The ASC-Inclusion Perceptual Serious Gaming Platform for Autistic Children

Erik Marchi, Björn Schuller, Alice Baird, Simon Baron-Cohen, Amandine Lassalle, Helen O'Rielly, Delia Pigat, Peter Robinson, Ian Davies, Tadas Baltrušaitis, Ofer Golan, Shimrit Fridenson-Hayo, Shahar Tal, Shai Newman, Noga Meir-Goren, Antonio Camurri, Stefano Piana, Sven Bölte, Metin Sezgin, Nese Alyuz, Agnieszka Rynkiewicz, Aurelie Baranger

Abstract—'Serious games' are becoming extremely relevant to individuals who have specific needs, such as children with an Autism Spectrum Condition (ASC). Often, individuals with an ASC have difficulties in interpreting verbal and non-verbal communication cues during social interactions. The ASC-Inclusion EU-FP7 funded project aims to provide children who have an ASC with a platform to learn emotion expression and recognition, through play in the virtual world. In particular, the ASC-Inclusion platform focuses on the expression of emotion via facial, vocal, and bodily gestures. The platform combines multiple analysis tools, using on-board microphone and web-cam capabilities. The platform utilises these capabilities via training games, text-based communication, animations, video and audio clips. This paper introduces current findings and evaluations of the ASC-Inclusion platform and provides detailed description for the different modalities.

Index Terms—Autism Spectrum Condition, inclusion, virtual computerised environment, emotion recognition, AI in games

I. INTRODUCTION

The field of artificial and computer intelligence is beginning to play a fundamental role in gaming, particularly with reference to computational narrative, believable agents, and commonly within commercial games [1]. The market for 'serious games' (i.e. games of which the purpose is not exclusively entertainment) is continually growing, with new intelligent methods for monitoring and interaction being introduced more frequently [2]. Recently, a focus on computer gaming as a support for intensive rehabilitation [3], as well as for social inclusion of children with particular needs such as those with an Autism Spectrum Condition (ASC) has

E. Marchi is now with Apple Siri. The work was done while at the Machine Intelligence & Signal Processing group, MMK, Technische Universität München (TUM), and at audEERING GmbH (Germany) – erik.marchi@tum.de. This work does not contain Apple proprietary information. B. Schuller and A. Baird are with the ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, (Germany). B. Schuller is also with the Department of Computing, Imperial College London (UK) – bjoern.schuller@imperial.ac.uk. S. Baron-Cohen, A. Lassalle, H. O'Rielly, and D. Pigat are with the Autism Research Centre, University of Cambridge (UK) – sb205@cam.ac.uk. P. Robinson, I. Davies, and T. Baltrušaitis are with the Computer Laboratory, University of Cambridge (UK) – peter.robinson@cl.cam.ac.uk. O. Golan and S. Fridenson-Hayo are with the Department of Psychology, Bar-Ilan University (Israel) – ofer.golan@biu.ac.il. S. Newman is with Compedia Ltd (Israel) – newmans@compedia.net. A. Camurri and S. Piana are with InfoMus Lab, University of Genoa (Italy) – antonio.camurri@unige.it. S. Bölte is with the Center of Neurodevelopmental Disorders, Karolinska Institute (Sweden) – sven.bolte@ki.se. A. Baranger is with Autism Europe (Belgium) – aurelie.baranger@autismeurope.org. M. Sezgin is with the Intelligent User Interfaces Lab, Koç University (Turkey) – mtsezgin@ku.edu.tr. A. Rynkiewicz is with the Neurodevelopmental Disorders Research Lab at Faculty of Medicine, University of Rzeszów and SPECTRUM ASC-MED (Poland) – rynkia@spectrumascmed.com.

shown promise. Children with an ASC, have shown to have a significant amount of difficulty perceiving and portraying the emotions and mental states of others [4]. An ASC can include an array of neurodevelopmental conditions, and can be characterised by social communication difficulties and restricted or repetitive behaviours. The difficulties become apparent during childhood, when individuals with an ASC are unable to recognise the emotions portrayed through the face [5], voice [6] and body [7]. Often those with an ASC will have a difficulty utilising the correct facial expression during interaction with others [8], as well as their vocal intonation and body language. Combining non-verbal communication cues along with speech is another common road block [9], and all of these social communication difficulties, make socialising (e.g. with care-givers, or family members) a daily challenge for those with an ASC [10]. Recent findings suggest that 1% of the global population might meet the criteria for an ASC diagnosis [11]. With this in mind, the study presented in [11] was the beginning of an increased focus on novel-cross cultural methodologies for ASC, as along with the recent increase in ASC diagnoses comes a need to assist individuals across ages bands in a more representative and inclusive way.

Due to the rapid growth in internet-based communication throughout the last decade, social networking is being utilised as a way to improve socialisation for individuals with an ASC. There are anecdotal reports of emerging on-line 'autistic communities', making use of forums, chat rooms, and virtual environments (VE), which could offer communication improvements for adolescents and adults with an ASC [12]. VE show great promise for aiding the social difficulties prevalent with an ASC. VE are artificially generated three-dimensional simulations and can come in either single or multi-user forms. Either format allows the user to operate real-life scenarios, to train skills, to stimulate conversations, and to solve social problems. Moore and colleagues investigated the use of VE for children with ASC. The results obtained in the study showed that 90% of the participants used the VE to recognize basic emotions [13]. There are also other studies which show the potential for individuals with an ASC to use VE social-emotional skills training, as well for other learning scenarios [14].

Another domain which has become popular in recent years is the utilisation of Information Communication Technology (ICT) for those with an ASC. In fact, computerised environments offer consistency, free from the social elements which individuals with an ASC can find overwhelming, as well as

offer the users the ability to work at their own pace, select their own level of ability and lessons can be repeated to improve understanding. Computerised rewards can also be incorporated into the system which may add to the level of interest and motivation needed to complete the task [15]. For these reasons, many ICT programs are being developed with the specific aim to teach a variety of skills to individuals with ASC. Commonly these programs focus on a single element (e.g., the recognition of facial expressions from a moving video). *Emotion Trainer* is one such tool which teaches the user to recognise four types of facial emotion [16]. *Let's Face It* is another which teaches emotion and identity recognition from facial expressions [17], and similarly the *Junior Detective* program combines ICT with group training as a way of helping to improve social skills [18]. These tools are examples of how modality-focused, current ICT solutions are for emotion recognition from facial expressions and contextual situations, and it would seem that little attention is being made to improving emotional expressions via the voice and gestures. It has been shown that people with an ASC have social and communication difficulties, but a superior ability in other non-social areas such as systematising [19]. This improved ability for systematising can be harnessed to create interventions aimed at enhancing social and communication skills. Previous attempts at intervention which capitalise on this superior systemising, include the *Mind Reading* [20] and the *Transporter* [21]. These training programs have an interactive guide to emotions and teach recognition of emotions and mental states, systematically placing those into emotion groups of developmental levels (from the age of 4 to adulthood). Using those game-like interventions, children with ASC made significant improvement in recognising emotions [21]. These training programs, however, only focus on improving the recognition of emotions and to the best of our knowledge there is no ICT tool which teaches emotional expressiveness.

Within this scenario, the ASC-Inclusion project [22], [23], [24] is proposing advanced ICT-enabled solutions and serious games for the empowerment of children on the autism spectrum (aged 5 to 10 years), particularly those who are at high risk of social exclusion. In order to enhance socio-emotional communication skills of children with an ASC (and those involved in their inclusion), an internet-based platform was created as part of the ASC-Inclusion project. The platform focuses particularly on the recognition and expression of socio-emotional cues and on the understanding and practice of conversational skills. The ASC-Inclusion platform combines state-of-the-art technologies, and consists of one comprehensive game environment, analysing a users' gestures, facial, and vocal expressions. Despite these innovative technologies, the ASC-Inclusion platform is very much aimed at home-use. Designed to assist children with ASC, ASC-Inclusion could also serve other population groups with difficulties in emotional understanding and socialising such as children with attention deficit or hyperactivity disorder (ADHD) [25].

