Developing Technology for Autism needs an interdisciplinary approach

K. Porayska-Pomsta[1], C. Frauenberger[2], T. Rajendran[3], T. Smith[4], H. Pain[4], R. Menzies[5], M. E. Foster[6], A. Alcorn[4], S. Wass[1], S. Bernadini[1], K. Avramides[1], W. Keay-Bright[7], J. Chen[4], A. Waller[5], K. Guldberg[8], J. Good[2], O. Lemon[6]

[1]London Knowledge Lab, [2]University of Sussex, [3]University of Strathclyde Glasgow, [4]Edinburgh University, [5]University of Dundee, [6]Heriott Watt University, [7]University of Wales Institute, [8]University of Birmingham

K.Porayska-Pomsta@ioe.ac.uk, C.Frauenberger@sussex.ac.uk, Thusha.Rajendran@strath.ac.uk, Tim.Smith@ed.ac.uk, Helen@inf.ed.ac.uk, RachelMenzies@computing.dundee.ac.uk, M.E.Foster@hw.ac.uk, A.Alcorn@sms.ed.ac.uk, SamWass@gmail.com, S.Bernardini@ioe.ac.uk, K.Avramides@ioe.ac.uk, WKBright@uwic.ac.uk, JChen3@staffmail.ed.ac.uk, AWaller@computing.dundee.ac.uk, K.K.Guldberg@bham.ac.uk, J.Good@sussex.ac.uk, O.Lemon@hw.ac.uk

http://echoes2.org

Abstract

In this article we argue for an interdisciplinary approach to designing interactive technology for young children on the Autistic Spectrum. We believe it key for the design process to embrace perspectives from diverse fields to arrive at a methodology, and consequently technology, that delivers satisfactory outcomes for all stakeholders involved. The ECHOES project has provided us with the opportunity to work on a technology-enhanced learning environment that supports acquisition and exploration of social skills by typically developing children and children with high-functioning autism and Asperger Syndrome. ECHOES' research methodology and the learning environment relies crucially on multi-disciplinary expertise including developmental and clinical psychology, visual arts, human-computer interaction, artificial intelligence, education and many other cognate disciplines. In this article, we reflect on the methods needed to develop a technology-enhanced learning environment for young users with Autism Spectrum Disorder. We identify key benefits, challenges and limitations of this approach. Although the context of ECHOES is very specific, we believe that there are a number of guidelines for the desing of technology-enhanced intervention for autism that can benefit a wider community of researchers in this emerging discipline.

Keywords: Autism, technology-enhanced intervention, interdisciplinary research, social interactions, social signal processing, autonomous agents.

1. Introduction

Technology is increasingly recognised as a new, motivating, and cost effective way of delivering a variety of interventions to people on the autistic spectrum.

Insofar as the bewildering heterogeneity of neurological and histochemical abnormalities associated with Autism Spectrum Disorder (ASD) (Herbert & Anderson, 2008; Pardo & Eberhart, 2007; Persico & Bourgeron, 2006) might combine to cause a single, autistic cognitive phenotype, this is found in the tendency of individuals with ASD to prefer systematisable, rule-based situations to unpredictable situations in which empathy is required (Baron-Cohen, 2002; Cohen, 2007). A common favourite here is technology, which requires attention to detail, an ability to derive and implement abstract rules, and which is affect-free (e.g. Kaliouby et al., 2006; Picard, 2009). Technology, as we know it, lends itself immediately to affording individuals with ASD the worlds over which they can have control, which they can explore on their own terms without the risk of failure, ridicule and without the social anxiety that often accompanies their experiences in the real-world contexts. It is not surprising therefore that most individuals with ASD, including children, have a natural affinity with technological devices (Brown & Murray, 2001; Murray et al., 2009).

The fact that many individuals with ASD enjoy technology and technology-mediated interaction makes it a perfect medium for providing interventions (Bishop, 2003). The field of technology-enhanced learning (TEL) provides some evidence that computer-assisted learning can be an effective medium for individuals with ASD (Murray & Lesser, 1999). An important point that is increasingly being reinforced in the field of neuro-psychology is that intervention delivered at an early stage in an individual's development (i.e. in early childhood), has the best chance of succeeding (Myers, 2007).

The recent interest in technology as a means for scaffolding people with ASD into the real-world of social interactions is further motivated by natural affordances of technology, which, at its best, is a manipulable medium: carers and practitioners can tailor the technology according to the individual needs of the intended users, for example by switching some of its features, changing the nature of the activities (e.g. Golan & Baron-Cohen, 2006) or changing the pace of the interaction. The heterogeneity of symptom severity in people with ASD is a headache for care providers (Myers et al, 2007). If appropriately designed, technology is perfectly situated to provide interventions that are suitable for a wide variety of different abilities (Bishop, 2003). Technology also allows the users to work at a variety of different speeds, and will not lose patience with the endless repetition that many people with ASD desire (Wilkinson et al., 2008). Finally, there are economic reasons for investing in technology as means for delivering intervention. With human care provisions being hugely expensive (Myers et al., 2007) and given growing numbers of ASD diagnosis every year (Insel, 2009), technology presents itself as a ubiquitously available and affordable means of care provision of the future.

While the arguments for the use of technology in the context of autism intervention are compelling and great advances have been made towards understanding the potential value, affordances and design requirements of technology in this context, delivering effective socio-cognitive intervention by means of technology presents significant methodological challenges. Arguably, the biggest challenge pertains to evaluating the success of an intervention. Detecting the impact of technology on the way interventions are delivered and assessing their persistent effects outside of the treatment environment is extremely difficult, not least, because of the possible attenuation effect of the intervention. Specifically, as one moves from proximal (i.e. within a specific environment) to distal (i.e. outside of a specific environment, or generalised) transfer (Green et al., 2010) the effect will get weaker and therefore harder to detect. There are also important ethical issues that raise the question of what a successful intervention means for the users themselves.

In this article we argue that research in this area is necessarily interdisciplinary and that the biggest challenge of all for developing technology for people on the Autistic spectrum is to manage the diverse and at times divergent perspectives of all the disciplines involved. Theories, practices, methods and scientific traditions in psychology, human-computer interaction, education, social-signal processing and artificial intelligence differ significantly, but are equally important in the process. We believe that establishing common ground and drawing on the strengths of each of those fields is fundamental to enabling successful development of technology that is truly able to support people on the Autistic spectrum. However, the interdisciplinary nature of the process provides scope for misunderstanding, the need for compromise and the need for team members to leave comfort zones of their respective expertise.

We use our work within the ECHOES project to illustrate some of the interdisciplinary tensions that occur in the context of developing intervention technology for individuals on the autism spectrum. The paper is structured as follows: section 2 introduces the ECHOES project and the ECHOES TEL environment. Section 3 provides research background that motivates the development of the ECHOES environment and presents the theoretical foundations of ECHOES. Section 4 presents the interdisciplinary design methodology we have adopted and discusses how it draws on Participatory Design and methods from AI in education. Section 5 provides details on the implementation of the environment and discusses the development of learning activities, intelligent reasoning and planning and the social-signal processing involved. Subsequently, section 6 discusses measures of success and evaluation frameworks in this interdisciplinary context. This leads to a broader discussion on ethics in section eight. We finish by summarising our main arguments and attempt to generalise over our experience and provide a list of recommendations as a contribution to a generic methodology to the field of Autism and Technology.

2. The ECHOES project

ECHOES is an interdisciplinary project whose main goal is to develop a technology-enhanced learning (TEL) environment to support young typically developing children and children with high-functioning Austism Spectrum Disorder (ASD), aged between 5 and 7 years, in exploring, acquiring and using social interaction skills. The project also aims to develop tools that would facilitate research in this area in ecologically valid situations, i.e. outside the laboratory, for example in the classroom.

ECHOES builds on recent activities in a number of traditionally independent research areas, each of which contributes an important insight into the design of the TEL environment. Psychology provides crucial theoretical background and guidelines as to social interaction and current diagnoses and remediation practices in relation to Autism Spectrum Disorder (ASD). Human Computer Interaction (HCI) and Artificial Intelligence (AI) provide both the methodology for conducting interdisciplinary research and sophisticated technologies for making the virtual world increasingly more tangible, explorable and readily manipulable. Finally TEL offers guidelines for underpinning technology design by real educational theory and practice in order to make it viable and educationally effective in the real world.

The ECHOES technology-enhanced learning environment allows a child to interact with intelligent virtual characters and socially realistic environments. Virtual characters are presented within a rich, multi-modal 3D environment via a large (42") display. The ECHOES system allows the child to explore numerous scenarios in which they interact with objects and agents in the environment such as collecting flowers, exploring object properties, and collaborating on building a structure.

The learning activities allow the children to explore, interact and manipulate objects in a rich multi-modal environment. Given that young children are often pre-literate and no robust solutions to the problem of children's speech recognition exist, multi-modal technology is currently the most enabling of child-computer interactions and carries a promise of rich, and relatively accurate input and reliable inference about the child's real-time behaviour. Such richness and accuracy are facilitated by the use of physiological sensors that are increasingly used to detect mental states of users in multi-modal settings.

ECHOES monitors the child's actions through a range of sensors, including computer vision and multi-touch gestures on the display. Input from these multiple channels is combined into composite multi-modal events, which are sent to an intelligent engine that selects the appropriate behaviour for the system to execute in response. The requested character actions and updates to the state of the environment are sent to a multi-modal rendering engine, which combines three-dimensional graphics with sound to present the actions to the world.

