Strathprints Institutional Repository

Alcorn, Alyssa and Pain, Helen and Rajendran, Gnanathusharan and Smith, Tim and Lemon, Oliver and Porayska-Pomsta, Kaska and Foster, Mary Ellen and Avramides, Katerina and Frauenberger, Christopher and Bernardini, Sara (2011) *Social communication between virtual characters and children with autism.* In: Artificial Intelligence in Education. Lecture Notes in Computer Science, 6738 . Springer, pp. 7-14. ISBN 978-3-642-218682

Social Communication between Virtual Characters and Children with Autism

Alyssa Alcorn[1], Helen Pain[2], Gnanathusharan Rajendran[3], Tim Smith[4], Oliver Lemon[1], Kaska Porayska-Pomsta[5], Mary Ellen Foster[1], Katerina Avramides[5], Christopher Frauenberger[6], and Sara Bernardini[5]

[1] Heriot Watt University
[2] University of Edinburgh
[3] Strathclyde University
[4] Birkbeck College
[5] London Knowledge Lab
[6] University of Sussex
A.Alcorn@hw.ac.uk
http://echoes2.org

Abstract. Children with ASD have difficulty with social communication, particularly joint attention. Interaction in a virtual environment (VE) may be a means for both understanding these difficulties and addressing them. It is first necessary to discover how this population interacts with virtual characters, and whether they can follow joint attention cues in a VE. This paper describes a study in which 32 children with ASD used the ECHOES VE to assist a virtual character in selecting objects by following the character's gaze and/or pointing. Both accuracy and reaction time data suggest that children were able to successfully complete the task, and qualitative data further suggests that most children perceived the character as an intentional being with relevant, mutually directed behaviour.

Keywords: autism spectrum disorder, virtual environment, virtual character, joint attention, social communication, technology-enhanced learning, HCI.

1 Introduction

The autism spectrum encompasses a group of pervasive developmental disorders characterised by notable difficulties in communication and social interaction, plus the presence of repetitive behaviours and interests [1]. Virtual environments and characters are a promising method for supporting social communication in children on the autism spectrum due to the potential for skills to be practiced repeatedly, in a way that may be less threatening, less socially demanding and more controllable than a face-to-face interaction with a human partner [2, 3]. There is potential for supporting skill generalisation by changing the virtual setting of tasks, or introducing multiple characters. To date, the use of VEs in interventions for those with ASD has been often narrowly focused on specific social situations, such as adolescents navigating through a cafe [4], rather than on supporting foundation skills like joint attention. Also, few have targeted young children (though see [5] for an exception).

G. Biswas et al. (Eds.): AIED 2011, LNAI 6738, pp. 7–14, 2011.

Many general questions about the abilities of young children with ASD to interact with VEs and virtual characters remain unanswered. The ECHOES Technology Enhanced Learning (TEL) project is developing an intelligent multi-modal VE for supporting and scaffolding social communication in children aged 5-8 years, with and without an ASD. It comprises a range of touch screen-based learning activities focused on joint attention initiation and response. AI software modules direct an autonomous virtual character and are capable of intelligent tutorial planning [6]. Learning is embodied [7], with the child an active participant and collaborator, creating an emergent narrative along with a child-like virtual character [8].

The current empirical study used a simplified version of ECHOES as a research tool, with children completing a single session of an object-selection task in which a virtual character, Paul, varied his strategies for initiating and directing joint attention. The character's behaviours were hard-coded rather than generated by the AI planner, in order to attain the necessary control over the interaction. Qualitative data yielded insight into more general questions about the children's interaction with the system. The current study also provides crucial formative evaluation of ECHOES with the target user group [9] which fed back to the design of the full system. It was not intended as an intervention in its own right, but as a means to explore how joint attention and gaze-following skills might be elicited and supported in a VE.