The ASC-inclusion platform also incorporates a users caregiver. This thereby: a) increases engagement of those involved in the children's daily life such as parents, and siblings; b) integrates carers input into the system, and monitors/corrects and re-trains the system; and c) enables didactic interactions. The younger ASC individuals are then able to play the games along with their carers and have an opportunity to see examples of the emotions in a real person next to them. The system also has the ability to collect emotional displays made by the carer or parent, this additionally can help to further train the system. Further ASC-Inclusion supports the ability to provide formative assessments as part of the user interaction in a semantically comprehensible and appealing way to the children. The value for the child users is substantial: They relate to targets and feedback which enable them to adjust their facial, vocal and gesture behaviour to that of prototypical manifestations.

While in [22], [23], [24] we described the platform during its early stages, here we provide a final update on the current findings and clinical evaluation of the platform that confirmed the impact of this serious game as an effective educational intervention. We provide a detailed description of the platform (Section II), before giving a thorough analysis of the user requirements and the procedures that were adopted to retrieve feedback from the users in Section III; then, we describe the three modalities namely face, voice, and body gesture (Sections IV, V, VI), and the formative assessment module (Section VII). Subsequently, we describe the content creation (Section VIII), and adult-child cooperative playing (Section IX). We then comment on the psychological evaluation and its results (Section X) before concluding in Section XI.

II. PLATFORM

The ASC-Inclusion platform integrates the emotion recognition systems described below in Sections IV, V, VI, and VII. The platform incorporates a structured comprehensive program to help children recognise and understand emotions. The program consists of: **A virtual environment** (cf. Figure 1a) themed as a research-camp in the jungle, with animated characters, and a smart reward system, all designed to enhance the children's motivation. **A learning management application** (LMA) to track and monitor the child's progress (cf. Figure 1b). **Games and activities** including basic single-modality practice games, advanced cross-modal games with test games, and interactive stories with related activities (cf. Figure 1d). **47 interactive lessons**, presenting information about the emotions and cues to identify them via tone of voice, facial expression and body language. Each of the 12 selected emotions has an introduction lesson, and a lesson for each modality.

The platform relies on a number of components which are responsible for the communication between different services. The main component (subsystem control component) sends control messages to the different services. The messages are handled via Apache ActiveMQ[1] and are implemented following the EmotionML[2] standard. In fact, different ActiveMQ queues receive control messages for each of the services (voice, face, and body). The ActiveMQ process acts as a

---

[1]http://activemq.apache.org (last access: 23.06.2018. Note that this applies to all the other links mentioned in this article).
[2]https://www.w3.org/TR/emotionml

Fig. 1. From left to right: The VE themed as a research-camp: The research centre (a). The LMA to track and monitor the child's progress: The main LMA interface to units and sessions per modality (b), the avatar with different emotional expressions (c). Games and activities including basic single/multi-modality practice games: The "robots training mission" game (d), and the multi-modal expression game (e).

coordinator to send data from the different subsystems to the game engine. The architecture of the platform is described with details in [22], [23], [24].

The platform is managed by the LMA (cf. Figure 1b) which controls, personalises, and presents the learning material to the user. The navigation through the program and the different activities and games are handled by the LMA. The LMA also tracks the user's behaviour and collects relevant information for later analysis and improvement of the system. The reward system is also managed by the LMA via the 'monetary system' which is the basis for the VE 'economy'. In fact, this is designed to encourage and motivate the child in the gaming experience. Virtual money can be earned in the VE by actively interacting with the 'research sessions' and playing with the different practice games and activities. To this end, the child advances in the game and – as a result – earns virtual money needed to do extra activities such as buy goodies, access fun locations out of the camp, and play with non-curricular games. The reward system was carefully designed to keep a suitable balance between learning and practising with the emotions content, and having pure fun with the recreation activities and areas. This is to avoid and prevent situations in which the user is stuck on a particular lesson or activity, allowing them to switch to further motivational elements such as (but not limited to) the personalisation of the avatar (cf. Figure 1c), the virtual wallet, and the camp-square. A more thorough analysis of each of the main components will now be described in turn.

A. The research camp and the practice games

In the theme narrative selected for the VE, the child is given the role of a young researcher who joins a research group, researching human behaviour and emotions. The aim of this was to put the child in the position of someone who is curious about emotions and interested in learning about them. The graphical design located the research camp in the jungle in order to give the VE a fun, adventurous flavour, to enhance the child's motivation, and create a positive attitude to learn about emotions, which are challenging and often very frustrating to children with an ASC. This theme was selected after a series of focus groups which included children, parents and professionals. Indeed, anecdotal responses to the VE later showed the theme and design were well liked and led to children's desire to engage in the subject.

Several practice games were designed to help the child memorise, repeat and practice the use of the emotion cues in the face, voice and body. The games used a colourful animated environment (cf. Figure 1). The gamification of the training

tasks aimed to help the children overcome their difficulties and engage in the training, by turning the learning goals into game goals. For example, in the "Robot training mission" the child is presented with short video clips depicting the expression of emotions from the body language. Rather than answering questions, the child is required to train robots to recognise emotions. During each turn, the child should select the robots who correctly identified the emotions expressed by the characters in the clips. The games presented 3–5 levels of difficulty, reflecting the pedagogical need for an easy start with a gradual advance in the task related requirements. Basic games focused on one modality while advanced games integrated two, three, or all modalities including social context.

B. Progress and personalisation

The LMA controlled the progress of the child and level to which new content should be displayed. The first time the child entered, all the content was visible in order to make them feel informed, secure, and confident about the scope of the game. The plan for the game was defined as an important clinical requirement for children with an ASC. At the same time, it was required that the child should learn and practice the emotions in a certain order as recommended by the clinician team. This was achieved with a linear design for the research plan in which only the first lesson was open when the child first entered, and each time they completed a learning task the next lesson/activity was unlocked. To unlock the next activity, the lessons needed to be fully completed and the practice games were required to meet a pre-set criterion established by the clinicians.

Personalisation is known to be one of the most powerful tools in VE's, used to evoke initial interest as well as maintain long-term engagement. In the ASC-Inclusion platform, the child has a virtual character for which he/she can choose the characteristics of facial and body appearance, cloths, and other elements. The child has a virtual home in which he can design a large variety of items, and virtual collections. The child earns virtual coins for each learning task, with which he can buy 'goodies' for his virtual character, home and collections. The child thus becomes "emotionally invested" in the VE which motivates him to continue and advance despite the growing level of difficulty in the emotion content.