3. Multidisciplinary Background and Motivation

Much of the effort in the field of Autism Spectrum Disorder focuses on understanding the causes of the condition and discovering the ways in which to alleviate the symptoms. The field is a largely fragmented space with clear divisions between educational and clinical approaches to intervention and a multitude of theories each of which has something to offer in terms of understanding the cognitive and social development of children in general and the deficits associated with ASD in particular. However, it is now commonly recognised that whilst each theory may relate especially well to some of the phenomena, individually they are unlikely to give a full explanation of ASD deficits and are therefore unlikely to provide a sufficient basis for learning scenarios for social engagement.

In this section we review the specific theoretical approaches that motivate ECHOES research and technology, and position it in the context of educational practice as well as within the-state-of-the-art in technology-enhanced intervention for autism. We also introduce ECHOES' overall methodological approach that draws in the first instance on the well established approaches in the fields of Artificial Intelligence, Human Computer Interaction and Education to developing technology for learning. Through this we show why and how interdisciplinarity is crucial to enabling the development of technology-enhanced learning environments for children on the autism spectrum.

3.1. Psychology, Autism and developmental theory

The theoretical foundations of ECHOES draw from the major theories of child development. More specifically, the design of the ECHOES learning environment is motivated by Developmental Psychopathology which views atypical development as a lens through which the norm can be better understood (Cicchetti, 1984). This means that ECHOES is based on theory that is appropriate for all children and therefore that the technology based on it carries the promise of being also suitable for children of all abilities.

Viewed broadly, development involves the transition from understanding physical causality to psychological causality. In their first year of life, children begin to understand the physics of interacting with objects. However, by the time they reach their fourth birthday, they are well on their way to understanding psychological causality: i.e. understanding that people are unlike physical objects in that they have minds. This understanding allows the child to learn that people act on the basis of their mental states, and crucially that these states may be different from the child's own mental states, or may be based on information that is incongruent with reality. These abilities to reason about one's own and others' mental states (known broadly as 'theory of mind') is fundamental to many other social, cognitive, and linguistic skills. The constellation of persistent socio-cognitive difficulties experienced by individuals with autism has been thought to stem from this inability to impute others' mental states.

Closely related to theory of mind is the group of skills and behaviours known as joint attention. Frequently conceptualised as a triadic social coordination between two persons and an object or event in the environment, it requires the ability to monitor another person's attention in relation to one's own (Charman, 2003). Others have described it more simply as the ability to follow and direct another person's focus of attention (Vismara & Lyons, 2007). Joint attention is considered a key developmental building block for theory of mind and some have argued that joint attention is in fact a necessary precursor for theory of mind, rather than its consequence (Tomasello, 1995).

Broadly, joint attention can be divided into three types, depending on what role a person plays in the interaction: (1) Joint Attention Initiation: pointing or looking between an object or event in the environment and another person to confirm joint awareness or to accomplish another social goal (e.g. show interest, inform); (2) Joint Attention Response: responding to another person's initiation of joint attention by following their direction of gaze or their gesture to a location in space; (3) 'Social Referencing': a special type of joint attention initiation, describing an infant's tendency to look towards a parent for information when faced with an ambiguous event. For instance, an infant may approach an unknown object if the parent smiles in response, but not if the parent looks afraid.

Apparently, uniquely among animals, humans frequently initiate joint attention (e.g. point to show) simply as a form of "social sharing" rather than to direct others' attention to danger or desirable resources. The critical relationship between joint attention and theory of mind becomes clear upon considering that the partner who initiates the interaction will always do so for some reason, whether it is to indicate interest, influence or to inform the recipient, or to accomplish some other goal. Alone, a gaze or a pointing gesture is highly ambiguous and draws its import only from the relevant shared context of the people involved in the interaction, whether this context is as concrete as a physical space or as abstract as what one believes that the other does or does not know. In every instance, one must infer the other's intentions in directing his or her attention to some referent. Without some limited grasp of theory of mind, the simplest joint attentional act may be confusing, or worse - it can be meaningless. This ability to establish a shared focus of attention is an important component of conversation or any other prolonged social interaction, and the centrality of joint attention to a range of other social skills and behaviours seems closely linked to the wide range of difficulties observed in many persons with autism. In many cases, those with a better grasp of joint attention also show a better grasp of other social behaviours. For example, several studies have demonstrated strong positive correlations between the level of joint attention skills in young autism spectrum children and their language use ability up to ten years later (Lord, Floody, Anderson, & Pickles, 2003; Sigman & Ruskin, 1999).

Joint attention also has a strong visual component, and the ability to follow and monitor others' eye gaze is key to many types of interactions. If you do not (or cannot) look where someone is directing your attention, you are far less likely to infer their intentions correctly and make an appropriate response. Conversely, if you do not understand how to direct someone else's gaze (or are unaware that it might be important to do so), you are severely limited in your ability to share relevant information about the environment. For many persons with autism, it seems that the problem is less inability to follow or direct another's gaze than it is the difficulty in understanding gaze as a communicative tool.

3.2. Educational practice

The central social role of joint attention has made it a clear target for many current intervention practices. The Social Communication and Emotional Regulation Transactional Support (SCERTS) framework (Prizant et al., 2009) constitutes one of the most coherent recent approaches to assessing children with social difficulties and to delivering intervention in this context. It is founded on research and evidence based practice in the field of autism, combining many of the major theoretical approaches with a number of well established intervention practices including contemporary ABA (e.g., Pivotal Response Treatment, LEAP), TEACCH, Floortime, RDI, Hanen, and Social Stories. SCERTS provides extensive guidelines in relation to assessment of individual children by trained practitioners and, based on such assessment, to selecting and organising intervention activities for the children. While, the activities pertain primarily to the different forms of joint attention, the elegance of the SCERTS framework lies in the fact that they provide a fertile ground for activities that target specific developmental precursors such as the ability of a child to imitate others, to understand the properties of objects as well as more advanced social skills related to turn-taking, initiating interactions and recognition of intentionality (agency). As such, the framework is very much in line with the current and emerging understanding of Autism and good intervention practice adopted in ECHOES, whereby every child is treated as following its own developmental trajectory. In this view any pre-requisite or a set of pre-requisites for the child's "typical" development may be broken in different ways preventing an easy categorisation of different individuals with ASD in terms of a uniform and fixed set of behavioural and neurological characteristics.

It has been argued that Autism at least partially results from an insufficient attunement to the social world beginning very early in life, a cascading effect which means that the child lacks the "right" type of social experiences to form a foundation for typical social-cognitive development (Klin, Jones, Schultz & Volkmar, 2003). Many social occurrences that are of high or "overriding" salience to a typically developing infant or child may not register as salient at all to one with autism. In theory, a child's positive social-communicative development could be facilitated by providing the "right" type of social experiences that require joint attention and other key skills. This goal forms the driving motivation for the learning activities in ECHOES. For instance, ECHOES can make the key joint attentional relation of self, other, and object explicitly salient and interactive by requiring the child and agent to cooperate on a task based around a digital object; this is precisely the type of experience that the autistic brain may not necessarily flag as interesting and significant in the wider social world. Of course, it is conceivable that in some children with autism their social-cognitive system is 'broken' and unlikely to benefit from even the most intensive intervention. For others, the positive shift in their ability may be minimal. If the interactive social experience in ECHOES does successfully facilitate learning, there is likely to be a wide range of outcomes. Furthermore, by focusing more broadly on social-cognitive development, rather than on difficulties very specific to autism, ECHOES has the potential to shift the developmental trajectories of typically developing children as well.

As well as joint attention, another key question for ECHOES within social-cognitive theory is that of 'agency', specifically 'intentionality', i.e. the understanding that an inner, mental state can lie behind observed behaviour. Klin et al. (2003) argue that from the very outset the autistic mind is not (or is minimally) attuned to the social world; for example, the gaze and gaze following patterns of individuals with autism are different from neurotypicals and most notably the eye region does not capture attention as strongly in those with autism. In stark contrast, the neurotypical mind seems to be constantly prepared to interpret social meaning, arguably overextending this capacity to such an extent that social meaning is interpreted amongst non-living entities (Rajendran & Mitchell, 2007). In ECHOES the central challenge, from the point of view of both intervention and technology design, relates to whether it can deliver a learning experience that supports recognition of intentionality by children.

As an educational intervention framework, SCERTS focuses its guidelines on the different, crucial precursors and skills needed for successful social interaction. Its creators emphasise that the framework "is most concerned with helping persons with autism to achieve "Authentic Progress", which is defined as the ability to learn and spontaneously apply functional and relevant skills in a variety of settings and with a variety of partners" (SCERTS website). With its focus on joint attention, social interaction initiation and recognition of intentional behaviours, SCERTS encapsulates much of the state-of-the-art in the field of Autism intervention and as such it forms the main practical underpinning for the learning activities within ECHOES. The visual emphasis within a virtual environment such as ECHOES and its capacity for interactivity offer a unique opportunity for children on a range of developmental trajectories to learn about the triadic relationship of self, other, and about different objects in the world through active participation rather than passive viewing, such as watching television which has been found, not only to have limited educational value (Courage & Howe, 2010), but also to produce 'video deficit effect' (Barr, 2010), whereby children learn less from television than they do from live demonstrations until they are at least 3 years old.

3.3. State-of-the-art in the technology-enhanced intervention for Autism

Developing technology for autism intervention is a relatively new, but fast emerging field. The flurry of recent activity both in America and Europe is motivated by the ubiquitousness of different technologies and increased power of computer technology, making it possible to create complex environments that can be manipulated through a variety of different modalities (touch, voice, text, etc. or a combination thereof). Furthermore, the increased interest in the potential of technology in the context of autism is motivated by the recognition of autistic people's affinity with computers and crucially, by the recent reports of dramatic increase in population of individuals diagnosed with ASD. The associated impact on children and families and continuous monitoring of ASD remains an urgent public health priority (Kogan et al., 2007). For example, ASD prevalence in the United States is growing with some reports showing ratio of individual with ASD to the rest of the population in 1992 as 1 in 1500, 1 in 500 in 2002, 1 in 110 in 2006 (Insel, 2009). This last figure places over 600,000 children in the US into this category.

