

A Gaze Contingent Environment for Fostering Social Attention in Autistic Children

Rameshsharma Ramloll[♥]
National Rehabilitation Hospital
102 Irving St., NW,
Washington DC 20010

Cheryl Trepagnier[♣], Marc Sebrechts[♣], Andreas Finkelmeyer[♣]
Department of Psychology
The Catholic University of America,
4001 Harewood Rd., NE, DC 20064

Abstract

This paper documents the engineering of a gaze contingent therapeutic environment for the exploration and validation of a proposed rehabilitative technique addressing attention deficits in 24 to 54 months old autistic subjects. It discusses the current state of progress and lessons learnt so far while highlighting the outstanding engineering challenges of this project. We focus on calibration issues for this target group of users, explain the architecture of the system and present our general workflow for the construction of the gaze contingent environment. While this work is being undertaken for therapeutic purposes, it is likely to be relevant to the construction of gaze contingent displays for entertainment.

Keywords: Autism, attention, eye tracker calibration, gaze contingent environment, design workflow

1 Introduction

This project establishes a platform for exploring and validating a proposed therapeutic intervention for infants with autistic spectrum disorder. Recent investigations revealed that the visual behaviors of children suffering from autistic spectrum disorder show significant differences when compared to controls [Leekam et al. 1998][Schultz et al. 2000][Peart 2002][William et al. 2001][Trepagnier et al. 2002]. Researchers contend that a gaze contingent environment could be used to reinforce appropriate gaze behaviors in autistic children [Trepagnier et al. 1999]. By training young children with autistic disorder to attend to, interpret, and respond to visual communication, during a critical period of language (including non-verbal) acquisition, it is hoped that this process will decrease their inadequacy in interpreting social stimuli. The clinical foundations of this hypothesis are described elsewhere [Trepagnier et al. 1998].

[♥] rameshsharma.ramloll@medstar.net

[♣] Cheryl.Trepagnier@medstar.net

[♣] sebrechts@cua.edu

[♣] 25finkelmeye@cua.edu

Copyright © 2004 by the Association for Computing Machinery, Inc.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.
© 2004 ACM 1-58113-825-3/04/0003 \$5.00

The gaze contingent environment being developed consists of an amusement ride customized to accommodate a display, sound speakers, a gaze tracker and cameras to observe the behavior of the user. A tracked gaze behavior, if deemed adequate, is reinforced by positive visual, auditory and vestibular responses. In this paper, we focus on (1) the motivation, (2) the experimental set up, (3) the calibration issues and (4) the production workflow of this environment in order to provide enough information to allow the replication of our work and to illustrate current challenges.

1.1 Fostering Social Attention in Autistic Children

The novelty of this gaze contingent environment is in its application as a means to foster social attention in autistic children. One of the basic behaviors to be reinforced is the act of looking at the informative area of the face from which communicative, as well as identity, gender and age information can be obtained. The training starts with simple subject-virtual human interactions that gradually grow in complexity. For example, we begin by monitoring the time taken for the subject to select the face of the virtual-human as an object of interest and the duration of his or her fixations on the face. We then attempt to entice the subject to modify these parameters to meet typical gaze parameters for such behaviors in controls by providing various levels of rewards ranging from congratulatory jingles, to vestibular rewards.

1.2 Gaze Training Scenarios

Developmental psychologists have been particularly interested in the emergence of gaze following as a basic indicator of the capacity to share attention and learn about the world [Scaife et al. 1975]. Our training environment attempts to gradually lead an autistic child through tasks, organized according to difficulty levels:

1. Looking at the virtual human's eyes
2. Looking at two virtual humans interacting with each other, from the eyes of one to the other
3. Following gaze direction: looking at the virtual human's eyes and inferring the object of interest.
4. Following gaze direction: then looking back again at the virtual human's eyes when the latter is verbally and non verbally expressing a reaction to what s/he is looking at.

It is of course anticipated that the content of interactions will need further modifications in the light of iterative testing. Detailed specification of interaction steps, thresholds for interaction triggers, frequency with which already acquired skills will be probed for retention, levels of rewards and the stories in which these will be implemented constitute a major component of content production.

2 User View Of Amusement Ride

Figure 1 shows a side view and two cross-sectional inside views of the amusement ride. In a typical session, a cartoon that the child is known to be fond of is started before s/he comes into the

play area. Three non-coercive approaches are used to lead the child into the ride. At the minimal level of guidance, the sounds and events happening on screen are expected to raise the curiosity of the child and entice the latter into the cubicle. At an intermediate level, the experimenter speaks to the child via a video link through the display to attract the child's attention. Finally, at the maximal level, parents who accompany the child are allowed to intervene voluntarily to encourage their child to get into the ride.

