================ ЭКСПЕРИМЕНТАЛЬНЫЕ СТАТЬИ ================

ACOUSTIC AND PERCEPTUAL FEATURES OF THE EMOTIONAL SPEECH OF ADOLESCENTS AGED 14−16 YEARS

© 2020   A. S. Grigorev[1], *, V. A. Gorodnyi[1], O. V. Frolova[1], A. M. Kondratenko[1], V. D. Dolgaya[1], and E. E. Lyakso[1], **

[1]The Child Speech Research Group, St. Petersburg State University, St. Petersburg, Russia
*E-mail: a.s.grigoriev89@gmail.com
**E-mail: lyakso@gmail.com

The goal of the study was to describe perceptual and acoustic characteristics of emotional spontaneous speech of adolescents aged 14−16 years. 18 adolescents aged 14−16 years (9 boys and 9 girls, three people in each age) were participants. Perceptual analysis of speech of children by adults (70 Russian speakers and 70 speakers of other languages) was conducted to define the emotional state via voice and speech characteristics. Instrumental and linguistic analysis was used to determine acoustic features and linguistic information on which listeners rely while recognizing the adolescents' state. It was shown that listeners had difficulty in recognizing the emotional state of adolescents, but Russian-speaking listeners recognized the emotional state better (26% of phrases recognized with probability 0.75−1.0) than speakers of other languages. Adults recognized comfort better than discomfort, Russian listeners better than foreigners. Russian listeners are likely to rely on acoustic and linguistic information, speakers of other languages − on acoustic features of adolescents' speech. Phrases uttered in comfort are characterized by a wider pitch range vs. phrases uttered in the neutral state. It was revealed that the phrases of girls uttered in comfort have significantly higher average pitch values vs. discomfort. There were no significant differences between boys and girls in acoustic characteristics of emotional speech. Phrase duration and pitch range values can be included in automatic recognition systems of the speaker's emotional state via speech as additional informative features. The paper discusses the specificity of emotion manifestation in adolescents and the possibility of further use of the results.

Keywords: speech, emotions, acoustical features, adolescents, typical development, perceptual analysis

DOI: 10.31857/S086981392003005X

Emotional manifestations and states of children and adolescents are studied at different levels of the organization − from cellular to system-behavioral, which requires the use of different analysis methods. The electrophysiological (EEG and EP indicators) [for example, 1] and hormonal [2] mechanisms of different emotional states are widely studied.

We addressed the problem of reflecting the emotional state in the characteristics of voice and speech, which has a long history starting with the works of Ch. Darwin [3], and chose two main approaches to research it − methods of auditory perceptual and instrumental spectrographic analysis.

The ability to recognize the speaker's state according to voice characteristics is shown on the material of different languages [4−6]. It is known that the assignment of speech patterns to different emotional states is affected by duration and pitch (F0) of the speech sig-

nal. It has been found on the material of the English language that emotional speech can be characterized by changes in the speech rate compared with calm speech [7]. A number of studies indicate that cultural and language peculiarities of a person who perceives emotions influence the emotional state perception [8, 9]. It was shown that values of the pitch, phrase duration, and speech rate affect the efficiency of recognition of the emotional state [10, 11]. Cross-linguistic studies revealed that listeners are able to recognize emotions spoken in a non-native language, but the recognition of emotions in their native language is more effective than in other languages [4—6, 12]. However, in most studies, speech uttered by professional actors is used as emotional speech. There are a few works in which colloquial language is used as emotional speech.

Our work is part of research aimed at studying the formation of the emotional sphere in ontogenesis and the reflection of the emotional state in child speech. The research conducted in the Child Speech Research group of St. Petersburg university showed that adults are able to recognize the state of infants by characteristics of their vocalizations from the first months of life [13, 14]. With the age of children, lexical information contained in a speech message has an increasing influence on the child's emotional state recognition [15, 16]. We have created the first database of emotional child speech for the Russian language [17]. We have revealed that adult Russian native speakers can correctly recognize the emotional state of children aged 4—7 years, with the recognition efficiency increasing with the child's age.

This study is dedicated to the problem of determining the emotional state of adolescents aged 14—16 years by Russian native speakers and speakers of other languages. In this study, we used spontaneous speech of typically developing adolescents. We chose a very contradictory age of 14—16 years to study the emotional state reflection in the child speech characteristics. At this age, adolescents often restrain their emotions while their ability to manifest and recognize emotional states does not always reach the adult level [18]. Emotions of adolescents of this age are mobile, changeable, imagination is actively developing [19]. Changes in the emotional sphere of teens are noted to continue until late adolescence [6]. The adolescent's emotional sphere is under the strong influence of neurohumoral processes linked with puberty, which proceeds under the control of the central nervous system and endocrine glands. In girls, puberty ends, and they demonstrate an "adult" type of behavior. In boys, under the influence of testosterone, the larynx grows, the vocal folds become longer and thicker, leading to voice changes [20]. This physiological process is referred to as a voice mutation (or a voice break) [19]. We previously showed that by the age of 14 years, the duration of stressed and unstressed vowels stabilizes, pitch values decrease, the vowel articulation index also becomes stable [21].

This work is part of the research of the emotional state reflection in the voice and speech characteristics of children [14, 22—24] and covers the period of 14—16 years.

The goal of the study is to describe perceptual and acoustic characteristics of emotional spontaneous speech of adolescents aged 14—16 years.