Given the current statistics and costs associated with providing intervention, it is not surprising that many interested parties are looking increasingly in the direction of technology-enhanced solutions. Already, there exists a multitude of computer systems that attempt to scaffold individuals with ASD in terms of the specific skills that different theories promote. We conducted an extensive review of different technologies for autism and other assisstive technologies (Wass and Porayska-Pomsta, in preparation; Parsons et al., 2009). Here we provide a brief summary of our findings that motivate the approach that we take in ECHOES.
The different uses of technology that provide behavioural interventions in autism fall into two categories. The first category encompasses interventions that aim to re-mediate one specific aspect of the autistic cognitive phenotype by providing explicit tutoring at that skill set. For example tutoring packages have been developed that target face recognition (e.g. Faja et al., 2008), emotion recognition (Golan & Baron-Cohen, 2006) and understanding the mental states of others (Grynszpan et al., 2008; Rajendran & Mitchell, 2000). Most of these are CD-deliverable software packages that can be used on any home computer (e.g. Golan & Baron-Cohen, 2006), and encourage active (i.e. user-driven) learning (see Chi et al., 2001). Educational games are often accompanied by features such as 'emotions databases' that can be freely browsed (Golan & Baron-Cohen, 2006). The crucial outcome of our investigation is that the success of these software packages has been mixed, with some studies reporting that trained improvements within the computer tutor fail to generalise to 'real-world' environments (e.g.

Swettenham, 1996; Golan & Baron-Cohen, 2006). It is a rather depressing finding considering that many of the systems reviewed took significant amount of time and effort to develop.

The second category of technology in this context is assistive technologies - interventions that aim not to re-mediate any one particular aspect of the autistic behavioural phenotype, but rather to help the subject to cope with the world as it feels to them. In this category we find robots that are equipped with infrared sensors allowing them to imitate a few human movements. Such robots have been used to provide 'robot friends' for children with severe, low-functioning autism (LFA) who often shun human-to-human contact entirely (Billard et al., 2007; Kozima et al., 2005; Duquette et al., 2008). Small-scale studies (Billard et al., 2007; Duquette et al., 2008) have shown that children with LFA will engage in shared attention and turn-taking with a robot more willingly than they will with a human. While robots are still relatively expensive, they clearly offer the potential that they can be used as a stepping-stone to human-human interaction.

Digital play environments have been used to provide affect-free, audio-visually stimulating digital play environments, which are extremely popular with subjects with ASD (Keay-Bright, 2007). Recent developments in affective computing, such as electro-dermal activity sensors (Picard, 2009; Poh et al., 2010) and wearable cameras featuring automated facial affect recognition (el Kaliouby et al., 2006) are being developed as 'emotional hearing aids' that can be used both by subjects with ASD and their care-givers. Virtual reality (VR) has been used to provide training at tricky social situations that many people with ASD find overwhelming, such as finding a place to sit in a crowded canteen (Mitchell et al., 2007) and going shopping (Lanyi & Tillinger, 2004).

In summary, evidence that technology can be used successfully to scaffold individuals with ASD to a more successful social existence is still very limited - not least because the findings are often based on limited numbers of learners. Current state-of-the-art shows which different modalities and forms of interaction may have a better chance of succeeding: technology that encourages and facilitates active rather than vicarious participation tend to show better results; tools that aim to create authentic social situations such as those involving robots, seem to provide an acceptable stepping stone to individuals who find interaction with other humans challenging; play environments for children where their imaginative interaction can be encouraged and where they can be given the opportunity to be in control of their environment and their actions are also very popular. One of the prevalent trends in the design of technology for autism that shows a great promise is the realisation that such technology needs to be "aware" of its user in order to facilitate authentic social interactions. This recognition is reflected in the increased investment that many researchers make in the physiological sensor technology, which although still in its infancy, provides a possibility of monitoring the user with respect to the social cues and behaviours that they may manifest. The ability to reason about the observed user-behaviours and the ability of a system to respond appropriately to those behaviours is also high on the technology developers' agendas. The ability of a system to observe, to reason and to act accordingly to the observations and inferences is the defining feature of intelligent technology in the sense introduced within the field of Artificial Intelligence.

4. An Interdisciplinary Design Methodology for Developing TEL Intervention for Autism

The methodology advocated in ECHOES derives from a combination of Action Research (from Education), Participatory Design (from Human Computer Interaction) and Applied Artificial Intelligence. Common to them all is an emphasis on the need to move the locus of design and development closer to the user's community of practice, viewing design as a dynamic, incremental process that both changes and is changed by the context of practice. The methodology is informed by, and contributes to, theory.

Conventional educational research distinguishes the roles of researchers from practitioners, and separates the activities of observation, interpretation, planning change, and implementing change. Action research seeks to combine these activities within a single framework, and both stimulates and is stimulated by the growth of theory. Typical Action Research (Cohen & Manion, 1980) involves small-scale interventions in ecologically valid educational contexts, and a close examination of the effects of such interventions. The proponents of action research tend to emphasise its practioner-led, 'democratic' character. It requires collaboration between researchers and practitioners, the ultimate objective being to improve practice. Fullan (1991) repeatedly demonstrates that when innovation is attempted without the active participation of the community that is expected to practice the change, its success is extremely limited. In contrast, an iterative, practice-driven approach should ensure that the systems and practices that emerge are those that have a real chance of taking root within the culture of schools. The outcome of any project is expected to be some combination of evolved practice with developed theory. The origin of this approach, which Schön (1983) has applied to other professions in his highly influential account of the 'reflective practitioner', is commonly attributed to Stenhouse (1975).

The following two sections describe the methodological basis of ECHOES aiming to highlight benefits and potential areas of conflict with this approach.

4.1. Participatory Design

Participatory Design approach is one that is grounded in the perspectives, practices and needs of the target user group. There is more to this approach than simply matching the look and feel with users' preferences, however. Participatory Design (PD) was born out of a political context in Scandinavia and sought to democratise working environments by involving workers as stakeholders in the decision making (e.g., Bjerknes, 1995). This was motivated mainly by an ethical argument that promotes empowerment and inclusion. It is strongly related to Human Computer Interaction (HCI) approaches such as user-centred design (e.g. Landauer 1995). PD has been adopted by the field of HCI as a method of achieving end-user involvement in the design of interactive artifacts (e.g., Muller, 2003). Thus, PD is not just about acquiring requirements for system developers, but also about the more fundamental ethical argument of giving users a voice in the design of technology they will use. Arguably, the less expressive user groups tend to be, the more important it is to actively facilitate their inclusion. This is certainly true for children, users on the Autistic spectrum and their carers, teachers or parents who often are marginalised in the design process - technology tends to be designed *for* them, rather then *with* them. By applying participatory methods, mutually respectful relationships with all stakeholders can be built. This naturally leads to a deep immersion of the designer into the world of their users and subsequently to an understanding that allows for an empathetic and mindful interpretation of the users' contribution to the design process.

If the design aspect of interactive environments plays a crucial role for the engagement of users in general, this is even more true for users who are on the Autistic spectrum. Monotropism (Lawson, Murray & Lesser, 2005) and obsession with detail, common traits of Autism, mean the aesthetics, the look and feel and the flow of the interaction can make all the difference whether technology can engage and play a role as a gateway to a more successful social interaction in the real world. Furthermore, if, as is the case ECHOES, the users are children, their perspective on the world around them differs significantly from an adult designer-researchers' view. As Good (2006) put it: "what children want and expect is likely to be different from what adults think children want and expect".

In ECHOES we implemented a participatory design process that involves close collaboration with a small number of primary schools and specialised units working with young children on the autistic spectrum. We organised a series of workshops that facilitated sensory exploration and idea generation for the design of the environment and its elements (Frauenberger, Good, Keay-Bright, in preparation). In the process, we encountered several challenges upon which we reflect in the following subsections.

4.1.1. Balancing responsibility

When designing with children, the level of involvement can vary from purely testing ideas to equally involved design partners (Druin, 2002). While, for the reasons outlined above, we aim for maximum involvement, our experiences also have shown that too high expectations can result in dis-engagement as our young participants can become overburdened with creative responsibility (see also Jones et.al., 2003). This effect is amplified with participants with ASD who often struggle with social interaction, including during design workshops and other PD activities, and require more scaffolding and guidance to unlock their creative potential. This means, in ECHOES, participants mainly play the role of informants rather than fully fledged design partners.

4.1.2. Mindful interpretation

The activities in which we asked children to engage produced a wealth of ideas and inspiration for the ECHOES system design. However, our experience suggests that the way in which children, particularly those with ASD, expressed their ideas was difficult to translate into actual design. Too often our participants became absorbed in details and were driven by recent experiences. We have developed an approach informed by phenomenology (Frauenberger, Good, Keay-Bright, submitted) that allows us to look beyond the literal meanings and take the expressed experience as the starting point for the interpretation of input. For example, when we explored possible magical transformations of objects - an ability with which we wanted to provide playful engagement in ECHOES - a child showed us how a playground slide turned into a boomerang which had the same shape. Looking at the phenomenological qualities of the described experience, we derived a generic design concept that allows us to induce magic into digital objects in our environment: by using similar shapes for objects with very different functionality, we can use scaling to transform one into the other. For example, an arch over the gate to a magic garden can be scaled by the child and gradually turns into a rainbow. This approach adopts the concept of mindfulness as an approach that is non-judgemental and pertains to the nature of experience that unfolds through experience in the here and now. (Kabat-Zinn, J., 2003)

4.1.3. Engagement & learning

Many aspects of the system and the interaction have been pre-determined, narrowing the design space for PD activities. A systematic tension that emerged from this, relates to the child's learning and their engagement with the environment. While the SCERTS framework that informs the design inevitably provides very clear guidelines as to the learning goals that the ECHOES learning activities may aim to achieve, PD activities aim to inform the delivery on enjoyable experiences within the ECHOES environment. In the context of Autism, PD runs the risk of reinforcing existing traits of the children with ASD, because it naturally draws the design into the comfort zone of the user. For example, an exaggerated focus on detail might be the most engaging feature for a child with ASD, but also hinder the progression in terms of the development of the child's social skills. The opposite is equally true: over-emphasising the achievement of learning goals is likely to disengage a child with ASD from the experience and while learning goals might be achieved the associated skills are not internalised. An inter-disciplinary dialogue between education, design and psychology, is an essential requirement for a design of a balanced and flexible interaction that affords a learning experience that is equally engaging as it is effective.

