

# Affect bursts to constrain the meaning of the facial expressions of the humanoid robot Zeno

Bob R. Schadenberg, Dirk K. J. Heylen, and Vanessa Evers

Human-Media Interaction, University of Twente, Enschede, the Netherlands  
{b.r.schadenberg, d.k.j.heylen, v.evers}@utwente.nl

**Abstract.** When a robot is used in an intervention for autistic children to learn emotional skills, it is particularly important that the robot’s facial expressions of emotion are well recognised. However, recognising what emotion a robot is expressing, based solely on the robot’s facial expressions, can be difficult. To improve the recognition rates, we added affect bursts to a set of caricatured and more humanlike facial expressions, using Robokind’s R25 Zeno robot. Twenty-eight typically developing children participated in this study. We found no significant difference between the two sets of facial expressions. However, the addition of affect bursts significantly improved the recognition rates of the emotions by helping to constrain the meaning of facial expression.

**Keywords:** emotion recognition, affect bursts, facial expressions, humanoid robot.

## 1 Introduction

The ability to recognise emotions is impaired in individuals with Autism Spectrum Condition [1], a neurodevelopmental condition characterised by difficulties in social communication and interaction, and behaviour rigidity [2]. Recognising emotions is central to success in social interaction [3], and due to impairment in this skill, autistic individuals often fail to accurately interpret the dynamics of social interaction. Learning to recognise the emotions of others may provide a toehold for the development of more advanced emotion skills [4], and ultimately improve social competence.

In the DE-ENIGMA project, we aim to develop a novel intervention for teaching emotion recognition to autistic children with the help of a humanoid robot – Robokind’s R25 model called Zeno. The intervention is targeted at autistic children who do not recognise facial expressions, and who may first need to learn to pay attention to faces and recognise the facial features. Many of these children will have limited receptive language, and may have lower cognitive ability. The use of a social robot in an intervention for autistic children is believed to improve the interest of the children in the intervention and provide them with a more understandable environment [5].

The emotions that can be modelled with Zeno’s expressive face can be difficult to recognise, even by typically developing individuals [6–8]. This can be

partly attributed to the limited degree's of freedom of Zeno's expressive face, resulting in emotional facial expressions that may not be legible, but more importantly because facial expressions are inherently ambiguous when they are not embedded in a situational context [9]. Depending on the situational context, the same facial expression can signal different emotions [10]. However, typically developing children start using the situational cues to interpret facial expression consistently around the age of 8 or 9 [11]. Developmentally, the ability to use the situational context is an advanced step in emotion recognition, whereas many autistic children still need to learn the basic steps of emotion recognition. To this end, we require a developmentally appropriate manner to constrain the meaning of Zeno's facial expressions during the initial steps of learning to recognise emotions.

In the study reported in this paper, we investigate whether a multimodal emotional expressions lead to a better recognition rates by typically developing children than unimodal facial expressions. We tested two sets of facial expressions, with and without non-verbal vocal expressions of emotion. One set of facial expressions was designed by Salvador, Silver, and Mahoor [7], while the other set is Zeno's default facial expressions provided by Robokind. The latter are caricatures of human facial expressions, which we expect will be easier to recognise than the more realistic humanlike facial expressions of Salvador et al. [7]. Furthermore, we expect that the addition of non-verbal vocal expressions of emotion will constrain the meaning of the facial expressions, making them easier to recognise.

## 2 Related Work

Typically developing infants initially learn to discriminate the affect of another through multimodal stimulation [12], which is one of the first steps in emotion recognition. Discriminating between affective expressions through unimodal stimulation develops afterwards. Multimodal stimulation is believed to be more salient to young infants and therefore more easily draws their attention. In the design of legible robotic facial expressions, multimodal expressions are often used to improve recognition rates. Costa et al. [6] and Salvador et al. [7] added emotional gestures to constrain the meaning of the facial expressions of the Zeno R50 model, which has a face similar to the Zeno R25 model, and validated them with typically developing individuals. The emotions joy, sadness, and surprise seem to be well recognised by typically developing individuals, with recognition rates of over 75%. However, the emotions anger, fear, and disgust were more difficult to recognise with recognition rates ranging from 45% to the point of guessing (17%). While the recognition rates improved by the addition of gestures for Costa et al. [6], they showed a mixed result for Salvador et al. [7], where the emotional gestures improved the recognition of some emotions and decreased the recognition in others.

