

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/280563052>

# Recent developments and results of ASC–Inclusion: An Integrated Internet–Based Environment for Social Inclusion of Children with Autism Spectrum Conditions

CONFERENCE PAPER · MAY 2015

CITATION

1

READS

163

11 AUTHORS, INCLUDING:



**Björn Schuller**

Imperial College London

427 PUBLICATIONS 5,043 CITATIONS

SEE PROFILE



**Erik Marchi**

Technische Universität München

27 PUBLICATIONS 117 CITATIONS

SEE PROFILE



**Simon Baron-Cohen**

University of Cambridge

605 PUBLICATIONS 44,972 CITATIONS

SEE PROFILE



**Helen O'Reilly**

University College London

15 PUBLICATIONS 46 CITATIONS

SEE PROFILE

# Recent developments and results of ASC-Inclusion: An Integrated Internet-Based Environment for Social Inclusion of Children with Autism Spectrum Conditions

Björn Schuller<sup>1</sup>,  
Erik Marchi<sup>2</sup>

<sup>1</sup> Complex and Intelligent  
Systems, University of Passau

<sup>2</sup> MISIP group, TU München  
Germany

bs@fim.uni-passau.de  
erik.marchi@tum.de

Simon Baron-Cohen,  
Amandine Lassalle,  
Helen O'Reilly  
Delia Pigat

Autism Research Centre  
University of Cambridge  
Cambridge, UK  
sb205@cam.ac.uk

Peter Robinson,  
Ian Davies,  
Tadas Baltrusaitis,  
Marwa Mahmoud  
Computer Laboratory  
University of Cambridge  
Cambridge, UK  
peter.robinson@cl.cam.ac.uk

## ABSTRACT

Individuals with Autism Spectrum Conditions (ASC) have marked difficulties using verbal and non-verbal communication for social interaction. The ASC-Inclusion project helps children with ASC by allowing them to learn how emotions can be expressed and recognised via playing games in a virtual world. The platform assists children with ASC to understand and express emotions through facial expressions, tone-of-voice and body gestures. In fact, the platform combines several state-of-the-art technologies in one comprehensive virtual world, including analysis of users' gestures, facial, and vocal expressions using standard microphone and web-cam, training through games, text communication with peers and smart agents, animation, video and audio clips. We present the recent findings and evaluations of such a serious game platform and provide results for the different modalities.

## Author Keywords

Autism Spectrum Conditions, inclusion, virtual worlds, computerised environment, emotion recognition

## INTRODUCTION

Autism Spectrum Conditions (ASC) are neurodevelopmental conditions, characterised by social communication difficulties and restricted and repetitive behaviour. Three decades of research have shown that children and adults with ASC experience significant difficulties recognising and expressing emotions and mental states [4]. These difficulties are apparent when individuals with ASC attempt to recognise emotions from facial expressions [18], from vocal intonation [17], from gestures and body language [31], and from the integration of multi-modal emotional information in context [17]. Limited emotional expressiveness in non-verbal communication is also

characteristic in ASC, and studies have demonstrated individuals with ASC have difficulties directing appropriate facial expressions to others [21], modulating their vocal intonation appropriately when expressing emotion [26] and using appropriate gestures and body language [1]. Integration of these non-verbal communicative cues with speech is also hampered [13]. The social communication deficits, characteristic of ASC, have a pervasive effect on the ability of these individuals to meet age appropriate developmental tasks, from everyday negotiation with the teacher or the shopkeeper to the formation of significant relationships with peers. Individuals with ASC lack the sense of social reciprocity and fail to develop and maintain age appropriate peer relationships [9]. Current findings suggest 1% of the population might fit an ASC diagnosis [5]. This study has been replicated across cultures and age bands, stressing the importance of accessible, cross-cultural means to support this growing group.

The use of Information Communication Technology (ICT) with individuals with ASC has flourished in the last decade for several reasons: the computerised environment is predictable, consistent, and free from social demands, which individuals with ASC may find stressful. Users can work at their own pace and level of understanding, and lessons can be repeated over and over again, until mastery is achieved. In addition, interest and motivation can be maintained through different and individually selected computerised rewards [28]. For these reasons, and following the affinity individuals with ASC show for the computerised environment, dozens of ICT programs, teaching various skills to this population were created. However, most of these tended to be rather specific (e. g., focusing only on recognition of facial expressions from still photos) and low budget, and have not been scientifically evaluated [19]. ICT programs that teach socio-emotional communication have been evaluated including *I can Problem-Solve*; others aim to teach social problem solving [10], such as *FEFFA*, e. g., by emotion recognition from still pictures of facial expressions and strips of the eye region [11]. *Emotion Trainer* teaches emotion recognition of four emotions from facial expressions [37]; *Lets Face It* is teaching emotion and identity recognition from facial expressions [39], and *Junior Detective* program combines ICT with group training in order to teach social skills to children with ASC [8]. These examples demonstrate

how focused most ICT solutions are in their training, focusing mostly on emotion recognition from facial expressions and contextual situations, with very little attention given to emotional gestures or emotional voices. In addition, there were no reports of ICT programs teaching emotional expressiveness. Further training programs such as *Mind Reading* implement an interactive guide to emotions and teaches recognition of 412 emotions and mental states, systematically grouped into 24 emotion groups, and 6 developmental levels (from the age of four years to adulthood). The evaluations of *Mind Reading* resulted in limited generalization when adults with ASC used the software [19]. However, when 8-11 year old children with ASC had used it, improved generalization was found.

