Recognition by the Voice and Speech Features of the Emotional State of Children by Adults and Automatically

Authors : Elena E. Lyakso, Olga V. Frolova, Yuri N. Matveev, Aleksey S. Grigorev, Alexander S. Nikolaev, Viktor A. Gorodnyi

Abstract : The study of the children's emotional sphere depending on age and psychoneurological state is of great importance for the design of educational programs for children and their social adaptation. Atypical development may be accompanied by violations or specificities of the emotional sphere. To study characteristics of the emotional state reflection in the voice and speech features of children, the perceptual study with the participation of adults and the automatic recognition of speech were conducted. Speech of children with typical development (TD), with Down syndrome (DS), and with autism spectrum disorders (ASD) aged 6-12 years was recorded. To obtain emotional speech in children, model situations were created, including a dialogue between the child and the experimenter containing questions that can cause various emotional states in the child and playing with a standard set of toys. The questions and toys were selected, taking into account the child's age, developmental characteristics, and speech skills. For the perceptual experiment by adults, test sequences containing speech material of 30 children: TD, DS, and ASD were created. The listeners were 100 adults (age 19.3 ± 2.3 years). The listeners were tasked with determining the children's emotional state as "comfort – neutral – discomfort" while listening to the test material. Spectrographic analysis of speech signals was conducted. For automatic recognition of the emotional state, 6594 speech files containing speech material of children were prepared. Automatic recognition of three states, "comfort – neutral – discomfort," was performed using automatically extracted from the set of acoustic features - the Geneva Minimalistic Acoustic Parameter Set (GeMAPS) and the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS). The results showed that the emotional state is worse determined by the speech of TD children (comfort – 58% of correct answers, discomfort – 56%). Listeners better recognized discomfort in children with ASD and DS (78% of answers) than comfort (70% and 67%, respectively, for children with DS and ASD). The neutral state is better recognized by the speech of children with ASD (67%) than by the speech of children with DS (52%) and TD children (54%). According to the automatic recognition data using the acoustic feature set GeMAPSv01b, the accuracy of automatic recognition of emotional states for children with ASD is 0.687; children with DS – 0.725; TD children – 0.641. When using the acoustic feature set eGeMAPSv01b, the accuracy of automatic recognition of emotional states for children with ASD is 0.671; children with DS – 0.717; TD children – 0.631. The use of different models showed similar results, with better recognition of emotional states by the speech of children with DS than by the speech of children with ASD. The state of comfort is automatically determined better by the speech of TD children (precision – 0.546) and children with ASD (0.523), discomfort – children with DS (0.504). The data on the specificities of recognition by adults of the children's emotional state by their speech may be used in recruitment for working with children with atypical development. Automatic recognition data can be used to create alternative communication systems and automatic human-computer interfaces for social-emotional learning. Acknowledgment: This work was financially supported by the Russian Science Foundation (project 18-18-00063).

Keywords : autism spectrum disorders, automatic recognition of speech, child's emotional speech, Down syndrome, perceptual experiment

Conference Title : ICCPT 2022 : International Conference on Clinical Psychology and Testing

Conference Location : Amsterdam, Netherlands

Conference Dates : January 21-22, 2022