METHODS

The participants in the study were 18 typically developing adolescents aged 14—16 years (9 boys and 9 girls) attending a high school in St. Petersburg. For all adolescents, we checked thresholds of hearing using clinical audiometry (the automated AA-02 audiometer) and the formation of phonemic hearing. According to the pediatricians, children did not have chronic diseases, neurological or psychiatric disorders. The speech material of children was recorded in the standardized situation of the dialogue between the experimenter and the child. All speech material was annotated by gender, age and emotional state of the adolescent and included in the created database "AD-Child.Ru" [25].

The study included a perceptual experiment and an instrumental analysis. The main purpose of the perceptual experiment was to determine the emotional state of adolescents by adults while listening to speech material. The instrumental analysis of speech was conducted to identify acoustic characteristics, based on which listeners classified speech material into three emotional states.

On the base of listening to audio recordings and watching video materials and research protocol, three experts carried out selecting phrases uttered by adolescents in different emotional states. The material was annotated on three states: "comfort — neutral (calm) — discomfort", previously used when studying the reflection of the emotional state of 4—7-year-old children in the characteristics of their voice and speech [15]. For each adolescent, 5 phrases were selected (2 in the state of comfort, 2 in discomfort, 1 in the neutral state), the total number of phrases is 90. The phrases did not contain linguistic information clearly indicating the speaker's gender or age.

The phrases were included in three test sequences, each test sequence contained 30 phrases. Each phrase was presented once, the interval between the phrases was 5 s. The presentation of test sequences was carried out in an open field to groups of listeners, the number of listeners in each group was no more 10 people.

The listeners were adult Russian native speakers ($n = 70$, 18.2 ± 1.0 years old) and speakers of other languages who are not Russian native speakers (foreigners) ($n = 70$, 21.0 ± 2.9 years old). All the listeners did not have diagnosed hearing impairment, and some ($n = 20$) were checked thresholds of hearing using tonal audiometry.

The speech intensity level during playback was 60—70 dB. The experiment was carried out in a classroom, the noise level in which did not exceed 20 dB, in an open field through LOGITECH S120 speaker set (power 2.3 W, frequency range 50 Hz—20 kHz). There were no significant differences between male and female listeners in recognizing in accordance with the objectives, and therefore the data are presented together.

The task to the listeners was to recognize the speaker's emotional state, determine the meaning of what the adolescent said, the speaker's gender and age.

Mimic facial movements of adolescents manifested during the verbal response were analyzed in the FaceReader 8.0 software (Noldus Information Technology, the Netherlands). The results obtained for every adolescent were summed and averaged, then the emotional states presented in the software were resolved into 4 states – comfort, neutral, discomfort, and other.

Analysis of speech material was carried out in the Cool Edit Pro sound editor. The phrase duration, the pause before the child's response to the experimenter's question, the pause in the phrase, average pitch values in the phrase, minimum and maximum pitch values, the number of words and syllables in the phrase, speech rate were determined for each phrase. Speech rate was determined as the number of syllables uttered per second without excluding pauses in speech [26].

Linguistic analysis of dialogues between adolescents and the experimenter was conducted. In the speech of every adolescent, the frequency of words reflecting comfort and discomfort was calculated. The ratio of words reflecting comfort and discomfort to the total number of words with emotional coloring was defined.

Statistical data analysis was performed in the STATISTICA 10.0 software using the nonparametric Mann—Whitney test and Discriminant analysis.

All procedures were approved by the Health and Human Research Ethics Committee (HHS, IRB 00003875, St. Petersburg State University) and written informed consent was obtained from parents of the child participant.

Table 1. Confusion matrices for the recognition of the emotional state of adolescents by native Russian-speaking and foreign listeners, % correct answers

	Russian speakers			Speakers of other languages		
	comfort	neutral	discomfort	comfort	neutral	discomfort
Comfort	**62**	22	16	**48**	33	19
Neutral	46	**37**	17	44	**41**	15
Discomfort	8	41	**51**	17	36	**47**

Table 2. Confusion matrices for the recognition of the emotional state of girls and boys by Russian and foreign listeners, % correct answers

	Russian speakers					
	girls			boys		
	comfort	neutral	discomfort	comfort	neutral	discomfort
Comfort	**69**	21	10	**55**	22	23
Neutral	45	**36**	19	47	**38**	15
Discomfort	10	34	**56**	7	52	**41**
	Speakers of other languages					
	girls			boys		
	comfort	neutral	discomfort	comfort	neutral	discomfort
Comfort	**56**	29	15	**41**	37	22
Neutral	52	**34**	14	35	**48**	17
Discomfort	18	37	**45**	15	36	**49**

RESULTS

The perceptual experiment showed that native Russian speakers and speakers of other languages were able to assess correctly the emotional state of the adolescent. The success in recognizing emotional states was higher among Russian speakers compared to speakers of other languages. Russian-speaking listeners defined correctly the state of comfort in 62% (speakers of other languages in 48%), the state of discomfort in 51% (speakers of other languages in 47%). The neutral state of the adolescent was better recognized by the speakers of other languages − 41% (Russian speakers − 37%) (Table 1).

Russian listeners recognized the state of comfort and discomfort in girls better than in boys: 69 and 56% of correct answers for phrases uttered by girls in the state of comfort and discomfort, respectively; 55 and 41% − by boys. There were no differences in recognizing the neutral state; the listeners correctly identified the neutral state in 36% in girls and 38% of cases in boys. Speakers of other languages better recognized the state of comfort in girls (56%, boys − 41%); the discomfort and neutral state − in boys (49 and 48% respectively, girls − 45 and 34%) (Table 2).