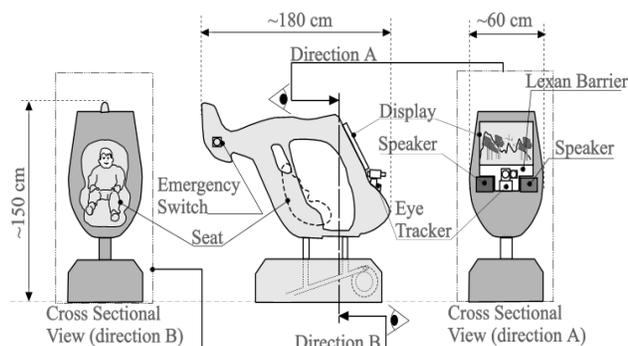


Figure 1 Custom ride with gaze contingent environment

Once the child is seated, a gaze calibration procedure is initiated and the gaze contingent behavior-training environment is started. An adequate visual behavior is rewarded with a variety of system responses that range from congratulatory events on screen and various types of vestibular rewards. For example a reward can be an invitation for a ride followed by the ride moving one cycle (forward and up-backwards and down) every two seconds with an appropriate animation on screen. Color plate 1 shows in further details the customized amusement ride.

2.1 Making the Environment Safe and Robust

Designing a safe gaze contingent environment for this special group of users is an important concern and significant efforts were spent to meet the strict clinical research guidelines of research projects carried out in a hospital environment. Engineers customizing the ride have introduced modifications to minimize the number of pinch points inside and around the vehicle. The inside environment is padded with preformed sponge/foam in order to minimize risks of injury. The inner surfaces are cleared of as many distractions, e.g. strippable or peelable objects, as possible. The eye tracker is hidden from the child by a transparent Lexan® barrier.

The ride can be stopped in three ways. (1) The experimenter can stop the ride by issuing an instruction to the ride via software. (2) S/he can also do so by hitting a hardware switch on the tail of the ride. (3) The child can stop the ride by sitting up causing a pressure switch in the seat to break contact. Thus, the child is in control of when to cease the experiment immediately.

2.2 Experimenter Activities Behind the Scenes

There are significant activities by two experimenters taking place behind the scenes to create and record the interactive experience offered to the child. Cameras with microphones installed at strategic locations provide continuous information about the child's activities. This information provides in context interaction opportunities between one experimenter and the child. One experimenter who is also a clinician experienced in dealing with

autistic children may interact with the child through a popup window on the ride display.

Eye tracker calibration and its maintenance are critical aspects of this environment more so because of our constraint of having to perform the process without any instructions given to the subject. While the ISCAN eye tracker does allow servo positioning of the camera, it does so in a range that is too small to be useful for our purposes. We achieve better results controlling its orientation manually. One experimenter is dedicated to the task of setting the direction, focus and zooming level of the ISCAN camera in order to obtain a sharp centered eye image of the subject. Meanwhile, the other is responsible for executing the calibration process and running the contingent gaze display applications on the content PC (Figure 2). The position of the child's gaze overlaying the content displayed on the display of the amusement ride is recorded throughout the experiment (Figure 2, VCR and Content with Tracked Gaze Position Monitor).

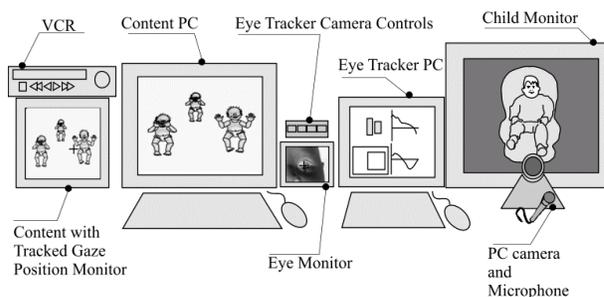


Figure 2 Experimenter equipment behind the scenes

Figure 3 describes the experiment environment organized in order to mask distractions due to the flurry of activities if any behind the scenes. Suffice to mention that some space is allocated for a parent whose presence is likely to be comforting to the child subject. The parent is allowed to hit the safety switch on the tail of the amusement ride to stop the ride if s/he wishes to. The inside of the amusement vehicle are constantly monitored from two cameras. One camera, which is installed next to the eye tracker, points at the child through the Lexan® barrier. The other points to the child through the window found opposite to the entrance of the amusement vehicle. The experimenter then has the choice of seeing either views or any one of them on the child monitor shown in Figure 2.