In the last decade, with the rapid development of internet-based communication, web applications have been increasingly used for social interaction, forming online communities and social networks. Anecdotal reports of the emergence of online autistic communities, and the use of forums, chat rooms and virtual-worlds, show the great promise the internet holds for better inclusion and social skills training for adolescents and adults with ASC [20]. However, there has been no scientifically documented attempt to use the internet for structured training of socio-emotional communication for individuals with ASC. Furthermore, since intervention into ASC has been shown to be more effective when provided early in life, using the internet as a platform for the support of children with ASC could significantly promote their social inclusion.

Virtual Environments (VE) form another domain with immense possibilities for those with ASC and related social difficulties. VE are artificial computer generated three-dimensional simulations and come in single- or multi-user forms. In either format, the user can operate in realistic scenarios to practice social skills, conversations, and social problem solving. Moore and colleagues investigated VE and children and youth with ASC and found that over 90 % of their participants used the VE to recognise basic emotions [27]. Other studies have also shown the potential for those with ASC to use VE for socio-emotional skills training and for other learning purposes [30].

The ASC-Inclusion project<sup>1</sup> [36, 35] – dealt with herein – suggests advanced ICT-enabled solutions and serious games for the empowerment of children with ASC who are at high risk of social exclusion. The project created an internet-based platform that assists children with ASC, and those interested in their inclusion, to improve their socio-emotional communication skills, attending to the recognition and expression of socio-emotional cues and to the understanding and practice of conversational skills. ASC-Inclusion combines several state-of-the-art technologies in one comprehensive game environment, including analysis of users' gestures, facial, and vocal expressions, training through games, text chatting, animation, video and audio clips. Despite the innovative technologies involved, the ASC-Inclusion is aimed for home use. Though designed to assist children with ASC, ASC-Inclusion could serve other population groups characterised by deficient emotional understanding and social skills, such as children with

learning difficulties [7], attention deficit and hyperactivity disorder (ADHD) [12], behavioural and conduct problems [38], or socio-emotional difficulties [33].

Additionally, the project integrates the carers into the platform, and thereby: a) increase engagement of parents, siblings and ASC children; b) incorporate carers inputs to the system, also monitor/correct (and by this also train) the system; c) enable new kind of interaction and hence benefit from didactic activities. The children are then able to play with the carers cooperatively, and have the chance to directly see emotion and social behaviour from another person next to them. In addition, the system platform is able to collect displays of emotion from the carers and parents that can be used for further training of the system and for the "corrective" technology as it enables carers input to the corrective system and to the emotion recognition system. ASC-Inclusion further support the ability to provide formative assessment as part of the user interaction in a semantically comprehensible and appealing way to the children. The gain for the children is highly significant: They receive understandable and targeted corrective feedback enabling them to adjust their facial, vocal and gesture behaviour to match prototypical manifestations.

The remainder of this paper is structured as follows: first, a detailed description of the user requirements and specification is given (Section User Requirements); then we describe the three modalities namely face, voice and body gesture (Sections Face Analysis, Voice Analysis, Gesture Analysis) and the formative assessment module (Section Formative assessment), before commenting the platform (Section Platform). We next describe content creation (Section Content Creation) and adult-child cooperative playing (Section Adult & cooperative playing). We then comment on the psychological evaluation and results (Section Psychological experiments and evaluation) before concluding in Section Conclusions.

## USER REQUIREMENTS

Previous literature reports social communication difficulties in individuals with ASC as well as enhanced abilities in other non-social areas such as systemising [6]. This provided us with a greater insight into the social, cognitive and behavioural abilities of these potential end users. Review of the design, content and effectiveness of the *Mind Reading* DVD as an intervention and *The Transporters* DVD [16] which have shown to enhance the ability of children with ASC to recognise emotions, permitted identification of a method through which learning material can be presented in an engaging manner which capitalises on the enhanced systemising skills seen in individuals with ASC.

A number of focus groups with ASC children and families have been carried out to provide qualitative feedback in order to specify user requirements in terms of game preferences and interests, game concepts and design, content specification, user behaviour and cognition, and learning styles. Targeted focus groups regarding teaching about vocal expression of emotions have also been run. An experts focus group was also formed which included specialists in the field of ASC, the purpose of the experts focus group was to provide feedback regarding the VE concept and implementation, as well as to highlight difficulties common to ASC children which may affect their

<sup>1</sup><http://www.asc-inclusion.eu>

ability to interact with the VE. The results of these activities have fed into the platform development. These focus groups have permitted identification of the learning requirements of our platform user and have been used to specify the learning structure that is adopted by the platform. Furthermore both the user and expert focus groups have identified technological opportunities and constraints that have fed into the platform development to ensure a user centric design.

Focus groups and user testing of subsystem prototypes and the emotion modality analysers (cf. Sections Face Analysis, Voice Analysis, Gesture Analysis, Platform) have been carried out across the clinical sites in United Kingdom, Sweden and Israel. Specifically the platform tutorials, games, quizzes to assess the users learning and emotion analysers have been tested. Feedback from these user testing sessions has been used to update the subsystems to ensure they are designed with the end user in mind and are comprehensible to our users. Furthermore as the project has evolved, new ethical considerations arose which we have addressed to ensure the project follows strict ethical guidelines.