The ability of emotional gestures to help constrain the meaning of facial expressions of emotions is dependent on the body design of the robot. Whereas

the Zeno R50 model can make bodily gestures that resemble humanlike gestures fairly well, the Zeno R25 model is very limited in its bodily capabilities due to the limited degrees of freedom in its body and the joints rotate differently from human joints. This makes it particularly difficult to design body postures or gestures that match humanlike expressions of emotion.

In addition to expressing emotions through facial expressions, bodily postures, or gestures, emotions are also expressed using vocal expressions [13]. In human-human interaction, these vocal expressions of emotions can constrain the meaning of facial expressions [14]. A specific type of vocal expressions of emotions are affect bursts, which are defined as “short, emotional non-speech expressions, comprising both clear non-speech sounds (e.g. laughter) and interjections with a phonemic structure (e.g. “Wow!”), but excluding “verbal” interjections that can occur as a different part of speech (like “Heaven!”, “No!”, etc.)” [15, p. 103]. When presented in isolation, affect bursts can be an effective means of conveying an emotion [15, 16].

### 3 Design Implementation

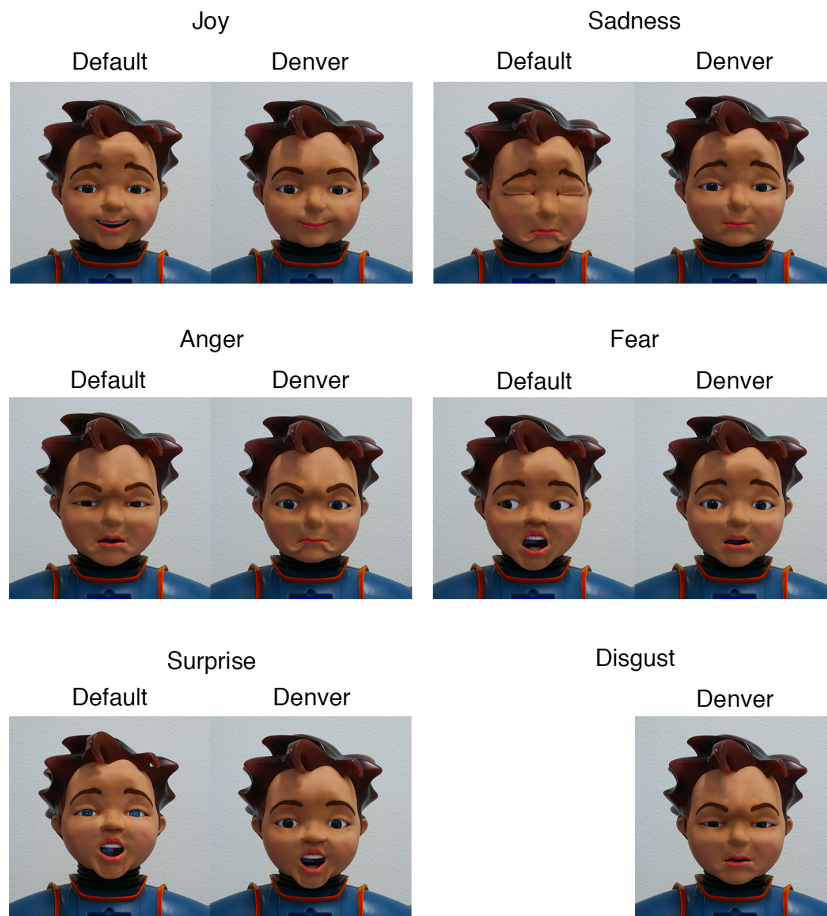
#### 3.1 Facial expressions

In this study, we used Robokind’s R25 model of the child-like robot Zeno. The main feature of this robot is its expressive face, which can be used to model emotions. It has five degrees of freedom in its face, and two in its neck.

For the facial expressions (see figure 1), we used Zeno’s default facial expressions provided by Robokind, and the facial expressions developed by Salvador et al. [7], which we will refer to as the Denver facial expressions. The Denver facial expressions have been modelled after the facial muscle movements underlying human facial expressions of emotions, as defined by the Facial Action Coding System [17], and contain the emotions joy, sadness, fear, anger, surprise, and disgust. Although the Denver facial expressions have been designed for the Zeno R50 model, the R25 has a similar face. Thus we did not have to alter the facial expressions.