2 Background and Project Objectives

2.1 The Autism Spectrum and Joint Attention

Joint attention is a key skill targeted by many intervention programmes for ASD, as its improvement seems to lead to lasting benefit in many areas, including language [10]. Joint attention is defined as the triadic coordination of attention between two persons and an object, and requires the ability to follow and direct another person's focus of attention [11]. A response involves following the initiator's gaze direction or gesture to a location in space and, perhaps, acting accordingly. Typically developing (TD) individuals frequently initiate joint attention for the purpose of *social sharing*, finding the reciprocal interest and affect strongly motivating. Closely related is *social referencing*, an attentional initiation in which an infant or young child looks towards a parent for information when faced with a novel event or object.

Attentional initiations through gaze and pointing are inherently ambiguous, and the two social partners must share a context in order for the respondent to understand the motivation for that initiation and infer the appropriate response (if any). Without joint attention, two or more persons have difficulty in establishing a shared focus of activity or communication. Individuals with ASD often show extreme difficulty with the gaze following and social inference necessary for successful joint attention, attaining proficiency with a great deal of effort, if at all.

2.2 Virtual Environments for Intervention and Social Scaffolding

Very little is currently known about how young children with ASD interact with VEs, or how they perceive and interact with virtual characters. Previous research is unclear about whether they might respond to joint attention initiations in such a context, and about which *specific* behaviours might be effective for directing and eliciting responses to attention. Before we can consider developing an intervention, we must first test the assumptions and prerequisites about the interaction between the child and the virtual character. A social-skills-focused programme supported by a virtual character crucially depends on the user's perception of the character as an *intentional being*, with agency, intentions and desires. He must not be perceived as an inanimate object nor as a cartoon to be passively watched, but as intending to communicate with the child and behaving in mutually relevant ways [12]. The literature suggests that *mutual gaze*, or gazing at the joint attention respondent before gazing at an object, may be a crucial method for establishing *mutuality* between the initiator and respondent [13]. These findings lead to our assumption that interactions which begin with the virtual character establishing this mutuality should increase a user's impressions of his intentionality.

2.3 The Present Study

The goals of this study were a) to investigate how young children with ASD interact with the ECHOES environment, and b) to analyse which combination of the character's mutual gaze and pointing gestures were most successful for eliciting the gaze-following behaviour necessary to complete the joint attention task.

Observational and video data collected in the course of the study illuminate the more general, exploratory questions regarding how children with ASD interact with the interface and the virtual character, and whether they perceive him as intentional and mutually-directed. Given the lack of existent research in this area, no specific predictions were made about these questions. The study results will inform the design of the full ECHOES system by highlighting which character behaviours are effective at directing attention, lead to perceived intentionality, and are fun and engaging for young users with ASD.

In relation to b), this population was predicted to exhibit at least *some* gaze-following, even if infrequent or inconsistent, with mutual gaze (engagement) conditions predicted to produce gaze-following to an object in the environment more often and more rapidly than in non-engagement conditions. The character's pointing cues were predicted to increase accuracy and decrease reaction times on all trials. These two behaviours (mutual gaze and pointing) could potentially interact; more rapid, more accurate, or more frequent gaze following may require both.

3 Methodology

3.1 Design of the Joint Attention Task

Users' gaze-following behaviours were measured during a simple selection task. Each trial involved three flowers[1] (two distractors and a target) to which Paul tried to direct the child's attention. A virtual character can initiate attention with the child and the object in the same way as would a human partner, by first looking into their partner's eyes (*mutual gaze*). Paul looking out from the screen gives an illusion of looking "at"

[1] One flower of each colour (red, yellow, blue) was presented per trial, with colour and screen position of the target counterbalanced across trials.

the viewer, similar to [13]. Varying the character's mutual gaze created 2 levels of the gaze following task: *engagement* and *non-engagement*. In the former, Paul established "mutual gaze", whereas in the latter condition he never gazed at the child, only directly to the object. Paul also varied his use of pointing, creating two levels within each gaze condition for a total of four trial types (mutual gaze + point, non-mutual gaze + point, mutual gaze + no point and non-mutual gaze + no point). Pointing shares several important features with gaze: both actions direct the respondent to a location in space, but are ambiguous and take their meaning from context [14]. Pointing also provides a visual cue in the form of directed motion, with greater potential for capturing the child's attention than gaze alone (see Figure 1, right).

