A Comparison of Acoustic Features of Speech of Typically Developing Children and Children with Autism Spectrum Disorders

Elena Lyakso[(✉)], Olga Frolova, and Aleksey Grigorev

Saint Petersburg State University, Saint Petersburg, Russia
lyakso@gmail.com

Abstract. The goal of this study is to find out the acoustic features specific for ASD children vocalizations and speech. Three types of experiments were conducted: emotional speech, spontaneous speech, and the repetition of words. Participants in the study were children with ASD (F 84.0 according to ICD-10), biologically aged 5–14 years (n = 25 children) and typically developing (TD) children aged 5–14 years (n = 60). We compare acoustic features that are widely used in speech recognition and speech perception: pitch values, max and min values of pitch, pitch range, formants frequency, energy and duration. Formant triangles were plotted for vowels with apexes corresponding to the vowels [a], [u], and [i] in F1, F2 coordinates, and their areas were compared. For all children with ASD voice and speech is characterized by high values of pitch, abnormal spectrum, and well-marked high-frequency. Stressed vowels from the words of children (TD & ASD), spoken in discomfort, have higher values of pitch and the third (emotional) formants than spoken in a comfortable condition. ASD children showed higher values of pitch in spontaneous speech than in repetition speech. The current results are a first step toward developing speech based bio-markers for early diagnosis of ASD.

Keywords: Acoustic features · Pitch · Formants frequency · Energy · Children typically developing · ASD children · Speech

1 Introduction

Autism spectrum disorders (ASD) are pervasive developmental disorders that have been defined as a triad of impairment: atypical development of reciprocal social interaction, atypical communication, and restricted, stereotyped, and repetitive behaviors. Since the first delineation of the autistic syndrome [1], abnormal prosody has been identified as a core feature for individuals who speak [2]. Monotonic or machine-like intonation, varying from flat and monotonous to variable, sing-song, or pedantic [1]. Problems in vocal quality, the control of volume, and use of aberrant stress patterns have also been widely reported [3–5]. Abnormal prosody production is a consistent feature of the ASD communication profile [6–8].

Studies of the acoustic properties of prosody production in height functional autism (HFA) [e.g. 3], generally showing that participants with ASD produce longer utterance durations even when their prosody is perceived as appropriate by listeners. In another study it was found larger spectral variability in the ASD group, which "blurs" or averages out the harmonic structure [6]. These authors note that the ASD children had a significantly larger pitch range and variability across time. The increased pitch range was found in speakers with HFA during both conversation and structured communication. Although the HFA group demonstrated an increased acoustic pitch range, listeners did not rate speakers with HFA as having increased pitch variation [5]. The same was noted in an earlier study [7]. In the studies of ASD pre-school children living in the Japanese language environment the negative correlation between the pitch variation and the domain of social reciprocal interaction scores of Japanese Autism Screening Questionnaire was revealed. Monotonous speech in school-aged children with ASD was detected [9].

So, given literature data indicate contradictory data regarding the prosodic speech features of children with ASD. The goal of this study is to find out the acoustic features specific for ASD children vocalizations and speech.

2   Method

2.1   Data Collection

Participants in the study were children with ASD (F84.0 according to ICD-10), biologically aged 5–14 years (n = 25 children) and typically developing (TD) children (coevals, n = 60). For this study the ASD sample was divided into two groups according developmental features: presence of development reversals at the age 1.5–3.0 years (first group - ASD-1) and developmental risk diagnosed at the infant birth (second group - ASD-2). For these children, the ASD is a symptom of neurological diseases associated with brain disturbed. Mean Child Autism Rating Scale [10] total scores was calculated for each group. In order to assess whether differences in autism severity varied across groups, a one-way ANOVA was conducted for two groups. The groups don't differ significantly.

Three types of experiments were conducted: emotional speech, spontaneous speech, and the repetition of words. For first experiment recording conditions for a TD the model experiment included playing with toys (a standard set of toys); repetition of words from a toy-parrot in the game store situation; watching the cartoon and the retelling of the story; for ASD children - playing with toys, and the pictures description [11]. Child speech recording of the second experiment was carried out in a situation of child dialogue with the experimenter (neutral theme as possible as), in the third - when the child repeated the words of the experimenter or parent (for ASD children). Places of recording were at home, in the laboratory and kindergarten for TD children and in the laboratory and the swimming pool for ASD children. The recordings were made by the "Marantz PMD222" recorder with a "SENNHEIZER e835S" external microphone.