Zeno’s default facial expressions include joy, sadness, fear, anger, and surprise, but not disgust. Compared to the Denver facial expressions, the default facial expressions are caricatures of human facial expressions of emotion. Additionally, the default expressions for fear and surprise also include a temporal dimension. For fear, the eyes move back and forth from one side to the other, and surprise contains eye blinks.

Both the Denver and default facial expressions last 4 seconds including a ramp-up of 0.5 seconds and returning back to the neutral emotion in 0.5 seconds. This leaves the participants with enough time to look at and interpret the facial expression.



**Fig. 1.** Zeno's default and Denver facial expressions for joy, sadness, anger, fear, surprise, and disgust. The default facial expressions do not cover disgust.

### 3.2 Affect bursts

The affect bursts<sup>1</sup> were expressed by an adult Dutch-speaking female actor. After the initial briefing, the Denver facial expressions were shown to the actor to make it easier for the actor to act being the robot. Furthermore, showing the facial expressions provided the actor with the constraints posed by the expressions. After each facial expression, the actor would express an affect burst that matches the emotion and Zeno's facial expression. The affect bursts were recorded using the on-board microphone of a MacBook Pro Retina laptop and last 0.7 to 1.3 seconds. To improve the audio quality, the affect bursts were played through a Philips BT2500 speaker placed on Zeno's back.

<sup>1</sup> <https://goo.gl/ztbMxw>

## 4 Methodology

### 4.1 Participants

The study took place during a school trip to the University of Twente where the participants could freely choose in which of several experiments to participate. The study took place in a large open room where each experiment was separated by a room divider on two sides. Of the children who joined the school trip, 28 typically developing children (19 female, and 9 male) between the ages 9 and 12 ( $M = 10.1$ ,  $SD = 0.9$ ) participated in the experiment.

### 4.2 Research design

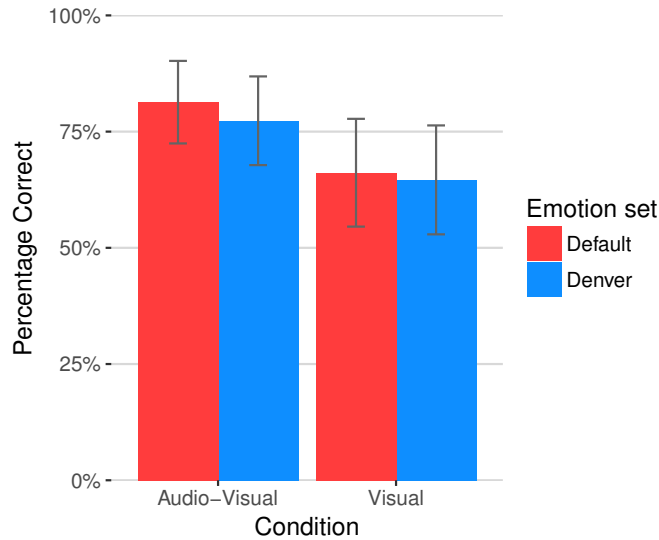
This study used a 2x2 mixed factorial design, where the set of facial expressions is a within-subject variable and the addition of affect bursts a between-subjects variable. The control (visual) condition consisted of 13 participants who only saw the facial expressions. The 15 participants in the experimental (audio-visual) condition saw the facial expressions combined with the corresponding affect bursts. All participants saw both the Denver facial expressions and the default facial expressions.

### 4.3 Procedure

The study started with the experimenter explaining the task and the goal of the study. If there were no further questions, Zeno would start by introducing itself. Next, the experiment would start and Zeno would show one emotion, which was randomly selected from either the default facial expressions or the Denver facial expressions. After the animation, Zeno returned to a neutral expression. We used a forced-choice format where the participant could choose between six emoticons, each depicting one of the six emotions, and select the emoticon they thought best represented Zeno's emotion. The emoticons of the popular messaging app *WhatsApp* were used for this task, to make the choices more concrete and interesting to children [18]. The corresponding emotion was also written below each emoticon. The same process was used for the remaining emotions, until the participant evaluated each emotion. We utilised the robot-mediated interviewing method [19] and had Zeno ask the participant three questions regarding the experiment. These questions included the participant's opinion on the experiment, which emotion he or she thought was most difficult to recognise, and whether Zeno could improve anything. Afterwards, the experimenters debriefed the participant.