Each participant was assigned a uniquely ordered trial script of 36 possible trials divided into three blocks of 12. Trials were randomly ordered and counterbalanced within each block; a child completing only 12 or 24 trials would see the same number of trials from each of the four types.

Fig. 1. Paul uses gaze only (left) and gaze plus gesture (right) to indicate his target flower

3.2 Participants and Procedure

Prior to the study, the study design and virtual environment were tested by 4 TD children aged 4-7 years (1 male, 3 female). This resulted in a number of changes to the environment, including adding a trial counter, adding background garden sounds and adjusting the timing between actions. Additionally, testing identified the need for prompting and support (e.g. reminding the child to wait for the character's indication). Experimental participants in the main study were primary-aged pupils at a specialised school for children with ASD (n= 32, 29 males, 3 females), aged 5 to 14 years (mean age 10.67 years, SD= 2.46 years) who represented a range of ability. All had previously received an autism spectrum diagnosis by a senior paediatrician or child psychiatrist, with evaluation of communication, reciprocal social interaction, and repetitive behaviours, using observational assessments including the *Autism Diagnostic Observation Schedule* [15].

Each child individually completed a single session of the flower-selection task in the ECHOES virtual environment, working in a quiet room with two experimenters present. Their interaction with ECHOES was video recorded. Each child followed the same order of events but heard a pre-recorded greeting personalised with their names. Paul introduced himself and asked: "*Will you help me pick some flowers for my mum? I will show you which ones to pick.*" The experimenters repeated these instructions, informing the child that they could pick a flower by touching it onscreen. After a correct choice, the flower flew across the screen to Paul's vase with a fanfare; an incorrect choice led him to say: *"Not that one"*, and indicate the target again. This simple narrative provides a framework to support joint attention and to motivate a shared (and repeated) activity between the child and the character.

In the initial trials, many participants had difficulty self-regulating and often touched the screen *before* Paul had indicated a flower. Consequently, most received additional verbal prompts from the experimenters to help them use the touch screen interface at the appropriate time (e.g. *"Touch the flower if you think that's the one that Paul wants"* or *"Wait until Paul shows you"*).[2] There were no training trials, as the participants' behaviour when they were still unfamiliar with Paul's cues was of prime interest. After each trial block Paul thanked the child and invited him or her to pick more flowers. When the child chose to end participation or completed all 36 trials, they heard a final goodbye and thank-you message. Each child's total participation lasted 10-30 minutes, varying with the number of trials completed.

4 Results and Discussion

4.1 Reaction Time and Accuracy Data

Each child in the ASD group (N=32) completed an average of 23.12 trials (range 4-36 trials, SD=10.21). A response was classified as *accurate* if the first touch after the character's indication[3] was to the target flower. An *error* was a touch to any non-target area, or a trial which timed out before the child responded (64 trials or 8.68%). Mean accuracy was very high, at 88.12% (SD=20.22%, median accuracy 95.14%). Accuracy did *not* correlate with age (r=0.23, p=0.21). Contrary to predictions, accuracy did not vary significantly between the four trial types.

Both the high percentage of correct trials and the *pattern* of errors strongly suggest that most children learned to complete the task accurately, despite their brief period of interaction with the VE. Many participants made several errors in the first 6 trials, but after this point most appeared to have completely grasped the task and responded correctly. A small number of participants had occasional errors until early in the second block of 12 trials.[4] Only 3 participants made repeated errors, and did not appear to have learned any kind of causal relationship between Paul's various actions, their own touch screen interaction and the environment's subsequent response.

A 2 (mutual gaze) x 2 (gesture) repeated measures analysis of variance (ANOVA) examined reaction times from the character's indication of the chosen flower to the user's first touch.[5] There were 699 correct trials across 30 participants.[6] The ANOVA revealed a significant interaction of mutual gaze and pointing cues, p<.01 (F=1, 30), with a strong effect size (Cohen's f= 0.477) (see Figure 2). The lack of a strong main

[2] Children were never prompted to attend to the virtual character's face, gaze, or gesture.