2.2   Data Analysis

The child's emotional state was revealed based on recording situation and video fragment analysis by 5 speech experts. The test sequences were presented to 140 adults (native Russian speakers) for perceptual analysis. Spectrographic analysis of speech was carried out in the Cool Edit (Syntrillium Soft. Corp. USA) sound editor. We analyzed and compared pitch values, max and min values of pitch, pitch range, formants frequency (F1 - first formant, F2 - second formant, F3 - third formant), energy and duration. Along with the absolute values of the F1 and F2 formants, their relative values, /F2 – F1/, were compared. Formant triangles were plotted for repetition vowels with apexes corresponding to the vowels /a/, /u/, and /i/ in F1, F2 coordinates, and their areas were compared. Vowel formant triangle areas were calculated as described in [12] modified for Russian [13]. To consider word stress development the vowel duration and its stationary part duration were compared in the stressed versus the unstressed vowels, as well as the pitch and formants values in the stationary parts. The same parameters were compared using the Mann-Whitney criterion in /a/, /i/ and /u/ after the following consonants: /k/ and /d/ for /a/, /b/ and /g/ for /u/ and /t'/ for /i/. These consonants cause the minimal articulator and hence acoustic influence on the corresponding vowels in Russian. The values of the amplitudes (energy) of pitch and the first three formants of vowels by the dynamic spectrogram were determined. The normalized values of formants amplitude concerning to the amplitude of the pitch (E0/En, where E0 – amplitude of pitch, En – amplitude of formants, where $n = 1$ for F1, $n = 2$ for F2, $n = 3$ for F3). All statistical tests were conducted using Statistica 10.0.

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.

3   Result

3.1   Acoustic Features of TD and ASD Child Emotional Speech

Different emotional states were used for comparing TD children and RAS children that allowed finding the variable characteristics of the voice. Both discomfort and comfort conditions in the speech of TD children were recognized by adults with the perception rate of 0.75–1.0 better compared to the neutral condition. Positive correlation between TD age and recognition of discomfort state $r = 0.9747$ ($p < 0.05$ Spearman) was revealed. Discomfort state in the vocalizations and speech of ASD children, adults recognized better ($p < 0.01$ Mann-Whitney test) than comfort and neutral state. Spectrographic analysis revealed that speech interpreted by listeners as discomfort, neutral and comfort is characterized by a set of acoustic features. Discomfort TD children's speech sample are characterized by highest maximum pitch values ($p < 0.01$), average pitch values ($p < 0.05$) and pitch variation values (F0max–F0min) ($p < 0.05$) vs. neutral speech sample. Discomfort state significantly don't differ from comfort state on the base of average pitch values of stress vowels from words. Correctly recognized by adults discomfort and comfort speech do not differ in pitch variation values. Changes of

comfort and neutral state recognition with a child's age are bonded together: positive correlations between recognition of comfort and neutral test samples were revealed $r = 0.9$. Discomfort state is mostly characterized by falling pitch contour type, comfort state – by rising and neutral – by flat pitch contour.

Discomfort ASD children's speech sample are characterized by vowels' highest average pitch values, pitch range, and third formant frequency of vocalizations and words ($p < 0.001$) than comfort and neutral speech samples (Fig. 1A, B).
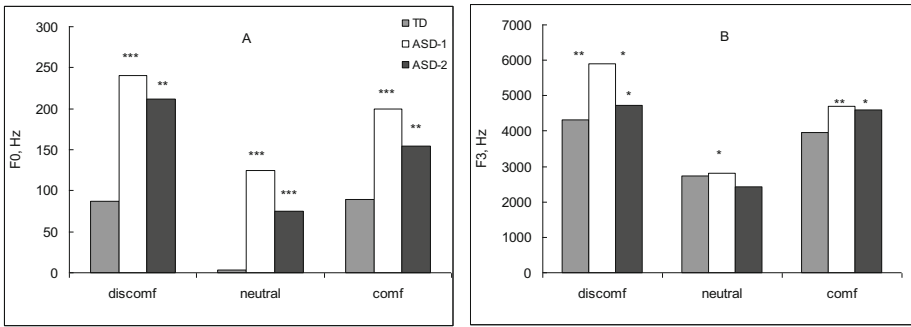

Fig. 1. Vowel's pitch range value (F0max–F0min) (A) and third formant frequency of vowels in discomfort, neutral and comfort (B). * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$ Mann-Whitney test.

Pitch variation values (F0max–F0min) in ASD-1 child's discomfort, neutral and comfort speech significantly higher ($p < 0.001$) than in ASD-2 child's speech. Pitch contour type does not change depending on the emotional state of ASD children. The F3 values in discomfort speech of ASD-1 children significantly higher than in corresponding voice features in ASD-2 children ($p < 0.01$) and TD peers ($p < 0.01$).