4.1.4. Mapping to Curriculum

While the participatory design activities are intended to be fun, engaging and playful for the users, it is imperative that they are conducted within a school environment in order to optimise the opportunities for contextual design. Furthermore, for ECHOES to be effective as a future intervention, issues that impact on its development and deployment such as the curriculum, class dynamics, and technical support must be considered from the outset.

4.1.5. Practicalities

Although working with children and with people with ASD in designing technology is very inspiring and rewarding, a strong and sustained collaboration throughout the course of the project requires significant amount of time, commitment and resources. One of the main difficulties is ongoing and timely access to participants, practitioners and parents. In ECHOES, schools were the primary point of access to our participants and this resulted in an additional level of complexity. We have always aimed to develop any activities in a way so that they cause the least disruption and teachers, parents and the school itself could benefit from our visits. This could be in the form of additional motivation for children, activities that played into the course of the current curriculum or sharing outcomes to include in the schools track record for inspections. It is important to recognise that schools normally gain very little else by engaging with research projects, so these incentives help to balance this relationship. Approached in this way, we have found that schools were very willing in collaborating with us, but intrinsically, work in schools is unpredictable and flexibility is required.

Parents are another dimension to this collaboration, which requires careful and empathetic management. Firstly, it is the parent's decision to give consent on behalf of their children and thus it is key to keep parents fully informed and make the process as transparent as possible. Furthermore, while the parents of children with ASD are generally keen to participate in a research that may bring benefit to their children in the long term, access to parents of typically developing children, and therefore to children themselves can be more challenging.

In ECHOES, we addressed these difficulties by developing a wide and committed network of different stakeholders, willing to act as informants as well as advocates of the research. Advantages of this approach, include contribution to continuing professional development of researchers and practioners; an approach that emphasises ecological validity and provides a developing model for working on classroom contexts; increased likelihood of greater impact at all levels and future uptake.

4.2. Artificial Intelligence in Education

he field of Artificial Intelligence (AI) provides a well established and coherent research and technology design methodology in the context of both emulating intelligent behaviour, including human cognitive processes, and designing technology to support learning. Bundy (1986) describes AI research as being of three different kinds, namely: applied AI which aims to build products; cognitive science which aims to model human or animal intelligence; and basic AI which seeks to explore techniques that have the potential for simulating intelligent behaviour. In relation to autism, technology can be developed with similarly related objectives: to build intelligent technology-enhanced learning environments to provide interventions; as a means to explore theoretical research questions of importance to the understanding of autism and its effects (e.g. in relation to joint attention); to facilitate technology-mediated interaction between children and virtual agents through multi-media technology.

Crucially, the AI methods are tacitly interdisciplinary. In the context of Artificial Intelligence in Education, this methodology is explicitly stated in terms of the Persistent Collaboration Methodology (PCM - Conlon and Pain, 1996), which advocates active and continuing (persistent) collaboration between researchers, practitioners and technology experts in both the design and evaluation of TEL. The methodology that we advocate in ECHOES is therefore derived from Education, HCI and AI, and in this respect is a specialisation of Participatory Design, applied to designing interactive technology for young children on the Autistic Spectrum. Our goal is successful development of technology that results in long-lasting change in both practice and in the evolving methods and beliefs of the collaborating partners.

PCM involves phases of four (unordered) cycles: observation, reflection, design and action. There may be a number of iterations of such cycles, that may stop and start anywhere within the process. In reality the division between them is fuzzy. Each of the collaborators contributes distinctive knowledge and skills to the process, and can influence, and be influenced by, the contributions of other stakeholders. As discussed in section 4.2, in ECHOES such contributions are facilitated through a establishing an network of researchers, experts, practitioners and end users and are obtained through workshops, focus groups, demonstrations and training.

All the activities are essential to producing intelligent technology that is educationally viable. In addition to these four activities, and in line with the wider AI methodology, the PCM advocates that any technology should have theoretical underpinnings and that as well as fulfilling a primary goal, for example to provide tutoring support to users in a specific domain, it should also function as a research tool capable of contributing back to the theory and practice. Figure 1 shows a schematic representation of the PCM methodology.

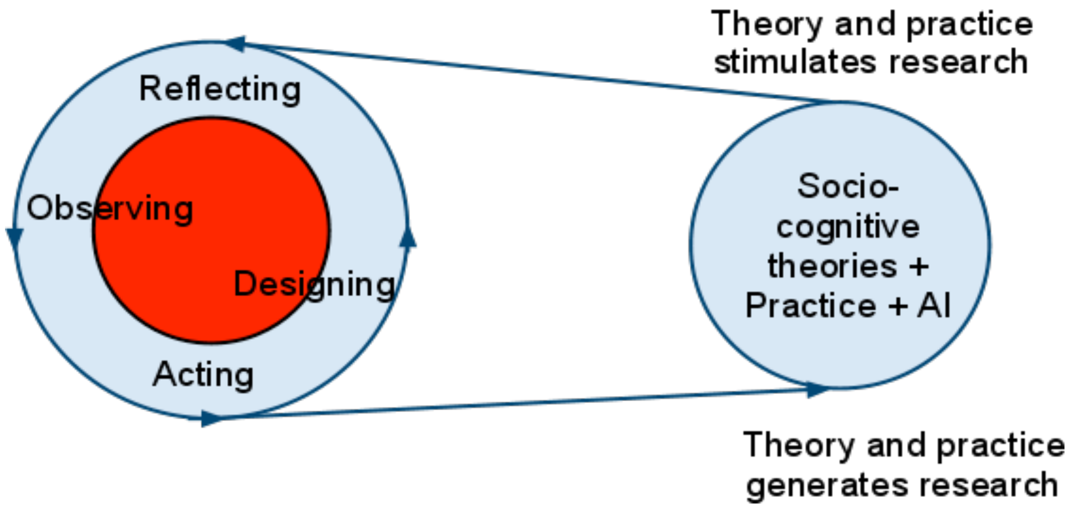

Figure 1: Persistent Collaboration Methodology

PCM is at the core of ECHOES, in which the participation of the individual stakeholders provides a crucial basis for the ECHOES environment's design and whose goals are (i) to provide children with an environment through which they can learn about social interaction, and (ii) to enable researchers and practitioners with a tool through which to study the autistic children's behaviour in this context.

In terms of the PCM methodology, in the observation phase, the design of the ECHOES environment relies crucially on significant amount of data about children who are involved in ECHOES-like activities and who are using the environment or its specific aspects as they are being developed. The nature of the data capture pertains to the design of the interface that children find enticing and the low-level actions and behaviours that the children engage in during the activities and during their using the environment. The actions and behaviours of interest will include the information about where the individual children are gazing, what objects they are touching, their facial expressions and verbal behaviours in specific situations. Such data also informs the high-level inferences that can be made about the children's underlying states, such as whether they are happy, focused, or frustrated, as these states will inform the pace and the nature of the intervention facilitated by the environment; see section 4 for more details on how this information is used in the system.

The reflection phase in ECHOES takes place in a multitude of smaller and frequent cycles involving different practitioners and clinicians who provide their interpretation of the children's behaviours in the context of their using ECHOES, based on their knowledge of the ASD condition and their experience of working with children with ASD. The reflection phase generates further data and corresponds to knowledge elicitation and acquisition phase in traditional AI. It provides crucial information for the design of the several aspects of the environment including the design of the user modelling tools, as well as the pedagogic and communication components. The action phase is multi-faceted in ECHOES also involving a number of smaller cycles that primarily inform the implementation and evaluation of the specific components of the system; all components are individually evaluated before being combined into the overall system. ECHOES evaluation happens at both formative and summative levels, with the individual formative evaluation trials and studies being used as the basis for further observation, reflection, design and implementation phases.

5. ECHOES' System Design and Implementation

Figure 2: A child interacting with the ECHOES environment

The system design reflects choices that were made in an attempt to portray a plausible technological infrastructure in classrooms of the near future. Interactive white boards are already common in our schools and we carefully extended this setup by adding sensing capabilities such as multiple video cameras and a multi-touch surface, and a high quality sound environment. We have deliberately chosen to stay close to existing technologies in the classroom, because although more sophisticated technologies are available, widespread change is likely to take a very long time.

In this section, we concentrate on two specific aspects of the ECHOES system: the processing of social signals and other input produced by the child, and the behaviour of the intelligent reasoning engine (See Foster et al. (2010) for a full technical description of the ECHOES system).