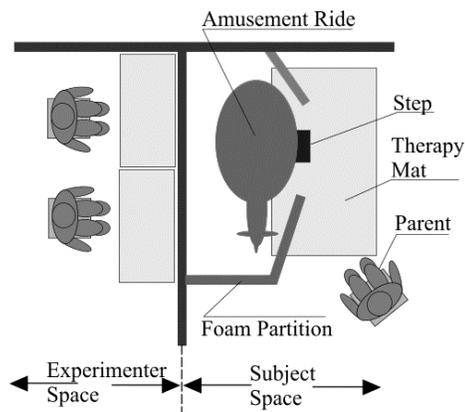


Figure 3 Top view showing the organization of the experiment space

3 Search For A Better Calibration Method

The success of any gaze contingent environment is to a large extent based on the quality of calibration achieved. Our calibration method is informed by an earlier study looking at characteristic gaze responses to dynamic visual stimuli. The study described in further details elsewhere [Sebrechts et al. 2003] examined the tendency to look at sudden onset stimuli in the absence of instruction, as well as examining the time course of that process. In short, stimuli were presented on a 27" diagonal monitor with a resolution of 640x480 pixels. Participants' faces were approximately 30 inches from the screen, resulting in a total horizontal viewing angle of 40°, or 16 pixels per degree. Eye tracking was conducted using an ISCAN desk-mounted tracker placed below the monitor. An ISCAN data collection software on a separate computer captured gaze data. Stimulus presentation and eye-tracking data were synchronized to yield a sample rate of 50 Hz (1 sample every 20 milliseconds). We describe 2 seconds of each stimulus's behavior according to 4 characteristics in Figure 4 (the box represents the stimulus; in this particular experiment, the stimuli were (1) a butterfly flapping wings and (2) a green box; the maximum length and breadth of the box is approximately 1.2°).

Onset-collapsing:

During the first second, the stimulus shrinks and during the next second, the stimulus stays put

Onset-no change:

During both seconds the stimulus stays put and does not change in size

Pursuit-collapsing:

During the first second, the stimulus stays small; the next half while traveling to the next point, zooms to its maximum size and during the final half second, shrinks again

Pursuit-no change:

In the first second, the stimulus stays put, and in the next second the stimulus travels from one point to the other.

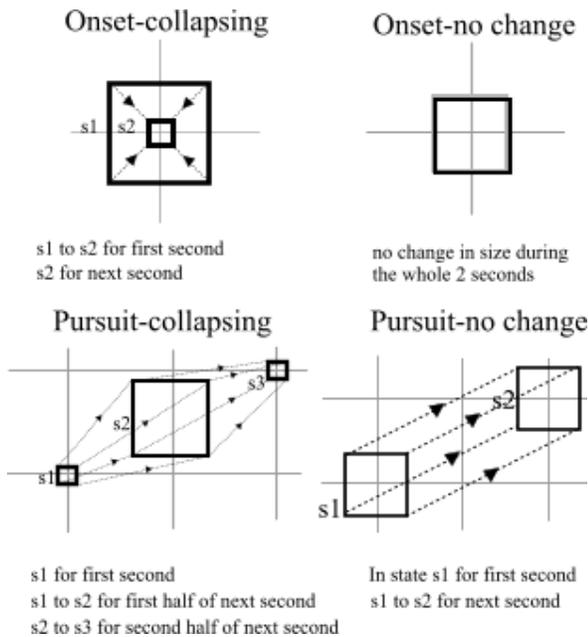


Figure 4 Animations used to study characteristic gaze responses to dynamic stimuli

The various stimuli while behaving in ways just described during the two seconds, also progress from one point to another as described in Figure 5. In this example, the stimulus proceeds along: random, N, C1, random, P, C2, random, B, C3, random, E, C1, random, G, C2, random, I, C3, random, S, C3, random, E, C4, random, G, C5, where random is any point that is not a calibration point and C1-C5 are typical calibration positions. The progressions may thus appear as discrete jumps from point to point or smooth transitions as dictated by the specifics of the animation. All point-to-point (e.g. C1-F) transitions take 2 seconds.

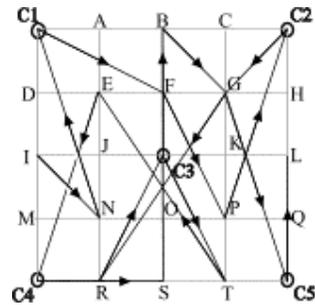


Figure 5 Example stimulus progression: I, N, C1, F, P, C2, G, R, C3, T, E, C4, S, B, G, C5, L

3.1 Observed Characteristic Gaze Responses

The average result for six non-autistic participants is illustrated in Figure 8 (see appendix). Results show little variation between the individual graph patterns. Data reliability is satisfactory. Tracking, checked before and after stimulus presentation, is found to be good. The minimum discrepancy between target and gaze in each condition is on average below 16 pixels or less than 1°, the stated precision of the eye-tracker.