### FACE ANALYSIS

The Facial Affect Computation Engine (FACE) [35] was enhanced in order to enable it running alongside the game and other inference engines on a modes domestic computer. This has also benefited from an improved version of the face tracker [3] which uses continuous conditional neural fields for structured regression. Initial pilot trials with users in the field revealed various difficulties. In particular, users were not clear when the face tracker was working and when it had lost registration. They were therefore confused as whether then inferences were meaningful or not. A small window showing the video from the camera with tracking information superimposed has now been included in the game to give immediate feed-back. Additionally a real-time analyser of hand-over-face occlusion detection has been developed by designing an avatar that mimics hand-over-face occlusions displayed by the user in real-time. The methodology described in [22] was implemented and the output of the classifier was mapped to an avatar head. For simplicity, we extracted only spatial features: Histograms of Oriented Gradients (HOGs) and likelihood values of facial landmarks detection. Principal Component Analysis (PCA) was then used to reduce the dimensionality of HOG features. Using the extracted features, we trained binary linear Support Vector Machine (SVM) classifiers for the following classification tasks: 1) Occlusion/ no-Occlusion, 2) Chin occlusion, 3) Lips occlusion, and 4) Middle face (nose/cheek) occlusion. A simple avatar was built with a face and a hand using Blender: an open source 3D animation suite. Using this avatar head, we then created a set of images representing different hand-over-face positions to display them according to the output of the classifier. Figure 1 shows a sample screen shot of the systems real-time output displaying two windows: One window is displaying the captured webcam image of the user with the head and facial landmarks highlighted, and the second window is displaying the output avatar image corresponding to the hand-over-face gesture detected.

### VOICE ANALYSIS

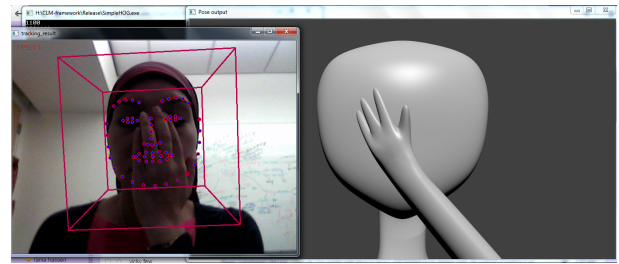


Figure 1. Hand occluding chin, lips and middle face detected..

According to the feedback from the evaluation carried out at the clinical sites, the vocal expression evaluation system [35] was refined and enhanced with new features. The children liked the activity with the voice analyser software; in particular they had fun playing with their voices and influencing the plots. For this reason, we decided to introduce the representation of the tracked acoustic features [25, 23, 24] and the display of the appropriateness of the childs expression in the analyser. Applying the openSMILE audio feature extractor [15, 14], features are extracted and tracked over time: 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 vocal expression evaluation system is shown in Figure 2. In the "Target emotion" box the player chooses the emotion she or he wants to play 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 on the "Recognised Emotion" box showing the detected emotion is expressed. Besides providing the classification result, the analyser shows a confidence measure that gives a measure on how much the system is certain about the recognised emotion against the remaining emotions. The confidence is represented by the probability estimate derived from the distance between the point and the hyper-plane. According to the correctness of the expressed emotion virtual coins are earned and on the bottom right part of the GUI and a corrective feedback is shown. The 'traffic light' on the top of the gauge bars indicates if the extracted parameters are distant or close to the reference values. If a green light is shown the childs expression is close to the reference; the red light indicates high distance between the reference value and the extracted one.

The new data collected at the clinician sites have been used to ameliorate the accuracy of the emotion recognition system. For the training of a voice analyser, a data set of prototypical emotional utterances containing sentences spoken in Hebrew, Swedish and English by children with ASC ( 8 per language) and typically developing ( 10 per language) has been created. The recordings mainly focused on the six 'basic' emotions except disgust: happy, sad, angry, surprised, and afraid plus three other states: ashamed, calm and proud. In the case of the English dataset a total of 19 emotions plus neutral and calm were collected. In order to enable adult and child cooperative playing, the enhanced version of the voice analyser is now capable to handle adult-child models and language-dependent

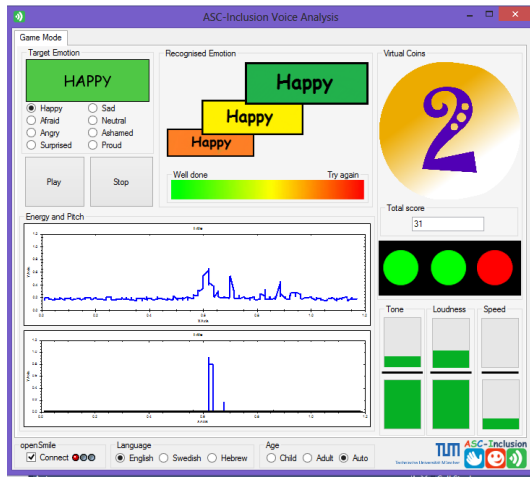


Figure 2. Enhanced vocal expression evaluation system.

models for emotion recognition. The acoustic models for adults were trained on the voice recordings collected by the clinical teams, as part of the EU-Emotion Stimulus set.

The output from the voice analyser is encoded in EmotionML [34] and delivered through the ActiveMQ communication infrastructure, allowing integration with the face analysis, the gesture analysis and formative assessment module to provide multi-modal inference in the central platform.

