J. Aichi Med. Univ. Assoc.
Vol. 45, No. 4, pp. 23–31 (December 15, 2017)

A Re-Analysis of Facial Expression and Voice Recognition Abilities in Children with Autism Spectrum Disorder

KEITA MATSUMOTO[1)2)], KOUSUKE KANEMOTO[2)]

[1)] *Department of Child and Adolescent Psychiatry, Osaka City General Hospital*
[2)] *Department of Psychiatry, Aichi Medical University* (*Chief Professor: Kousuke Kanemoto*)

Introduction: We elucidated the processing mechanisms of recognizing emotions in individuals with Autism Spectrum Disorders (ASD) with the aim of contributing to improvement in their social adaptability. In our previous report, we found that individuals with ASD have similar abilities as those with typical development (TD) in discriminate emotions when an emotional voice is recognized but have significantly slower reaction times, indicating a difficulty in recognizing emotions. In the present study, we re-analyzed our previous results on emotionally-toned voices with the addition of results from a facial expression recognition task. Here, we report our findings.

Materials and Method: The behavioral characteristics of facial expression and voice cognition in ASD were investigated in 12 individuals with the ASD (mean age=11.83 years) and 12 matched typical development (TD) controls (mean age=11.58 years). Participants were asked to judge the emotional valence of facial expression (the facial expression task), prosody (the emotion task) and verbal content (the semantic task) stimuli.

Results: Accuracy was not significantly different between the two groups in any of the tasks regardless of the emotion stimulus. Regarding the reaction time, the two groups were not significantly different in the semantic task but were significantly different in the emotion recognition and facial expression recognition tasks. Moreover, correlation analysis comparing the reaction times for each task showed that reaction times were significantly correlated between all tasks in TD but were not significantly correlated between the facical expression task and the semantic task in ASD.

Conclusion: To communicate smoothly with individuals with ASD, it is important to understand their recognition of emotions and to pay careful attention to the speed of conversation and the amount of information presented.

Key words: Autism Spectrum Disorder, recognition of emotions, reaction time, accuracy

INTRODUCTION

When we communicate with other individuals, we are not simply exchanging words or semantics. In addition to the literal verbal message, we utilize nonverbal messages with various emotions through different facial expressions or tone of voice to deepen mutual understandings. Fisher (2004) stated that social relationships cannot form without emotions, indicating the necessity to appropriately feel other people's emotions and to match one's own emotions in order to maintain a good relationship[1)]. Our lives are not complete without emotion; some may even claim that our lives are ruled by emotions.

Autism Spectrum Disorder (ASD) is a neuro-

Received January 30, 2018; Accepted February 15, 2018

logical condition characterized by 1) persisting impairment in mutual social communication or interpersonal interaction, and 2) limited and repetitive forms of behavior, interest, or activity[2]. It has been demonstrated that an individual with ASD, even if he/she is highly intellectual, may misread other people's intentions and may present problems in communicating; for example, he/she may interpret a euphemistic expression verbatim[3].

Difficulty in recognizing and empathizing with the intentions and emotions of others through their voices or reactions has been shown to be one of the main factors that causes issues when individuals with ASD communicate with others[2]. It has been indicated that such problems in communication abilities may prevent social success and obstruct the formation of friendships, causing individuals with ASD to become isolated[4].

Previous studies on the ability of individuals with ASD in identifying and recognizing emotions primarily involved the recognition of facial expressions[5]. Tracy et al. reported that individuals with ASD and individuals with typical development (TD) have similar abilities in discriminating different emotions[6] while Uljarevic reported that individuals with ASD have greater difficulty compared to those with TD in recognizing facial expressions[7], demonstrating that there is no consensus on this subject.

On the other hand, while there are only a few studies on emotion recognition using voice[8], it has been reported that children with ASD experience greater difficulty in recognizing different emotions in voice compared to children with TD[9] and that individuals with ASD and individuals with TD have similar cognitive abilities[10], also demonstrating that there is no consensus on this topic.

In our previous report "Behavioral Study on Emotional Voice Perception in Children with Autism Spectrum Disorder", we not only examined the ability of individuals with ASD in discriminating emotions, but also focused on the reaction times from recognition to processing of emotionally-toned voices, and compared the results between children with ASD and children with TD. We found that even though individuals with ASD have similar abilities as those with TD in discriminating different emotions, their reaction time for responding is significantly delayed, indicating the difficulties they encounter in recognizing emotions[11]. In the present study, we included data on facial expression recognition in addition to the previously-reported voice data. Here, we report the results of this re-analysis.