5.1. ECHOES Learning Activities

ECHOES' learning activities correspond directly to the intervention goals specified in the SCERTS framework. As we discussed in section 3.2, SCERTS focuses primarily on goals related to different forms of joint attention by drawing on a multitude of different theories of child development and of the autism spectrum disorder. The framework emphasises the development of personalised intervention programs that target a few of the child's individual needs that are deemed most important to their everyday life. In SCERTS, the intervention activities are integrated within the child's existing everyday routine and are tailored to what motivates the individual child. SCERTS' activities are organised in terms of specific fine-grained skills related to social communication and emotional regulation. Social communication is defined in terms of the child's ability to engage in joint attention and use symbols to communicate. Emotional regulation is defined in SCERTS in terms of the child's ability to regulate their emotions through others (mutual regulation) and through themselves (self regulation). Transactional support is defined in terms of interpersonal support (how the carer responds to the child) and learning support (what materials are used to structure the interaction. The framework provides explicit guidelines as to how to engage the child in an interaction. For example, in order to provide interpersonal support, the carer must follow the child's focus of attention, they must attune to the child's emotion and pace, whereas in order to provide the child with learning support, it might be necessary for the practitioner to define steps within a task as well as to define steps and time for completion of activities.

It is important to bear in mind that the SCERTS framework has been developed for human-human intervention context only. In this context the practitioner uses their long-term expertise and experience in assessing the child's needs and in deciding on what activities may benefit the child the most at any given point. The fact that the framework is designed for human-human intervention context presents ECHOES with several challenges with respect to how the framework can be adapted to the human-computer interaction context.

The first challenge relates to our access to the practitioners' knowledge and expertise. Practitioner's understanding of the possible behavioural manifestations that may be indicative of the child's affective states (e.g. boredom, happiness, joy, frustration etc.), cognitive states (e.g. focus of attention, curiosity, understanding and desire to pursue the specific goals of an activity) as well as the level of attainment of the goals by the child are all crucial to the delivery of the intervention that works for the individual child. It is this kind of knowledge that the practitioners rely on in deciding what specific skills to target, in choosing the appropriate activities for the child and in selecting the appropriate way in which to facilitate these activities to the child. Unfortunately, such knowledge is not easily accessible to others and therefore its formalisation into a computer system is equally difficult. Even if all of the knowledge and expertise of the practitioners could be represented explicitly, current technologies limit us in what information about the user a technology-enhanced environment is able to capture in real-time - an ability that is crucial to reproducing some of the human practitioners' intervention skills. As we discuss in section 5.3, in ECHOES we invest a lot of our efforts in enabling the environment with an ability to detect the child's actions within the environment, through a multi-modal detection system. However, given that that technologies that facilitate such sophisticated information harvesting in real-time are still in their infancy and are therefore not always reliable, the learning activities in ECHOES are designed in such a way as to allow for a variety of different modalities to be used together and individually (should some information be unavailable at any given time) as the basis for progressing the interaction between the child and system.

The second challenge relates to the fact that the affordances of a digital environment are different to those of a human-human context. Specifically, unlike the physical, tangible environment of traditional intervention situations, a digital environment offers the possibility of creating magical worlds, where children can play with the different objects in a way that the real world does not afford. Objects can transform into other objects and exploration of normally inaccessible worlds such as underwater world or the inside of a cloud can be made possible. In ECHOES we strive to exploit the ability of a digital environment to play to the child's imagination in order to encourage the child to explore and thereby to engage the child in the active generation of knowledge. Since digital objects have different affordances to physical ones, the child will need to discover them through exploration of the environment at their own pace. Exploration of these digital objects leads to more complex actions where objects are combined, for example, stacked to create a tower, or used to produce further effects within the virtual world. From the child's perspective ECHOES' activities are not defined as tasks that relate to social interaction skills, but to the objects themselves.

The context of interaction in ECHOES is provided in the form of a sensory garden. Based on a series of participatory design workshops with children and practitioners, the garden environment was selected for ECHOES as a versatile setting for children's exploration and for initiating social interactions with others. The social interaction skills are targeted by the presence of a virtual agent that engages in the activity with the child and adapts its behaviour according to the targeted skill and the child's needs. The agent performs a role of a practitioner / peer and are the main means through which the learning activities in ECHOES are facilitated to the child. Figure 3 shows the implemented the ECHOES garden environment together with the ECHOES agent – Paul – who is able to engaging the child in joint attention through different means including gesture, verbal request, eye-gaze and a combination thereof.

Figure 3: The magical garden

A third challenge relates to whether the child's perceives the agent as an intentional being or merely an inanimate object. For the child that interacts with ECHOES to treat and interact with the agent as an intentional being is crucial to our being able to facilitate a technology-enhanced learning experience in the context of believable social interactions. One test bed for the children's perception of agent's intentionality is 'mutuality', i.e. the degree to which a user views the agent's communicative acts and intentions as being relevant to them (Behne, Carpenter & Tomasello, 2005) and, consequently, whether children act/respond differently when and if they view the agents' communication as being relevant to them?

With the sensory garden as the overarching environment, ECHOES' activities are organised around its different elements, for example flowers are objects of interest, desire or admiration and can serve as triggers for the joint attention episodes between the child and the agent. In order to support the coherence between the activities, we link the activities through narratives/stories that motivate the existence of the agent in the environment and its specific actions. For example, the agent may justify to the child its desire for a flower by saying that it is collecting flowers for its mum. But its mum only likes red flowers and therefore it needs the child to help it pick only the red flowers. Paul, the agent, may encounter different obstacles in the environment, for example, a pond, or a wall, or its ability to notice the objects of desire may be occluded by other objects. Paul may get upset if it does not manage to get what he needs, etc. In ECHOES all of these different scenarios contribute to the story that unfolds in real-time, based on what the child does and based on the different possible behaviours of the objects and of the agent. The different scenarios also provide opportunities for exploring and improving the child's specific skills of interest.

5.2. ECHOES' intelligence, planning and reasoning

ECHOES draws from the AI philosophy, whereby a system that facilitates naturalistic interaction and learning should itself be equipped with some of the human characteristics such as knowledge of the domain, planning and reasoning abilities with respect to the domain as well as the users, and the ability to deliver feedback that is appropriate to the individual user's immediate and long-term needs. In ECHOES, we adopt this philosophy as an essential pre-requisite for addressing the issues raised in psychology literature and of providing users with an environment in which they are agents that can engage in a purposeful, active generation of knowledge and skills. Active generation of knowledge is deemed as the most effective form of learning (Chi et al., 2001). Personalisation is core to the success of the intervention, where ECHOES adapts the interaction to the specific child, in the immediate context.

The core interaction between ECHOES and its users takes placec between the users and ECHOES' agents. Because, the effectiveness of ECHOES as a pedagogical system depends crucially on the success of such interaction. The agents need to have the ability to adapt their behaviours according to each specific child's needs and requirement at any give point in the interaction. Ideally, in order to support the authenticity of the different social situations, the agents should also be able to emulate human behaviour in similar interactions. This ambitious goal requires that the agents exhibit a number of key AI features. First, they need to be autonomous, which means that their behaviour should be synthesized in real time by the characters themselves on the basis of the events that they perceive in the virtual and real world around them. Second, since ECHOES is a pedagogical environment focused on supporting social skills acquisition, the agents need to show a repertoire of emotions in order to relate empathically to the users. This requires that the agents have an internal model of their own goals, beliefs and desires. In addition, they need to have continuous access to a dynamic model of the user's cognitive and emotional states (user model) to enable them to adapt their behaviour to the user's current needs and mood. Third, the interaction with the agents needs to be as seamless for the user as possible, so that a flow can emerge from the interaction. Such a seamless interaction is facilitated by the agents' ability to use and react to both verbal and non-verbal communication.

Developing an agent that brings together all these properties is currently considered the prototypical AI problem because it requires a strong integration between a number of AI features such as automated reasoning, autonomy, natural communication, emotion modelling and user modelling (Swartout, 2010). Although, in ECHOES we are aware that the full implementation of such a character is not feasible in the short-term because substantial improvements in many of these areas are still necessary, we believe that significant pieces of the required technology are already available in the AI community for creating an initial prototype of this kind of technology. In particular, in ECHOES, we focused on bringing together automatic reasoning, autonomy, emotional modelling, user modelling, artificial vision, non-verbal communication and animation.

5.2.1. Autonomy, intelligence and emotional modelling

In ECHOES, autonomous agents control the decision-making processes of the embodied virtual characters. The model of each agent is characterised by: (1) a set of internal goals; (2) a set of strategies to achieve these goals; and (3) an affective system regulating the agent's emotional tendencies. In our current implementation, each intervention session has a set of agent models associated with it that correspond to the characters acting in that session. The internal goals of an agent reflect the overall goals of a specific intervention session in the same way as the acting strategies of an agent demonstrate and/or promote those specific social behaviours and cues that are the focus of a specific learning activity or intervention session. Therefore, in ECHOES, the actual intervention is delivered by the autonomous agents that interact with the children.

The architecture of the ECHOES agents is based on the FAtiMA system (Dias and Paiva, 2005), which was designed to control the behaviour of emotionally intelligent virtual characters and has been successfully used in other educational systems (Figueiredo et al., 2008). The emotional model behind FAtiMa is based on OCC theory of emotions (Ortony et al., 1988) and on the "appraisal theory" (Smith et al., 1990). Based on those theories, virtual characters "experience" emotions as valenced (i.e. good/bad) reactions to external events. These reactions are triggered by comparing the external events with the characters' internal goals, beliefs and desires. So, if external circumstances appear to facilitate the characters in achieving their goals, they will be happy or satisfied, whereas if the characters' efforts are opposed by the surrounding environment, they will be sad or angry. The exact emotion experienced by a character depends not only on its appraisal of the current external events, but also on its subjective tendencies to "feel" certain emotions instead of some other ones (emotional thresholds). The repertoire of emotions that can be exhibited by a FAtiMA character is quite sophisticated encompassing twenty two different affective states.