## GESTURE ANALYSIS

An on-line gesture analyser based on the EyesWeb<sup>2</sup> XMI platform was improved based on its previous version [35]. To enhance the overall module performances, we introduced an adaptive feature representation algorithm that helps in generalization and in situations where the amount of available data is not big, this technique is called dictionary learning [29]. In addition, in order to make the software available on a variety of hardware configurations where depth camera are 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 system is based on the method proposed in [32]: a set of low level movement features are extracted from motion capture data, then high level representations of movement qualities (i.e., fluidity, impulsiveness, etc.) are built. These representations are observed through time and an histogram-based representation is created, starting from the histograms a further refinement of the representation is computed using sparse coding and finally the data is classified through SVMs to get an infer of the expressed emotion. The module is fully integrated in the game platform and it can communicate with the other modules – via activeMQ and EmotionML messages – to record streams of data and log outputs that can be then processed offline. For training the SVMs, a dataset of 15 people expressing the six basic emotions (happiness, anger, sadness, disgust, fear, and surprise) was collected. Each person was asked to express one of the six emotions with their body from three to seven times in front of

<sup>2</sup>[http://www.infomus.org/eyesweb\\_ita.php](http://www.infomus.org/eyesweb_ita.php)



Figure 3. Body gesture analyser.

a Microsoft Kinect sensor. On average, each recording session lasted about 10 minutes. After the recording sessions, each video was manually segmented to find the prototypical expressive gestures of each emotion. Those segments were used to evaluate the data with human subjects. 60 judges labelled 10 segments, each. The material was used to train SVM for the classification modules.

Based on the body gesture analyser two games were developed [32]. Both games perform a real-time automatic emotion recognition, and interact with the user by asking to guess and to express an emotion with the body (cf. Figure 3). During the games the user interacts with the GUI through body gestures.

## FORMATIVE ASSESSMENT

The face, voice, and body gesture analysers recognize the performed emotion through the respective modality, where the output is restricted to either the discrete emotion categories or the continuous arousal-valence values. However, using this output to display a binary correct/incorrect response for a target emotion is not informative for children who are trying to learn how to perform emotions correctly. The aim of the formative assessment is to complement each of the classification modules, by including a formative assessment module that provides corrective instructions when an emotion is performed incorrectly. A candidate listing for forms of semantically meaningful mid-level features considering a variety of affective states for each of the three modalities was elaborated. The list was based on modality-specific characteristics given for different emotions. These attributes serve as a dictionary between the game engine and the formative feedback module, where the feedback generated as either increase/decrease (for voice and gesture attributes) or extraneous/missing (for facial action units) are converted to audio-visual output through the game engine. The feedback generation is based on explanation vector generation [2]. This generalised technique generates 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. Using explanation vector generation in an iterative manner, the feature set is modified so that it resembles the target emotion better. For the audio and gesture modalities, the original and the modified feature sets. For the face modality, attribute classifiers are utilized to carry out the comparison process in a semantically meaningful level: The attribute values, which are stated as either active or passive for the attributes of action units, of original and modified features sets are compared to state which of the action units are missing or incorrectly present. The module is also

integrated with the other modules, where the communication is handled via ActiveMQ.

## PLATFORM

The general architecture presented in [35] has been updated and extended including the formative assessment module. An overview of the new integrated platform architecture is given in Figure 4. The container of the game engine in flash is ZINC<sup>3</sup>; the container loads the game environment and the subsystems services. The flash game engine is lying under the game logic and the user interface. It is responsible for the communication with the different services through Apache ActiveMQ by applying Streaming Text Oriented Messaging Protocol<sup>4</sup> (STOMP). The communication with the different services are based on three components. The subsystem control component sends control messages to the different services – the control messages are sent to a unique ActiveMQ queue for each of the services (face, body, voice). When a message is sent to the activeMQ queue, the subsystem service reads it and processes the command. The STOMP message receiver gets XML text messages from ActiveMQ topic based on STOMP protocol. The component connects to ActiveMQ through STOMP and starts receiving messages from the server. Messages are then used by the game engine in order to provide a certain feedback to the user. As above, the XML messages are encoded using the EmotionML standard. The ActiveMQ process is used as the pipe to push the information from the different subsystems to the game engine. ActiveMQ includes one Topic (named asc) which contains all classification results and parameters from the different modalities (based on the running services). ActiveMQ also includes four queues (one per subsystem) which contain control messages produced by the platform (start, stop, shutdown, etc.). The four subsystems provide output provide the actual values of the recognised emotion and the platform compares them according to the target space.