C. The expression game

The expression game is another important part of the platform (cf. Figure 1e) which integrates the face (Section IV), voice (Section V), and body language (Section VI) analysis
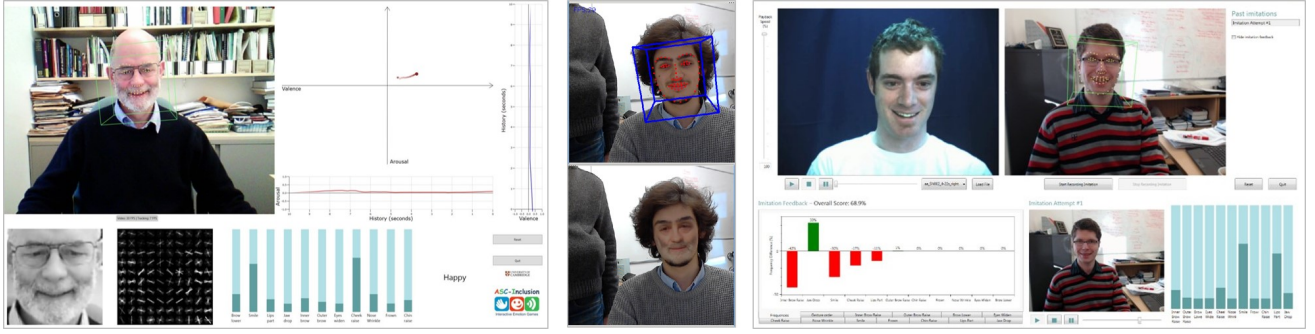
Fig. 2. From left to right: The facial affect computation engine, the facial affect mapping engine, and the teaching tool.

technologies. At the end of each learning unit the child can practice with the expression game and try out his emotional expressions that he has learnt. The game is designed as a 'race' board game. In each turn, the child is asked to express an emotion in a chosen modality. If he has expressed it to a level that is seen as adequate based on prior information, the racing robot advances one step on the board. Whoever gets to the ending point first wins. We introduced a mechanism to prevent situations in which the emotion analysers do not recognise the target emotion even though the child correctly expresses it. In fact, each modal analyser provides a confidence measure indicating the degree of confidence of the recognition result. In the case of low confidence values, no feedback is shown to the child and he is asked to play the target emotion again.

An important phase in the design and implementation of the platform was defining the needs of the user. Next section gives a detailed analysis of the user requirements.

III. USER REQUIREMENTS

Several focus groups of children with and without ASC were recruited throughout the ASC-Inclusion project and provided critical qualitative feedback. This feedback was used to determine user requirements in terms of game preferences and interests, game concepts and design, content specification, and learning styles. In the last year of the project, the focus groups also tested (and provided feedback on) the voice, face, and body analysers, as well as the voice games. Finally, a focus group composed of autism experts (autism researchers, autism teachers, parents of children with autism) was formed with the purpose of providing feedback regarding the realisation and implementation of the VE, and to highlight the difficulties that ASC children may experience with the VE. The results of these activities have now fed into the development of the VE. For instance, sixteen families rated the intervention 8.14 out of 10 on average. In addition, all families who went through with the intervention thought the game was suitable for their child, noticed positive changes in the behaviour of their child (although 2 parents were not sure whether to attribute those changes to the VE) and saw definite benefits to using VE. We also obtained feedback from parents by contacting them every two weeks and asking them a few questions. This feedback revealed that many children liked the features of the VE and were motivated to play with it. For the interested reader we have compiled a document [3] that contains these suggestions

[3]http://asc-inclusion.eu/wp-content/uploads/2018/06/appendix.pdf

to refine the platform and a list of recommendations defined by the team of clinicians and researchers.

These focus groups allowed for the identification of user learning requirements, and based on those we were able to specify a suitable learning structure for the platform. Furthermore, the feedback of both the user and expert focus groups allowed us to identify technological opportunities and constraints that have fed into the platforms development ensuring a user centric design. The various sub-system prototypes and the emotion modality analysers (cf. Sections IV, V, VI, II) were tested by focus groups at the participating clinical sites in United Kingdom, Sweden, Israel, and Poland. In addition, the tutorials, games and quizzes assessing the retention of the users were tested in these focus groups. For example, feedback was retrieved from 18 families and the expression game was rated 3.06 out of 5 while the voice game was rated 4.04 out of 5 [3].

Feedback from these user testing sessions were used to update the sub-systems to ensure they are easy to use and designed with the end user in mind. These analysers are described in the next four sections.

IV. FACE ANALYSIS

ASC-Inclusion investigated ways of teaching children with ASC how to read emotional content in facial expressions and to display responses with their own faces.

There are four main components: a) the face tracker, b) the facial affect inference system, c) the facial affect mapping system, and d) the teaching tool.

a) The face tracker: The face tracker evolved through several versions, initially using a Constrained Local Model that could incorporate depth information [26] and later moving to Continuous Conditional Random Fields [27] and Continuous Conditional Neural Fields [28]. Points are also superimposed on the input image to show the fitting of feature points. The resulting system has now been publicly released as OpenFace [29]. It is common during normal conversation that a person might partly cover their own face (e.g. through hand gestures). This can hinder the ability of facial tracking, but can also give useful affective information. A refinement of the system, allows for handling of hand-over-face gestures [30], [31]. The tracker copes with occlusion by the hand and the combined information is shown on the animated avatar.

b) The facial affect inference system: We developed a free-standing prototype Facial Affect Computation Engine

(FACE) that takes video from a standard webcam, infers probabilities for a set of discrete mental state conditions together with coordinates in a continuous valence-arousal space, and displays the results. This extends our earlier work on Mind Reading machines [32].

Figure 2 (left) shows the facial affect inference system in use. The video input is shown at the top left, with superimposed annotations showing the head pose and feature points around the face. A normalised monochrome image of the face is also displayed, together with an illustration of the Histograms of Oriented Gradients, the strengths of indivdual Action Units (AU), and a discrete measure of mental state. The animation at the top right shows a trace of the inferred valence and arousal in real time.

The dimensional inference system uses the face tracker to extract 34 features from live video. Support Vector Machines are then used to infer the occurrences of action units, and Support Vector Regression is used to infer the intensity [33]. In both cases, we used linear kernels as complex kernels did not improve performance and significantly slowed down training. Furthermore, we are especially interested in approaches that allow for an effective real-time applications. A variant of the system also incorporated voice analysis for multi-modal fusion [27].

The valence-arousal dimensional model of emotions is useful for illustration, but is too coarse for precise interpretation [34]. So a discrete set of categorical mental states (the six basic emotions: *anger, disgust, fear, happiness, sadness*, and *surprise*) is also inferred from the action units to give a more helpful interpretation. It is clear that there is a meaningful mapping from a dimensional model to a categorical one [34], so we did not attempt to calculate new one.

ActiveMQ was used to communicate output from the inference system with the game platform. We then implemented this message protocol in the FACE software, and demonstrated live inference results streaming to the main platform. Similarly, we also designed and implemented an ActiveMQ-based protocol allowing the game platform to control the sub-systems, starting and stopping inference. The communication between the systems is based on EmotionML standard.

c) The facial affect mapping system: A simple application was built to demonstrate the face tracking software at dissemination events. The Facial Affect Mapping Engine (FAME) takes continuous video as its input, tracks key features on the principal face in the video, and applies them as animation to a different face superimposed on the video stream [35], [36]. The input video can be taken from a webcam capturing a visitor to an exhibition, or it can be one of the standard videos of emotional expressions on the Mind Reading DVD [20]. The output face can be one of a number of well-known people, or a new face can be captured using the webcam.

Figure 2 (middle) shows the facial affect mapping system in use. The top side of the screen shows the tracked image from a webcam and a different face is being animated on the bottom. Although this demonstration was designed to help explanation of facial affect analysis at dissemination events, it also can be used as an educational tool. The face of the child

can be animated to show emotional expressions taken from the library on the Mind Reading DVD.

d) The teaching tool: The teaching tool uses a webcam to capture the facial expressions and head movements of the user [37]. The user selects an emotion category and then selects a target video to imitate from the provided database of videos belonging to that emotion category.

Figure 2 (right) shows the teaching tool in use. Two animated bar graphs are displayed. The first analyses the target video, and the second analyses the webcam video. The values in these animated bar graphs change frame-by-frame to correspond with the changing action unit (AU)[4] intensities in the videos. Thus, the user is able to visually compare his own AU intensities with those in the target video.

When ready, the user presses "Record" and then makes appropriate facial expressions and head gestures either (a) to portray the selected emotion (in any way desired), or (b) to imitate the target video precisely. When "Stop"is pressed, the recorded frames are analysed to gather aggregate intensity and frequency statistics for each AU. Based on the difference between the statistics of the imitation and those of the target emotion, an overall score for the imitation is given to the user, along with detailed feedback.