ECHOES' agents use coping strategies in order to deal with their own emotions. In particular, they use problem-focused coping strategies when they try to reduce the dissonance between their goals and the external events by acting on the external world and changing it. An emotion-focused strategy is used when an agent tries to adapt its own emotions to the external events by changing its goals and beliefs on the basis of external circumstances. Both, the appraisal and the coping processes work at two different levels: the reactive level, which affects the short term horizon of the agent's behaviour, and the deliberative level, which pertains to the long term goal-oriented behaviour of the agent. The implementation of the deliberation layer is based on automated planning techniques, which are traditionally used to produce the intelligent behaviours for autonomous agents (Russell and Norvig, 2003).

5.2.2. User modelling

The user model, which we call the "child model" since our system targets young children, assesses in real-time the goals and cognitive and affective states experienced by the child during interaction with the system, using a combination of supervised and unsupervised learning techniques. This assessment is based on: (1) static information about the child such as age, gender, and preferences; (2) information about their previous interactions with the system; and (3) real-time information coming from the multi-modal communication stream, as processed by the fusion component.

Automatically detecting social signals produced by humans in interactive situations is a topic that has received an increasing amount of interest in recent years; see (Vinciarelli et al., 2009; Castellano et al., 2010) for overviews of the area. Our approach to this topic is similar to that employed by Kapoor et al., 2007. We begin by analysing the recorded behaviour of children interacting with ECHOES prototypes. We then annotate those recordings to indicate relevant features such as engagement and affective/emotional state. Finally, we use the resulting annotated data together with the system logs to train supervised-learning models that are able to estimate the child's engagement and affective state while he or she is interacting with the system. This information is critical to allow the ECHOES intelligent engine to make appropriate decisions about the ongoing behaviour of the intelligent agents, the selection of learning activities, and the features of the virtual world.

5.3. Input processing

In order to adapt the interaction and intervention to the individual child, ECHOES's agent needs to be able to detect the same social cues that a human would in such an interaction. Non-verbal signals such as facial expressions, glances to a social partner or an object, goal-directed interactions such as touching, manipulating and offering objects will all contribute to a human practitioners' assessment of the child in the specific situation. Thus, in the ECHOES environment, this information is essential for the construction of the child model and for the system to decide how to respond to the child at any given moment. In real-world interactions, children will indicate their interest in an object by verbally referring to the object, gazing/pointing at it or touching it. Given the technical difficulties in reliably processing the spontaneous speech of children and the low verbal communication ability of children with ASD, in ECHOES we decided to focus on detecting social cues through touch and vision.

5.3.1. Touch

Touch screens (in the form of interactive whiteboards) have been shown to be highly motivating and engaging for children with ASD (Keay-Bright, 2007). It is hoped that the use of multi-touch in ECHOES will draw the child into the interaction and engage them in constructive interaction with objects and social agents within the environment. Touch also bypasses the problem of forcing the child to learn a new mode of interaction, e.g. a controller, which can be difficult for some ASD children who have low motor skills.

When a child touches the screen, the multi-touch server publishes the time and x/y coordinates of the touch to multi-modal fusion engine, which interprets the touch information in terms of objects from the rendering engine. A touch on an object is then registered which can trigger new actions in the environment, such as a bubble popping or a flower growing, via the Action Engine. Socially relevant touch information is also logged by the Child Model and used to inform inferences about the child's progression through learning activities.

5.3.2. Visual Input Processing

The ECHOES Visual Input Processor is designed to recognize social signals expressed in the visual channel. The processor supports head pose estimation, gaze/eye tracking and expression detection simply using three low-cost web cameras without the need for any special hardware such as goggles (Bardins et al 2008), head mounted equipment (Arrington Research 2010), image processing board (Matsumoto and Zelinsky 2000) or infrared sensitive cameras equipped with infrared LEDs (Prez 2003).

In ECHOES, a large viewing volume is required as children should be free to move around in front of the screen instead of being locked down to a particular position. Most existing eye tracking systems require users to stabilise their head on a chin rest (Duchowski, 2007). Such a restriction would impede the naturalistic interaction desired within the ECHOES environment and be impossible with some ASD children. Therefore, we devised a method was needed to keep track of the child's face as they moved freely in front of the screen. We used multiple cameras to capture a large viewing volume in front of the screen. The camera arrangement in our system is shown Figure 1. Two Logitech Quick Cams are placed on the sides of the multi-touch screen and verged (angled towards the center of the screen) at about 45 degrees. One Minoru 3D webcam is positioned on the top center of the screen. This is designed to offer the best measurements for a child's interaction with the 42" multi-touch screen. All the cameras are calibrated individually to get the intrinsic and extrinsic parameters of each camera. The screen center is used as the origin for the system. All pose measurements are interpreted with respect to this origin.

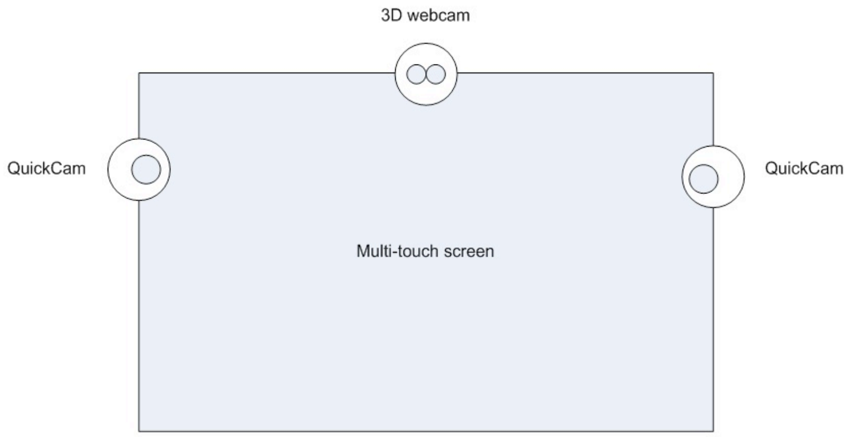

Figure 4: The multiple camera arrangement

The system works automatically including two stages: modelling and tracking. First, a 3D facial feature model is built using the 3D webcam, and then the tracking process is activated using one of the three cameras. We estimate the head pose from six features (inner eye corners, nostrils, and mouth corners) and a facial feature model (the 3D locations of the six features) using the POSIT algorithm (DeMenthon and Davis, 1992). Head pose estimation (HPE) provides a rough indication of the direction of the child's attention as the head is usually oriented in the direction of eye gaze in order to provide a more comfortable viewing position, i.e. with the eyes centred in their ocular orbit.

HPE also enables the identification of meaningful gestures including head nodding and shaking. The system is able to detect tracking failures using constraints derived from a facial feature model. Once the tracking failure has been detected from one camera, another camera will be activated. For example, when the system fails to track the facial features from the top webcam, the left or right webcam will be used to track the features. In this way, computation time is saved, making the system more efficient than using the three cameras simultaneously. The system works as long as there is a face visible in any of the cameras.

To detect eye gaze direction, the movement of the pupils relative to the inner eye corners are used to calculate gaze displacement relative to the head. The combination of head and gaze direction provides 2D screen coordinates of the child’s attention. This information is then combined with rendering information by the Multi-modal Fusion engine to identify object-based attention. This information is then used to update the Child Model and to choose appropriate action within the learning activity by the Action Engine.

In addition to head and gaze estimation, the vision system also uses the facial features to detect smiles. Smile detection is based on a cascade of boosted tree classifiers with Haar-like features (Chen and Lemon, 2009). Expression detection is important within ECHOES in order to assess the child’s emotional state to understand his/her responses in the environment.

The vision system is designed to be robust and resistant to the unpredictable behaviour of children when interacting with the ECHOES environment. As such, the ECHOES vision system represents a novel combination of face detection, head pose estimation, eye tracking and facial expression recognition which can provide a robust platform for future developments and applications.

6. Evaluation

In order to assess the success of ECHOES, evaluation of many aspects of the environment is required. Aspects internal to the ECHOES environment, such as testing the accuracy of the child model; assessing the suitability of the actions selected for an individual child; confirming the appropriateness of a particular interaction and validation of the gaze detection, will be evaluated as part of the developmental cycles of the respective relevant technologies, within the PCM methodological framework. Input from the various interdisciplinary groups involved in the design and development team is central to this.

Broader formative aspects of evaluation related to the development process, such as the initial testing of the learning activities; usability of the environment in the various stages of development, and fine tuning of the environment (e.g. in relation to appropriate timing and duration of activities) to the target populations, is addressed through task-based, exploratory, formative evaluation studies with small groups of typically developing and children on the Autistic spectrum, and with input from expert practitioners. Such studies both provide feedback at each stage of development of the ECHOES environment, and inform the design of the next stage. Evaluation of these aspects requires input from all stakeholders, primarily led by the participatory design team. In view of its classroom-centred, practice-driven basis, the informal and exploratory methods described by, e.g., Murray (1993) and Twidale (1993) are of relevance here.

The most significant effort within the evaluation, however, is in relation to assessing the impact of the technology-enhanced learning environment on a child's learning and performance, and assessing other consequences of the intervention. The general approach to this, in relation to the role of the various stakeholders in this process, will be considered in the following section.

6.1. The overall approach to assessing the impact of intervention through ECHOES

In order to assess the impact of the ECHOES learning environment, performance can be evaluated by looking at an individuals' change in performance over time. Key questions are evaluating children's performance within ECHOES, within environment generalisation and also whether performance within ECHOES generalises to behaviour and development outside ECHOES.