[3] Touches prior to the character's indication were ignored.

[4] Twelve participants, mostly older students (aged 11 to 14), made no errors.

[5] Trials with reaction times less than 200 ms were excluded: such responses may be due to the user touching the screen repeatedly before, during and after the character's flower indication.

[6] Two participants were excluded from the reaction time analysis, one for not completing at least one trial of each type, and a second for making some responses with feet, elbows, etc.

effect suggests that it is the *conjunction* of gaze and pointing cues which creates significantly different reaction times, and that this user group may more rapidly process combined gaze and gesture cues than single cues, emphasising the importance of including both types of cues in future virtual environments.

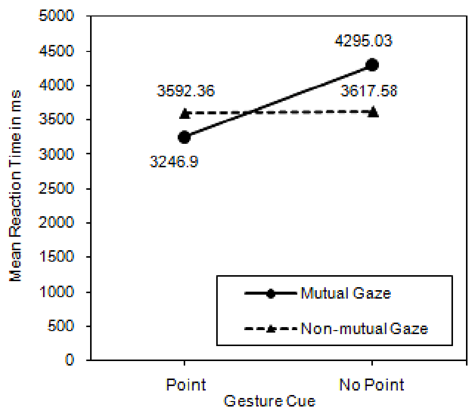

Fig. 2. The effect of character's mutual gaze and gesture on participants' mean reaction time

4.2 Qualitative Analysis and Observation

The video data collected in this study has yet to be fully analysed, but preliminary analysis has been fruitful. The combination of mean accuracy and the experimenters' qualitative observations indicate that young children with ASD successfully—often enthusiastically—engaged with Paul and followed his joint attentional bids to complete the flower selection task. Examples included spontaneously greeting him, answering him directly when he posed questions such as: *"Would you like to help me pick some more [flowers]?"* and expressing surprise or curiosity when Paul did *not* respond to being poked or could not "hear" them. Paul greeting each child by name seemed to be a major factor in generating liking for his character, and in indicating that his actions were both responsive and directed specifically to the child. As far as a perception of mutuality could be said to constitute intentionality, a large proportion of the children across the age range treated Paul as an intentional being.

There were numerous observations of participants spontaneously directing social behaviours to the experimenters and other adults. A large proportion of participants spontaneously gazed to adults, for example after Paul had indicated a flower, but before touching the screen (a possible instance of social referencing, as task demands may still have been unclear in early trials). Also common was spontaneous gaze to adults *after* the child made a flower choice (see Figure 3), when there was no further action demanded by the environment, a possible example of social sharing. Video data shows children concurrently smiling when looking to the adult, or engaging in behaviour regulation such as waving arms or jumping in excitement.[7] Some children

[7] For this population, such behaviours are often considered a means of emotional regulation.

Fig. 3. Left: A child (age 5) watches Paul and waits for his flower indication. Right: The same child turns to look excitedly at his classroom aide (not shown) in an instance of social sharing.

verbally commented on their own success, exclaiming *"I did it!"* or pointing out that they had progressed to a "new level" (e.g. the trial counter had incremented).

5 Conclusions and Further Work

These results are highly encouraging and suggest that young children with ASDs can learn to follow a virtual character's gaze and gesture cues, and to respond through the touch screen interface.[8] The degree and variety of the children's reactions to the character and the high, rapidly-achieved mean accuracy exceeded the experimenters' expectations based on previous literature.[9] We interpret these results as evidence that the children read Paul's actions as mutually relevant and directed toward them specifically, i.e., they perceived him as an intentional being— a "rich" interpretation in developmental psychology terms.