MATERIALS AND METHODS

1) Participants

Twelve right-handed children diagnosed with ASD (age 10〜15 years old, six boys and six girls) participated in this study. These participants were randomly selected from patients who were receiving care at the Aichi Children's Health and Medical Center on an outpatient basis, were diagnosed with high-functioning (intelligent quotient (IQ)＞70) ASD, were not experiencing major problems in daily life, and were not undergoing treatment with medication. The author interviewed all children, and Diagnostic and Statistical Manual of Mental Disorders-forth edition and the Pervasive Developmental Disorders Rating Scale (PARS) were used to diagnose ASD anew[12]. PARS is the standard scale for evaluating ASD in Japan with established reliability and validity, and it has also been reported to be positively correlated with Autism Diagnostic Interview-Revised[13]. Twelve right-handed children were

Table 1. Participant demographics

	ASD	TD	Statistical group differences (Student t-test)
Participants (Male)	12 (6)	12 (6)	
Age (SD)	11.8 (1.8)	11.5 (1.6)	$p=0.736$
FIQ (SD)	96.9 (13.6)	106.7 (9.9)	$p=0.057$
VCI (SD)	97.5 (13.7)	106.0 (17.8)	$p=0.209$
PRI (SD)	97.4 (14.6)	103.6 (11.5)	$p=0.258$

Values represent mean (Standard Deviation). *ASD*, Autism Spectrum Disorder; *TD*, Typical Development; *FIQ*, Full Scale IQ; *VCI*, Verbal Comprehension Index; *PRI*, Perceptual Reasoning Index.

recruited from nearby residents as a typical development group. The author interviewed all participants in both groups to confirm that they had normal vision with or without correction, and that they did not have any comorbid hearing impairments or other mental disorders. There were no statistically significant differences between the two groups in the male-to-female ratio, IQ, or age ($p=0.736$) (Table 1). IQ was measured using Wechsler Intelligence Scale for Children (WISC) III or IV. There were no significant differences in Full scale IQ (FIQ) or in subtests of the Verbal Comprehension Index (VCI) and Perceptual Reasoning Index (PRI) between the two groups (FIQ: $p=0.057$, VCI: $p=0.209$, PRI: $p=0.258$).

This study was designed based on the Declaration of Helsinki and was approved by the Ethics Committee at the Aichi Children's Health and Medical Center. Prior to the experiment, detailed explanations of the study were provided to all participants and their parents/guardians, and the study was conducted after obtaining written consent from both parties.

2) Creation of Auditory Stimulus Sets

Using the Nippon Telegraph and Telephone Corporation (NTT) Database Series "Nihongo-no Goi-Tokusei (Lexical Properties of Japanese)"[14], 60 nouns with three moras and six verbs with six moras with similar degrees of familiarity were selected. The NTT Database Series comprehensively manages nearly all Japanese words and is a Japanese language database that is used in research and information analysis. Familiarity is a subjectively assessed value pertaining to the degree in which humans are familiar with a specific word. Familiarity scoring was established by Amano et al. and is categorized into four levels (Max 7.0; 7.0〜5.5, 5.5〜4.0, 4.0〜2.5, 2.5〜1.0)[15]. In the present study, words and phrases that are thought to be highly recognizable especially by children were selected from words with the greatest familiarity (score of 7.0〜5.5). Mora is a unit of phonology in the Japanese language, and one kana (syllabic script) character equates to one mora. The above words were read by a professional female announcer whose native language is Japanese with three different emotions (happy, angry, and neutral) and were recorded. The recorded vocal audio files were processed on a personal computer with Sound Engine Free 4.60 (Coderium, Sapporo, Japan) to equalize the volume. Furthermore, each word was extracted as a 1100-msec fragment by aligning the audio (voice) start times. To validate whether or not words are accurately recognized, healthy adults without hearing impairment (mean age of 21.1 years old, four men and 17 women) were asked to write down the emotion and semantics of the stimuli presented. Forty-five words with an accuracy of ≥95% were ultimately chosen as the voice stimuli. In addition, illustrations [400 (height)×600 (width) pixels] that matched the semantics of the word were created as visual stimuli.