Child membership to an ASD group $F(5.13) = 8.536$ $p < 0.0009$ associated with average pitch values (Beta $= -0.364$, $R^2 = 0.7665$), values of third formant (emotional) (Beta $= -0.743$, $R^2 = -0.7665$), the level of speech (Beta $= -0.484$, $R^2 = 0.7665$ – Multiple Regression analysis). The relation between the heaver child disease, the higher pitch values and third formant values, and the lower speech level was revealed.

3.2   Acoustic Features of TD and ASD Spontaneous Child Speech

The purpose of this experiment is to examine the process of the acoustic features of the vowel from ASD spontaneous speech approaching corresponding values in the TD speech. For all children with ASD voice and speech is characterized by high values of the pitch, abnormal spectrum, and well-marked high-frequency. Pitch values of spontaneous speech of ASD children higher ($p < 0.001$) than pitch values of TD children, pitch values of ASD-1 children higher ($p < 0.01$) than in ASD-2 were shown. Comparison of formant frequency values showed differences for the vowel /a/ by F1 and [F2–F1] values between ASD-1 and ASD -2 ($p < 0.05$), ASD-1 and TD ($p < 0.001$); for the vowel /u/ between ASD-1 and TD ($p < 0.05$); for the vowel /i/ - between [F3–F2]

values between ASD-1 and TD ($p < 0.05$). Comparison of vowel formant triangle areas showed that areas were greatest for the vowels of ASD-2 children than ASD-1 ones. Decrease the area of the vowel formant triangle of TD child speech was shown.

In our study, in the ASD-1 were more boys than girls. Therefore additionally only boy's spontaneous speech was analyzed. Belonging to a group $F(6.443) = 57.861$ $p < 0.0000$ was a predictor of the acoustical features of vowels: vowels duration (Beta $= 0.1175$, $R^2 = 0.4393$), average pitch values (Beta $= -0.6811$, $R^2 = 0.4393$), values of F1 (Beta $= -0.1237$, $R^2 = 0.4393$), and values of F2 (Beta $= -0.1024$, $R^2 = 0.4393$ – Multiple Regression analysis). Boy's age $F(6.443) = 11.455$, $p < 0.0000$ was a predictor of the same acoustical features of vowels as a group. Child's membership to a groups (data for all ASD & TD children) $F(4.59) = 43.902$ $p < 0.0000$ correlated with average pitch values (Beta $= -0.4027$, $R^2 = 0.7485$), values of third formant (Beta $= -0.5647$, $R^2 = -0.4027$– Multiple Regression analysis).

A specific characteristic of the dynamic spectrum of the vowels of the ASD –1 child is the intensity of the third formant (Fig. 2). The intensity of the vowel formants ASD-2 children was not significantly different from the TD corresponding data, except for F3/E0 for vowel /a/.
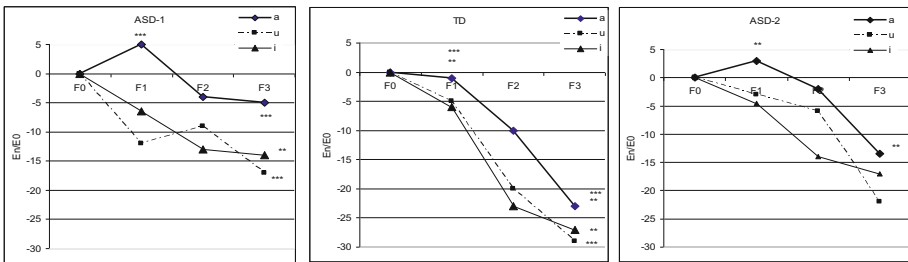

Fig. 2. The distribution of the three first format's amplitudes normalized to the amplitude of pitch for vowels /a/, /u/, /i/ is spontaneous speech. Vertical axis – En/E0 (normalized amplitude), horizontal axis – F0 and formants (F1, F2, F3).

3.3   Acoustic Features of TD and ASD Repetition vs. Spontaneous Child Speech

In speech task - words repetition child membership to an ASD group (first & second) $F(18.163) = 2.7161$ $p < 0.0004$ associated with child sex (Beta $= 0.4168$, $R^2 = 0.2307$), and stress vowel duration (Beta $= 0.1804$, $R^2 = 0.2307$).

At 5 years of age in all the TD children the stressed vowel and its stationary part duration, as well as their difference, is higher in the stressed vowels than in unstressed ones. This is unusual for Russian language where the stress is expressed by the duration of vowels. For ASD children stressed vowels don't differ significantly from unstressed vowels on the base of the vowel duration was shown. Stressed vowel extracted on the high pitch values or vowel duration and pitch with a typical allocation of each child was revealed. The context (/a/, /i/ and /u/ after the following consonants) influence on the characteristics of vowels in speech repeated was not significant for ASD and TD children.