For the “onset” stimuli cases, there are on average 150 milliseconds between the onset of the target stimulus and initiation of eye movements. It then takes about 250 milliseconds for the eye to move to the stimulus location. Gaze is then within 0.8° of the target, where it remains relatively stable for about 800 milliseconds. Target stimuli in this experiment appeared on average at about 13° from their previous location. The distance between successive targets did not appear to influence the lag time or the subsequent saccade.

For “pursuit” stimuli cases, it takes about 150 milliseconds for the eye position to match the target location, suggesting that there is a lag of about 150 milliseconds during pursuit. As the target position is approached there is often a slight overshoot, represented in the figure as a temporary increase in the discrepancy. The gaze location then settles within about 0.8° of the actual target location, where it remains for 700 milliseconds while the target stimulus is in a fixed location. After 1 second, the stimulus starts moving again. It takes about 200 milliseconds before the eye begins its pursuit. Optimal pursuit of within about 1.5° is achieved within 250 milliseconds.

For both onset and stimuli cases, by 500 milliseconds gaze is as close to the target as it will get. After 1 sec, gaze starts to drift. Having someone pursue a target continuously prior to arriving at its location ensures that the eye is closer to the new target location. There is a lag of about 150 milliseconds and a tendency to overshoot the target, so that optimal matching of eye gaze to target location still occurs at about 500 milliseconds.

3.2 Optimal Time Interval for Calibration Data Capture

We have shown that eye gaze follows systematic patterns in response to specific stimulus presentation sequences. Since the patterns are predictable, we attempt to use this effect in an automated calibration procedure. A calibration window can be defined as the time interval during which the difference between the target position and gaze position is less than 20 pixels. It may be viewed as a way to specify the time interval during which valid calibration data can be obtained. This 20 pixels value is likely to depend on the quality of the tracker. Figure 8, in the appendix, shows the calibration windows (T1: pursuit; T2 & T3: onset) relevant to the various types of visual stimuli.

3.3 Towards Instruction Free Calibration

The typical calibration procedure for the ISCAN eye tracker and probably others involves asking the subject to look at a location, then visually confirming that location, and repeating this process five or more times. Entering a key press, can easily take more than a second, leading to imprecision in calibration. By automatically recording gaze 500 ms after the onset of calibration stimuli, our data shows that one is likely to capture data at the appropriate time. This observation also implies that the whole calibration process can be automated and done without human intervention as will be explained further. Suffice to mention here that the quality of the eye tracker calibration for early trials involving lab members instructed to keep a steady head position using this method appears to be high especially when the acquisition of a good eye image has been maintained. We plan to carry out formal experiments to confirm this observation.

3.4 Challenges for Calibrating Toddlers

However, automating the calibration process for our target user group, to whom no instructions are dispensed, currently faces a number of problems. The main one being that it is not always possible of obtaining a sharp eye image of such subjects for a reasonable time interval especially if they are moving about exploring their new environment inside the ride.

The calibration process as it stands begins with a cartoon show that the subject enjoys. One experimenter keeps a close watch on the tracked eye and follow the latter as it moves. When the child stops moving and if the eye image is satisfactory, the experimenter hits a key to trigger a white screen with a small animation, e.g. a dancing rabbit with music, a fire engine with sound, at the location of one of the calibration points. The ocular and pupil reflection data are captured 500 ms after the onset of the stimulus. The calibration stimulus stays on only for one second followed by the continuation of the interrupted cartoon. This step is repeated five times whenever the experimenter decides that a good eye image has been obtained. Once calibration is completed, the calibration is checked. This is achieved by having one experimenter hit a key to interrupt the current cartoon, clear the screen and produce an object that moves about on the screen. If the gaze of the child follows the object to within 1 degree, a good calibration has been obtained. This process can be repeated whenever there are concerns that calibration has been lost. It is to be noted that this process has consistently produced good calibrations whenever a good and stable eye image is obtained. At this point in our iterative design, we have successfully calibrated four autistic toddlers using this method and their gaze behavior has been captured on video.

3.5 Towards Total Automation of the Calibration Process

In order to fully automate the calibration process, we need to achieve the following goals:

Firstly, we need a better servo positioning of the eye tracker camera. The range, focus and zoom should be servo controlled over a range that will allow the user to move anywhere within the cubicle while facing the display without degrading the quality of the captured and centered eye image. Secondly, we need a method to detect whether the quality of the eye image obtained is good enough to trigger a calibration operation. We are currently consulting with the developers of ISCAN in order to explore solutions to achieve these goals.