Besides the platform architecture, the Learning Manager Application (LMA) was improved (cf. Figure 5). This application is the main tool to control, personalise, and present the learning material to the user. It also enables the navigation within the program including access to the practice games. The LMA enables to track the user's play patterns for data collection and for system improvement purposes. The LMA also manages the rewards granted to the child while going through the program. The 'monetary system' is the basis for the VE 'economy', planned to motivate the child and encourage long-term engagement. Advancing in the program by actively participating in 'research sessions' (lessons) and playing the practise games to a pre-defined achievement, is the only way to earn virtual money in the VE, and virtual money is needed for virtually anything the child might want to do in the VE apart from 'doing research' (learning) - namely: play non-curricular games, buy goodies, go to fun locations out of the camp, etc. This way we maintain an efficient balance between coping with the possibly challenging emotions content, and having the opportunity for recreation and pure fun. This is also meant to prevent a situation of ignoring the program or 'getting stuck' on a certain activity or comfort zone. Additionally, further

<sup>3</sup><http://www.multimedia.com/software/zinc/>

<sup>4</sup><http://stomp.github.io>

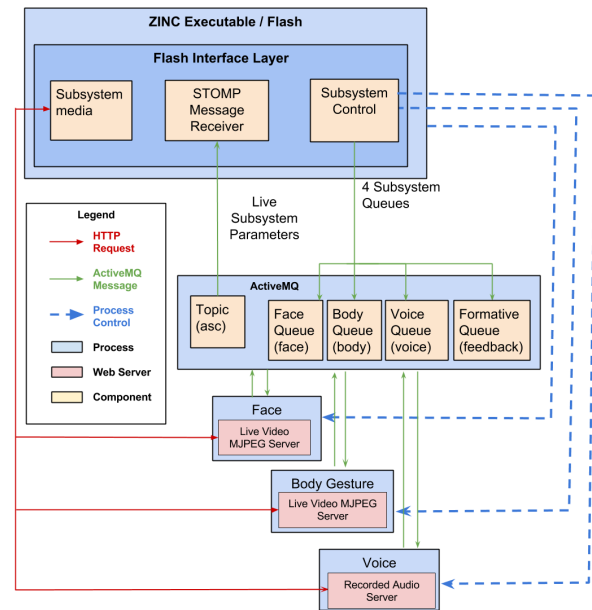


Figure 4. Integrated platform architecture.



Figure 5. The new learning program manager application user interface.

motivational elements were added, such as the avatar, the camp-home, and the virtual wallet. Four new characters have been designed to serve and support the learning sessions and add interest and fun. Further, four fun locations were designed as out-of-camp 'tourist' recreation destinations. The new locations contain fun games with themed content that also has an enrichment element. The child needs to earn the ticket to go there, and usage is limited.

An improved expression game was developed (cf. Figure 6). It integrates the face, voice, and body language analysis technologies. This game is the main activity in the 'expression sub-unit' at the end of each of the learning units. The game is designed as a 'race' board game. Each turn the child is asked to express an emotion in a chosen modality. If she/he expressed it well enough, the racing robot advances one step on the board. Whoever gets to the ending point first wins.

## CONTENT CREATION

In order to determine the most important emotions for social interaction in children with ASC, an emotion survey was developed and completed by parents, typically developing adults and parents in particular of children with ASC. 20 emotions – happy, sad, afraid, angry, disgusted, surprised, excited, interested, bored, worried, disappointed, frustrated, hurt, kind, jealous, unfriendly, joking, sneaky, ashamed, and proud – were



**Figure 6.** The improved ‘Expression’ game with face, voice, and body gesture modalities.

identified by the survey as being the most important for social interaction and selected to be taught through the platform. Additionally, 496 face stimuli, 82 body gesture stimuli and 95 social scene stimuli previously recorded with professional actors were validated. The stimuli were divided into 21 validation surveys with each survey containing 30-45 emotion stimuli. Each survey was completed by 60 typically developing adults (20 per country), at minimum age of 18 years. For each stimulus validated a percentage of correct recognition score and a chance corrected percentage recognition score was calculated. A further 20 UK-based voice surveys have been created and validated along side voice recordings in Hebrew and Swedish: As outlined above, to assist with the development of the voice analysis module, a series of emotion phrases were recorded by in UK, Israel, and Sweden. A total of 20 children (10 typically developing and 10 children with ASC) were recorded at each site, in English, Hebrew and Swedish. In UK, 19 emotions plus neutral were recorded. Audio content into four additional European languages: French, German, Italian and Spanish was recorded and later selected, and validated. A total of 84 Sentences were selected for translation. The selection procedure was as follows: a total of four sentences were chosen per emotion. Sentences that showed the highest mean passing rates for any given emotion in the previous voice validation study (Swedish data) were selected.

#### ADULT & COOPERATIVE PLAYING

In order to have an understanding of parent’s needs on the context of participating on an intervention like ASC-Inclusion, a thorough literature review has been carried out, focusing at previous work with parents of children on the spectrum. In Poland, 40 parents of children with ASC and ADHD, did psychological experiments to identify the needs of the adults as active users of the VE environment, analysed what unique needs of parents were in learning recognition and expression of emotions for themselves, as well as on the best way for adults to communicate and tutor their children on these skills. A pilot study with 3 games of ASC-Inclusion program was conducted and defined adult users requirements and specification for the program. An analysis of the assesment of parents broader autism phenotype, including the measurement of the magnitude of stress in the parent-child system, the assesment of how ASC children understand emotions before and after usage of the VE (Virtual Enviroment), and the assesment of adult-child interaction during collaborative playing with the VE with successful completion of joint tasks. Psychological experiments and evaluations were conducted with 39 Polish children and adults. In Sweden, 20 children in special school settings in different municipalities in the Stockholm area were recruited. This group of children is using the ASC Inclusion program in a school setting for 6 months. During the program,

three measurement points (pre, post, follow-up) are set in order to measure the results of quantitative changes of symptoms, impairment, and quality of life on the Strengths and Difficulties Questionnaire (SDQ) and Social Responsiveness Scale (SRS) (Teacher report), the Global Assessment Scale, and Global Clinical Impression (Severity Scale) (expert ratings), and the Kids Screen (student self-report). In addition, school attainment data (school performance, attendance) is collected as outcome variables.