Analysis of mimic facial movements of children showed that all children most often demonstrated the neutral state (Table 3). There were no significant differences between adolescents depending on gender and age.

In this study, listeners identified only three speaker's states via speech − comfort, neutral, and discomfort. Therefore, the results of analysis of mimic facial movements are also presented for three states: comfort−neutral−discomfort. The states "sad", "scared", "angry", and "disgusted" are combined into the state of discomfort, the state "happy" is selected as the state of comfort. The state "surprised" is assigned to the other group as surprise can accompany both the comfort and discomfort state (Fig. 1).

Table 3. Definition of the emotional state of adolescents by their mimic facial movements, %

	Neutral	Happy	Sad	Scared	Angry	Surprised	Disgusted	Other
Female	68.7	7.5	5.9	4.7	1.8	2.0	0.8	8.7
Male	67.3	7.0	5.7	4.2	5.2	2.6	0.0	8.0

Table 4. Recognition of age of adolescents by native Russian-speaking and foreign listeners

Gender	Girls			Boys		
real age	14	15	16	14	15	16
Russian speakers	14.7 ± 0.6	15.1 ± 0.4	15.0 ± 0.4	15.0 ± 0.4	15.4 ± 0.2	15.4 ± 0.3
Speakers of other languages	14.6 ± 0.5	15.0 ± 0.3	14.9 ± 0.4	14.7 ± 0.4	15.1 ± 0.3	15.2 ± 0.4

The speaker's emotional state is recognized by listeners via speech more effectively than via mimic facial movements. This may indicate that data on mimic facial movements are not enough to recognize the emotional state and they should be added by a speech message.

In addition to recognizing the emotional state, the listeners were given the task of recognizing the adolescent's gender and age. All the listeners with probability 0.75−1.0 were revealed to recognize correctly the adolescent's gender for all phrases used in the test, but in general, Russian speakers identified the gender better ($p < 0.0001$, Mann−Whitney test) than speakers of other languages − 97.2 ± 5.0 and $91.6 \pm 5.9\%$, respectively. When recognizing the adolescent's age, there were no differences between Russian speakers and speakers of other languages, there was no relationship between the adolescent's state and age determined by the listeners (Table 4). Meanwhile, it was shown that speakers of other languages more often found it difficult to choose the adolescent's age (13.8%) vs. Russian speakers (3%).

Additionally, Russian-speaking listeners had to indicate whether they understood the lexical meaning of the phrase. With probability of 0.75−1.0, listeners correctly recognized 88.9% of the phrases contained in the test sequences.

Instrumental analysis showed that there were no significant differences in duration of phrases uttered by adolescents in different emotional states and in duration of phrases uttered by boys and girls. The duration of phrases uttered by boys aged 14−16 years is 3454 ms in the state of comfort, 3907 ms in the neutral state, 5041 ms (median values) in the state of discomfort; by girls aged 14−16 years is 3920.5 ms in comfort, 3544 ms in the neutral state, 6974 ms in discomfort (Fig. 2).
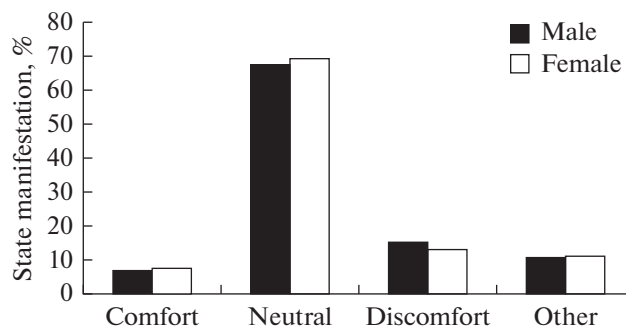
Fig. 1. Manifestation of the emotional state in mimic facial movements of adolescents, % Horizontal axis − state, vertical − state manifestation, %. Black columns − data for boys, white − girls.
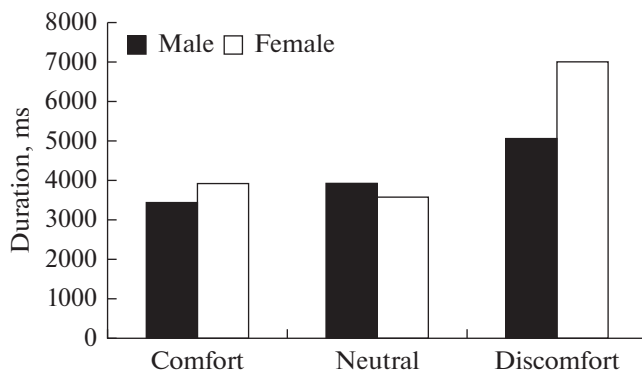
Fig. 2. Duration of phrases uttered by boys and girls in different emotional states. Horizontal axis − state, vertical − phrase duration, ms. Black columns − data for boys, white − for girls.
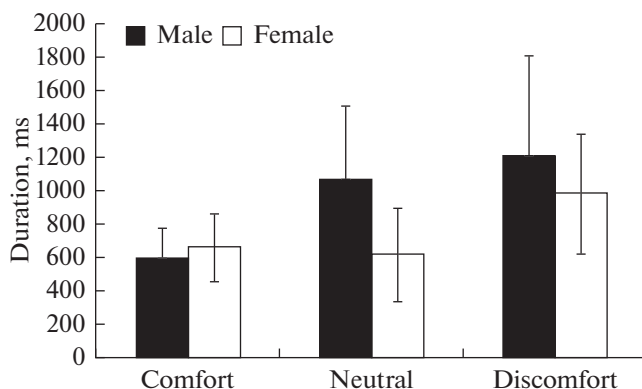

Fig. 3. Pause duration in phrases uttered by boys and girls in different emotional states. Horizontal axis − state, vertical − pause duration, ms. Black columns − data for boys, white − for girls.