In order to avoid re-calibrations, the position of the tracked gaze can be adjusted in software to compensate for systematic or random errors. Unfortunately, our early attempts did not produce simple functions that could be used to reliably reposition accurately a drifted gaze position. This is probably because the compensating functions are non-linear and distinguishing between situations where errors are systematic or random is in itself a difficult task. Some fairly complex inference mechanism needs to be integrated in a system capable to detect such situations automatically.

Reliable software based calibration degradation checks and error control remains an outstanding challenge. But we are confident these problems can be solved as we learn about what is essential for calibration to take place, identify the specific areas that can be automated and then piecing all the various automated parts to yield a whole process that will hopefully not require human intervention.

3.6 Eye Tracking During Ride Motion

An important observation made during calibration and which addresses an earlier popular concern is that stable, sharp, centered eye images can be obtained even while the ride is moving, keeping in mind that currently the cubicle moves slowly but steadily. We observe that, even with our current eye tracker arrangement, eye tracking a subject in a slowly and steadily moving ride is not more difficult than if the ride were stationary.

4 System Architecture

This section describes how the various modules of the amusement ride interact and the more technical aspects of the set up.

4.1 Software Development Platform

There are four homegrown applications on the content PC shown earlier in Figure 2. These are (1) the auto calibration application, (2) a mouse driver for the ISCAN eye tracker, (3) an application for creating and editing the gaze contingent content and (4) an application for rendering the content. (1), (2) and (4) are fully implemented, (3) is under construction. We defend our choice of using Flash MX as a development environment mainly because it is graphics centric and allows easy manipulations of graphical elements from a programming point of view. However, the downside is that it has a relatively slow rendering speed. Fortunately this does not outweigh its advantages significantly to discourage us from using it. Macromedia's Flash has also evolved from being just an application that produces 'cool' interactive web graphics to a more powerful object-oriented development

environment that can be used to create stand-alone applications with appropriate third party tools such as FlashStudio Pro. In addition, Flash MX has now evolved to Flash MX 2004 Pro which is reported to allow still faster rendering speeds and is increasingly viewed as a serious competitor to the .NET and Java environments.

	Application	Functionality	Tools
1.	Auto-calibration	Sends timed instructions to the ISCAN eye tracker and for the presentation of calibration stimuli according to a predefined script	VC++, Flash MX, FlashStudio Pro
2.	Mouse driver for the ISCAN eye tracker	Uses gaze data to position mouse cursor, controls cursor visibility and allows simulations of mouse events	VC++
3.	Training Level Editor	Allows clinician or psychologist to create interaction scenarios and save a training file describing content and behaviors	Flash MX, FlashStudio Pro, VC++
4.	Training Player	Loads and renders the training file	Flash MX, FlashStudio Pro, VC++

Table 1 Summary of the development environment

4.2 Design Philosophy: Gaze Code - Content Code Separation

We factor out the code responsible for the gaze capture and control from the code responsible for the presentation of gaze contingent content. This separation allows us to split the design and implementation tasks to enable the programming of the two modules separately in order to provide better opportunities for each module to assimilate more complex functionalities efficiently and independently. This separation is achieved trivially in our case. The mouse driver listens to the gaze position data from the eye tracker through a separate serial connection and positions the mouse cursor accordingly. Thus, the cursor is controlled via gaze and the environment, designed to respond to its position and behavior, now appears to respond directly to gaze behavior. In this case of course, the cursor is also made invisible since it would be a distraction otherwise. This approach while simplifying development in an almost trivial way also allows the designer to try out the interactivity of his or her creation immediately without requiring the starting up of the whole gaze contingent system.

4.3 ISCAN Data Acquisition Software Control Protocol

We provide here the necessary background for explaining the implementation of our custom eye tracking procedure. The ISCAN eye tracking system provides ample opportunities for customizing the calibration procedure. Our Auto-calibration application sends instructions to the eye tracker application (ISCAN DAQ 3.58) running on a separate PC through a serial null modem connection. These instructions (Table 2) are specified by

the ISCAN data acquisition protocol and trigger appropriate remote procedure calls in the client program to complete the calibration process. In short, the ISCAN DAQ provides hooks for software control. The most relevant instructions are described here. The <Reset> command indicates the beginning of a new calibration process. The <Enter> command triggers the capture of eye data for a given calibration point and brings the next calibration point into consideration. The <Run> command marks the end of the calibration process and instructs the software to output calibrated gaze data.