#### PSYCHOLOGICAL EXPERIMENTS AND EVALUATION

Psychological evaluation of the system took place through an Open Trial (OT). About 20 children in each site went through the intervention. The different sites are currently on different stages of the OT: in UK the OT is finished running all its participants (and finalized with 15 participants on the intervention group and 3 controls), In Israel the OT ran half of its participants and intends to finish with 15 participants in each group, and in Sweden the OT is started after finishing task validation stage, and also aiming at 15 participants per group. In Israel, preliminary results reveal significant improvement of children in the intervention group, compared to the waiting-list control group, on some of the emotion recognition tasks, and potentially significant results on other tasks, when a larger sample is collected. 12 children with ASC, participated in the open trial, also tested the technological systems with their parents. Children and parents played a game of emotional charades, in which children were asked to express an emotion in a selected modality (face, voice, or integration) from a number of choices of emotions provided by the computer. The child’s parent was asked to guess the emotion the child intended to produce, and the computer collected the data, and gave feedback to the child on the matching between child’s choice and the parent’s guess (match/mismatch). Most children (11/12) found this activity to be very entertaining. An Analysis of system logs revealed that the system correctly recognized 80% of the children’s expressions, out of the trials that did not result in technical errors. In United Kingdom, a sample of 18 children (7F) participated to the open trial. Data was collected from 15 children (4F) in the intervention group and 3 children (all females) in the control group. The intervention group had an average age of 8 years and half (102.8 months,  $SD=13.59$ ), an average ADOS score of 13.93 ( $SD=4.68$ ) and an average IQ score of 99.73 ( $SD=6.74$ ). The intervention that was evaluated involved two main tools: the prototype of Camp exploration and parent child activities around emotions, aiming at consolidating the concepts learned in the Camp exploration each week and at encouraging the transfer of the teaching of Camp Exploration to real life. The intervention took 73.83 days on average ( $SD: 18.71$ ) to complete. The emotion recognition ability of children with autism was assessed across various modalities (face, body, voice and context) in two visits, before and after they received the intervention. Results show that the group of children with ASC in the intervention group performed significantly better on the integration survey (recognizing emotion from context) and from the body survey. Additionally the group of children with ASC was also rated significantly higher (by their parents) on the Socialization scale of the Vineland (although this finding is to be taken cautiously, as the sample size is only 11 for this analysis) while their ratings did

not improve on the SRS. Our results show that children with autism improved on emotion recognition and socialization after undergoing the intervention, which suggests that, as a whole, the intervention was effective in helping children learn socio-emotional skills. Those results echo the feedback from the parents of children with ASC who were part of the intervention group. Indeed, they rated the platform 8.14/10 and all noticed positive changes in the socio-emotional behaviour of their child after using Camp Exploration. This is very encouraging. Thus feedback from both British and Israeli sites suggest the VE is largely liked by users and their parents and perceived as effective in improving children's emotional repertoire. Parents and children have provided useful suggestions for improvement, which could be utilized in future versions of the VE.

## CONCLUSIONS

We described the recent findings and developments of the gaming platform ASC-Inclusion targeted to children aged 5 to 10 years with ASC. User requirements and specification have been refined and further analysed, tailoring the needs of the target audience and creating user scenarios for the platform subsystems. The integration of the three subsystem into the game platform was refined. The technical partners collaborated to improve communication protocols and platform architecture for the main game engine. The face analyser implements a automatic facial analysis system that can be used to provide feedback to children as they learn expressions. The vocal expression evaluation system has been adapted and refined in a collaborative loop process with the clinical sites. The voice analyser was improved to support formative assessment and adult-child cooperative playing. A refined version of the body gesture analyser was developed including a new version of the module that works on 2D webcam data. The body gesture analyser and the designed game demonstrations were evaluated. In parallel, the platform was improved yielding more advanced versions of the game. The game design and flow were improved according to the clinician's recommendations and following usability tests results. Visual and vocal feedbacks were refined to fit the clinical teams' recommendations. New contents were created, including the EUemotion Stimulus set and additional voice material in French, German, Italian and Spanish which have been validated. The additional functionalities recently introduced in the ASC-Inclusion platform, namely the adult and child cooperative playing and the formative assessment to generate corrective feedback were developed and integrated in the system.

In conclusion, results from the evaluation at the clinical sites are promising, indicating children with ASC improved on emotion recognition and socialization after undergoing the intervention: children with ASC in the intervention group performed significantly better on the integration survey (recognizing emotion from context) and from the body survey. The group of children with ASC was also rated significantly higher (by their parents) on the Socialization scale of the Vineland. Itasks, and potentially significant results on other tasks, when a larger sample is collected. Thus, feedback from both British and Israeli sites suggest the VE is largely liked by users and their parents and perceived as effective in improving children's

emotional repertoire. Parents and children have provided useful suggestions for improvement, which could be utilized in future versions of the VE.

## Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 289021 (ASC-Inclusion).