3) Creation of Visual Stimulus Sets

Seventeen staff members at the author's work place were recruited and their headshots were taken with careful consideration to protect their privacy and with the condition that the photos will not be used for any purpose other than research. Headshots were taken with different facial expressions corresponding to three emotions (happy, angry, and neutral), similarly to the auditory stimuli. A total of 51 photos were processed with Photoshop using a personal computer to adjust the face in each photo to the same size, and were saved as bmp files. To validate whether or not the emotions of the headshots taken recognized accurately, eight healthy adults (four men and four women) without visual impairment were asked to verbally discriminate the emotion presented. Forty-five photos with accuracy of 100% were used as visual stimuli.

4) Procedure

The experiment was conducted in a quiet laboratory with one individual at a time. A 23-inch liquid crystal display (LCD) monitor (RDT233WX, Mitsubishi) was placed on a desk, and participants were asked to sit approximately 40 cm from the monitor at a viewing angle of approximately 40°.

The experiment consisted of a task to identify different facial expressions, a task to determine voices spoken with different emotions, and a task to decipher the semantics of the words/sentences spoken with different emotions.

The presentation of stimuli was controlled by a personal computer using Presentation (Neurobehavioral Systems, Inc., USA), and the voices were presented using audio speakers (Companion® 2 Series II Multimedia Speaker System, BOSE Corporation) placed next to the LCD monitor. The volume was set such that the voices could be heard clearly but did not cause discomfort to the participants.

Three custom-made buttons were used to measure the reaction, and participants were instructed to press the buttons with their right hand. Prior to the actual experiment, practice sessions were conducted thoroughly using stimuli different from the ones used in the experiment to familiarize the participants with appropriately maneuvering the buttons.

Emotion task

Sentences were heard in a sequence such that the different emotions were presented in a random order. Participants were asked to discriminate the emotion expressed by the voice and respond using pre-assigned buttons that were arranged in the following manner: "happy" on the left, "neutral" in the middle, and "angry" on the right. Reaction time was measured with the start of voice presentation as the zero point, and the participants were asked to respond at the time in which they determined the answer, even if that occurred during the voice presentation.

Semantic task

Similarly to the emotion task, sentences were presented using an emotional tone, and three illustrations that correspond to the semantics of the sentence were displayed on the monitor 0.3 s later. Participants were asked to select the illustration that matched the voice using the button placed in front of each illustration. Reaction time was measured as the time between visual stimuli presentation (zero point) and when the participants pressed the button

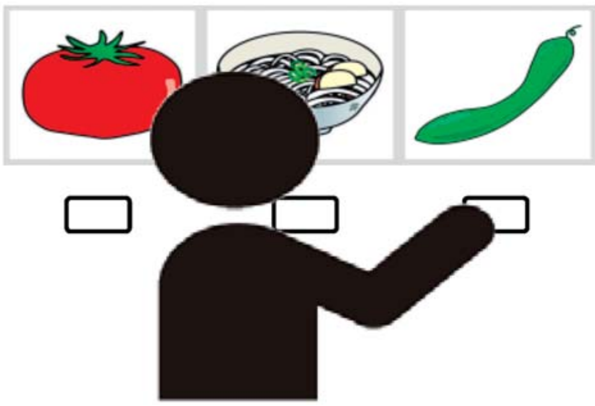


Figure 1. Semantic Task (presented illustrations)
Ex) cucumber (presented with emotional voice)
A word was spoken with one of 3 emotions (happy, neutral, angry). At the same time, an illustration showing the meaning of the word or SVO sentence was displayed on a computer screen, and subjects were asked to press the button of the picture that they thought was correct (3 options).

(Figure 1).

Facial expression task

Headshots with different facial expressions were displayed on the monitor in a sequence such that the different emotions were presented in random order. Participants were asked to discriminate the emotion for the stimuli presented on the monitor and answer accordingly using the buttons pre-labeled with three types of emotions and placed in front of the monitor. Reaction time was measured as described above.

5) Analysis

Reactions to word stimuli from the emotion task and semantic task were extracted, combined with the facial expression task, and compared between the two groups. Regarding the reaction time, three-way analysis of variances (ANOVAs) were performed using three factors—task, emotion, and condition (ASD or TD)—and the three-way interaction as well as simple interaction and main effects were determined. Sphericity was also validated at the same time using Mauchly's test, and Greenhouse-Geisser corrections were applied when the results were significant. The Bonferroni correction was used for post-hoc analysis. The effect size was evaluated with η^2_p. SPSS statistics 21 (IBM, New York, USA) was used for analyses.

RESULTS

The results on accuracy and reaction time are shown in Table 2. Mean accuracy was ≥80% regardless of the emotion stimuli in any of the tasks, and there were no significant differences between the ASD group and TD group.