Pause duration in phrases uttered by adolescents in different emotional states was determined. The pauses in the phrases uttered by boys are 597 ± 173 ms in comfort, 1063 ± 442 ms in the neutral state, 1213 ± 594 ms in discomfort. The pauses in the phrases uttered by girls are 654 ± 204 ms in comfort, 616 ± 280 ms in the neutral state, 981 ± 359 ms in discomfort. There were no significant differences in the pause duration in boys and girls, in the pause duration in the phrases uttered by adolescents in different emotional states (Fig. 3).

Speech rate, which is defined as the number of syllables uttered per second, in boys was 4.69 syllables per second for phrases spoken in the state of comfort, 5.25 syllables per second in the neutral state, and 4.87 syllables per second in the state of discomfort. In girls, speech rate was 5.24 syllables per second for phrases uttered in comfort, 5.54 syllables per second in the neutral state, 5.53 syllables per second in discomfort. When comparing the speech rate, no significant differences were revealed between boys and girls, between different emotional states, and between adolescents of different ages (Fig. 4).

Average, minimum, and maximum pitch values for the phrase were determined. It was found that average pitch values in the phrases uttered by boys are 212 ± 48 Hz in comfort, 187 ± 53 Hz in the neutral state, 187 ± 59 Hz in discomfort; by girls − 229 ± 33 Hz in comfort, 217 ± 23 Hz in the neutral state, 209 ± 23 Hz in discomfort. The average pitch values in
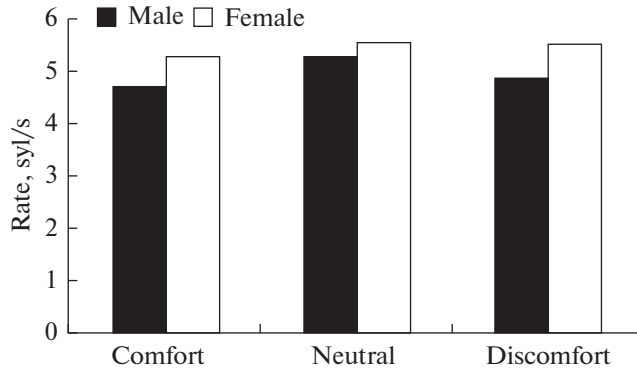
Fig. 4. Speech rate in phrases uttered by boys and girls in different emotional states. Horizontal axis — state, vertical — speech rate, syllables per second (syl/s). Black columns — data for boys, white — for girls.
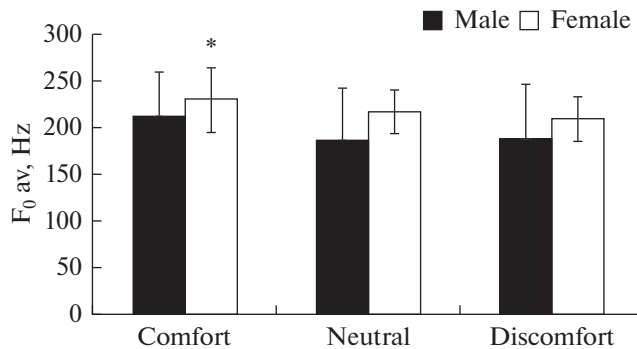

Fig. 5. Average pitch values in phrases uttered by boys and girls in different emotional states. Horizontal axis — state, vertical — pitch, Hz. Black columns — data for boys, white — for girls. * $p < 0.05$, Mann–Whitney test.

phrases uttered by girls in the state of comfort are revealed to be significantly higher ($p < 0.05$) than the average pitch values in phrases uttered in discomfort. No significant differences in pitch values in boys and girls, in pitch values in the phrases uttered by boys in different emotional states were found (Fig. 5).

The pitch range values in phrases uttered by boys and girls in different emotional states were calculated. In boys, the pitch range values are 180 ± 19 Hz in comfort, 70 ± 31 Hz in the neutral state, 132 ± 33 Hz in discomfort. In girls, the pitch range values are 138 ± 25 Hz in comfort, 59 ± 13 Hz in the neutral state, and 96 ± 19 Hz in discomfort.

It was found that, in both boys and girls, the pitch range values in phrases uttered in the state of comfort are significantly higher ($p < 0.01$ in boys, $p < 0.05$ in girls — Mann–Whitney test) than in the neutral state. There were no significant differences between the states of comfort and discomfort, between discomfort and the neutral state, between boys and girls (Fig. 6).

Discriminant analysis revealed the relationship between the emotional state and minimum pitch values $F(32,144) = 2.7350$ $p < 0.001$ (Wilks' Lambda: 0.386 $p < 0.00001$): in the states of comfort and discomfort, minimum pitch values are higher than in the neutral state; between gender and minimum pitch values $F(16,73) = 4.2881$ $p < 0.0001$ (Wilks' Lambda: 0.515, $p < 0.00001$): minimum pitch values in girls are higher than in boys.