Arguably, the 'Holy Grail' of any intervention in ASD is that learning generalises to everyday functioning. However, generalisation has proved very hard to achieve even for large scale, resource intensive studies with very specified outcomes (e.g. Green et al., 2010). Theoretically this may be because we are fighting against 'Reduced Generalisation' (Plaisted, 2001) in ASD, i.e. the reduced processing of the similarities that are held between stimuli and situations - which in essence means that it is more difficult for people with ASD to generalise from one context to another than for non autistic individuals. Recently, it has been suggested (e.g. Rao, Beidel & Murray, 2008) that before expensive large scale randomised controls (which measure effectiveness) are rolled out, the efficacy of an intervention must first the established; e.g. through more single-case and open trial designs (e.g. Walen & Schreibman, 2003). So, ECHOES falls under the umbrella of looking for efficacy. The approach taken in ECHOES is therefore one of exploratory small-group case-based research, rather than large-scale longitudinal studies; using a single participant, multiple baseline design across participants, staggered over time, across multiple sites.

Within the framework of intervention there are also issues of proximal versus distal effects, with the possibility of the attenuation of any intervention as we move from proximal to distal (Green et al., 2010). So, for example, within ECHOES a proximal effect of intervention would be any 'within ECHOES' environment change. Whereas a distal effect would be whether the child showed any improvement in their everyday social cognitive understanding. One standard method of evaluating an intervention is to have an intervention group and a non-intervention group and compare the two groups pre- and post intervention (e.g. Tanaka et al., 2010). However, this may not be suitable for looking at more fine-grained individual performance. Within ECHOES, individual level performance can be evaluated by looking at an individuals' change in performance over time (e.g. any improvement across trials) within a particular learning activity. However, this only tells us about learning within that particular learning activity.

A key indicator of proximal change - as a result of experiencing ECHOES - would be to show that children could transfer or apply their learning to a novel, hitherto previously unfamiliar, 'test' environment. A yet even stronger case for proximal learning could be made, for example, if the child has the experience of one type of joint attention (e.g. following the virtual agent's gaze) and is then put in a test environment in which the child shows that he or she has learned to use another type of joint attention, using the conventions of ECHOES (e.g. directing the virtual character's attention) (c.p. Golan et al., 2010, for within and outside training environment change). Other advantages for looking at proximal learning is that 1) there is probably an increased likelihood of finding a change at the proximal level and 2) if there is distal change then identifying promixal change means that we can identify the actual mechanism for that distal change. In the ECHOES project any distal effects can be evaluated by asking children, teachers and parents about their perception of any difference ECHOES has made.

Success within ECHOES could be measured in terms of the levels of generalisation. For example, improvement across trials within a one learning activity would constitute success to some degree. A greater level of success would be for this learning to be shown in a novel 'test' environment. A still greater level of success would be if this learning could be shown in a novel ‘test’ environment that the children had no prior experience of (see above for joint attention exemplars). If it could be shown that children’s experiences of ECHOES had influenced the child’s behaviour outside the ECHOES environment, then this would constitute arguably the highest level of success. Overlaid upon this, however, is individual variability and for some children simply interacting and engaging with the environment could be seen as a success: if they show learning across trials in a single environment this would be seen as impressive learning. It seems likely, therefore, that success is relative and this will very much depend on the individual child’s starting point and expected capability. One of the strengths of ECHOES is arguably that it has no prior assumptions about the child and so success can be deemed on a case-by-case level.

6.2. Participation and collaboration across disciplines and stakeholders in ECHOES evaluation

The collaborative and participatory design approach taken in ECHOES further extends to the evaluation of the impact of ECHOES. The research design proposed involves experienced practitioners, from a range of backgrounds, from the outset. Evaluation is and will be undertaken in partnership with both mainstream primary and special schools, grounding it in practice and clarifying its contribution to practice. The selection of schools and teacher partners is informed by expert practitioners. Training is offered to those schools that choose to participate as partners. In addition to standard means of dissemination, outcomes will be reported and feedback provided directly to school and to parents.

Advantages of this approach include enhanced ecological validity and a starting point for developing models for working with and in schools. It will also provide increased opportunities to shape evidence-based practice, and to support practitioners in developing the skills needed to implement this. Hence it will contribute to multi-professional working, continuing professional development and will promote interdisciplinarity. The likelihood of greater impact at all levels and future uptake will also be increased. By promoting research and practice partnerships, the impact of the research may be extended beyond the life of the project.

7. Ethics

7.1. A model of disability

By providing an individualised intervention for children with ASD to scaffold their social skills development we already have taken a fundamental ethical stance. A stance that is ultimately founded in a medicalised individual model of disability where an impairment is treated by an intervention delivered to the individual in order to improve the "patient's" well-being. In contrast, the social model of disability aims to facilitate inclusion through changes within the society and provisions in the environment (Oliver, 1990). While this is common place and relatively easy to implement in some cases, in others change is very hard to achieve and often not desirable for its far reaching consequences. For example, barrier-free access to official buildings for wheelchair users became the norm, but redefining social norms to effectively reduce anxiety of people with ASD in any given social situation might not be possible.

For developing technology in this area, these lines of thought have fundamental implications. The ethical stance taken in this argument determines the user groups the technology is directed towards, the way it is delivered and what constitutes success. We believe firmly that it is important to have an ongoing discourse about defining the stance in multi-disciplinary projects like ECHOES. Ongoing, because in our experiences, the argument does not go away once the main parameters of any such project are defined; every design decision has the potential for providing openings that would allow the system to serve aspects of either view. For example, while ECHOES primarily targets the development of social skills of an individual, the system also recognises the roles of practitioners, teachers or carers within its context of use. It thereby mediates an understanding of behaviours that is directed towards the environment and mitigates effects of the disability through a change in the people around the individual.

7.2. Collaborating with people with ASD

The ECHOES project, through the focus on Participatory Design, is focused on social inclusion, particularly of vulnerable or minority groups such as children with Asperger Syndrome. In previous work, the role of those with Asperger Syndrome in Participatory Design has been marginalised and thus as a population are under-represented in the development of technology for use in social learning. The inclusion of this user group alongside their Typically Developing peers has led to a positive research environment of equal representation which is reflected in the various skills of each discipline represented.

Within the research, the child users are involved as 'Design Informants' through continuous access and participation as opposed to being fully-fledged 'Design Partners'. This position is as far along the design spectrum as is most ethically sound (Olsson 2004) due to difficulties experienced in Theory of Mind by those with Asperger Syndrome. This means that the children find it difficult to imagine other situations and contexts, possibly leading to increased anxiety and social stress. This is mediated by the research team assuming responsibility for the design, making inferences from data gathered in design workshops.

The principle of empowerment is also prominent within the ECHOES research. Previous researcher experience has shown that children, particularly those with Autism Spectrum Disorders enjoy using new technology. This enjoyment is enhanced when the children realise that they are involved in the creation of technology to help both themselves and others. This gives rise to an increased confidence. Additionally, children with Asperger Syndrome benefit from the provision of new activities that extend any current therapy, while, alongside their TD peers, they gain extra experience in and support in their development of social interaction and communication. This leads to an increase in knowledge, skills and abilities that can enable them to develop their own social learning.

There are a number of potential risks or burdens to participants that have been mediated in the research design. Since children are socially at risk, particularly those with Asperger Syndrome, research is conducted in a comfortable and familiar environment with informed consent. There is a possibility that children may feel stressed or anxious during sessions, particularly with new researchers previously unknown to them. For this reason, introductory activities are conducted with a familiar adult such as a teacher or therapist present. Sessions are as playful as possible, with the focus being on positive feedback rather than the correct answers. The nature of the design activities is that creativity is encouraged and there is often not a 'correct' answer.

Since the design sessions are conducted primarily within a school environment, it is imperative that there is not a negative impact on the child's schooling or ongoing therapy. The timing and location of the design sessions is determined to avoid this negative impact. On the contrary, the design workshops frequently enhance the school curriculum and allow for additional time to be spent on the research project. This also serves to increase the engagement of the children as the research is ongoing and a regular part of their school structure. Where possible, design workshops are conducted on a regular basis for a period of time. This ensures that there is a degree of predictability in the timetable for children with Asperger Syndrome in order to avoid anxiety. Furthermore, this reduces any disruption to usual school routines and practices.

8. Conclusions

In this article we presented the ECHOES project as one of the most recent exemplars of an interdisciplinary approach to designing technology for users on the autistic spectrum. We argue that adopting an interdsciplinary perspective is critical to developing technology that has a chance to deliver effective intervention in this context. With reference to the specific aspects of the ECHOES environment we aimed to show that if technology design is viewed through the prism of interdisciplinary research, it can not only serve as a means of delivering intervention in situ, or in addition to the increasingly overstretched care provisions by humans, but it can actually provide an extension to human-human intervention that is adaptive, intelligent and engaging. However, conducting interdisciplinary research presents huge challenges for the individual stakeholders involved. These range from fundamental ethical questions to technological feasibility and measuring success.

The ECHOES project has served here as a case-study to present the interdisciplinary research methodology that are proposing and applying. Through its uniquely diverse composition, the members of ECHOES brought a great number of different perspectives, skills, scientific traditions and personal presumptions to this project and throughout this paper we have described how we, collectively, have approached the challenge of developing technology for people with ASD and what we have learnt from the process. In the following we attempt to distill from these experiences some guidelines that we hope will be able to contribute to the practice, theory and culture of research in this field.