Results from the three-way ANOVA with third factors (task, emotion stimuli, condition) on reaction time are shown in Table 3. Although a three-way interaction between task × emotion stimuli × condition was not observed ($p = 0.494$, $\eta^2_p = 0.036$), a two-way interaction between task × condition was observed ($p = 0.003$, $\eta^2_p = 0.242$). Significant differences were not observed between the ASD group and TD group in the semantic task ($p = 0.308$, $\eta^2_p = 0.047$). However, significant differences in reaction time were observed between the two groups in the emotion

Table 2. Accuracy [%(SD)] and Reaction times [msec(SD)] during the tasks

		Facial expression			Emotion			Semantic		
		Neutral	Happy	Angry	Neutral	Happy	Angry	Neutral	Happy	Angry
Accuracy	ASD	92 (8)	95 (4)	80 (9)	97 (4)	99 (3)	97 (4)	100 (0)	100 (0)	100 (0)
	TD	87 (8)	95 (6)	81 (11)	97 (4)	100 (0)	99 (4)	99 (2)	100 (0)	99 (2)
Reaction time	ASD	1465 (442)	1413 (221)	1603 (370)	1322 (281)	1250 (206)	1283 (221)	1264 (196)	1230 (196)	1211 (199)
	TD	1117 (244)	1077 (151)	1163 (256)	1072 (184)	1031 (179)	1055 (249)	1192 (191)	1137 (184)	1132 (204)

Values represent mean (SD). SD: standard deviation; ASD: autism spectrum disorder; TD: typical development

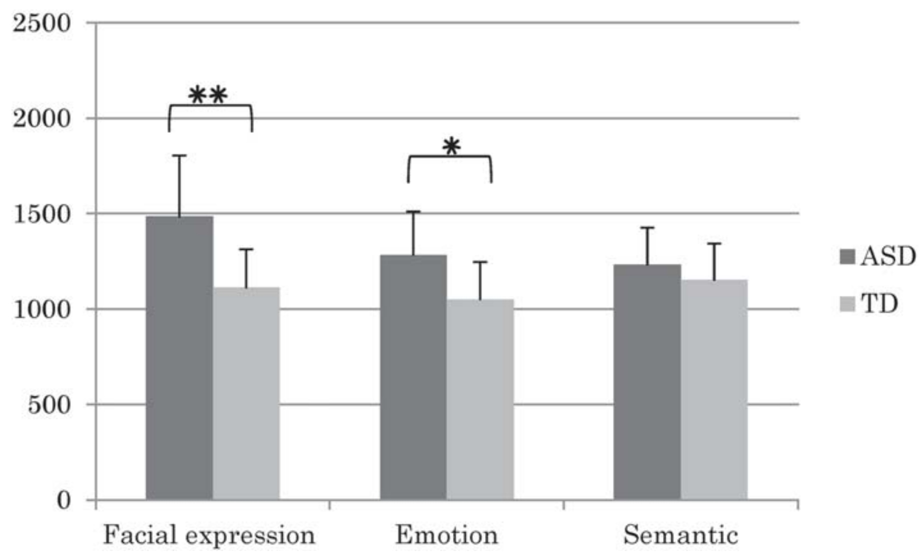
Table 3. ANOVA result

	Degree of freedom	F	p value	η^2_p
Main effect				
Condition	1	8.148	0.009**	0.270
Task	1.876	6.939	0.003**	0.240
Emotion	1.902	4.229	0.023*	0.161
3-way interaction				
Condition × Task × Emotion	3.397	0.830	0.494	0.036
2-way interaction				
Condition × Task	1.876	7.006	0.003**	0.242
Condition × Emotion	1.902	0.383	0.673	0.017
Task × Emotion	3.397	0.830	0.002	0.185

*< 0.05 **< 0.01

recognition task and in the facial expression recognition task (emotion recognition, $p = 0.014$, $\eta^2_p = 0.246$; facial expression recognition, $p = 0.002$, $\eta^2_p = 0.348$) (Figure 2). The differences were 233 msec for the emotion task (ASD mean, 1284 ± 225 msec; TD mean, 1050 ± 195 msec) and 372 msec for the facial expression task (ASD mean, 1485 ± 318 msec; TD mean, 1113 ± 199 msec).