Acoustic features of emotionally colored phrases recognized by listeners with low (0–0.49) and high probability (0.5–1.0) were determined. Significant differences ($p < 0.05$,
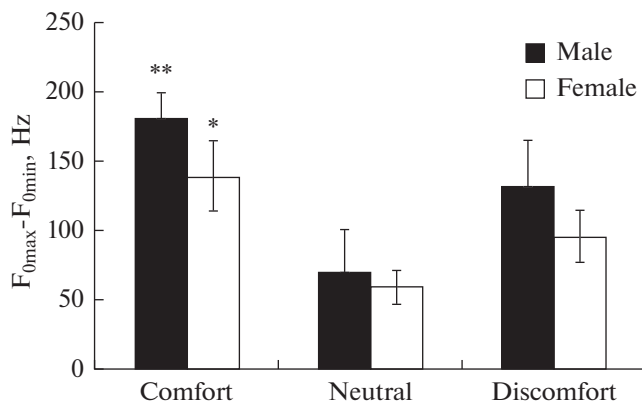
Fig. 6. Pitch range in phrases uttered by boys and girls in different emotional states. Horizontal axis − state, vertical − pitch, Hz. Black columns − data for boys, white − for girls. * − $p < 0.05$, ** − $p < 0.01$, Mann−Whitney test.

Mann−Whitney test) in duration of phrases recognized with low and high probability were found. In boys, the duration of phrases uttered in discomfort and the neutral state and recognized with high probability is lower ($p < 0.05$) than the duration of phrases recognized with low probability; no significant differences for the state of comfort were found. In girls in the state of comfort, the phrases recognized with high probability were longer ($p < 0.05$) than the phrases recognized with low probability; no significant differences for discomfort and the neutral state were found (Fig. 7).

Significant differences in average pitch values, pitch range values, speech rate, and the sum of pauses in phrases between phrases recognized by listeners with high and low probability were not found.

Linguistic analysis of utterances of adolescents included in test sequences for the perceptual experiment was conducted to select emotionally colored words. It was shown that words reflecting the state of comfort prevailed in the speech of adolescents of all ages (Fig. 8). In the speech of boys, the number of words reflecting the state of comfort was 65.2% of the total number of emotionally colored words, the state of discomfort − 34.8%. In the speech of girls, the number of words reflecting the state of comfort was 53.8%, the state of discomfort − 46.2%. The most frequent words reflecting the state of comfort in adolescents of both genders were "like" (with frequency 1.34 and 2.14 in boys and girls, respectively), "love" (1.11 and 0.66), "good" (0.89 and 0.66), "favorite" (0.45 and 0.33). The most frequent words reflecting the state of discomfort were "dislike" (with frequency 0.67 and 0.82 in boys and girls, respectively), "hate" (0.45 and 0.33).

DISCUSSION

In this study, for the first time, the data on peculiarities of recognition of the emotional state of adolescents aged 14-16 years via speech features were obtained. The situation of a dialogue with the experimenter was used in the work to compare the speech material of adolescents of different age and gender. Participants in the study were two groups of listeners − native Russian speakers and speakers of other languages. This was done in order to assess the contribution of linguistic and acoustic information in the recognition of the emotional state of adolescents. The Russian-speaking listeners recognize the emotional state of adolescents better than speakers of other languages. It may indicate that, while recognizing the emotional state, they are likely to rely on acoustic and linguistic information vs. speakers of other languages who rest only on acoustic information. Our findings that listeners recog-
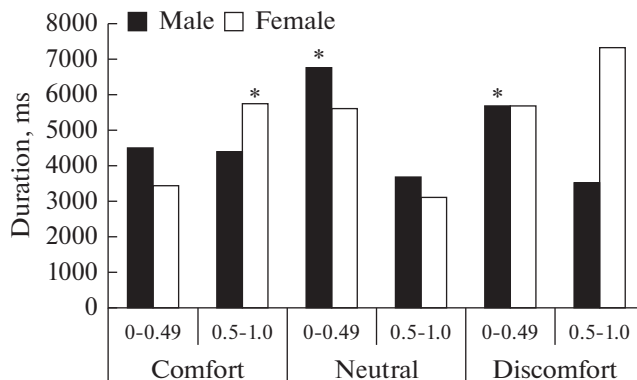
Fig. 7. Duration of phrases recognized by listeners with high and low probability. Horizontal axis — speaker's state and recognition probability, vertical — duration, ms. Black columns — data for boys, white — for girls. * — $p < 0.05$, Mann–Whitney test.
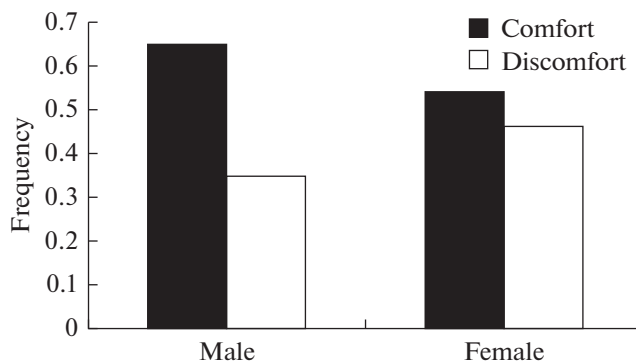

Fig. 8. Frequency of words with different emotional coloring in the speech of boys and girls. Black columns — the state of comfort, black — discomfort. Horizontal axis — adolescents' gender, vertical — frequency.

nize the emotional state in the native language better than speakers of other languages are confirmed by the literature data [5, 12].

