhatla M., Mahler L., Risoli Th. Jr., Jones C.W., Kishnani P. Auditory-perceptual speech features in children with Down syndrome // American Journal on Intellectual and Developmental Disabilities. 2019. — N. 124 (4). — P. 324–338.
78. Juremi N.R., Zulkifley M.A., Hussain A., Zaki W. Inter-rater reliability of actual tagged emotion categories validation using Cohen’s Kappa coefficient // Journal of Theoretical and Applied Information Technology. 2017. — V. 95. N. 2. — P. 259–264.
79. Kang W., Kim G., Kim H., Lee S.H. The Influence of Anxiety on the Recognition of Facial Emotion Depends on the Emotion Category and Race of the Target Faces // Experimental neurobiology. 2019. — N. 28 (2). — P. 261–269.
80. Kanner L. Autistic Disturbances of Affective Contact // Nervous Child. 1943. — N. 2. — P. 217–250.
81. Kaya H., Salah A.A., Karpov A., Frolova O., Grigorev A., Lyakso E. Emotion, age, and gender classification in children’s speech by humans and machines. Computer Speech & Language. 2017. — N. 46. — P. 268–283.
82. Kelley E., Paul J.J., Fein D., Naigles L.R. Residual language deficits in optimal outcome children with a history of autism // Journal of Autism and Developmental Disorders. 2006. — N. 36 (6). — P. 807–828.
83. Kent R.D. Perceptual sensorimotor speech examination for motor speech disorders // Clinical management of sensorimotor speech disorders. — Thieme, 2009. — P.19–29.
84. Kent R.D, Vorperian H.K. Speech impairment in Down syndrome: a review // Journal of Speech, Language, and Hearing Research. 2013. — N. 56. — P. 178–210.

85. Kimura D. Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology. 1961. — N. 15 (3). P. 166–171.
86. Koizumi A., Tanaka A., Imai H., Hiramatsu S., Hiramoto E., Sato T., de Gelder B. The effects of anxiety on the interpretation of emotion in the face-voice pairs // Experimental Brain Research. 2011. — N. 213 (2−3). — P. 275−282.
87. Kuhl P. K., Meltzoff A.N. Infant vocalizations in response to speech. Vocal imitation and developmental change // Journal of the Acoustical Society of America. 1996. — V. 100. — N. 4. — P. 2425–2438.
88. Landis J.R., Koch G.G. The measurement of observer agreement for categorical data // Biometrics. 1977. — V. 33. N. 1. — P. 159–174.
89. Lausen A., Schacht A. Gender differences in the recognition of vocal emotions // Frontiers in Psychology. 2018. — N. 9. — P. 882.
90. Lee S. Iverson G.K. The development of monophthongal vowels in Korean: age and sex differences // Clinical Linguistics & Phonetics. 2008. — V. 22 (7). — P. 523–536.
91. Lee S. Potamianos A., Narayanan S. Acoustics of children’s speech: Developmental changes of temporal and spectral parameters // The Journal of the Acoustical Society of America. 1999. — V. 105. — P. 1455–1468.
92. Liberman A., Cooper F., Shankweiler D., Studdert-Kennedy M. Perception of the speech code // Psychological Review. 1967. — V. 74. — P. 31–61.
93. Loveall S.J., Hawthorne K., Gaines M. A meta-analysis of prosody in autism, Williams syndrome, and Down syndrome // Journal of communication disorders. 2021. — V. 89: 106055.
94. Lyakso E.E., Frolova O.V. Early development indicators predict speech features of autistic children // ICMI '20 Companion: Companion Publication of the 2020 International Conference on Multimodal Interaction. 2020. — P. 514–521.
95. Lyakso E., Frolova O., Grigorev A. Perception and acoustic features of speech of children with autism spectrum disorders // Lecture Notes in Computer Science. 2017. — V. 10458. — P. 602–612.
96. Lyakso E., Frolova O., Grigorev A., Gorodnyi V., Nikolaev A. Strategies of Speech Interaction between Adults and Preschool Children with Typical and Atypical Development // Behavioral Sciences. 2019. — V. 9 (12): 159.
97. Lyakso E.E, Frolova O.V., Grigorev A., Gorodnyi V., Nikolaev A., Kurazhova A.V. Speech Features of 13–15-Year-Old Children with Autism Spectrum Disorders // Lecture Notes in Computer Science. 2020. — V. 12335. — P. 291−303.

98. Lyakso E., Frolova O., Grigorev A., Gorodnyi V., Nikolaev A., Matveev Y. Speech Features of Adults with Autism Spectrum Disorders and Mental Retardation // Lecture Notes in Computer Science. 2018. — V. 11096. — P. 357−366.