Next, the results on the correlation between task and reaction time by condition are shown in Table 4. In the TD group, extremely strong correlations in reaction times were observed between the facial expression task and emotion task ($r = 0.8954$, $p < 0.001$) and between the emotion task and semantic task ($r = 0.8404$, $p < 0.001$), and a strong correlation in reaction time was also observed between the facial expression task and semantic task ($r = 0.7654$, $p < 0.05$). However, in the ASD group, while there



**< $p = 0.01$ *< $p = 0.05$

Figure 2. Difference in reaction time between both groups in tasks of facial expression, emotion, semantic
The averaged reaction times to all emotions for each tasks are displayed.

Table 4. Pearson's Correlation Analysis results from ASD and the TD

	ASD			TD		
	Facial expression	Emotion	Semantic	Facial expression	Emotion	Semantic
Facial expression	1.0000 —	0.6208 (p=0.0312*)	0.4565 (p=0.1357)	1.0000 —	0.8954 (p=0.00008**)	0.7654 (p=0.0037**)
Emotion	0.6208 —	1.0000 —	0.6269 (p=0.0291*)	0.8954 —	1.0000 —	0.8404 (p=0.0006**)
Semantic	0.4565 —	0.6269 —	1.0000 —	0.7654 —	0.8404 —	1.0000 —

*＜0.05 **＜0.01

were strong correlations in reaction times between the facial expression task and emotion task ($r=0.6208$, $p<0.05$) and between the emotion task and semantic task ($r=0.6269$, $p<0.05$), there was no significant correlation in reaction time between the facial expression task and semantic task ($r=0.4565$, $p=0.135$).

DISCUSSION

The purpose of the present study was to clarify the processing mechanism for recognizing emotions in individuals with ASD with the aim of contributing to improvements in their social adaptability. Although researchers have previously studied emotion recognition in individuals with ASD, there are very few reports that have compared facial expression and auditory recognition at the same time. We therefore believe that this report will facilitate expanding new knowledge on the tendencies of ASD-specific emotion recognition.

In the present study, we re-analyzed the results from the word sessions in the emotion task and semantic task from our previous report "Behavioral Study on Emotional Voice Perception in Children with Autism Spectrum Disorder" by adding results from a facial expression task. In all of the tasks, there were no differences between the groups in accuracy. These results suggested that individuals with ASD and individuals with TD have similar abilities in discriminating emotions, corroborating the report by Tracy[6]. However, when examining the reaction time, although there were no differences in the semantic task between the two groups, the reaction time was significantly slower in those with ASD compared to those with TD for the facial expression task and emotion task. This suggested that individuals with ASD have difficulties in "reading" emotions, both through facial expression and through voice. This result was consistent with the report by Uljarevic[7] and may be associated with the "mindblindness" hypothesis presented by Baron-Cohen[16]. The mindblindness hypothesis claims that individuals with ASD present a developmental delay in the ability to interpret the intentions or emotions of others through their actions and behaviors, causing social and communication impairment[16]. Furthermore, an association with impairment in mirror neurons was also indicated. As typically-developed individuals, we have a tendency to naturally imitate the behavior of others when we observe them (mirror system), and this characteristic reportedly can also apply to emotions. It has been asserted that individuals with ASD present an impairment to this mirror system and therefore cannot empathize with the intrinsic intentions or emotions of others, causing social problems[17]. Our results from the present experiment showed no difference in reaction time between the two groups for the semantic task, which does not involve deciphering emotions, signifying that individuals with ASD do not present difficulties in recognizing verbal messages.

The comparison of reaction times between each task using correlation analysis showed that all tasks were significantly correlated in TD, but the facial expression task and semantic task were not significantly correlated in ASD. In TD, the individual differences in reaction time were significantly correlated regardless of the type of task. In ASD, however, the individual differences in reaction time were different depending on the type of task. Individual differences in reaction time in cognitive tasks have been drawing attention as Intra-Individual Variability (IIV) in recent years. IIV is related to the activity of frontal cortex[18], Children with developmental disorders, such as ADHD and ASD show increased levels of intra-individual variability[19]. Our results also supported this indication. There is a report that the reaction time and IIV correlated positively with social problems[20], it is expected that accumulation of research.

There are several significant limitations to this study. (1) The sample size was small. We determined η^2_p and verified that the effect size was statistically adequate. Nonetheless, further investigation with a greater sample size is desirable. (2) The mean accuracy was ≥80% for all tasks, and the sensitivity of accuracy evaluation may not be sufficient due to a ceiling effect. (3) In the present study, we were not able to use ADI-R and the Autism Diagnostic Observation Schedule (Lord et al. 2000), which are standard instruments for the confirmatory diagnosis of ASD, as their usage for research purposes is restricted within Japan. Also, the participants were not administered the same IQ test, as both WISC-III or WISC-IV were used.