The following describes in steps (S#) an example Auto-calibration script running on the Content PC.

```
S1: send <Reset>, wait <200 ms>
S2: send <Calibration>, wait <200 ms>
S3: show <Stimulus at current Calibration Point>, <wait 500ms>, send <Enter>, wait <500 ms>,hide < Stimulus at current Calibration Point>
Repeat S3 once for each calibration position
S Final: <Run>
```

Note that the 200 ms delay between instructions S1 and S2, and S2 and S3, has been determined by trial and error and is required to take into account any serial data transfer delays between the content PC and the eye tracker PC together with a software idiosyncrasy that prevents the eye tracker software to interpret instructions sent to it in rapid succession. In step S3, we capture the calibration data, 500 ms after the stimulus is shown in order to make sure that capture is achieved within the calibration window as defined earlier in section 3.2.

While it is possible to have this script run through to completion with a cooperative subject willing to sit fairly still, this is not the case with toddlers who have not been provided with any instructions. In the latter case, step S3 is triggered with a single key press whenever the experimenter sees a stable sharp eye image in the eye monitor (shown in Figure 2), for example when the child is pausing between exploratory head turns. Thus, while our approach achieves total automation for a cooperative subject, this is not the case for an autistic toddler or for that matter, any child not provided with any instructions.

Recording Commands	
1.	Begin Recording Data <Begin>
2.	Stop Recording Data <Stop>
POR Calibration Commands	
1.	Reset Mode <Reset>
2.	Calibration Mode <Calibration>
3.	Run Mode <Run>
4.	Select calibration point <Select>
5.	Enter calibration point <Enter>
6.	Move Selected calibration point UP <Up>
7.	Move selected calibration point DOWN <Down>
8.	Move selected calibration point LEFT <Left>
9.	Move selected calibration point RIGHT <Right>
10.	Toggle between Fine and Coarse Cursor Movement <Toggle>

Table 2 ISCAN Data Acquisition (DAQ 3.58) Protocol

4.4 Ride Motion Control Protocol

The application responsible for managing the gaze contingent content provides facilities for sending instructions through a parallel port to a relay switch that controls the motor of the amusement ride. Table 3 describes the instructions that can be understood by the ride. Mainly because of time constraint, the ride currently can only be turned on (Table 3, case 1) and off but in the next iteration, the other vestibular responses will be implemented.

1.	Turns amusement ride ON (clockwise, looking from the entrance side) (slow, 0.5 Hz)
2.	Turns amusement ride ON (fast, 1 Hz)
3.	Turn amusement ride OFF
4.	Turn amusement ride ON (anticlockwise, looking from the entrance side) (slow, 0.5 Hz)
5.	Turn amusement ride ON (anticlockwise, looking from the entrance side) (fast, 1 Hz)

Table 3 Amusement Ride Control Protocol

Implementation of this protocol can potentially give rise to a greater set of possible vestibular responses by combining the various types of motions to produce more complex sequences.

5 Creating Content for Gaze Training Environment

After describing (1) how the calibration of the children subjects is performed, (2) how the interaction between various subparts of the system is achieved, we now discuss (3) the behavior triggering mechanism and (4) the content production workflow of the gaze contingent application. These issues, beyond the main clinical research goals, are important from an engineering perspective.

5.1 First Gaze Contingent Prototype

The earliest prototypical scene of our gaze contingent display consisted of three babies who are able to look at the position of gaze of the child subject. They would also start dancing when the child would look at them. In addition, the virtual babies could look at objects in motion in their environment, perhaps to simulate attention sharing with the child subject.

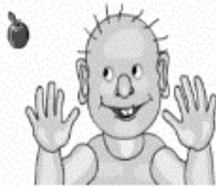


Figure 6 Gaze behavior of baby following moving object of interest

5.2 Dwell-threshold Event-trigger Mechanism

The basis of our gaze contingent display so far is based on a simple dwell threshold-event triggering mechanism. Each element of the scene (e.g. hand, eye, head, torso, fore arm each being a Flash MX MovieClip object) keeps track of the time the subject continuously spends gazing at it. If this time exceeds a threshold level, then the element sends an event to a reaction-broadcaster whose function is to collect events posted, identify their source, look up a table of rules and broadcast other events to target elements that will respond accordingly. Another advantage of having gaze-aware graphical elements, as it were, is that the dwell

mechanism will still be valid even if the graphical element concerned is geometrically transformed to meet the demands of animation. While this is trivial in a Flash environment, a significant amount of development effort has to be spent implementing efficient object tracking algorithms if the animation was implemented in a non-object-oriented fashion. While the look up table of rules is a good starting point for building reactivity into our prototype more complex behaviors can only be achieved with better artificial intelligence approaches. In the future, we plan to integrate inference mechanisms that are informed by physiological and psychological models into the behavior engines that will drive our virtual characters [Badler et al 1999].