The results obtained in the work about the level of recognition of the emotional state of children correlate to the results obtained for 4—7-year-old children [15]. However, in the speech of adolescents, the neutral state is recognized worse than in the speech of 4-7-year-old children. Our results on the speaker's emotional state recognition level are correlated with the results obtained on the material of the English and Swedish languages, for which the recognition probability of the emotional state in speech is about 60% [27] — in our work, the recognition probability of the state of comfort was 61%, discomfort — 51%. The Russian-speaking listeners unambiguously (with probability 0.75—1.0) recognize the emotional state in 26% of the phrases uttered by adolescents. The speakers of other languages with probability 0.75—1.0 did not recognize a single phrase. These difficulties can be explained, on the one hand, by the peculiarities of culture and education of Russian-speaking adolescents [18] who try to restrain their emotions in the dialogue with the experimenter, on the other hand, by the inability of speakers of other languages to recognize linguistic information, contained in a speech message.

Listeners better recognize the gender while listening to the phrases of adolescents than listening to the speech of 4—7-year-old children (with recognition probability of 64—71% for different emotional states). However, our study revealed no influence of the adolescent's emotional state on the recognition of gender. An increase in the recognition accuracy of age compared with the results of recognition of the age of preschoolers has been shown.

Recognizing the emotional state by mimic facial movements revealed that the majority of emotional states in adolescents are recognized as a neutral emotional state. The absence of vivid emotion manifestations may indicate the peculiarities of the behavior of adolescents due to the social environment or specificities of the experiment (the recordings were carried out in the school building) [14].

In our earlier works, we showed the dynamics of changes in duration and pitch with the age of children [28]. It was shown that the duration of stressed and unstressed vowels in children does not significantly change, starting from the age of 13 years [21]. The acoustic parameters of vowels from a neutral (calm) speech of 14—16-year-old teens are described [29]. We found that phrases uttered in the state of discomfort are characterized by a longer duration than phrases uttered in comfort and the neutral state, but these differences are not significant. For the Swedish and Spanish languages, it was shown [7] that emotional speech differed from speech uttered in the neutral state in the speech rate; our study did not reveal this correlation. It was found that, for adult speech, there were differences in the speech rate between men and women depending on their emotional state [10]. We did not identify these distinctions for speech of adolescents. Our findings correspond to the speech rate values, which are described for adult Russian speech [30]. Significant differences in the average pitch values, observed in the speech of younger children, between phrases uttered by adolescents in different emotional states were not found. The correlation between the emotional and neutral speech of adolescents according to pitch range values was revealed, which is confirmed by the results previously obtained for child speech [15]. These results correspond to the results on the material of the Japanese language. It is shown that both pitch parameters and parameters of the vowel formant frequencies F1 and F2 influence the speaker's emotion perception, but pitch range provides key input; the formant frequencies values affect the emotional coloring perception in case of fixed pitch values. It is also indicated that emotion perception can be influenced by the individual characteristics of the speaker, but even in this case pitch and its range make a statistically more significant contribution to the emotion determination than the variability of formant frequencies or vowel duration [31].

The determination of characteristics that are sufficient for recognition of the emotional state is necessary to develop automatic speech recognition systems [11], to study recognition of the emotional state of people with developmental disabilities [32]. Automatic recognition of emotions is important for creating human-computer interfaces, the learning process, monitoring for older adults, interactive entertainment [33]. Artificial intelligence systems based on various parameters of a speech message (for example, pitch, spectral characteristics, linguistic information) are extensively used in automatic speech recognition systems [34]. Different approaches to machine learning of emotion recognition only according to the paralinguistic information contained in speech are exploited [35]. It is shown as well that not all of the existing corpora of sounding emotional speech are suitable for learning artificial intelligence systems.

It was shown that, for the English language, the particular difficulty in recognizing the speaker's emotions is the distinction of those emotional states that have similar acoustic features [34]. The use of a comprehensive assessment of the speaker's acoustic characteristics is thought to be necessary for assessing the emotional state [11]. Some works [36] indicate that Russian speakers use verbal means of expressing emotions in combination with non-verbal ones.

Most of the works are now dedicated to the problems of emotions formation and manifestation in adolescents with various disorders, for example, with obsessive-compulsive

disorder [37], manifestations of depression [38], and experiencing the difficulties with control and emotions manifestations [39].

Therefore, the obtained results may be useful for studying the development of emotions in ontogenesis in typical (control) and atypical development.

CONCLUSIONS

In the work, original data on the recognition of the emotional state of adolescents via their speech characteristics by Russian native speakers and speakers of other languages were obtained. Listeners have difficulty in recognizing the emotional state of adolescents. Nevertheless, Russian-speaking listeners recognize the emotional state better than foreigners. It is shown that adults recognize the state of comfort better than discomfort. Russian listeners cope with the emotion recognition better than foreigners. In recognizing, Russian listeners are likely to rely on acoustic and linguistic information, speakers of other languages rest on acoustic features of adolescents' speech. Phrases uttered in comfort are characterized by a wider pitch range vs. phrases uttered in the neutral state. It was revealed that the phrases of girls uttered in comfort have significantly higher average pitch values vs. discomfort. The phrases for which the emotional state is recognized with high probability differ in duration from phrases recognized with low probability. Phrases uttered by boys, for which the emotional state is recognized with high probability, are shorter, phrases uttered by girls are longer. The Russian-speaking listeners recognize the emotional state of girls better than the emotional state of boys. It was revealed that all the listeners recognize correctly the gender of adolescents.