## 5 Results

To calculate the main effect of the addition of affect bursts, we aggregated the emotions for the visual and for the audio-visual condition, and ran a chi-squared



**Fig. 2.** Recognition rates of Denver and default set of facial expressions, excluding disgust, for both conditions. The error bars represent the 95% confidence interval.

test which indicates a significant difference ( $\chi^2(1, N = 280) = 6.16, p = .01, \phi = .15$ ). The addition of affect bursts to the facial expressions improved the overall recognition rate of the emotions, as can be seen in figure 2. To calculate the main effect of the two sets of facial expressions, we aggregated the emotions from both sets and ran a chi-squared test. The difference was not significant ( $\chi^2(1, N = 280) = 0.16, p = .69$ ). The emotion disgust is omitted from both chi-squared tests, because only the Denver facial expressions covered this emotion.

### 5.1 Visual condition

Table 1 shows the confusion matrix for the facial expressions shown in isolation. The mean recognition rate for Zeno’s default facial expressions was 66% ( $SD = 29\%$ ). The emotions joy and sadness were well recognised by the participants with recognition rates of respectively 100% and 92%. Anger was recognised correctly by eight participants (62%), but was confused with disgust by four participants. Fear and surprise were both recognised correctly by five participants (38%). Seven participants confused fear with surprise, and surprise was confused with joy six times.

For the Denver facial expressions ( $M = 62\%, SD = 25\%$ ) both anger and joy had high recognition rates, respectively 100% and 85%. Whereas the default facial expression for surprise was confused with joy, the Denver facial expression for surprise was confused with fear instead. Vice versa, fear was confused with surprise by seven participants. Surprise and fear were correctly recognised by respectively 54% and 38%. The recognition rate for sadness was 46%, and four

**Table 1.** Perception of the facial expressions in isolation ( $n = 13$ ).

<i>Stimulus</i>	<i>% correct</i>	<i>Response</i>					
		<i>Joy</i>	<i>Sadness</i>	<i>Anger</i>	<i>Fear</i>	<i>Surprise</i>	<i>Disgust</i>
<i>Default</i>							
Joy	100%	<b>13</b>	-	-	-	-	-
Sadness	92%	-	<b>12</b>	-	-	-	1
Anger	62%	-	1	<b>8</b>	-	-	4
Fear	38%	-	-	-	5	<b>7</b>	1
Surprise	38%	<b>6</b>	-	-	1	5	1
<i>Denver</i>							
Joy	100%	<b>13</b>	-	-	-	-	-
Sadness	46%	-	<b>6</b>	1	1	1	4
Anger	85%	-	1	<b>11</b>	-	-	1
Fear	38%	-	-	1	5	<b>7</b>	-
Surprise	54%	1	-	-	4	<b>7</b>	1
Disgust	46%	-	-	<b>7</b>	-	-	6

**Table 2.** Perception of the facial expressions combined with affect bursts ( $n = 15$ ).

<i>Stimulus</i>	<i>% correct</i>	<i>Response</i>					
		<i>Joy</i>	<i>Sadness</i>	<i>Anger</i>	<i>Fear</i>	<i>Surprise</i>	<i>Disgust</i>
<i>Default</i>							
Joy	100%	<b>15</b>	-	-	-	-	-
Sadness	87%	-	<b>13</b>	-	-	-	2
Anger	87%	-	-	<b>13</b>	-	-	2
Fear	80%	-	-	-	<b>12</b>	3	-
Surprise	53%	5	-	-	1	<b>8</b>	1
<i>Denver</i>							
Joy	93%	<b>14</b>	-	-	-	-	1
Sadness	73%	-	<b>11</b>	-	-	2	2
Anger	87%	-	-	<b>13</b>	1	-	1
Fear	47%	-	1	-	<b>7</b>	7	-
Surprise	87%	-	-	-	2	<b>13</b>	-
Disgust	80%	-	2	1	-	-	<b>12</b>

confused it with disgust. Lastly, seven participants confused disgust with anger. The recognition rate for disgust was 46%.

## 5.2 Audio-visual condition

In the audio-visual condition, the facial expressions were combined with corresponding affect bursts. With the exception of surprise, all default facial expressions combined with affect bursts were recognised correctly 80% of the time or

more (see table 2). The mean recognition rate was 81% ( $SD = 17\%$ ). Surprise was recognised correctly by eight participants (53%), and confused with joy by five participants.