99. Lyakso E., Frolova O. Gorodniy, Grigovev A., Nikolaev, A., Matveev, Y. Reflection of the emotional state in the characteristics of voice and speech of children with Down syndrome // 10th International Conference on Speech Technology and Human-Computer Dialogue, SpeD. 2019. — P. 1−6.

100. Lyakso E., Frolova O., Karpov A. A New Method for Collection and Annotation of Speech Data of Atypically Developing Children // Proceedings — 2018 International Conference on Sensor Networks and Signal Processing. SNSP 2018. 2019. — P. 175–180.

101. Lyakso E., Frolova O., Matveev A., Matveev Y., Grigorev A., Makhnytkina O., Nersisson R. Recognition of the Emotional State of Children with Down Syndrome by Video, Audio and Text Modalities: Human and Automatic // Speech and Computer: 24th International Conference, SPECOM 2022. Proceedings. 2022. — P. 438–450.

102. Lyakso E., Frolova O., Nikolaev A. Voice and speech features as a diagnostic symptom // Psychological Applications and Trends 2021. 2021. — P. 359–363.

103. Lyakso E., Frolova O., Nikolaev A., Grechanyi S., Matveev A., Matveev Y., Makhnytkina O., Nersisson R. Emotional state of children with ASD and Intellectual Disabilities: perceptual experiment and automatic recognition by video, audio and text modalities // Lecture Notes in Computer Science. 2023. — V. 14338. — P. 535–549.

104. Lyakso E.E., Frolova O.V., Nikolaev A.S., Grigorev A.S. Perceptual Analysis by Adults of the Speech of Children with Autism Spectrum Disorders, Down's Syndrome, and Intellectual Disabilities // Neuroscience and Behavioral Physiology. 2022. — V. 52 (4). — P. 582–590.

105. Lyakso E., Frolova O., Nikolaev A., Kleshnev E., Grave P., Ilyas A., Makhnytkina O., Nersisson R., Mekala A.M., Varalakshmi M. Recognition of the Emotional State of Children by Video and Audio Modalities by Indian and Russian Experts // Lecture Notes in Computer Science. 2023. — V. 14338. — P. 469–482.

106. Lyakso E., Frolova O., Nersisson R., Mekala A.M. Child's Emotional Speech Classification by Human Across Two Languages: Russian & Tamil // Lecture Notes in Computer Science. 2021. — V. 12997. — P. 384–396.

107. Lyakso E., Gromova A. The acoustic characteristics of Russian vowels in children of 4 and 5 years of age // Psychology of Language and Communication. 2005. — V. 9. N. 2. — P. 5–14.

108. Lyakso E., Nersisson R., Frolova O., Mekala M.A. The children's emotional speech recognition by adults: Cross-cultural study on Russian and Tamil language // PLoS ONE. 2023. — N. 18 (2): e0272837.

109. Madison A., Vasey M., Emery C.F., Kiecolt-Glaser J.K. Social anxiety symptoms, heart rate variability, and vocal emotion recognition in women: evidence for parasympathetically-mediated positivity bias // Anxiety, stress, and coping. 2021. — N. 34 (3). — P. 243–257.

110. Makhnytkina O., Grigorev A., Nikolaev A. 2021. Analysis of Dialogues of Typically Developing Children, Children with Down Syndrome and ASD Using Machine Learning Methods // Lecture Notes in Computer Science. 2021. — V. 12997. — P. 397–406.

111. Markova D., Richer L., Pangelian M., Schwartz D., Leonard G., Perron M., Pike G.B., Veilette S., Chakravarty M., Pausova Z., Paus T. Age- and sex-related variations in vocal-tract morphology and voice acoustic during adolescence // Hormones and Behavior. 2016. — P. 84–96.

112. Mattys S.L., Seymour F., Attwood A.S., Munafò M.R. Effects of acute anxiety induction on speech perception: Are anxious listeners distracted listeners? // Psychological Science. 2013. — N. 24 (8). — P. 1606–1608.

113. Matveev A., Matveev Y., Frolova O., Nikolaev A., Lyakso E. A Neural Network Architecture for Children's Audio–Visual Emotion Recognition // Mathematics. 2023. — V. 11 (22): 4573.

114. Matveev Y., Lyakso E., Matveev A., Frolova O., Grigorev A., Nikolaev A. Automatic Classification of the Emotional State of Atypically Developing Children // Proceedings of the 24 th International Congress of Acoustics. October 24 to 28. 2022 in Gyeongju, Korea (ABS-0338). 2022. — P. 1–7.