The frequency with which many children initiated social sharing while interacting with the VE is also noteworthy, given that children with ASD are typically impaired in such behaviours. The participants appeared to find the virtual environment novel, exciting, or rewarding enough that they were motivated to share some aspect of that experience with an additional social partner. The observed instances of spontaneous and socially-directed gaze are particularly positive, especially if some are interpreted as instances of social referencing: gazing to another person in an ambiguous situation only makes sense if the child believes that person could be a source of support. This study did not collect baseline video (e.g. in the home or classroom) or questionnaire data documenting each child's verbal ability or social skills, so it is impossible to say whether their social initiations during the experiment notably deviated from their usual behaviour. Overall, the results of this study are an affirmation of the potential for virtual characters as engaging and motivating tools to support social interaction, both within an environment and between child and additional social partners. They inform the full ECHOES system design, and form the basis of future interventions.

---

[8] What *cannot* be claimed is that these participants have demonstrated any skill learning.
[9] As well as those of the Head Teacher when shown video of the children's interactions.

Acknowledgements. The ECHOES project is funded by ESRC/EPSRC TEL (RES-139-25-0395). This work was undertaken as part of a Masters' dissertation by the first author, at the University of Edinburgh. In addition to the authors, other members of the ECHOES team contributed to the design and development of the environment (see www.echoes2.org). Special thanks to the participating school staff and students.

References

1. DSM-IV, A.P.A.T.F.: DSM-IV: Diagnostic and statistical manual of mental disorders. American Psychiatric Association, Washington, DC (1994)
2. Rajendran, G., Mitchell, P.: Text Chat as a Tool for Referential Questioning in Asperger Syndrome. J. of Speech, Language, and Hearing Research 49, 102–112 (2006)
3. Schmidt, C., Schmidt, M.: Three-dimensional virtual learning environments for mediating social skills acquisition among individuals with autism spectrum disorders. In: IDC 2008: Proceedings of the 7th International Conference on Interaction Design and Children, pp. 85–88. ACM, New York (2008)
4. Parsons, S., Mitchell, P., Leonard, A.: The use and understanding of virtual environments by adolescents with autistic spectrum disorders. Journal of Autism and Developmental Disorders 34(4), 449–466 (2004)
5. Tartaro, A., Cassell, J.: Playing with virtual peers: bootstrapping contingent discourse in children with autism. In: ICLS 2008: Proceedings of the 8th International Conference for the Learning Sciences, pp. 382–389. International Society of the Learning Sciences (2008)
6. Foster, M., Avramides, K., Bernardini, S., Chen, J., Frauenberger, C., Lemon, O., Porayska-Pomsta, K.: Supporting Children's Social Communication Skills through Interactive Narratives with Virtual Characters. In: Proc. of the ACM Multimedia Conference (2010)
7. Goldman, A., Vignemont, F.: Is social cognition embodied? Trends in Cognitive Sciences 13(4), 154–159 (2009)
8. Porayska-Pomsta, K., Bernardini, S., Rajendran, G.: Embodiment as a means for Scaffolding Young Children's Social Skill Acquisition. In: Proc. IDC 2009 (2009)
9. Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T.J., Menzies, R., Foster, M.E., Alcorn, A., Wass, S., Bernadini, S., Avramides, K., Keay-Bright, W., Chen, J., Waller, A., Guldberg, K., Good, J., Lemon, O.: Developing Technology for Autism: an interdisciplinary approach. Personal and Ubiquitious Computing (in press)
10. Sigman, M., Ruskin, E.: Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. Wiley-Blackwell, Malden (1999)
11. Charman, T.: Why is joint attention a pivotal skill in autism? Philosophical Transactions: Biological Sciences 358, 315–324 (2003)
12. Behne, T., Carpenter, M., Tomasello, M.: One-year-olds comprehend the communicative intentions behind gestures in a hiding game. Developmental Science 8(6), 492–499 (2005)
13. Pellicano, E., Macrae, C.N.: Mutual eye gaze facilitates person categorization for typically developing children, but not for children with autism. Psychonomic Bulletin & Review 16(6), 1094–1099 (2009)
14. Tomasello, M., Carpenter, M., Liszkowski, U.: A new look at infant pointing. Child Development 78(3), 705–722 (2007)
15. Lord, C., Rutter, M., DiLavore, D., Risi, S.: Autism Diagnostic Observation Schedule (ADOS). Western Psychological Services, Los Angeles (1999)