## ADDITIONAL AUTHORS

Ofer Golan, Shimrit Fridenson, Shahar Tal (Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel, email: ofer.golan@biu.ac.il) Shai Newman, Noga Meir, Roi Shillo (Compedia Ltd, Ramat-Gan, Israel, email: newmans@compedia.net) Antonio Camurri, Stefano Piana, Alessandra Staglianò (InfoMus Lab, DIST, University of Genoa, Genoa, Italy, email: antonio.camurri@unige.it) Sven Bölte, Daniel Lundqvist, Steve Berggren (Center of Neurodevelopmental Disorders, Karolinska Institute, Stockholm, Sweden, email: sven.bolte@ki.se) Aurelie Baranger and Nikki Sullings (Autism-Europe aisbl, Brussels, Belgium, email: aurelie.baranger@autismeurope.org), Metin Sezgin, Neşe Alyuz (Intelligent User Interfaces Lab, Koç University, Istanbul, Turkey, email: mtsezgin@ku.edu.tr), Agnieszka Rynkiewicz, Kacper Ptaszek, Karol Ligmann (Spectrum ASC-Med, Poland, email: rynkia@spectrumascmed.com).

## REFERENCES

1. Attwood, T. *Asperger's syndrome: A guide for parents and professionals*. Jessica Kingsley Pub, 1998.
2. Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., and Müller, K.-R. How to explain individual classification decisions. *The Journal of Machine Learning Research 11* (2010), 1803–1831.
3. Baltruaitis, T., Robinson, P., and Morency, L.-P. Continuous conditional neural fields for structured regression. In *Computer Vision ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8692 of *Lecture Notes in Computer Science*. Springer International Publishing, 2014, 593–608.
4. Baron-Cohen, S. *Mindblindness: An essay on autism and theory of mind*. MIT press, 1997.
5. Baron-Cohen, S., Scott, F. J., Allison, C., Williams, J., Bolton, P., Matthews, F. E., and Brayne, C. Prevalence of autism-spectrum conditions: Uk school-based population study. *The British Journal of Psychiatry 194*, 6 (2009), 500–509.
6. Baron-Cohen, S., Wheelwright, S., Lawson, J., Griffin, R., Ashwin, C., Billington, J., and Chakrabarti, B. Empathizing and systemizing in autism spectrum conditions. *Handbook of autism and pervasive developmental disorders 1* (2005), 628–639.
7. Bauminger, N., Edelsztein, H. S., and Morash, J. Social information processing and emotional understanding in children with ld. *Journal of learning disabilities 38*, 1 (2005), 45–61.
8. Beaumont, R., and Sofronoff, K. A multi-component social skills intervention for children with asperger syndrome: The junior detective training program. *Journal of Child Psychology and Psychiatry 49*, 7 (2008), 743–753.
9. Bell, C. C. Dsm-iv: Diagnostic and statistical manual of mental disorders. *JAMA: The Journal of the American Medical Association 272*, 10 (1994), 828–829.
10. Bernard-Opitz, V., Sriram, N., and Nakhoda-Sapuan, S. Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of Autism and Developmental Disorders 31*, 4 (2001), 377–384.