With the exception of fear, the Denver facial expressions combined with affect bursts had high recognition rates ranging from 73% to 93%. Taken together, these emotions had a mean recognition rate of 78% ( $SD = 17\%$ ). Fear was recognised correctly by seven participants (47%), but was confused with surprise by seven participants as well.

## 6 Discussion and Conclusion

In the study presented in this paper, we set out to determine whether affect bursts can be used effectively to help constrain the meaning of Zeno’s facial expressions. Compared to the facial expressions shown in isolation, the addition of the affect bursts increased the recognition rates by 15% on average. This constraining effect is well illustrated by the default facial expression for anger and the Denver facial expression for disgust, which look very similar to each other as can be seen in figure 1. The participants often confused these facial expressions with either anger or disgust. However, with the addition of the affect bursts, the participants were able to disambiguate the facial expressions.

Not all facial expressions were recognised well. The default facial expression for surprise was not well recognised, neither with nor without the affect burst. Surprise was often confused with joy, possibly because the facial expression also use the corners of Zeno’s mouth to create a slight smile. Additionally, the Denver facial expression for fear was often confused with surprise, regardless of the addition of the affect burst. In human emotion recognition, fear and surprise are also often confused (e.g., [20, 21]). While the affect burst for fear did help constrain the meaning of the default facial expression of fear, it failed to do so in combination with the Denver facial expression of fear. Salvador et al. [7] also reported low recognition rates for the Denver facial expression of fear. However, with the addition of an emotional gesture, they were able to greatly improve the recognition rate of fear.

While we expected that caricatured default facial expressions of emotion would be more easy to recognise than more humanlike Denver facial expressions, we did not find such a difference. Nevertheless, there are differences between the sets on specific facial expressions. Of the six emotions, only the facial expression for joy was well recognised in both sets. As well as joy, the default facial expressions for sadness was well recognised, along with the Denver facial expressions of anger. The other facial expressions were ambiguous in their meaning and require additional emotional information to be perceived correctly.

In light of an intervention that aims to teach autistic children how to recognise emotions, there is also a downside to expressing emotions using two modalities. The autistic children may rely solely on the affect bursts for recognising emotions, and not look at Zeno’s facial expression. If this is the case, they will not learn that a person’s face can also express emotions and how to recognise them. For



those children, additional effort is needed in the design of the intervention to ensure that they do pay attention to Zeno’s facial expressions.

For future research, we aim to investigate whether the addition of affect bursts also helps constrain the meaning of the facial expressions for autistic children. While typically developing children can easily process multimodal information, it may be difficult for autistic children [22, 23], which may reduce the effect of the addition of the affect bursts found in our study. Conversely, Xavier et al. [24] reported an improvement in the recognition of emotions when both auditory and visual stimuli were presented.

While we found differences in recognition rates for specific facial expressions between the default facial expressions and the Denver facial expressions, we did not find an overall difference in recognition rate between these two sets of facial expressions. We conclude that when Zeno’s facial expressions are presented in isolation, the emotional meaning is not always clear, and additional information is required to disambiguate the meaning of the facial expression. Affect bursts can provide a developmentally appropriate manner to help constrain the meaning of Zeno’s facial expressions, making them more easy to recognise.

## Acknowledgement

We are grateful to Michelle Salvador, Sophia Silver and Mohammad Mahoor for sharing their facial expressions for Zeno R50 with us. This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 688835 (DE-ENIGMA).

## References

1. Uljarevic, M., Hamilton, A.: Recognition of Emotions in Autism: A Formal Meta-Analysis. *Journal of Autism and Developmental Disorders* 43(7), 1517–1526 (2013). doi: 10.1007/s10803-012-1695-5
2. American Psychiatric Association: Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author (2013)
3. Halberstadt, A. G., Denham, S. A., Dunsmore, J. C.: Affective Social Competence. *Social Development* 10(1). 79–119 (2001). doi: 10.1111/1467-9507.00150
4. Strand, P. S., Downs, A., Barbosa-Leiker, C.: Does facial expression recognition provide a toehold for the development of emotion understanding?. *Developmental Psychology* 52(8). 1182–1191 (2016). doi: 10.1037/dev0000144
5. Diehl, J. J., Schmitt, L. M., Villano, M., Crowell, C. R.: The clinical use of robots for individuals with Autism Spectrum Disorders: A critical review. *Research in Autism Spectrum Disorders* 6(1). 249–262 (2012). doi: 10.1016/j.rasd.2011.05.006
6. Costa, S. C., Soares, F. O., Santos, C.: Facial Expressions and Gestures to Convey Emotions with a Humanoid Robot. In: *International Conference on Social Robotics*, pp. 542–551 (2013). doi: 10.1007/978-3-319-02675-6\_54
7. Salvador, M. J., Silver, S., Mahoor, M. H.: An emotion recognition comparative study of autistic and typically-developing children using the zeno robot. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6128–6133 (2015). doi: 10.1109/ICRA.2015.7140059