The system provides the user with visual feedback on their expressions, either by showing the real-time video stream from the webcam or by replaying the videos of previous imitation attempts. Slowing down the presentation of stimuli has been found to be beneficial in prior facial expression imitation tasks used in ASC interventions, so the system can slow down the playback of the target videos. This change in speed does not negatively affect the imitation comparison mechanisms, since one depends on aggregate statistics (emotion imitation) and the other uses dynamic time warping (exact-expression imitation).

In particular, using the .NET Dynamic Time Warping (NDTW)[5] package the system performs dynamic time warping (minimising the total distance across all action units simultaneously) to determine the frame-by-frame differences in action unit intensities. These differences are presented to the user visually through line and bar graphs. A detailed description of the accuracy of the core engine integrated into the teaching tool is provided in Section III of [39].

V. VOICE ANALYSIS

Voice-enabled assistive technology has been shown as a potential modality to help children with an ASC [40], [41]. The ASC-Inclusion platform also focuses on teaching children with an ASC to learn how to express and recognise vocal expression. In order for the children to interact with the system and have fun, they play using their voices, influencing the plots based on speech acoustics, with a built-in graphical user interface for the voice analyser. The automatic vocal expression evaluation system [24], [42] provides corrective feedback while the children learn about emotional expressions. The system was designed and refined following an iterative

---

[4]Action units (AUs) are a set of muscular activations, which provoke changes in facial appearance [38].

[5]https://github.com/doblak/ndtw

TABLE I
CLASSIFICATION PERFORMANCES IN TERMS OF THE UNWEIGHTED AVERAGE RECALL (UAR) FOR AROUSAL, VALENCE, AND EMOTION TASKS, AND THE AVERAGE OVER ALL THE TASKS PER GROUP: TIPICALLY DEVELOPING (TD), AUTISM SPECTRUM CONDITIONS (ASC)

UAR [%]	Emotion	Valence	Arousal	Average
TD	49.8	69.1	80.1	66.3
ASC	43.9	63.9	74.8	60.9

evaluation process carried out at the clinical sites. First, the player selects the emotion that he wants to practice with. Once the emotion has been selected, a reference emotion expression is played back to the child. Then, the child is prompted to repeat the selected emotion. According to the expressed emotion the evaluation system is providing a visual feedback showing the detected emotion expressed. Besides providing the classification result, the analyser shows a confidence measure that shows how much the system is certain about the recognised emotion against the remaining emotions. The confidence is calculated by the probability estimate derived from the distance between the instance feature points and the hyper-plane in the used distance-based classification. The corrective feedback is provided graphically via gauge bars, which indicate if the extracted parameters are distant or close to the reference values for that specific vocal expression. In order to assess a child's performance in expressing emotions via speech, the extracted parameters are compared to the respective parameters extracted from pre-recorded prototypical utterances. The actual core engine behind the evaluation system is the voice analyser [43], [44], [45], which uses the openSMILE audio feature extractor [46], [47] to extracted acoustic features and track them over time. Table I shows a summary of the accuracy of the system in the three different classification tasks. The emotion task covers the recognition of the nine target classes (happy, sad, angry, surprised, afraid, ashamed, calm, proud, and neutral). We further evaluated the discrimination between high and low arousal as well as between positive and negative valence. Technical details about accuracy and evaluation are provided in [45].

The ASC-Inclusion platform also enables adult and child cooperative play, thus, the enhanced version of the voice analyser was made to handle adult-child models and language-dependent models for emotion recognition. The acoustic models for adults were trained on the voice recordings collected by the clinical teams, as part of the obtainable "EU-Emotion Stimulus" set [48].

The output from the voice analyser is encoded in EmotionML [49] and delivered through an ActiveMQ communication infrastructure, allowing for integration with the face and gesture analysis, and the formative assessment module provides a multi-modal inference for the central platform.

VI. GESTURE ANALYSIS

Our work in analysing body movements to infer emotions is inspired by previous studies in experimental and humanistic psychology [50]. Many authors focused on distinct patterns of emotional body expressions along a number of dimensions,

using terms such as speed, force, energy, expansiveness or directness [51]. There have been different descriptions of specific postures adopted in connection with different emotions [50]. Atkinson et al. recently showed that static form and dynamic motion cues provide distinct contributions to the communication of emotion through the body [52]. In other studies authors have demonstrated the importance of dynamic motion cues for the communication of emotions in different contexts, including interpersonal dialogues [53], dance [54], and infant expressions [55]. Most of these studies stress the importance of certain movement qualities such as jerkiness, energy or speed as discriminating attributes [51]. The result of such studies are emotion-related dimensions, that help in discriminating between different expressed feelings. Based on these findings we build a computational model and develop modules to automatically extract such movement qualities.

In this process we used body gesture clips (cf. Section VIII) to verify the state-of-the-art and to identify a set of relevant movement characteristics that help in the cognitive process of discriminating emotions. As a result of this analysis we identified a set of descriptive features belonging to a multi-layered framework [56] that includes biomechanic features (e.g., speeds, accelerations) at a small time scale (observable frame by frame), more complex qualities and representations, usually extracted on groups of joints or on the whole body, that requires substantially longer temporal intervals (e.g., rhythmic patterns typically require a range of 0.5–5s to be detected [57]). A summary of the extracted features, the points on which they are computed and the emotional dimension to which they relate, is given in [58].

Based on the available data and given the necessity to refine the recognition models, we decided to enlarge the available corpus involving non-actors adults and typically developing children (aged 8–11) freely expressing emotions with their body: starting from the protocols of the recordings collected with professional actors [48], we recorded a new dataset of expressive gestures with two different systems: a Qualisys[6] optical motion capture system, and a Microsoft Kinect[7]. In both cases we recorded sequences of 3D coordinates (or 3D skeletons), corresponding to the same body joints.

The resulting recognition system works as follows: input data consist of 3D coordinates from motion capture systems (either low-cost RGB-D sensors or more sophisticated optical motion capture devices). Given 3D data of the user, we can build a skeleton representation of their body. Each specific body joint is tracked over time, resulting in a trajectory that describes how the user moves and the corresponding part of the body. The captured data is then analysed to extract the set of movement features introduced above. An adaptive histogram-based representation is created. This representation is learnt over the input data. This further step leads to better performances as it improves the encoding of the large intra-class variability typical of the task considered here. Finally, the obtained representation is classified in one of N possible emotions with a N-class linear support vector machine. To

---

[6]http://www.qualisys.com
[7]https://www.microsoft.com/en-us/download/details.aspx?id=44561

see details on the pipeline of the system, performances and evaluation of the algorithms see [58].

The system was implemented using the EyesWeb[8] XMI platform. Additionally, in order to make the software available on a variety of hardware configurations where a depth camera is not available, we developed a 2D version of the recognition module that uses information coming from a standard webcam; this version uses a 2D based subset of the features used by the 3D based version of the module. The module is fully integrated in the game platform and can communicate with the other modules via ActiveMQ and EmotionML messages to record streams of data and log outputs that can then be processed offline.

To evaluate the performances of the body gesture analyser with children, two serious mini-games were developed [58] and tested. Both games perform a real-time automatic emotion recognition from body gestures, and interact with the user by asking him to guess and to express an emotion with the body. During the games, the user controls the GUI (a digital black board) by body gestures [59], and is also asked to perform body gestures to express certain emotions as part of the games.

A pilot evaluation study of the body gesture recognition system and related serious games was carried out with a group of ten children diagnosed with ASC (mean IQ 87.2; mean age 9.6 years old, 9 boys and 1 girl).