Type of speech task (spontaneous and repetition) was revealed as predictor for stress vowels duration $F(6.337) = 3,965$ $p < 0.007$ (Beta $= 0.1234$, $R^2 = 0.065$), and for average pitch values $p < 0.0001$ (Beta $= -0.2625$, $R^2 = 0.065$) – Multiple Regression analysis. Child's membership to a group (ASD & TD) was revealed as predictor for speech task realization $F(7.336) = 106.33$ $p < 0.0000$ (Beta $= -0.6932$ $R^2 = 0.6889$), words duration $p < 0.003$ (Beta $= -0.0092$, $R^2 = 0.6889$), stress vowels average pitch values $p < 0.0000$ (Beta $= -0.667$, $R^2 = 0.6889$), and F1 values $p < 0.019\,0000$ (Beta $= 0.0916$, $R^2 = 0.6889$ – Multiple Regression analysis).

Pitch values don't differ significantly in stress vowels from TD children's words in twice task. Pitch values of the ASD-1children was significantly higher ($p < 0.001$) then in the ASD-2 child's spontaneous speech. Pitch values variation (F0max–F0min) significantly higher in spontaneous speech ASD-1 children than ASD -2 и TD children, and in repetition words ASD-2 children were revealed.

The formant triangles of vowels from the words from the spontaneous speech of ASD children were shifted on the two-formant plane to the higher-frequency region (for vowels /a/ and /u/, and F1 for vowel /i/) as compared with the formant triangles of vowels from repetition words (Fig. 3A). The largest shifts in the values of the first two formants of the vowels /a/, /i/, and /u/, leading to displacement of the formant triangles into the higher-frequency area, were seen for the vowels of ASD -1 children vs. ASD-2 peers (Fig. 3B). The differences in the location on the two coordinate plot of formant triangles from the twice types of speech for TD children was not revealed. Comparison of vowel formant triangle areas showed that areas were greatest for the vowels of TD children's repetition speech (Fig. 3C) and ASD-1 children's repetition speech (Fig. 3D).
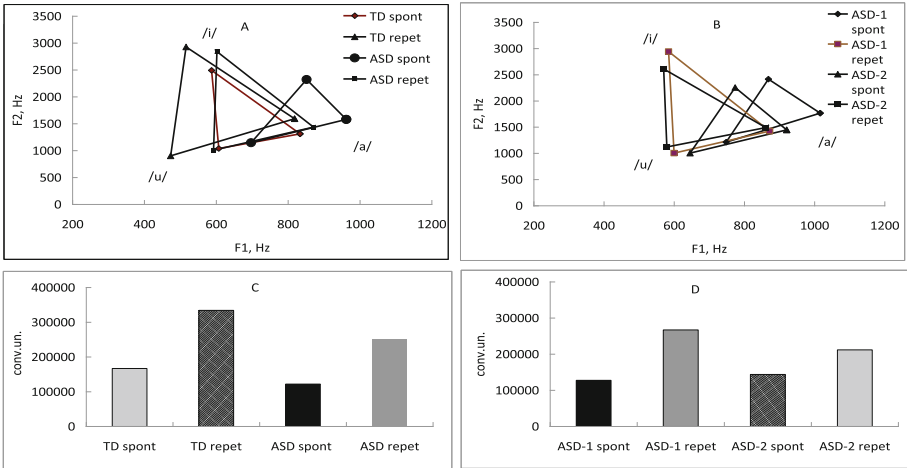

Fig. 3. The vowels formant triangles with apexes /a/, /u/, /i/: A - of ASD and TD children in twice speech tasks (spontaneous and repetition), B - of ASD children (ASD-1 - group 1, ASD-2 - group 2). Horizontal axis values are F1, Hz, vertical axis values are F2, Hz. Areas of vowels formant triangles (in conventional unit): C - data for TD and ASD child's areas of vowels formant triangles in twice speech tasks, D - ASD children (ASD-1 - group 1, ASD-2 - group 2).

These data indicate a clearer articulation of repetition words vs. word articulation in spontaneous speech.

The values of the normalized intensities of formants in repetition speech demonstrate a distribution pattern similar to that in spontaneous speech.