8. Chevalier, P., Martin, J.-C., Isablue, B., Bazile, C., Tapus, A.: Impact on sensory preferences of individuals with autism on the recognition of emotions expressed by two robots, an avatar, and a human. *Autonomous Robots* 41(3). 613–635 (2016). doi: 10.1007/s10514-016-9575-z
9. Hassin, R. R., Aviezer, H., Bentin, S.: Inherently Ambiguous: Facial Expressions of Emotions in Context. *Emotion Review* 5(1). 60–65 (2013). doi: 10.1177/1754073912451331
10. Barrett, L. F., Mesquita, B., Gendron, M.: Context in Emotion Perception. *Current Directions in Psychological Science* 20(5). 286–290 (2011). doi: 10.1177/0963721411422522
11. Hoffner, C., Badzinski, D. M.: Children’s Integration of Facial and Situational Cues to Emotion. *Child Development* 60(2). 411–422 (1989). doi: 10.2307/1130986
12. Flom, R., Bahrick, L. E.: The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology* 43(1), 238–252 (2007). doi: 10.1037/0012-1649.43.1.238
13. Scherer, K. R.: Vocal communication of emotion: A review of research paradigms. *Speech Communication* 50(1-2). 227–256 (2003). doi: 10.1016/S0167-6393(02)00084-5
14. Barrett, L. F., Lindquist, K. A., Gendron, M.: Language as context for the perception of emotion. *Trends in Cognitive Sciences* 11(8). 327–332 (2007). doi: 10.1016/j.tics.2007.06.003
15. Schröder, M.: Experimental study of affect bursts. *Speech Communication* 40(1-2). 531–539 (2003). doi: 10.1016/S0167-6393(02)00078-X
16. Belin, P., Fillion-Bilodeau, S., Gosselin, F.: The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods* 40(2). 531–539 (2008). doi: 10.3758/BRM.40.2.531
17. Ekman, P., Friesen, W. V., Hager, J. C.: Facial action coding system (FACS): A technique for the measurement of facial action. Palo Alto: Consulting Psychologist Press (1978)
18. Borgers, N., de Leeuw, E., Hox, J.: Children as Respondents in Survey Research: Cognitive Development and Response Quality. *Bulletin de Méthodologie Sociologique* 66(1), 60–75 (2000). doi: 10.1177/075910630006600106
19. Wood, L. J., Dautenhahn, K., Rainer, A., Robins, B., Lehmann, H., Syrdal, D. S.: Robot-Mediated Interviews - How Effective Is a Humanoid Robot as a Tool for Interviewing Young Children?. *PLoS ONE* 8(3). e59448 (2013). doi: 10.1371/journal.pone.0059448
20. Calder, A. J., Burton, A., Miller, P., Young, A. W., Akamatsu, S.: A principal component analysis of facial expressions. *Vision Research* 41(9). 1179–1208 (2001). doi: 10.1016/S0042-6989(01)00002-5
21. Castelli, F.: Understanding emotions from standardized facial expressions in autism and normal development. *Autism* 9(4). 428–449 (2005). doi: 10.1177/1362361305056082
22. Happé F., Frith, U.: The Weak Coherence Account: Detail-focused Cognitive Style in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders* 36(1). 5–25 (2006), doi: 10.1007/s10803-005-0039-0
23. Collignon, O., Charbonneau, G., Peters, F., Nassim, M., Lassonde, M., Lepore, F., Mottron, L., Bertone, A.: Reduced multisensory facilitation in persons with autism. *Cortex* 49(6). 1704–1710 (2013). doi: 10.1016/j.cortex.2012.06.001
24. Xavier, J., Vignaud, V., Ruggiero, R., Bodeau, N., Cohen, D., Chaby, L.: A Multi-dimensional Approach to the Study of Emotion Recognition in Autism Spectrum Disorders. *Frontiers in Psychology* 6. 1–9 (2015). doi: 10.3389/fpsyg.2015.01954