Supplementary Material for: Blending Human and Artificial Intelligence to Support Autistic Children's Social Communication Skills

KAŚKA PORAYSKA-POMSTA, ALYSSA M. ALCORN, and KATERINA AVRAMIDES,
UCL Institute of Education
SANDRA BEALE, Birkbeck, University of London
SARA BERNARDINI, Royal Holloway
MARY ELLEN FOSTER, University of Glasgow
CHRISTOPHER FRAUENBERGER, HCI Group
JUDITH GOOD, University of Sussex
KAREN GULDBERG, University of Birmingham
WENDY KEAY-BRIGHT, Cardiff School of Art and Design
LILA KOSSYVAKI, University of Birmingham
OLIVER LEMON, Heriot-Watt University
MARILENA MADEMTZI, Yale University
RACHEL MENZIES, University of Dundee
HELEN PAIN, University of Edinburgh
GNANATHUSHARAN RAJENDRAN, Heriot-Watt University
ANNALU WALLER, University of Dundee
SAM WASS, University of East London
TIM J. SMITH, Birkbeck, University of London

A SUPPLEMENTARY MATERIALS

A.1 Examples of ECHOES Activities and Child-ECHOES Interaction

In order to give readers a sense of the interaction engendered by ECHOES, we describe two contrasting learning activities. These activities highlight the types of actions that are available to the child, to Andy, and to the researcher/practitioner (via the GUI), and show how these actions relate to the socio-communicative behaviours targeted in ECHOES.

The first example, flower growing (Figure 10), involves shaking the interactive cloud to make it rain in order to grow flowers, and is one of the simplest activities in ECHOES. The activity is exploratory in nature; it is about fostering playful exploration and immersion in the environment. Many children were captivated by the flower growing and would shake the cloud for long periods without a break, appearing to enter into a flow-like state (Csikszentmihalyi 1991). Such interactions can continue indefinitely, as the activity has no explicit end point. Although it could be argued that such a set up might have encouraged obsessive behaviours in children, in ECHOES it was used explicitly to foster children's sense of calm. In line with the ethos of SCERTS, positive emotional states, especially a feeling of calm, provide a basis that is conducive to social interaction. During the periods of flow-like states, the rain sound effects would continue. Here Andy's role was to periodically give positive feedback such as 'Wow!' or 'Cool!', whereas the role of the human

1073-0516/2018/12-ART35 $15.00
https://doi.org/10.1145/3271484

1. The activity opens with the non-interactive garden background (trees and hills), the interactive cloud and a numerical counter (displaying zero). Ambient garden sounds, such as chirping birds, play in the background.
2. Andy enters, draws the child's attention to the cloud by saying "Look, a cloud, let's make it rain!" and jumps up to shake it.
3. A rain animation and sound effect are triggered, causing a small flower to grow from the ground, and incrementing the counter.
4. Andy draws the child's attention to the cloud-flower relationship by turning and exaggeratedly looking from the cloud to the flower. After a pause, he re-orients to the front and looks out 'at' the child, and invites him to take a turn.
5. The child can 'shake' the cloud by touching it and rapidly moving it from side to side. The child can move the cloud across the screen to any desired location using touch-and-drag.
6. The flowers can be picked (using an upward dragging movement) and moved around the screen; wiggling the flower heads makes them sway.
7. When the child grows a flower, Andy provides positive feedback.
8. Continuing to shake the cloud increases the size of a flower. Shaking and moving the cloud across the screen grows a whole line of flowers.
9. When the child stops shaking the cloud, and Andy has given feedback, he then announces "My turn" and proceeds to act.

Fig. 10. Example of Flower Growing activity.

practitioner was to manage the transitions between activities, by carefully gauging the child's state. The on-screen flower counter was a useful tool for the practitioner in helping them to create a goal or a predictable end-point, preparatory to transitioning to a new activity. It also provided very clear opportunities for the practitioners to initiate bids for interactions in ways that were contextually relevant to the children. For example, the practitioner might say 'Look, you grew 12 flowers! You can grow three more and then it is time for a new game'. This type of activity offers an important example for (a) how the AI and human intelligence and skills were blended in ECHOES and (b) why it is important to allow for such blending to take place in order to build on the specific strengths of both AI and human understanding of the child and the possible support.

The second example, sorting balls into boxes (Figure 11), is a very different type of activity to the flower growing one, as it is goal-oriented, involving a sequence of steps toward a clearly identifiable end goal. Teachers suggested this activity during one of the design sessions, having observed that many children with ASC enjoy sorting objects or helping to tidy the classroom. Sorting objects is also a typical activity used in autism intervention as it provides ample opportunities for modelling and practicing turn taking and joint attentional skills.