5.3 Design Philosophy: Code Data Separation

After completing the construction of our first prototype we learn that if the code is closely tied with the content then fewer opportunities are available for clinicians to control the production process. Changes to the content are, as expected, difficult in such conditions. In addition, the process of programming content based on storyboards or text scripts to describe interactions is found to be error prone and open to loose interpretations. This is inefficient both for the programmers and for the clinicians or psychologists who have the expertise to define content. While clinicians or psychologists have a clear idea of what they need as content, they lack the knowledge about whether their current development environment or development team can realistically satisfy their specifications. And the number of iterations to reach a satisfactory consensus is usually large or is really never reached. We attempt to apply lessons learnt from the game industry to address this problem.

5.4 Applying Lessons from the game industry

A decade ago, game developers have realized the importance of tools for optimizing computer game production. The once groundbreaking notions of game engines [Bishop et al. 2001] and level editors are now standard production tools [Rouse III 2000]. Once we realize early on that our gaze contingent training environment is indeed a game in its own right, it follows that in order to design a multi-level engaging and motivating experience we need to divide and conquer the problem by separating the content creation from the content rendering tasks. It is thus imperative for an environment editing application and a separate rendering system to be implemented. It is widely known that simplifying content creation leads to better interactive environments because designers have more time to make their levels fun instead of just making them work [Rouse III 2000].

5.5 Training Level Editor and Player

The purpose of the level editor is to allow the clinicians or psychologists to design and refine directly the interactive story according to lessons learnt from studying the interactions of the autistic children to the contingent display. The purpose of the Player is to render the environment containing objects and behaviors as defined by the clinicians and psychologists themselves using the Training Level Editor. By allowing the clinicians or psychologists to quickly jump into the experience and test out interactions on the fly while editing it, provides these stakeholders with a faster route to what they are really aiming at.

5.6 Developing Virtual Human Characters

While, our early prototypes involving 2D cartoon looking characters were good enough to test the basic functionality of our

system, a decision was made to outsource the creation of character models to 3d digital media experts from DigitalElite in order to boost the graphical quality and realism of the environment. Figure 7 shows a snapshot of the virtual baby to be used in the next version of our gaze contingent environment. The rendering of emotions, lip-synching and gesturing need to be accurate for a virtual character to be believable. By outsourcing these tasks to expert designers and artists, we hope to improve the aesthetic appeal and simulation accuracy of our virtual environment.

The workflow to produce the elements of the virtual character is briefly as follows. Firstly, the clinicians lists a range of emotions and reactions they are interested in and write down what they expect to happen in a range of situations. Actors performing the exhaustive set of behaviors are then filmed. Digital artists then use the videos as props to produce virtual humans with as realistic behaviors as possible to meet the requirements of the clinicians and psychologists using a state of the art procedure to animate models in real-time [Takacs et al., 2003]. The animations are then shipped to the designer responsible for the creation of gaze content in a form that can be imported and used in our development environment to create the virtual characters.



Figure 7 Model of a virtual baby to be used in ride.

6 Discussions And Future Work

We have described our work aimed at the design and implementation of a gaze contingent therapeutic environment for autistic children. This project faces hard challenges from a number of fronts. The workflow resulting from the coordination of the various system building tasks is still evolving. We make a lot of effort to get our production method right by applying lessons learnt from the design of popular virtual environments in the commercial arena. It is clear that the quality of the environment will have an impact on our primary goal: to test the new therapeutic technique for autistic toddlers. This evaluation will yield meaningful results if and only if subjects are exposed to the procedure long enough for outcomes to become tangible. Thus, it is evident that our gaze contingent content needs to be sufficiently varied in order to sustain its engaging quality. For example, an overly repetitive, uninteresting environment can quickly become uninteresting to the subject and this may have an adverse effect on the proposed therapy.

The validity of the proposed therapeutic technique cannot be tested unless the instruction-free-toddler-eye-tracking-calibration problem is solved. Thus this paper illustrates our efforts to make the eye tracker calibration process as minimally disruptive to the experience as possible. While we have not achieved total automation of the process, we have automated significant portions and have identified areas that need improvement to make this goal achievable. Attempts to make calibration stress free is a very popular undertaking in the eye tracking community. Some researchers have discussed the use of information such as the likelihood of gaze position just before a GUI element is

manipulated, as a possible means to automate the calibration process [Hornof and Halverson, 2002]. We have used our own observations of gaze responses to dynamic stimuli to implement our custom method. A critical investigation that needs to be undertaken is to verify whether these gaze responses are maintained consistently in most autistic toddlers. The utility of our calibration method will certainly be questioned if this population fails to show predictable gaze responses similar to the kind of dynamic stimuli described in the paper. However, this will not deter us from pursuing other gaze responses that seem typical in a given population by using an approach similar to that described in the paper. The key is to find the right type of visual stimulus that will produce a predictable response. Once this is known, we only have to use this knowledge to inform our calibration procedure. If no generic response is pinned down for the autistic toddler population, we can always chase subject specific ones if any.