115. Matveev Y., Matveev A., Frolova O., Lyakso E., Nersisson R. Automatic Speech Emotion Recognition of Younger School Age Children // Mathematics. 2022. — V. 10 (14): 2373.

116. Mazzaggio. G., Shield. A. The Production of Pronouns and Verb Inflections by Italian Children with ASD: A New Dataset in a Null Subject Language // Journal of Autism and Developmental Disorders. 2020. — N. 50. — P. 1425–1433.

117. McGregor K.K., Berns A.J., Owen A.J., Michels S.A., Duff D., Bahnsen A.J., Lloyd M. Associations between syntax and the lexicon among children with or without ASD and language impairment // Journal of Autism and Developmental Disorders. 2012. — N. 42 (1). — P. 35–47.

118. Moura C.P., Cunha L.M., Vilarinho H., Cunha M.J., Freitas D., Palha M., Pais-Clemente M. Voice parameters in children with Down syndrome // Journal of Voice. 2008. — N. 22. — P. 34–42.

119. Najmi S., Kuckertz J.M., Amir N. Attentional impairment in anxiety: inefficiency in expanding the scope of attention // Depression and anxiety. 2012. — N. 29 (3). — P. 243−249.

120. Nicholas J.S., Charles J.M., Carpenter L.A., King L.B., Jenner W., Spratt E.G. Prevalence and characteristics of children with autism spectrum disorders // Annals of Epidemiology. 2008. — N. 18 (2). — P. 130–136.

121. O'Leary D., Lee A., O'Toole C., Gibbon F. Perceptual and acoustic evaluation of speech production in Down syndrome: A case series // Clinical Linguistics & Phonetics. 2020. — N. 34 (1–2). — P. 72–91.

122. Pacheco-Unguetti A.P, Acosta A., Callejas A., Lupiáñez J. Attention and anxiety: different attentional functioning under state and trait anxiety // Psychological Science. 2010. — N. 21 (2). — P. 298−304.

123. Patel S.P, Nayar K., Martin G.E., Franich K., Crawford S., Diehl J.J., Losh M. An Acoustic Characterization of Prosodic Differences in Autism Spectrum Disorder and First-Degree Relatives // Journal of Autism and Developmental Disorders. 2020. — N. 50 (8). — P. 3032–3045.

124. Paul R., Augustyn A., Klin A., Volkmar F. Perception and production of prosody by speakers with autism spectrum disorders // Journal of Autism and Developmental Disorders. 2005. — N. 35. — P. 205–220.

125. Paulmann S., Pell M.D., Kotz S.A. How aging affects the recognition of emotional speech // Brain and Language. 2008. — N. 104. — P. 262–269.

126. Perrachione T.K., Chiao J.Y., Wong P.C.M. Asymmetriccultural effects on perceptual expertise underlie anown-race bias for voices // Cognition. 2010. — N. 114. — P. 42–55.

127. Purnell T., Idsardi W., Baugh J. Perceptual and phonetic experiments on American English dialect identification // Journal of Language and Social Psychology. 1999. — N. 18. — P. 10–30.

128. Rapin I., Dunn M.A., Allen D.A., Stevens M.C., Fein D. Subtypes of language disorders in school-age children with autism // Developmental Neuropsychology. 2009. — N. 34 (1). — P. 66–84.

129. Redford M.A, Kapatsinski V., Cornell-Fabiano J. Lay Listener Classification and Evaluation of Typical and Atypical Children's Speech // Language and Speech. 2018. — N. 61 (2). — P. 277–302.

130. Rockwell P. Lower, slower, louder: Vocal cues of sarcasm // Journal of Psycholinguistic Research. 2000. — V. 29. — P. 483−495.

131. Rossi N.F., Giacheti C.M. Association between speech-language, general cognitive functioning and behaviour problems in individuals with Williams syndrome // Journal of intellectual disability research. 2017. — V. 61(7). — P. 707–718.

132. Roy N., Nissen S.L., Dromey C., Sapir S. Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy // Journal of Communication Disorders. 2009. — N. 42 (2). — P. 124–135.

133. Sanders L.D., Neville H.J., Woldorff M.G. Speech segmentation by native and non-native speakers: the use of lexical, syntactic, and stress-pattern cues // Journal of Speech, Language, and Hearing Research. 2002. — N. 45(3). — P. 519−30.

134. Sharrer L., Christman U. Voice Modulations in German Ironic Speech // Language and Speech. 2011. — V. 54. N. 4. — P. 435−465.