Having multiple objects with which the child can interact allows Andy to direct the child's attention to those objects explicitly, providing an opportunity for the child to practice following referential pointing gestures and gaze. As before, the practitioner can use the GUI to make Andy repeat his prompting as needed. Children can also act by giving Andy a ball to sort by dragging it to him and holding it anywhere over his body, until he accepts and says 'Thanks. I'll put it in the box'. Some children discovered this action independently, but in other cases the practitioner had to draw the child's attention to this possibility, sometimes modelling the 'giving" action for the child.

1. Three coloured boxes (red, blue, and yellow) and nine bouncy balls, three of each colour appear at the start of the activity.
2. Andy enters and identifies the task, saying ``Let's tidy up these balls into the boxes''.
3. Andy models the ball-sorting action by walking to his target ball, and then dropping it into the box of the same colour.
4. Andy then turns and looks 'to' the child and invites him or her to take a turn. He indicates a specific ball through one of: gaze only, distal pointing (with arm and forefinger extended), or contact pointing, walking to the ball and touching it with his forefinger.
5. Children sort balls by touching one and dragging it up and around the boxes, dropping it in the top. Balls cannot be dragged across boxes, and can only go in the box of the same colour.
6. If incorrectly sorted, balls roll off the top and bounce to the floor.
7. When a box is filled, the child receives an attractive sensory reward. The red box and balls vanish to be replaced by three bubbles. The blue box becomes a fireworks display, and the yellow box yields three cartoon bumblebees, which buzz around the screen for several seconds before vanishing.

Fig. 11. Example of Sorting Balls into Boxes activity.

Even though ball sorting is much more structured than flower growing, it is important to emphasise its inherent flexibility. Balls can be sorted and boxes filled in any order, even if some children independently identified and followed a particular order (e.g., completing boxes one-at-a-time, left to right). The boxes vanish upon activity completion, providing an explicit reward, such as a release of fireworks, bumble bees or poppable soap bubbles. This gives clear cues to the child that they have completed the activity and it signals to them that they have been successful.

In all activities, Andy reliably gives positive verbal feedback for the child's actions. To this end he may use any of several phrases, such as 'Good job', 'Cool' and 'Woah', all of which are supported by appropriate gestures such as Andy extending his arms in the air or using a thumb up, the latter also being used in the Makaton language to mean 'Well done'. Andy's planning architecture prioritises positive feedback over other possible actions. The overall pacing of these actions is slow compared to what adults may initially expect, or what may be customarily designed in programmes for TD children, but this pacing gives children with ASC time to notice and process what is happening, particularly where it involves Andy directing attention or giving instructions.

Many children were able to take turns with Andy, either spontaneously or when given additional support by the practitioner (e.g., explicitly drawing the child's attention to the potential of turn-taking or asking him to consider whose turn it should be next). However, even where children agreed that they should take turns or it was Andy's turn next, they often found self-inhibition difficult, likely due to developmental age. Almost all children needed repeated prompting to wait while Andy took a turn. Aside from its importance to turn taking as a social skill, waiting was especially important in ECHOES due to the functioning of the agent's planner. For example, if the child began manipulating the cloud when it was supposed to be Andy's turn, Andy would then re-plan and give feedback, rather than take his turn, which sometimes resulted in higher than expected latency in his reactions. This latency was inherent in the planning architecture chosen and as such constitutes an interesting challenge for AI technologies in real-time applications such as ECHOES more generally. Nevertheless, this inconsistency between what the children were told