Children were enrolled in a training program consisting of two gaming sessions per week, for a total of ten gaming sessions. During each session participants played with the serious mini-games for about fifteen minutes. Each gaming session included different tasks where the children were asked to express (Expression Task) or recognise a certain emotion (Recognition Task).

Each child was assessed before and after the participation in the program. The assessment included:

- Leiter Scale [60] (if the IQ was not already available)
- The Italian version of Test of Emotions Comprehension (TEC) by Harris, Pons & De Rosnay [61];
- Go/No-Go task: is a well know paradigm that measures the ability to inhibit certain responses, the task required participants to press a button when a given target figure (a blue square) was displayed and to not press it if any other figure was displayed. Participants were instructed to withhold pressing the space-bar button every time a "No-Go" figure occurred (20% of the trials). The task was composed by a practice phase and a test phase. Participants were asked to answer as fast as they could every time the "Go" stimulus appeared and received a feedback every time they pressed the space-bar by reporting their reaction time so they could realise how fast or slow they were. Reaction time and accuracy at "No-Go" trials were the dependent measure.
- Emotional Go/No-Go task [62], it is a modified version of the previous Go/No-Go paradigm: children responded by pressing a button indicating a facial expression that expresses a particular emotion – they should avoid to respond to neutral expressions.

[8]http://www.infomus.org/eyesweb_eng.php

TABLE II
CORRELATION BETWEEN IN-GAME SCORES AND PRE-POST TEC TEST GAINS. *: SIGNIFICANTLY CORRELATED ($p < 0.05$).

	Recognition Task		Expression Task	
	Sess. 1-6	Sess. 7-10	Sess. 1-6	Sess. 7-10
Raw Scores	0.56	0.00	−0.21	0.47
Z-Scores	0.68*	−0.16	−0.19	0.53($p = .12$)
Percentiles	0.68*	−0.31	−0.40	0.71($p = .05$)

The comparison of the gains in TEC test score and in-game tasks results on the whole training period showed an overall correlation of 0.64 ($p = .08$) between the Expression in-game Task and the TEC scores, while no significant correlation between the Recognition Task and the scores was found. In order to better understand the evolution of the two cognitive processes we performed an additional analysis by separating the training into a first period (sessions 1–6) and a second period (sessions 7–10). The results of the two periods are shown in Table II. If we consider the first six training sessions the recognition task is significantly correlated ($p < 0.05$) with both z-scores and percentiles tendencies while the expression task is correlated with both the z-score and percentiles in the last training sessions. This may indicate that the two cognitive processes are characterised by different learning curves (the recognition being a faster process and the expression requiring a longer training period). In fact, we observed bigger gains in the recognition rates during the first period (training sessions 1–6) while the accuracy of the children in the expression task grows more during the second period (sessions 7–10).

VII. FORMATIVE ASSESSMENT

The face, voice, and body analysers can differentiate between accurate and inaccurate expression of emotion by children with an ASC. We considered that an important step for the analysers would be to provide feedback to the children regarding the aspects of the emotion expressions that were inaccurately expressed. In order for the analysers to be able to provide feedback, the system needs to be trained to identify the incorrect features of emotion expression, when emotion is expressed inaccurately. For this, the analysers would need to be fed with correct exemplars of emotion expression (e.g., expression of an emotion containing all the features of the particular emotion) and incorrect exemplars of emotion expression (e.g., expression of an emotion containing all but one features of the particular emotion). We thus created a database of correct and incorrect emotional stimuli in three modalities (face, voice, body). The adequate features of emotion expression for each of the eighteen emotions of interest (happy, angry, afraid, worried, sad, disgusted, interested, kind, joking, proud, sneaky, surprised, bored, hurt, frustrated, unfriendly, ashamed, and neutral) and each of the three modalities of interest (face, voice, body) were determined by a survey study conducted in Poland that involved a sample of 40 typically developing adults. The features that characterised correct facial expressions were largely overlapping with the AU that had been previously described by Ekman in his Facial Action Coding System (FACS) [38] as the signature facial expressions. We recruited an experienced actor and asked

them to memorise the features most commonly associated with correct emotion expression, and to practice expressing the emotion incorrectly by removing one of those features. Several instances of the correct and incorrect expressions were recorded for all emotions of interest, using a video camera (for the face and body modalities) and a microphone (for the vocal modality). The recordings were then examined by the team of researchers who selected the best correct and incorrect exemplars for each emotion and each modality. This resulted in the selection of 152 stimuli that were used to implement the feedback feature on the analysers.

The emotion analysers through the facial, vocal, and gestural modalities are the crucial steps for guiding a child who is learning how to perform emotions more effectively. However, the output of an analyser cannot be directly employed: considering the target emotion and displaying a binary response as correct or incorrect would not be meaningful for the child to understand what should be done differently. Instead, semantically meaningful higher-level instructions should be provided to direct the child in altering their face, voice, or gesture for expressing the target emotion. Therefore, we propose to complement each of the modality specific classifiers with a formative assessment module to provide corrective instructions when an emotion was performed incorrectly.

As an initial step for providing meaningful feedback for a modality, a semantically meaningful set of mid-level features (i.e., attributes) were defined. These attributes were either selected from the features or they were extracted as residing in a higher-level, and they were based on modality-specific characteristics of the emotions. These attributes serve as a look-up table between the formative assessment module and the human-computer interaction module: In the formative assessment module, a feedback is generated as either increase/decrease or extraneous/missing for the active attributes. These corrective feedbacks are provided to the child as an audio-visual output, through the human-computer interaction module.

We compiled a candidate list including dictionary elements for each modality: The face modality included a number of AUs such as but not limited to inner/outer brow raiser, brow lowerer, upper lid raiser, cheek raiser, nose wrinkler, upper lip raiser, and lip/lid tightener. The voice modality dictionary included pitch, pitch variation, loudness, and speech rate. The body gesture modality included direction of motion, kinetic energy, smoothness, and contraction index. These attributes define the dictionary elements to be fired as necessary modifications: For voice and gesture attributes, the feedback is in terms of the direction of necessary change (e.g., increase or decrease) in any of these elements; and for the face modality, the attributes highlight which of the action units are invalid (i.e., missing/extraneous). The feedbacks generated in terms of these attributes are provided to the human-computer interaction module, which are then displayed in terms of audio-visual outputs (e.g., displaying direction of necessary change with up/down arrows for voice modality; missing AU represented on face image with outward pointing arrows at lip edges).

In the formative assessment module, we have developed a corrective feedback generator that employs the generalised technique of Baehrens et al. [63] to generate instance-specific explanations about classifier decisions. When the feedback module is initiated by an unexpected emotion, the class of the target emotion is used to infer explanations about low-level features. The inference represented by means of explanation vectors, points out the features causing the misclassification and expresses how these values should be modified to obtain the desired affective state. In our formative feedback module, we have implemented an iterative feature modification approach, where the original input is modified to be classified as from the expected class. Instead of low-level features, we employed semantically meaningful attribute sets as defined for each modality.

In the formative assessment experiments for the voice modality, we have utilised a dataset collected from a group of typically developed children and considered a two-class classification problem to differentiate between happy and sad emotions. The generated feedbacks through the attributes in the form of increase or decrease were evaluated by human evaluators through a user study: evaluators listened to the original audio and evaluated the feedback generated for the target class. The results showed that on average 65% of the feedback generated was evaluated as correct. When attributes are investigated separately, pitch and pitch variation feedbacks were evaluated as 70% correct on average. This rate was lower for the loudness and speech rate attributes, showing that these feedbacks cannot be interpreted easily. Therefore, results indicate that constant pitch and pitch variation are the most important attributes, and feedback generated through these can significantly guide children towards the correct vocal intonations when performing emotions.