1. **Discussing ethics:** Throughout our work we have encountered ethical questions that impacted directly on the potential role of technology. Albeit driving the fundamental directions of projects, these questions are all to often neglected and replaced by technology induced requirements. Ethical issues range from underlying perspectives on disabilities to practices when collaborating with people with ASD and possible impacts on the wider society. They require constant attention and ongoing reflection.
2. **Marrying multiple methodologies:** Whilst different disciplines may bring with them tools and approaches that, at first glance, may seem disparate, it is important to explore where they overlap, in principle and in intent, and to examine ways in which the most pertinent aspects of each can be combined within a methodological framework that serves the multiple disciplinary perspectives. Whilst this may often be difficult, the in-depth discussion of the varying perspectives alone can serve to shape and meld a combined methodology that respects and serves the research in ways previously not considered. Developing of a coherent research framework that supports different stakeholders in understanding, appreciating and achieving the goals of the research is fundamental to the success of an interdisciplinary approach.
3. **Facilitating the participation of stakeholders:** We believe strongly that the inclusion of stakeholders, that is researchers, practitioners and most importantly people with ASD, is a key factor for success in this area. Besides the ethical obligations attached to the often marginalised roles of people with ASD in processes that impact onto their lives, their perspective on the world is vastly different from the assumptions we have developed. Only an empathetic dialogue with stakeholders can support designers and researchers in understanding what technology should be like for people with ASD, and how it might best be evaluated.
4. **Supporting personalisation and diversity:** "If you have seen one child with Autism, you have seen one child with Autism" is a common phrase amongst researchers and practitioners in the field. This pronounced diversity in behaviours, preferences and traits amongst people with ASD coupled with a tendency to monotropic attention means that any technology has to be designed to support personalisation.

5. **Providing adaptivity and interactivity:** The idea that a system should be able to emulate at least some of the human behaviours in order to support naturalistic interaction is a crucial one in a context where the goal is to support exploration and acquisition of social skills by users who are young children on the autism spectrum. In relation to this particular population, while it is important to adapt the environment to the individual child, at the same time this must be balanced with avoiding overspecialisation to what may be a narrow interest of the individual child. AI methods and techniques equip us with a starting point in relation to affect and cognition modeling. However, typically those methods and techniques have been applied and tested only in the context of older users with no identified neuro-cognitive disorders such as Autism, and often within well-defined interaction domains such as flight information or teaching mathematics. The differences in the domain of application and the special needs of the users with Autism and of young children, present new and exciting challenges and an opportunity to further test and extend the existing methods and techniques.
6. **Determining the impact of technology:** In relation to the effect of intervention on young children with ASD, it is often more appropriate to consider small scale, multiple baseline measures across participants, staggered over time, than large scale randomised control studies, i.e. focussing on establishing efficacy of an intervention, before considering effectiveness. If any generalisation of learning is to be demonstrated, it is vital to define and evaluate both proximal and distal indicators of change.

References

Arrington Research. (2010). http://www.arringtonresearch.com/headmountframe.html

Bardins, S., Poitschke, T. and Kohlbecher, S. (2008). Gaze-based interaction in various environments, in Proceeding of the 1st ACM workshop on Vision networks for behavior analysis, Vancouver, British Columbia, Canada, 2008, pp. 47-54.

Baron-Cohen, S., Golan, O., Chapman, E., and Granader, Y. (2007). Transported to a World of Emotions. The Psychologist, 20(2):76–77.

Bjerknes, G. and Bratteteig, T. (1995). User participation and democracy: a discussion of Scandinavian research on systems development. Scand. J. Inf. Syst., 7(1):73–98.

Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13:6:4–16.

Bundy, A. (1986). What kind of field is Artificial Intelligence? DAI Research Paper No. 305. Department of Artificial Intelligence, University of Edinburgh.

G. Castellano, K. Karpouzis, C. Peters, and J.-C. Martin. Special issue on real-time affect analysis and interpretation: closing the affective loop in virtual agents and robots. Journal on Multimodal User Interfaces, vol. 3, no. 1, pp. 1–3, 2010.

Charman, T. Why is joint attention a pivotal skill in autism? Philosophical Transactions: Biological Sciences, 358:315324, 2003.

Chen, J. and Lemon, O. (2009). Facial Feature Detection and Tracking in a New Multimodal Technology-Enhanced Learning Environment for Social Communication, in Proc. of the IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2009.

Chen, J. and Tiddeman, B. (2010). Multi-cue facial feature detection and tracking under various illuminations, International Journal of Robotics and Automation, 2010.

Cohen, L. & and Manion, L. (1980). Research Methods in Education. New York: Routledge.

Conlon, T. and Pain, H. (1996). Persistent collaboration: A methodology for applied artificial intelligence and education. International Journal on Artificial Intelligence in Education, 7(3/4):219–252.

DeMenthon, D. F. and Davis, L. S. (1992). Model based object pose in 25 lines of code," in Proc of 2nd European Conference on Computer Vision, Santa Margherita Ligure, May, pp. 335-343.

Dias and Paiva, A. (2005). Feeling and Reasoning: A Computational Model for Emotional Characters

Druin, A. (2002). The Role of Children in the Design of New Technology. Behaviour and Information Technology, 21(1):1–25.

Duchowski, A. T., Eye Tracking Methodology: Theory & Practice, Springer-Verlag, London, UK, 2nd edition, 2007.

Figueiredo, R., Brisson, A., Aylett, R., and Paiva, A. (2008). Emergent stories facilitated: An architecture to generate stories using intelligent synthetic characters. In Joint International Conference on Interactive Digital Storytel ling (ICIDS 2008). Erfurt, Germany.

Foster, M.E., Avramides, K., Bernardini, S., Chen, J., Frauenberger, C., Lemon, O., and Porayska-Pomsta, K. Supporting Children's Social Communication Skills through Interactive Narratives with Virtual Characters. Proceedings of ACM Multimedia 2010, Florence, Italy, October 2010.

Fullan, M. (1991). The New Meaning of Educational Change. London: Cassell.

Good, J. and J. Robertson, CARSS: A Framework for Learner-Centred Design with Children. Int. J. Artif. Intell. Ed., 2006. 16(4): p. 381-413.

Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., Howlin, P., et al. (2010). Parent-mediated communication-focused treatment in children with autism (PACT): a randomized controlled trial. Lancet, 375(9732), 2152-2160.

Jones, C., McIver, L., Gibson, L., and Gregor, P. (2003). Experiences Obtained from Designing With Children. In IDC '03: Proceedings of the 2003 conference on Interaction design and children, pages 69–74, New York, NY, USA. ACM.

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. Clinical Psychology: Science and Practice, 10(2):144–156.

A. Kapoor, W. Burleson, and R. W. Picard. Automatic prediction of frustration. International Journal of Human-Computer Studies, 65(8):724–736, 2007.

Keay-Bright, WE. (2007). The Reactive Colours Project: Demonstrating Participatory and Collaborative Design Methods for the Creation of Software for Autistic Children. Design Principles and Practices: An International Journal. 1:7–16

Landauer, T. (1995). The Trouble with Computers: Usefulness, Usability, and Productivity. Cambridge, MA: MIT Press.

Matsumoto, Y. and Zelinsky, A. (2000). An Algorithm for Real-Time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement, in Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 499-504.

Muller, M. J. (2003). The Human-Computer Interaction Handbook, chapter Participatory Design: The third Space in HCI, pages 1051–1068. Lawrence Erlbaum Associates, London, UK.

Murray, D. (1997). Autism and information technology: Therapy with computers. In S.Powell. S., J., editor, Autism And Learning: A Guide To Good Practice. David Fulton Publishers.

Murray, D. and Aspinall, A. (2006). Getting It: Using Information Technology To Empower People With Communication Difficulties. London: Jessica Kingsley Publishers.

Murray, D. and Lesser, M. (1999). Autism and Computing. In Proceedings of Autism99 online conference organised by the NAS with the Shirley Foundation.

Murray, T. (1993). Formative Qualitative Evaluation for "Exploratory" ITS Research.
Journal of Artificial Intelligence in Education, Vol 4 No. 2/3, 179-207.

Oliver, M. (1990). The politics of disablement. Palgrave Macmillan.

Olsson, E., What active users and designers contribute in the design process. Interacting with Computers, 2004. 16(2): p. 377-401.

Ortony, A., Clore, G. L., and Collins, A. (1988). The Cognitive Structure of Emotions. Cambridge University Press.

S Parsons, KK Guldberg, A MacLeod, GE Jones, A Prunty, T Balfe. (2009). International review of the literature of evidence of best practice provision in the education of persons with Autistic Spectrum Disorders, National Council for Special Education: Ireland.

Prez, A., Crdoba, M.L., Garca, A., Mndez, R., Muoz, M., Pedraza, J. and Snchez, F. (2003). A precise eye-gaze detection and tracking system, in: Proc. of the 11th International Conference in Central Europe of Computer Graphics, Visualization and Computer Vision'2003, Plzen, Czech Republic.

van Rijn, H. and Stappers, P. J. (2008). The Puzzling Life of Autistic Toddlers: Design Guidelines from the LINKX Project. Advances in Human-Computer Interaction, 2008.

Safyan, L. and Lagattuta, K. H. (2008). Grownups are not afraid of scary stuff, but kids are: young children's and adults' reasoning about children's, infants', and adults' fears. Child Development, 79(4):821–835

Schön, D. A. (1983). The Reflective Practioner: how professionals think in action. London: Temple Smith.

Smith, C. A. and Lazarus, R. S. (1990). Emotion and adaptation. In Pervin, L. A., editor, Handbook of personality theory and research, pages 609–637. New York: Guilford.

Stenhouse, L. (1975). An Introduction to Curriculum Research and Development. London: Heinmann.

Swartout W. (2010), Lessons Learned from Virtual Humans, AI Megazine, Vol. 31, No. 1, Spring 2010.

Twidale, M. (1993). Redressing the Balance: the Advantages of Informal Evaluation Techniques for Intelligent Learning Environments. Journal of Artificial Intelligence in Education, Vol 4 No 2/3 155-178.

A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. Image and Vision Computing, vol. 27, no. 12, pp. 1743–1759, 2009.

L.A. Vismara and G.L. Lyons. Using perseverative interests to elicit joint attention behaviors in young children with autism: Theoretical and clinical implications for understanding motivation. Journal of Positive Behavior Interventions, 9(4):214, 2007.