135. Scherer K.R., Scherer U. Assessing the ability to recognize facial and vocal expressions of emotion: construction and validation of the emotion recognition index // Nonverbal Behavior. 2011. — N. 35. — P. 305–326.

136. Schoen E., Paul R., Chawarska K. Phonology and vocal behavior in toddlers with autism spectrum disorders // Autism Research. 2011. — N. 4 (3). — P. 177–188.

137. Schölderle T., Haas E., Ziegler W. Speech Naturalness in the Assessment of Childhood Dysarthria // American journal of speech-language pathology. 2023. — N. 32(4). — P. 1633–1643.

138. Schopler E., Reichler R.J., DeVellis R.F., Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS) // Journal of Autism and Developmental Disorders. 1980. — N. 10 (1). — P. 91–103.

139. Sharda M., Subhadra T.P., Sahaya S., Nagaraja Ch., Singh L., Mishra R., Sen A., Singhal N., Erickson D., Singh N. Sounds of melody — Pitch patterns of speech in autism. Neuroscience Letters. 2010. — N. 478 (1) — P. 42–45.

140. Shriberg L.S., Paul Rh., Black L.M., van Santenc J.P. The Hypothesis of apraxia of speech in children with autism spectrum disorder // Journal of Autism and Developmental Disorders. 2011. — N. 41 (4). — P. 405–426.

141. Smith S.L., Pichora-Fuller M.K., Wilson R.H., Macdonald E.N. Word recognition for temporally and spectrally distorted materials: the effects of age and hearing loss // Ear & Hearing. 2012. — N. 33 (3). — P. 349–366.

142. Spielberger C.D., Gorsuch R.L., Lushene R.E. STAI Manual for the State-Trait Anxiety Inventory. — Palo Alto: Consulting Psychologists Press, 1970.

143. Spille C., Kollmeier B., Meyer B.T. Comparing human and automatic speech recognition in simple and complex acoustic scenes // Comput. Speech Lang. 2018. — V. 52. C. — P. 123–140.

144. Tek S., Mesite L., Fein D., Naigles L. Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders // Journal of Autism and Developmental Disorders. 2014. — N. 44 (1). — P. 75–89.

145. Terzi A., Marinis T., Zafeiri A., Francis K. Subject and Object Pronouns in High-Functioning Children with ASD of a Null-Subject Language // Frontiers in Psychology. 2019. — V. 10:1301.

146. Tseng H.H., Huang Y.L., Chen J.T., Liang K.Y., Lin C.C., Chen S.H. Facial and prosodic emotion recognition in social anxiety disorder // Cognitive neuropsychiatry. 2017. — N. 22 (4). P. 331–345.

147. Verkhodanova V., Trckova D., Coler M., Lowie W. More than words: Cross-linguistic exploration of Parkinson's disease identification from speech // Lecture Notes in Computer Science. 2020. — V. 12335. — P. 613–623.

148. Volkmar F.R., Nelson D.S. Seizure disorders in autism // Journal of the American Academy of Child and Adolescent Psychiatry. 1990. — V. 29. — P. 127–129.

149. Vorperian H. Kent R. Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data // Journal of Speech, Language, and Hearing Research. 2007. — V. 50 (6). — P. 1510–1545.

150. Walenski M., Europa E., Caplan D., Thompson C.K. Neural networks for sentence comprehension and production: An ALE-based meta-analysis of neuroimaging studies // Human Brain Mapping. 2019. — N. 40 (8). — P. 2275–2304.

151. Ward G., Hirschberg J. Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation // Language. 1985. — V. 61. N. 4. — P. 747–76.

152. Watson L.R., Patten E., Baranek G.T., Poe M., Boyd B.A., Freuler A., Lorenzi, J. Differential associations between sensory response patterns and language, social, and communication measures in children with autism or other developmental disabilities // Journal of Speech, Language, and Hearing Research. 2011. — N. 54 (6). — P. 1562–76.

153. Wilson S.M., Saygin A.P., Sereno M.I., Iacoboni M. Listening to speech activates motor areas involved in speech production // Nature Neuroscience. 2004. — N. 7 (7). — P. 701–702.

154. Wing L. Asperger syndrome: a clinical account // Psychological Medicine. 1981. — N. 11(1). — P. 115–129.

155. Wing L. The Definition and Prevalence of Autism: A Review // European Child & Adolescent Psychiatry. 1993. — N. 2(1). — P. 61–74.

156. Wolk L., Brennan C. Phonological difficulties in children with autism: an overview // Speech, Language and Hearing. 2016. — N. 19 (2). — P. 121–129.