VIII. CONTENT CREATION

This section contains details about data collection which was used as content for the platform and its localisation in different countries. In order to make the game available to children of other countries within the European Union (e.g., Spain, Italy, France, Germany), we needed validated emotional voice stimuli from the different native-languages of such regions. Thus, 84 sentences (4 per 21 emotions) were selected and translated to the different languages by a translation team, including a native speaker and a professional interpreter. At least three -native speaker- actors were recruited for each language (a child, an adolescent, and an adult) to enact the selected sentences. The best recording of each sentence was selected for each actor before being cut and edited. Two researchers who did not speak Spanish, Italian, French, or German listened to the audio-recordings and were subjected to a forced-choice task among four emotion labels to determine whether the recordings were accurately recognised. This validation process resulted in the selection of 229 French voice stimuli, 157 Spanish voice stimuli, 117 Italian voice stimuli, and a total of 152 voice stimuli were used to produce the game in different languages.

The data collected in the scope of the project was validated and is contained in the obtainable "EU-Emotion Stimulus"

Fig. 3. Adult-child cooperative playing: A session when a child and her parent are cooperatively playing with the ASC-Inclusion platform. Parents role was to build pleasant cooperative parent-child playtime at home with the software.

set [48]. Recently, an evaluation study pertaining to the emotional voice stimuli of the "EU-Emotion Stimulus" set, the so-called "EU-Emotion voice database" was conducted in [64]. This database contains a total of 2159 emotional voice stimuli depicting more than 20 emotional states in three different languages (Hebrew, Swedish, British), and is freely available for scientific use.

IX. ADULT & CHILD COOPERATIVE PLAYING

Focusing on previous works with parents of children with ASC, a series of user studies have been carried out, in order to define the requirements and the technology adjustments which can enable the adult-child cooperative play [24]. In fact, psychological experiments and evaluation with recruited adults-parents and children (cf. Figure 3) were conducted in Poland. During the period of the trial, the researchers were periodically monitoring the parents once a week to check on any queries. Parents were also encouraged to contact the research team with any questions or if any technical aspect of the program needed adjustment. The psychological experiments were carried out with 40 parents recruited in Poland. A preliminary study was conducted with a number of games in the ASC-Inclusion platform with the aim of defining requirements and specifications for the program. A subsequent evaluation was then conducted to measure the magnitude of stress in the parent-child system, and to assess how children with an ASC are able to understand emotions before and after usage of the VE (during adult-child collaborative play). This evaluation also enabled a focus on the lower ASC diagnosis rates of females and the differences between the two sexes in both verbal and non-verbal modes of communication. The study led to an important finding that high-functioning females with an ASC have greater self-awareness and make greater effort to camouflage their deficits [65], [66]. A follow-up study has also been conducted in Sweden with 20 children with ASC.

Based on the feedback obtained from the pilot studies, the platform and the sub-systems were adjusted to handle inputs from adults in order to enable adult and child cooperative play. Updates were applied to the body gesture module to track an additional users, and to the voice analyser to include automatic emotion recognition models for adult speech.

X. PSYCHOLOGICAL EXPERIMENTS AND EVALUATION

The effectiveness of the serious game designed as part of the ASC-Inclusion platform was cross-culturally evaluated through clinical and controlled trials. The evaluations were conducted with 6-9 year old children with a high-functioning ASC, who used the serious game for 8-12 weeks. Evaluations included face, voice, body, and integrative emotion recognition tasks [67], as well as parental reports on the socialisation and level of autism symptoms present in their child. Specifically, the evaluation was conducted following a cross-cultural study which evaluated the emotion recognition impact [67]. In that evaluation we have compared the emotions by dividing them into a basic emotion group (happy, sad, angry, disgust, afraid and surprised) and a complex emotion group. The study included two trials: The first clinical trial was conducted in the UK. Fifteen children with an ASC (4 female, 11 male), with an IQ within the normative range were tested pre and post use of the intervention for 8 weeks. This test focused on their ability to recognise emotions from body language and from integrative emotional scenarios. Paired sample t-tests were used to examine whether the children improved from pre to post intervention for the different modalities. The results revealed that 8 weeks of game use, significantly improved the ability for a user to recognise emotions from body language and from integrative emotional scenarios. In addition, the serious game improved socialisation, as reported by parents. Following the encouraging results from the UK trial, the Israeli and Swedish research teams conducted controlled trials, in which children with an ASC using the serious game were compared to a control group of children with an ASC, who took part in a conventional therapy session. A selection of children from Israel and Sweden (38 and 36 respectively), aged 6–9 years, with an ASC participated in this study. Their emotion recognition skills from faces, voices, body language, and integrative emotional scenarios were tested pre and post the intervention period (or conventional therapy for the control group). Repeated measures MANOVAs revealed substantial emotion recognition gains for the group that used the serious game, but not for the control group on all modalities, over and above country grouping. In addition, Israeli parents reported that their children showed fewer autism symptoms following the use of the serious game. Based on the effect sizes of the pre-post evaluation conducted in the Israeli and Swedish controlled trial, the greatest gains were made in the body language/gesture modality (partial eta squared = 0.51), followed by the voice (0.26) and the face (0.24). The findings from these studies are reported in detail in [68], and confirm that the serious game developed as part of ASC-Inclusion platform is an effective and motivating psycho-educational intervention. The platform cross-culturally teaches emotion recognition from faces, voices, body language and their integration in context to children with an ASC. In particular, participants that used the serious game were found to have a more generalised gain for socialisation and a reduction in other symptoms typical of an ASC. Note that the evaluation described here did not include an assessment of participants' emotional expressiveness, but rather an assessment of their ability to recognise emotions in others. Theoretically, the improvement in socialisation skills could be related to improvement in the ability for a participant to note the emotional cues of others, as well as their improved expressiveness. Although the latter has not been assessed, the

platform was designed and implemented to provide formative assessment and evaluate the emotional expressions, based on the relevant modal-dependent parameters.

XI. SUMMARY

We presented the findings and development of the perceptual serious game platform ASC-Inclusion, designed for children with an ASC aged from 5–10 years. The platform makes use of modal inference systems trained to provide automatic analysis and evaluation of facial, vocal, and body movement expressions. In particular, a module was specifically implemented to generate formative assessment and corrective feedback. The automatic system was refined to fit the clinical teams' recommendations, and adjusted to enable adult-child cooperative play. The clinical evaluation findings confirmed that the platform is an effective educational intervention, resulting in a substantial gain in emotion recognition for those within the focus groups, showing an evident generalised improvement in socialisation and other symptoms present in those with an ASC. Future efforts will focus on the introduction of other modalities (e.g. touch) relying on motion sensors and alternative gaming interfaces [69], as well as on the dynamic adjustment of game difficulty [70], and on a deeper analysis of game behavioural data [71].

REFERENCES

[1] G. N. Yannakakis and J. Togelius, "A panorama of artificial and computational intelligence in games," *IEEE Trans. on Computational Intelligence and AI in Games*, vol. 7, no. 4, pp. 317–335, 2015.

[2] M. Frutos-Pascual and B. G. Zapirain, "Review of the use of AI techniques in serious games: Decision making and machine learning," *IEEE Trans. on Computational Intelligence and AI in Games*, vol. PP, no. 99, 2016.

[3] M. Pirovano, R. Mainetti, G. Baud-Bovy, P. L. Lanzi, and N. A. Borghese, "Intelligent game engine for rehabilitation (IGER)," *IEEE Trans. on Computational Intelligence and AI in Games*, vol. 8, no. 1, pp. 43–55, 2016.

[4] S. Baron-Cohen, *Mindblindness: An essay on autism and theory of mind.* MIT press, 1997.

[5] O. Golan, S. Baron-Cohen, and J. Hill, "The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger Syndrome," *Journal of autism and developmental disorders*, vol. 36, no. 2, pp. 169–183, 2006.