11. Bölte, S., Hubl, D., Feineis-Matthews, S., Prvulovic, D., Dierks, T., and Poustka, F. Facial affect recognition training in autism: can we animate the fusiform gyrus? *Behavioral neuroscience* 120, 1 (2006), 211.
12. Da Fonseca, D., Seguier, V., Santos, A., Poinso, F., and Deruelle, C. Emotion understanding in children with adhd. *Child Psychiatry & Human Development* 40, 1 (2009), 111–121.
13. de Marchena, A., and Eigsti, I.-M. Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism Research* 3, 6 (2010), 311–322.
14. Eyben, F., Wenginger, F., Groß, F., and Schuller, B. Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor. In *Proceedings of the 21st ACM International Conference on Multimedia, MM 2013*, ACM, ACM (Barcelona, Spain, October 2013), 835–838.
15. Eyben, F., Wöllmer, M., and Schuller, B. openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor. In *Proceedings of the 18th ACM International Conference on Multimedia, MM 2010*, ACM, ACM (Florence, Italy, October 2010), 1459–1462.
16. Golan, O., Ashwin, E., Granader, Y., McClintock, S., Day, K., Leggett, V., and Baron-Cohen, S. Enhancing emotion recognition in children with autism spectrum conditions: an intervention using animated vehicles with real emotional faces. *Journal of autism and developmental disorders* 40, 3 (2010), 269–279.
17. Golan, O., Baron-Cohen, S., and Golan, Y. 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* 38, 8 (2008), 1534–1541.
18. Golan, O., Baron-Cohen, S., and Hill, J. The cambridge mindreading (cam) face-voice battery: Testing complex emotion recognition in adults with and without asperger syndrome. *Journal of autism and developmental disorders* 36, 2 (2006), 169–183.
19. Golan, O., LaCava, P. G., and Baron-Cohen, S. Assistive technology as an aid in reducing social impairments in autism. *Growing Up with Autism: Working with School-Age Children and Adolescents* (2007), 124–142.
20. Jordan, C. J. Evolution of autism support and understanding via the world wide web. *Intellectual and developmental disabilities* 48, 3 (2010), 220–227.
21. Kasari, C., Chamberlain, B., and Bauminger, N. 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. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 2001, 309–323.
22. Mahmoud, M. M., Baltrušaitis, T., and Robinson, P. Automatic detection of naturalistic hand-over-face gesture descriptors. In *Proceedings of the 16th International Conference on Multimodal Interaction, ICMI '14*, ACM (New York, NY, USA, 2014), 319–326.
23. Marchi, E., Batliner, A., Schuller, B., Fridenzon, S., Tal, S., and Golan, O. Speech, Emotion, Age, Language, Task, and Typicality: Trying to Disentangle Performance and Feature Relevance. In *Proceedings First International Workshop on Wide Spectrum Social Signal Processing (WS3P 2012), ASE/IEEE SocialCom 2012, ASE/IEEE, IEEE* (Amsterdam, The Netherlands, September 2012).
24. Marchi, E., Ringeval, F., and Schuller, B. Perspectives and Limitations of Voice-controlled Assistive Robots for Young Individuals with Autism Spectrum Condition. In *Speech and Automata in Health Care (Speech Technology and Text Mining in Medicine and Healthcare)*, A. Neustein, Ed. De Gruyter, Berlin, 2014. invited contribution.
25. Marchi, E., Schuller, B., Batliner, A., Fridenzon, S., Tal, S., and Golan, O. Emotion in the Speech of Children with Autism Spectrum Conditions: Prosody and Everything Else. In *Proceedings 3rd Workshop on Child, Computer and Interaction (WOCCI 2012), Satellite Event of INTERSPEECH 2012*, ISCA, ISCA (Portland, OR, September 2012).
26. McCann, J., and Peppe, S. Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders* 38, 4 (2003), 325–350.
27. Moore, D., Cheng, Y., McGrath, P., and Powell, N. J. Collaborative virtual environment technology for people with autism. *Focus on Autism and Other Developmental Disabilities* 20, 4 (2005), 231–243.
28. Moore, D., McGrath, P., and Thorpe, J. Computer-aided learning for people with autism – a framework for research and development. *Innovations in Education and Training International* 37, 3 (2000), 218–228.
29. Olshausen, B. A., and Field, D. J. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research* 37, 23 (1997), 3311–3325.
30. Parsons, S., Beardon, L., Neale, H., Reynard, G., Eastgate, R., Wilson, J., Cobb, S., Benford, S., Mitchell, P., and Hopkins, E. Development of social skills amongst adults with aspergers syndrome using virtual environments: the as interactiveproject. In *Proc. The 3rd International Conference on Disability, Virtual Reality and Associated Technologies, ICDVRAT* (2000), 23–25.
31. Philip, R., Whalley, H., Stanfield, A., Sprengelmeyer, R., Santos, I., Young, A., Atkinson, A., Calder, A., Johnstone, E., Lawrie, S., et al. Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological medicine* 40, 11 (2010), 1919–1929.
32. Piana, S., Staglianò, A., Odone, F., and Camurri, A. Emotional charades. In *Proceedings of the 16th International Conference on Multimodal Interaction, ICMI '14*, ACM (New York, NY, USA, 2014), 68–69.
33. Pollak, S. D., Cicchetti, D., Hornung, K., and Reed, A. Recognizing emotion in faces: developmental effects of child abuse and neglect. *Developmental psychology* 36, 5 (2000), 679.
34. Schröder, M., Baggia, P., Burkhardt, F., Pelachaud, C., Peter, C., and Zovato, E. Emotion markup language (emotionml) 1.0. *W3C Working Draft* 29 (2010), 3–22.
35. Schuller, B., Marchi, E., Baron-Cohen, S., O'Reilly, H., Pigat, D., Robinson, P., Davies, I., Golan, O., Fridenzon, S., Tal, S., Newman, S., Meir, N., Shillo, R., Camurri, A., Piana, S., Staglianò, A., Bölte, S., Lundqvist, D., Berggren, S., Baranger, A., and Sullings, N. The state of play of ASC-Inclusion: An Integrated Internet-Based Environment for Social Inclusion of Children with Autism Spectrum Conditions. In *Proceedings 2nd International Workshop on Digital Games for Empowerment and Inclusion (IDGEI 2014)*, L. Paletta, B. Schuller, P. Robinson, and N. Sabouret, Eds., ACM, ACM (Haifa, Israel, February 2014). 8 pages, held in conjunction with the 19th International Conference on Intelligent User Interfaces (IUI 2014).
36. Schuller, B., Marchi, E., Baron-Cohen, S., O'Reilly, H., Robinson, P., Davies, I., Golan, O., Fridenzon, S., Tal, S., Newman, S., Meir, N., Shillo, R., Camurri, A., Piana, S., Bölte, S., Lundqvist, D., Berggren, S., Baranger, A., and Sullings, N. ASC-Inclusion: Interactive Emotion Games for Social Inclusion of Children with Autism Spectrum Conditions. In *Proceedings 1st International Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI 2013) held in conjunction with the 8th Foundations of Digital Games 2013 (FDG)*, B. Schuller, L. Paletta, and N. Sabouret, Eds., ACM, SASDG (Chania, Greece, May 2013).
37. Silver, M., and Oakes, P. Evaluation of a new computer intervention to teach people with autism or asperger syndrome to recognize and predict emotions in others. *Autism* 5, 3 (2001), 299–316.
38. Stevens, D., Charman, T., and Blair, R. J. Recognition of emotion in facial expressions and vocal tones in children with psychopathic tendencies. *The Journal of genetic psychology* 162, 2 (2001), 201–211.
39. Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S., Kaiser, M. D., and Schultz, R. T. 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* 51, 8 (2010), 944–952.