CONCLUSION

Our results indicated that, in order to communicate smoothly with individuals with ASD, it is important to understand that those with ASD are better at recognizing verbal messages than nonverbal messages. Furthermore, it is necessary to pay thorough attention to the speed of the conversation and the amount of information presented. In the future, it is essential to increase the sample size and investigate complex emotions such as disdain and temptation as well as communication expressions that are likely to occur in real life situations.

Acknowledgments

We thanks Toshiro Sugiyama, Chiaki Saito, Shiho Kato and Akinori Nakamura for their helpful.

REFERENCES

1) Fischer AH, Rotteveel M, Evers C, Manstead A.S.R. Emotional assimilation: How we are influenced by others' emotions. Cahier de Psychologie Cognitive 2004; 22: 223–45.
2) American Psychiatric Association. Diagnostic and Statistical Manual of mental disorders (5th edition.). Washington, DC: APA Press, 2013.
3) Ozonoff S, Miller JN. An exploration of right-hemisphere contributions to the pragmatic impairments of autism. Brain Lang 1996; 52(3): 411–34.
4) Rutter M. Cognitive deficits in the pathogenesis of autism. Journal of Child Psychology and Psychiatry 1983; 24(4): 513–31.
5) Weigelt S, Koldewyn K, Kanwisher N. Face identity recognition in autism spectrum disorders: A review of behavioral studies. Neuroscience and Biobehavioral Reviews 2012; 36: 1060–84.
6) Tracy JL, Robins RW, Schriber RA, Solomon M. Is emotion recognition impaired in individuals with autism spectrum disorders? Journal of Autism and Developmental Disorder 2011; 41(1): 102–9.

7) Uljarevic M, Hamilton A. Recognition of emotions in autism: A formal meta-analysis. Journal of Autism and Developmental Disorders 2013; 43(7): 1517–26.
8) Lartseva A, Dijkstra T, Buitelaar JK. Emotional language processing in autism spectrum disorders: A systematic review. Front Hum Neurosci 2015; 8: 991.
9) Golan O, Baron-Cohen S, Hill JJ, Rutherford MD. The 'reading the mind in the voice' test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. J Autism Dev Disord 2007; 37(6): 1096–106.
10) Jones CR, Pickles A, Falcaro M, et al. A multimodal approach to emotion recognition ability in autism spectrum disorders. J Child Psychol Psychiatry 2011; 52(3): 275–85.
11) Matsumoto K, Sugiyama T, Saito C, Kuriyama K, Kanemoto K, et al. Behavioral study on emotional voice perception in children with autism spectrum disorder. Journal of Pediatric Neuro-Psychology 2016; 3: 108–18.
12) Autism Society Japan. Pervasive Developmental Disorders Autism Society Japan Rating Scale. Tokyo: Spectrum Publishing Company, 2008.
13) Ito H, Tani I, Yukihiro R, Adachi J, Hara K, Ogasawara M, et al. Validation of an interview-based rating scale developed in japan for pervasive developmental disorders. Research in Autism Spectrum Disorders 2012; 6(4): 1265–72.
14) Amano S, Kondo T. Nihongo-no Goi-Tokusei (Lexical properties of Japanese). Tokyo: Sanseido, 1999.
15) Amano, S, & Kondo T. (2000). On the NTT psycholinguistic databases "Lexical Properties of Japanese". Journal of the Phonetic Society of Japan 4(2): 44–50.
16) Baron-Cohen S. The Essential Difference: Male and Female Brains and the Truth about Autism. London: Penguin, 2004.
17) Frith U. Autism—A very short introduction, first edition. New York: Oxford University Press, 2008.
18) Karalunas SL, Geurts HM, Konrad K, Bender S, Nigg JT. Annual Research Review: Reaction time variability in ADHD and autism spectrum disorders: measurement and mechanisms of a proposed trans-diagnostic phenotype. The Journal of Child Psychology and Psychiatry 2014; 55(6): 685–710.
19) van Belle J. van Hulst BM, Durston S. Developmental differences in intra-individual variability in children with ADHD and ASD. The Journal of Child Psychology and Psychiatry 2015; 56(12): 1316–26.
20) Høyland AL, Nærland T, Engstrøm M, Lydersen S, Andreassen OA. The relation between face-emotion recognition and social function in adolescents with autism spectrum disorders: A case control study. PLoS One 2017; 12(10).