[6] O. Golan, S. Baron-Cohen, and Y. Golan, "The 'Reading the Mind in Films' task [child version]: Complex emotion and mental state recognition in children with and without autism spectrum conditions," *Journal of Autism and Developmental Disorders*, vol. 38, no. 8, pp. 1534–1541, 2008.

[7] R. Philip, H. Whalley, A. Stanfield, R. Sprengelmeyer, I. Santos, Young *et al.*, "Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders," *Psychological medicine*, vol. 40, no. 11, pp. 1919–1929, 2010.

[8] C. Kasari, B. Chamberlain, and N. Bauminger, "Social emotions and social relationships: Can children with autism compensate?" in *Development and autism: Perspectives from theory and research*, J. Burack, T. Charman, N. Yirmiya, and P. Zelazo, Eds. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 2001, pp. 309–323.

[9] J. McCann and S. Peppe, "Prosody in autism spectrum disorders: A critical review," *International Journal of Language & Communication Disorders*, vol. 38, no. 4, pp. 325–350, 2003.

[10] C. C. Bell, "Dsm-iv: Diagnostic and statistical manual of mental disorders," *The Journal of the American Medical Association*, vol. 272, no. 10, pp. 828–829, 1994.

[11] S. Baron-Cohen, F. J. Scott, C. Allison, J. Williams, P. Bolton, F. E. Matthews *et al.*, "Prevalence of autism-spectrum conditions: Uk school-based population study," *The British Journal of Psychiatry*, vol. 194, no. 6, pp. 500–509, 2009.

[12] C. J. Jordan, "Evolution of autism support and understanding via the World Wide Web," *Intellectual and developmental disabilities*, vol. 48, no. 3, pp. 220–227, 2010.

[13] D. Moore, Y. Cheng, P. McGrath, and N. J. Powell, "Collaborative virtual environment technology for people with autism," *Focus on Autism and Other Developmental Disabilities*, vol. 20, no. 4, pp. 231–243, 2005.

[14] S. Parsons, L. Beardon, H. Neale, G. Reynard, R. Eastgate, J. Wilson *et al.*, "Development of social skills amongst adults with Asperger's syndrome using virtual environments: The 'AS Interactive' project," in *Proc. 3rd Int. Conf. on Disability, Virtual Reality and Associated Technologies, ICDVRAT*, 2000, pp. 23–25.

[15] D. Moore, P. McGrath, and J. Thorpe, "Computer-aided learning for people with autism – a framework for research and development," *Innovations in Education and Training International*, vol. 37, no. 3, pp. 218–228, 2000.

[16] M. Silver and P. Oakes, "Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others," *Autism*, vol. 5, no. 3, pp. 299–316, 2001.

[17] J. W. Tanaka, J. M. Wolf, C. Klaiman, K. Koenig, J. Cockburn, L. Herlihy *et al.*, "Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Lets Face It! program," *Journal of Child Psychology and Psychiatry*, vol. 51, no. 8, pp. 944–952, 2010.

[18] R. Beaumont and K. Sofronoff, "A multi-component social skills intervention for children with Asperger syndrome: The Junior Detective Training program," *Journal of Child Psychology and Psychiatry*, vol. 49, no. 7, pp. 743–753, 2008.

[19] S. Baron-Cohen, S. Wheelwright, J. Lawson, R. Griffin, C. Ashwin, J. Billington *et al.*, "Empathizing and systemizing in autism spectrum conditions," *Handbook of autism and pervasive developmental disorders*, vol. 1, pp. 628–639, 2005.

[20] S. Baron-Cohen. (2002) Mind reading. DVD. J. Kingsley Publishers.

[21] O. Golan, E. Ashwin, Y. Granader, S. McClintock, K. Day, V. Leggett *et al.*, "Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces," *Journal of autism and developmental disorders*, vol. 40, no. 3, pp. 269–279, 2010.

[22] B. Schuller, E. Marchi, S. Baron-Cohen, H. O'Reilly, P. Robinson, I. Davies *et al.*, "ASC-Inclusion: Interactive Emotion Games for Social Inclusion of Children with Autism Spectrum Conditions," in *Proc. 1st Int. Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI)*. Chania, Greece: ACM, May 2013, p. 8.

[23] B. Schuller, E. Marchi, S. Baron-Cohen, H. O'Reilly, D. Pigat, P. Robinson *et al.*, "The state of play of ASC-Inclusion: An integrated internet-based environment for social inclusion of children with autism spectrum conditions," in *Proc. 2nd Int. Workshop on Digital Games for Empowerment and Inclusion (IDGEI), IUI*. Haifa, Israel: ACM, 2014, p. 8.

[24] B. Schuller, E. Marchi, S. Baron-Cohen, A. Lassalle, H. O'Reilly, D. Pigat *et al.*, "Recent developments and results of ASC-Inclusion: An integrated internet-based environment for social inclusion of children with autism spectrum conditions."

[25] D. Da Fonseca, V. Seguier, A. Santos, F. Poinso, and C. Deruelle, "Emotion understanding in children with adhd," *Child psychiatry and human development*, vol. 40, no. 1, pp. 111–121, 2009.

[26] T. Baltrušaitis, P. Robinson, and L. P. Morency, "3D constrained local model for rigid and non-rigid facial tracking," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2610–2617.

[27] T. Baltrušaitis, N. Banda, and P. Robinson, "Dimensional affect recognition using continuous conditional random fields," in *Proc. 10th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013, pp. 1–8.

[28] T. Baltrušaitis, P. Robinson, and L.-P. Morency, *Continuous Conditional Neural Fields for Structured Regression*. Zurich, Switzerland: Springer International Publishing, 2014, pp. 593–608.

[29] T. Baltrušaitis, P. Robinson, and L. P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *Proc. IEEE Winter Conf. on Applications of Computer Vision (WACV)*, 2016, pp. 1–10.

[30] M. Mahmoud, L.-P. Morency, and P. Robinson, "Automatic multimodal descriptors of rhythmic body movement," in *Proc. 15th Int. Conf. on Multimodal Interaction (ICMI)*. ACM, 2013, pp. 429–436.

[31] M. M. Mahmoud, T. Baltrušaitis, and P. Robinson, "Automatic detection of naturalistic hand-over-face gesture descriptors," in *Proc. 16th Int. Conf. on Multimodal Interaction (ICMI)*. ACM, 2014, pp. 319–326.

[32] R. el Kaliouby and P. Robinson, *Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures*. Boston, MA: Springer US, 2005, pp. 181–200.

[33] T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *Proc. 11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 06, 2015, pp. 1–6.

[34] P. Robinson and T. Baltruaitis, "Empirical analysis of continuous affect," in *Proc. Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2015, pp. 288–294.

[35] L. Impett, P. Robinson, and T. Baltrušaitis, "A facial affect mapping engine," in *Proc. 19th Int. Conf. on Intelligent User Interfaces (IUI)*. ACM, 2014, pp. 33–36.

[36] M. Mahmoud, T. Baltrušaitis, V. Gavrila, M. Rozycki, L. Impett, and P. Robinson, "Automatic face analysis tools for interactive digital game," in *Proc. 3rd Int. Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI)*. Atlanta, GA: ACM, 2015, p. 9.

[37] A. Adams, T. Baltrušaitis, and P. Robinson, "Expression training to convey complex emotions," in *Proc. 6th biannual Conf. on Affective Computing and Intelligent Interaction (ACII)*, AAAC. Xi'an, P. R. China: IEEE, 2015, pp. 778–780.

[38] P. Ekman and W. V. Friesen, *Facial Action Coding System: Investigatoris Guide*. Consulting Psychologists Press, 1978.

[39] A. Adams and P. Robinson, "Automated recognition of complex categorical emotions from facial expressions and head motions," in *Proc. Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2015, pp. 355–361.