ACKNOWLEDGEMENTS

REFERENCES

1. *Barker R.M., Bialystok E.* Processing differences between monolingual and bilingual young adults on an emotion n-back task. Brain Cogn. 134: 29−43. 2019.
2. *Bernhard A., van der Merwe C., Ackermann K., Martinelli A., Neumann I.D., Freitag C.M.* Adolescent oxytocin response to stress and its behavioral and endocrine correlates. Horm. Behav. 105: 157−165. 2018.
3. *Darwin Ch., Ekman P., Prodger P.* The Expression of the Emotions in Man and Animals. 3rd edn. London. Harper Collins. 1998.
4. *Thompson W.F., Balkwill L.-L.* Decoding speech prosody in five languages. Semiotica. 407−424. 2006.
5. *Paulmann S., Uskul A. K.* Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. Cognition and Emotion. 28: 230−244. 2014.
6. *Chronaki G., Wigelsworth M., Pell M.D., Kotz S.A.* The development of cross-cultural recognition of vocal emotion during childhood and adolescence. Scient. Rep. 8(1): 8: 8659. 2018.
7. *Abelin Å., Allwood J.* Cross linguistic interpretation of emotional prosody. ISCA workshop on Speech and Emotion. Newcastle. Northern Ireland. 110−113. 2000.
8. *Matsumoto D., Franklin B., Choi J.-W., Rogers D. Tatani H.* Cultural influences on the Expression and Perception of Emotion. In: W.B. Gudykunst, B. Moody (eds.) Handbook of International and Intercultural Communication. Sage Publications. 2002.
9. *Abelin Å.* Cross-cultural multimodal interpretation of emotional expressions—an experimental study of Spanish and Swedish. In: Proc. Speech Prosody. ISCA. March 23−26. Naran. 2004.
10. *Ververidis D., Kotropoulos C.* Emotional speech recognition: Resources, features, and methods. Speech Communic. 48 (9): 1162−1181. 2006.
11. *Yamazaki T., Nakayama M.* Extracting acoustic features of Japanese speech to clas-sify emotions. Communication papers of the Federated Conference on Computer Science and Information Systems. ACSIS. 13: 141−145. 2017.
12. *Jiang X., Paulmann S., Robin J., Pell M.D.* More than accuracy: Nonverbal dia-lects modulate the time course of vocal emotion recognition across cultures. J. Exp.Psychol.: Human Perception and Performance. 41: 597−612. 2015.
13. *Lyakso E.* Characteristics of infant's vocalizations during the first year of life. Inter-national J. Psychophysiol. 30: 150−151. 1998.
14. *Lyakso E., Frolova O.* Emotion state manifestation in voice features: chimpanzees, human infants, children, adults. Lecture Notes in Computer Science. 9319. 201−208. 2015.

15. *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. Comput. Speech Lang. 46: 268–283. 2017.

16. *Yildirim S., Narayanan S., Potamianos A.* Detecting emotional state of a child in a conversational computer game. Comput. Speech Lang. 25 29–44.2011.

17. *Lyakso E., Frolova O., Dmitrieva E., Grigorev A., Kaya H., Salah A.A., Karpov A.* EMOCHILDRU: emotional child russian speech corpus. Lecture Notes in Computer Science. 9319: 144–152. 2015.

18. *Иванова Е.С.* Половозрастные особенности эмоционального интеллекта и его структурных компонентов. Образование и наука. 7(86): 65–74. 2011. [*Ivanova E.S.* Gender and age characteristics of emotional intelligence and its structural components. Education and sci. 7(86): 65–74. 2011. (In Russ)].

19. *Ляксо Е.Е., Ноздрачев А.Д., Соколова Л.В.* Возрастная физиология и психофизиология: учебник для среднего профессионального образования. М. Юрайт. 2016. [*Lyakso E.E., Nozdrachev A.D., Sokolova L.V.* Vozrastnaya fiziologiya i psihofiziologiya: uchebnik dlya srednego professional'nogo obrazovaniya.[Age physiology and psychophysiology: the textbook for undergraduate academic.]Moscow. Urait. 2016. (In Russ)].

20. *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. 84–96. 2016.

21. *Grigorev A., Frolova O., Lyakso E.* Acoustic features of speech of typically devel-oping children aged 5–16 years. Communic. Comput. Informat. Sci. AINL. 930: 152–163. 2018.

22. *Ляксо Е.Е.* Вокально-речевое развитие ребенка в первый год жизни. Рос. физиол.журн. им. И.М. Сеченова. 89(2): 207–218. 2003. [*Lyakso E.E.* Vocal and speech development of a child in the first year of life. Russ. J. Physiol. 89(2): 207–218. 2003. (In Russ)].

23. *Pavlikova M.I., Makarov A.K., Lyakso E.E.* Acoustic characteristics of vocalization reflecting the states of discomfort and comfort in babies aged three and six months Neurosci. Behav. Physiol. 47. (1): 40–46. 2017.

24. *Ляксо Е.Е., Фролова О.В., Григорьев А.С., Соколова В.Д., Яроцкая К.А.* Распознавание взрослыми эмоционального состояния типично развивающихся детей и детей с расстройствами аутистического спектра. Рос. физиол. журн. им. И.М. Сеченова. 102(6): 729–741. 2016. [*Lyakso E.E., Frolova O.V., Grigorev A.S., Sokolova V.D., Yarotskaya K.A.* Recognition by adults of emotional condition of typically developing children and children with autism spectrum disorders. Russ. J. Physiol. 102(6): 729–741. 2016. (In Russ)].

25. *Lyakso E., Frolova O., Karpov A.* A new method for collection and annotation of speech data of atypically developing children. In: Proc. of Internat. IEEE Conference on Sensor Networks and Signal Processing (SNSP 2018). 175–180. Xi'an. China. 2018.

26. *Goldman-Eisler F.* Speech Analysis and Mental Processes. Language and Speech. 1(1): 59–75. 1958.