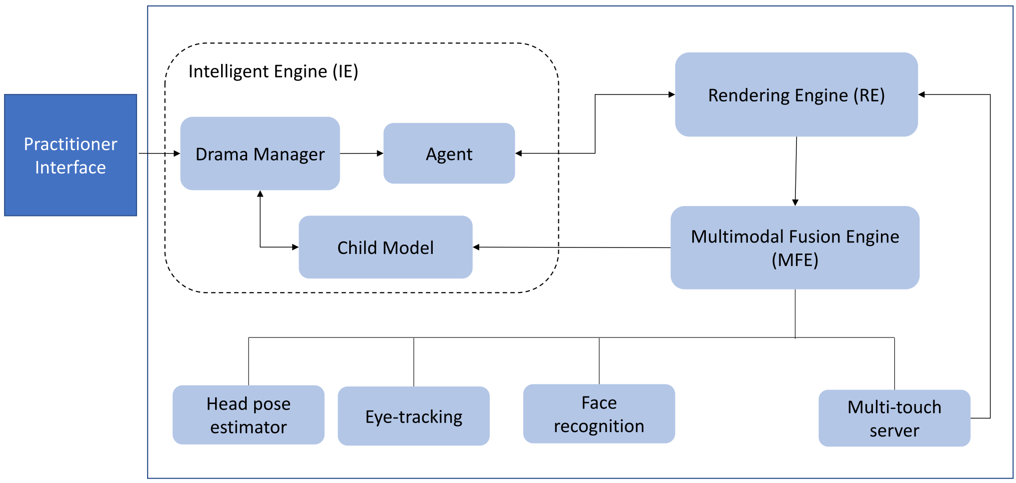

Fig. 12. ECHOES overall system architecture.

to expect and what occurred in the environment was potentially confusing for some of them. In situations where delays occurred, the practitioner could use dedicated buttons on the GUI to override the planner and to make Andy take a turn on demand. Other direct instructions included the command for Andy to repeat instructions/prompting, or to leave when it was time to end the activity.

When transitioning to a new activity, the practitioner would trigger the transition process from the GUI: in such cases Andy would walk out, and a red bubble would appear in the centre of the screen, slowly floating off to the right to signal the end of the current activity. This way of signalling transition between activities was the same in all of the ECHOES activities, and gave a clear visual cue that the activity was ending. To indicate the beginning of a new activity, a green 'transition' bubble would float on the screen from the left, shortly followed by Andy entering the scene. Again this signalling was used across all activities.

An important point to take away from both of these activity descriptions, and one that directly impacts the reporting of the results, is that the number of initiations (to which a child could potentially respond) varies across activities, sessions, and individuals due to the agent's reactive planning and the necessity for occasional interventions by the human researcher/practitioner (e.g., making the character repeat instructions). Thus, child responses to partner initiations must be compared as proportions, rather than as raw numbers, as explained in detail in Section 4.

A.2 ECHOES System Architecture

The ECHOES environment comprises three distinct software components, which communicate with one another to detect and process user actions and to select appropriate responses. These components (shown in Figure 12) include:

(1) Multimodal Fusion Engine (MFE), which combines low-level input events (e.g., touch-screen input) into higher-level composite multimodal events. The composite events allowed in ECHOES include gaze and eye tracking data, multi-touch events as well as face recognition. The MFE utilises the ICE middleware to allow for an open-source package with an active development and support community; industry-standard application development; use in a wide set of operating systems and programming languages;

publish-subscribe messaging and direct module-to-module communication; and use of structured, strongly typed messages.

(2) Intelligent Engine (IE), which selects actions for the virtual character and specifies changes to the state of the world based on the current learning objectives and the child's behaviours as represented by the composite events that are sent by the MFE. The composite events detected by the environment are recorded in the Child Model which is used as the basis for the determining the sequencing and duration of the learning activities within ECHOES. The sequencing and the duration of the activities within ECHOES can also be determined by the practitioner, through the Practitioner Interface, which is connected directly to the Drama Manager. The practitioner interface proved necessary to further ensure that the intervention/pedagogical decisions made during children's interactions with ECHOES were beneficial to the children, and to account for the fact that in the real-school (i.e., non-lab) contexts in which the studies reported have been conducted only the multi-touch events provided a consistently robust source of information about children's actions. Whilst this was a hindrance from the point of view of automating all of the decision-making processes within ECHOES, it is not an unusual difficulty encountered when advanced sensing technologies are used in the wild. With respect to the target population discussed in this article coupled with the use of the technology in physically non-restricted ways (e.g., children were not made to sit at a particular distance from the screen), the sensing technology did not prove up to the task, even though it worked robustly during lab tests. The key component of the IE is the agent which is underpinned with an emotional planning architecture called FAtiMA. This architecture allows the agent plan and execute its actions based on the information from other components in the system in a way that is both reactive in real-time and deliberative over the duration of each ECHOES activity. The emotional element is critical as it allows the agent's decisions to be driven by its emotional thresholds. Whilst in ECHOES these thresholds were set to low for positive emotions and high for negative emotions (Andy is a positive social partner), the architecture offers the flexibility for nuancing the emotional displays of the agent as needed by the different contexts, activities and user cohorts.

(3) Rendering Engine, which modifies its display and behaviour as necessary, based on the actions requested by the IE for the virtual character and on the world updates sent by the MFE.

Further details of the ECHOES' software architecture and individual components are given in (Foster et al. 2010) and (Bernardini and Porayska-Pomsta 2013).