[40] E. Marchi, F. Ringeval, and B. Schuller, "Voice-enabled assistive robots for handling autism spectrum conditions: An examination of the role of prosody," in *Speech and Automata in Health Care*, A. Neustein, Ed. Boston/Berlin/Munich: De Gruyter, 2014, ch. 8, pp. 207–236.

[41] E. Marchi, Y. Zhang, F. Eyben, F. Ringeval, and B. Schuller, "Autism and speech, language, and emotion – A survey," in *Evaluating the Role of Speech Technology in Medical Case Management*, H. Patil and M. Kulshreshtha, Eds. Berlin, Germany: De Gruyter, 2015.

[42] E. Marchi, B. Schuller, S. Baron-Cohen, A. Lassalle, H. O'Reilly, D. Pigat *et al.*, "Voice emotion games: Language and emotion in the voice of children with autism spectrum condition," in *Proc. 3rd Int. Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI), IUI*, ACM. Atlanta, GA: ACM, 2015, p. 6.

[43] E. Marchi, B. Schuller, A. Batliner, S. Fridenzon, S. Tal, and O. Golan, "Emotion in the speech of children with autism spectrum conditions: Prosody and everything else," in *Proc. 3rd Workshop on Child, Computer and Interaction (WOCCI), INTERSPEECH*. Portland, OR: ISCA, 2012.

[44] E. Marchi, A. Batliner, B. Schuller, S. Fridenzon, S. Tal, and O. Golan, "Speech, emotion, age, language, task, and typicality: Trying to disentangle performance and feature relevance," in *Proc. 1st Int. Workshop on Wide Spectrum Social Signal Processing (WS3P), SocialCom, ASE/IEEE*. Amsterdam, The Netherlands: IEEE, 2012.

[45] E. Marchi, B. Schuller, S. Baron-Cohen, O. Golan, S. Bölte, P. Arora *et al.*, "Typicality and emotion in the voice of children with autism spectrum condition: Evidence across three languages," in *Proc. 16th Annual Conf. of the International Speech Communication Association (INTERSPEECH)*. Dresden, Germany: ISCA, 2015, pp. 115–119.

[46] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE – The Munich versatile and fast open-source audio feature extractor," in *Proc. 18th ACM Int. Conf. on Multimedia*, ACM. Florence, Italy: ACM, 2010, pp. 1459–1462.

[47] F. Eyben, F. Weninger, F. Groß, and B. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in *Proc. 21st ACM Int. Conf. on Multimedia*, ACM. Barcelona, Spain: ACM, 2013, pp. 835–838.

[48] H. O'Reilly, D. Pigat, S. Fridenson, S. Berggren, S. Tal, O. Golan *et al.*, "The EU-emotion stimulus set: A validation study," *Behavior Research Methods*, vol. 48, no. 2, pp. 567–576, 2016.

[49] M. Schröeder, P. Baggia, F. Burkhardt, C. Pelachaud, C. Peter, and E. Zovato, "Emotion markup language (EmotionML) 1.0," *W3C Working Draft*, vol. 29, pp. 3–22, 2010.

[50] H. G. Wallbott, "Bodily expression of emotion," *European Journal of Social Psychology*, vol. 28, pp. 879–896, 1998.

[51] M. De Meijer, "The contribution of general features of body movement to the attribution of emotions," *Journal of Nonverbal behavior*, vol. 13, no. 4, pp. 247–268, 1989.

[52] A. P. Atkinson, W. H. Dittrich, A. J. Gemmell, and A. W. Young, "Emotion perception from dynamic and static body expressions in point-light and full-light displays," *Perception*, vol. 33, no. 6, pp. 717–746, 2004.

[53] T. J. Clarke, M. F. Bradshaw, D. T. Field, S. E. Hampson, and D. Rose, "The perception of emotion from body movement in point-light displays of interpersonal dialogue," *Perception*, vol. 34, no. 10, pp. 1171–1180, 2005.

[54] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques," *International journal of human-computer studies*, vol. 59, no. 1, pp. 213–225, 2003.

[55] L. A. Camras, J. Sullivan, and G. Michel, "Do infants express discrete emotions? Adult judgments of facial, vocal, and body actions," *Journal of Nonverbal Behavior*, vol. 17, no. 3, pp. 171–186, 1993.

[56] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe, "Multimodal analysis of expressive gesture in music and dance performances," in *Gesture-based communication in human-computer interaction*. Springer, 2004, pp. 20–39.

[57] P. Fraisse, "Rhythm and tempo," *The psychology of music*, vol. 1, pp. 149–180, 1982.

[58] S. Piana, A. Staglianó, A. Camurri, and F. Odone, "Adaptive body gesture representation for automatic emotion recognition," *ACM Trans. on Interactive Intelligent Systems*, vol. 6, no. 1, 2016.

[59] S. Piana, A. Staglianò, F. Odone, A. Verri, and A. Camurri, "Real-time automatic emotion recognition from body gestures," 2014.

[60] G. H. Roid, L. J. Miller, and C. Koch, *Leiter international performance scale*. Stoelting, 2013.

[61] F. Pons, P. L. Harris, and M. de Rosnay, "Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization," *European journal of developmental psychology*, vol. 1, no. 2, pp. 127–152, 2004.

[62] T. A. Hare, N. Tottenham, M. C. Davidson, G. H. Glover, and B. Casey, "Contributions of amygdala and striatal activity in emotion regulation," *Biological psychiatry*, vol. 57, no. 6, pp. 624–632, 2005.

[63] D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe, K. Hansen, and K.-R. Müller, "How to explain individual classification decisions," *The Journal of Machine Learning Research*, vol. 11, pp. 1803–1831, 2010.

[64] H. O'Reilly, D. Pigat, S. Fridenson, S. Berggren, S. Tal, O. Golan *et al.*, "The EU-emotion stimulus set: A validation study," *Behavior research methods*, vol. 48, no. 2, pp. 567–576, 2016.

[65] A. Rynkiewicz, "Autism spectrum disorders in females. sex/gender differences in clinical manifestation and co-existing psychopathology," 2016, PhD Dissertation. Medical University of Gdansk.

[66] A. Rynkiewicz, B. Schuller, E. Marchi, S. Piana, A. Camurri, A. Lassalle *et al.*, "An investigation of the 'female camouflage effect' in autism using a computerized ADOS-2, and a test of sex/gender differences," *Molecular Autism*, vol. 7, no. 10, 2016.

[67] S. Fridenson-Hayo, S. Berggren, A. Lassalle, S. Tal, D. Pigat, S. Bölte *et al.*, "Basic and complex emotion recognition in children with autism: Cross-cultural findings," *Molecular Autism*, vol. 7, no. 1, p. 52, 2016.

[68] S. Fridenson-Hayo, S. Berggren, A. Lassalle, S. Tal, D. Pigat, N. Meir-Goren *et al.*, "'Emotiplay': A serious game for learning about emotions in children with autism: Results of a cross-cultural evaluation," *European child & adolescent psychiatry*, vol. 26, no. 8, pp. 979–992, 2017.

[69] A. Y. Kaplan, S. L. Shishkin, I. P. Ganin, I. A. Basyul, and A. Y. Zhigalov, "Adapting the P300-based brain-computer interface for gaming: A review," *IEEE Trans. on Computational Intelligence and AI in Games*, vol. 5, no. 2, pp. 141–149, 2013.

[70] C. H. Tan, K. C. Tan, and A. Tay, "Dynamic game difficulty scaling using adaptive behavior-based AI," *IEEE Trans. on Computational Intelligence and AI in Games*, vol. 3, no. 4, pp. 289–301, 2011.

[71] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering game behavior data," *IEEE Trans. on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 266–278, 2015.