27. *Juslin P.N., Laukka P.* Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion. 1(4): 381–412. 2001.

28. *Lyakso E.E., Grigor'ev A.S.* Dynamics of the duration and frequency characteristics of vowels during the first seven years of life in children. Neurosci. Behavi. Physiol. 45 (5): 558–567. 2015.

29. *Григорьев А.С., Ляксо Е.Е.* Акустические характеристики гласных из слов детей 14–16 лет. Учен. зап. физического факультета московского университета. 5: 1750202. 2017. [*Grigorev A.S., Lyakso E.E.* Acoustic features of vowels from the words of children 14–16 years. Sci. notes Phys. Dep. Mosc. University. 5: 1750202. 2017. (In Russ)].

30. *Stepanova S.* Russian spontaneous speech – based on the speech corpus of Russian everyday interaction. In: ICPhS XVII. Hong Kong. 17–21 August 2011. 1902–1905. 2011.

31. *Li Y., Li J., Akagi M.* Contributions of the glottal source and vocal tract cues to emotional vowel perception in the valence-arousal space. J. Acoust. Soc. Am. 144(2): 908. 2018.

32. *Schelinski S., von Kriegstein K.* The relation between vocal pitch and vocal emo-tion recognition abilities in people with autism spectrum disorder and typical develop-ment. J. Autism and Development. Disorders. 68–82. 2018.

33. *Papakostas M., Siantikos G., Giannakopoulos T., Spyrou E., Sgouropoulos D.* Recognizing Emotional States Using Speech Information. GeNeDis 2016. 155–164. 2017.

34. *Jin Q., Li C., Chen S., & Wu H.* Speech emotion recognition with acoustic and lexical features. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4749–4753. 2015.

35. *Stolar M., Lech M., Bolia R., Skinner M.* Acoustic characteristics of emotional speech using spectrogram image classification. 12th Internat. Conference on Signal Proc. Commun. Systems (ICSPCS). Cairns. Australia. 17–19 December 2018. 1–6. 2018.

36. *Коновалова Ю.О.* Вьетнамцы и русские: отношение к эмоциям как проявление национальной культуры. Язык и культура. 38: 51–69. 2017. [*Konovalova Y.O.* Vietnamese and Russian: related to emotions as an expression of national culture. Lang. and Culture. 38: 51–69. 2017. (In Russ)].

37. *Yazici K.U., Yazici I.P.* Decreased theory of mind skills, increased emotion dysregulation and insight levels in adolescents diagnosed with obsessive compulsive disorder. Nord. J. Psychiatry. 73(7): 462–469. 2019.

38. *Durbeej N., Sörman K., Norén Selinus E., Lundström S., Lichtenstein P., Hellner C., Halldner L.* Trends in childhood and adolescent internalizing symptoms: results from Swedish population based twin cohorts. BMC Psychol. 2;7(1): 50. 2019.

39. *Zheng Y., Asbury K.* Genetic and Environmental Influences on Adolescent Emotional Inertia in Daily Life. J. Youth Adolesc. 48(9): 1849−1860. 2019.

АКУСТИЧЕСКИЕ И ПЕРЦЕПТИВНЫЕ ХАРАКТЕРИСТИКИ ЭМОЦИОНАЛЬНОЙ РЕЧИ ПОДРОСТКОВ 14-16 ЛЕТ

А. С. Григорьев[a, *], В. А. Городный[a], О. В. Фролова[a], А. М. Кондратенко[a], В. Д. Долгая[a], Е. Е. Ляксо[a, **]

[a]*Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*
*E-mail: a.s.grigoriev89@gmail.com
**E-mail:lyakso@gmail.com

Цель исследования − описание перцептивных и акустических характеристик эмоциональной спонтанной речи подростков 14—16 лет. В исследовании приняли участие 18 подростков 14—16 лет (9 мальчиков и 9 девочек, по три человека в каждый возрастной срез). Проведен перцептивный анализ речи детей взрослыми (70 носителей русского языка и 70 носителей других языков) с целью определения эмоционального состояния по характеристикам голоса и речи. Инструментальный и лингвистический анализ использован для определения акустических признаков и лингвистической информации, на которые опираются слушатели при распознавании состояния подростков. Показано, что слушатели испытывают трудности в распознавании эмоционального состояния подростков, но при этом русскоговорящие слушатели распознают эмоциональное состояние лучше (26% фраз, распознанных с вероятностью 0.75—1), чем носители других языков. Взрослые лучше распознают состояние комфорта, чем дискомфорта, русские слушатели лучше, чем иностранные. При распознавании русские аудиторы опираются на акустическую и лингвистическую информацию, носители других языков − на акустические параметры речи подростков. Фразы, произнесенные в комфортном состоянии, характеризуются большей вариативностью частоты основного тона по сравнению с фразами, произнесенными в нейтральном состоянии. Выявлено, что фразы девочек, произнесенные в комфортном состоянии, имеют значимо более высокую среднюю частоту основного тона по сравнению с дискомфортным состоянием. Не выявлено значимых различий между мальчиками и девочками по акустическим характеристикам эмоциональной речи. Значения длительности фраз и вариативности частоты основного тона могут быть заложены в системы автоматического распознавания эмоционального состояния говорящего по речи в качестве дополнительных информативных признаков. В работе обсуждается специфика проявления эмоциональных состояний подростками и возможности дальнейшего использования полученных результатов.

Ключевые слова: речь, эмоции, акустические характеристики, типичное развитие, перцептивный анализ