4   Discussion

We present the first data for Russian ASD children of acoustic measures of participant's speech. This study has shown that the ASD children differ from TD children on the base of higher values of pitch, pitch values variability and formant characteristics. These acoustic features and well-marked high-frequency in spectrum more clearly presented in the speech of the first group ASD children than the second group ASD children. ASD children from the first group have diagnosis ASD (F 84.0) as a primary. A common data about the high values of pitch and pitch variability of children with ASD was obtained on the base of three complementary experiments. Our data confirm other studies with similar results [5, 6, 14, 15]. Contrary to the common impression of monotonic speech in autism, the ASD children had a significantly larger pitch range and variability across time. These results indicate that speech abnormalities in ASD are reflected in their spectral content and pitch variability [6]. Paul et al. [15] reported prosodic deficits in only 47 % of the 30 adult speakers with HFA studied. They compared participants with HFA and a typical control group on both the perception and production of a range of specific prosodic elements. Results suggested between-group differences in both the perception and production of prosodic stress, suggesting that both understanding and producing appropriate stress patterns appear to be difficult for participants with HFA [15].

The current results are one of the first steps toward developing speech based biomarkers for early diagnosis of ASD. We believe that the acoustic features of speech of children with different neurological state are perspective for early diagnosis of developmental risk.

5   Conclusions

Differences between children with ASD and TD on the basis of higher values of pitch, pitch variability and formant characteristics of ASD children were revealed. In general, for all children with ASD voice and speech is characterized by high values of pitch and pitch variability. These acoustic features and well-marked high-frequency in spectrum were more clearly presented in the speech of the first group ASD children than the second group ASD children.

References

1. Kanner, L.: Autistic disturbances of affective contact. Nerv. Child **2**, 217–250 (1943)
2. Tager-Flusberg, H.: On the nature of linguistic functioning in early infantile autism. J. Autism Dev. Disord. **11**, 45–56 (1981)
3. Diehl, J., Paul, R.: Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems by children with autism spectrum disorders. Appl. Psychol. **34**, 135–161 (2013)
4. Grossman, R.B., Bemis, R.H., Plesa, S.D., Tager-Flusberg, H.: Lexical and affective prosody in children with high-functioning autism. J. Speech Lang. Hear. Res. **53**, 778–793 (2010)
5. Nadig, A., Shaw, H.: Acoustic and perceptual measurement of expressive prosody in high-functioning autism: increased pitch range and what it means to listeners. J. Autism Dev. Disord. **42**, 499–511 (2012)
6. Bonneh, Y.S., Levanov, Y., Dean-Padro, O., Lossos, L., Adini, Y.: Abnormal speech spectrum and increased pitch variability in young autistic children. Front. Hum. Neurosci. **4** (2011) doi:10.3389/fnhum.2010.00237
7. Shriberg, L.D., Paul, R., McSweeny, J.L., Klin, A., Cohen, D.J., Volkmar, F.R.: Speech and prosody characteristics of adolescents and adults with high functioning autism and asperger syndrome. J. Speech Lang. Hear. Res. **44**, 1097–1115 (2001)
8. Shriberg, L.S., Paul, Rh., Black, L.M., van Santenc, J.P.: The hypothesis of apraxia of speech in children with autism spectrum disorder. J. Autism Dev. Disord. **41**(4), 405–426 (2011)
9. Nakai, Y., Takashima, R., Takiguchi, T., Takada, S.: Speech intonation in children with autism spectrum disorder. Brain Dev. **36**(6), 516–522 (2014)
10. Schopler, E., Reichler, R.J., DeVellis, R.F., Daly, K.: Toward objective classification of childhood autism: childhood autism rating scale (CARS). J. Autism Dev. Disord. **10**(1), 91–103 (1980)
11. Lyakso, E., Frolova, O., Dmitrieva, E., Grigorev, A., Kaya, H., Salah, A.A., Karpov, A.: EmoChildRu: emotional child russian speech corpus. In: Ronzhin, A., Potapova, R., Fakotakis, N. (eds.) SPECOM 2015. LNCS, vol. 9319, pp. 144–152. Springer, Heidelberg (2015)
12. Vorperian, H., Kent, R.D.: Vowel acoustic space development: a synthesis of acoustic and anatomic data. J. Speech Lang. Hear. Res. **50**(6), 1510–1545 (2007)
13. 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. Behav. Physiol. **45**(5), 558–567 (2015)
14. Sharda, M., Subhadra, T.P., Sahaya, S., Nagaraja, Ch., Singh, L., Sen, A., Singhal, N., Erickson, D., Singh, N.: Sounds of melody — pitch patterns of speech in autism. Neurosci. Lett. **478**, 42–45 (2010)
15. Paul, R., Augustyn, A., Klin, A., Volkmar, F.: Perception and production of prosody by speakers with autism spectrum disorders. J. Autism Dev. Disord. **35**, 205–220 (2005)