So far we have only focused on gaze position as a parameter of interest. Previous research shows that other parameters such as blink rate and pupil size correlate with enjoyment [Yamada, 1998] and mental activity [Polt et al, 1964]. We plan to enhance our software to capture these parameters and to use these additional cues in our inference engine that dictates the behavior of the gaze contingent environment. ISCAN is also in the process of developing very small eye trackers with better servo tracking capabilities. Future versions of our system will certainly benefit from this emerging technology.

7 Acknowledgments

This work is supported by the National Science Foundation SGER (Small Grant for Exploratory Research) and the Assistive Technology Research Center of the National Rehabilitation Hospital. We wish to thank the following: Maya Coleman, Maiseley Paxton and Willie Stewart from the Catholic University of America, Linsey Barker from the National Rehabilitation Hospital, Matt Peterson and Corinna Lathan from Anthrotronix Inc and Barnabas Takacs from Digital Elite for their sustained and significant contributions.

8 References

1. BADLER, N., PALMER M.S., and R. BINDIGANAVALA, 1999, "Animation Control for Real-Time Virtual Humans," *Comm. ACM*, vol. 42, no. 8, 1999, pp. 64-73.
2. BISHOP, L., EBERLY, D., WHITTED, T., FINCH, M., AND M. SHANTZ, 2001, Designing a PC Game Engine: BISHOP L, *IEEE Computer Graphics and Applications*, pg 46-53
3. HORNOF, A. J., & HALVERSON, T., 2002, Cleaning up systematic error in eye tracking data by using required fixation locations. *Behavior Research Methods, Instruments, and Computers*, 34(4), 592-604.
4. LEEKAM, S. and E. HUNNISETT, 1998, Targets and Cues: Gaze-following in Children with Autism, *J. Child Psychol. Psychiat.* Vol. 39, No. 7, pp. 951±962.
5. PEART, K., 2002, Results of autism research may provide a key to determining severity of individual's condition, *Yale Bulletin & Calendar*, October 25, 2002, volume 31, Number 8
6. POLT, J.M., E.H. HESS, 1964, "Pupil size in relation to mental activity during simple problem solving ", *Science*, Vol. 132, pp.349-350.
7. ROUSE III, R., 2000, Gaming and Graphics: Toiling with Tools. *ACM SIGGRAPH Computer Graphics archive*, Volume

8. SCAIFE M. and J. S. BRUNER, 1975, The capacity for joint visual attention in the infant. *Nature*, 253:265-266, 1975.
9. SCHULTZ R. T. GAUTHIER I, KLIN A, FULBRIGHT R K et al., 2000, Abnormal Ventral Temporal Cortical Activity During Face Discrimination Among Individuals With Autism and Asperger Syndrome. *Arch Gen Psychiatry*. 2000;57:331-340
10. SEBRECHTS, M., FINKELMEYER, A., RAMLOLL, R., & C. TREPAGNIER, 2003, AutoCal: Assessing Attentional Capture for Automating Eye-Tracking Calibration. Paper presented at the 33rd Annual Meeting of the Society for Computers in Psychology, Vancouver, Canada, November 6th.
11. TAKÁCS, B., B. KISS, 2003, The Virtual Human Interface: A Photo-realistic Digital Human, in *IEEE Computer Graphics and Applications Special Issue on Perceptual Multimodal Interfaces*, September-October 2003.
12. TREPAGNIER, C., SEBRECHTS, M., & R. PETERSON, 2002, Atypical Face Gaze in Autism. *CyberPsychology & Behavior*, (5)3, 213-217.
13. TREPAGNIER, C., 1998, Autism etiology: A face-processing perspective, *Brain and Cognition* 37(1), 15.
14. TREPAGNIER, C., 1999, Virtual environments for the investigation and rehabilitation of cognitive and perceptual impairments. *NeuroRehabilitation* 12, 63-72.
15. WILLIAM, H., PORTIA, I., and V. S. RAMACHANDRAN, 2001, Autonomic responses of autistic children to people and objects, *Proceedings of the Royal Society London. B* , 268, pages 1883-1888 1883
16. YAMADA, F., 1998, Frontal midline theta rhythm and eyeblinking activity during a VDT task and a video game: useful tools for psychophysiology in ergonomics. *Ergonomics* 41(5), 678-88.

Appendix

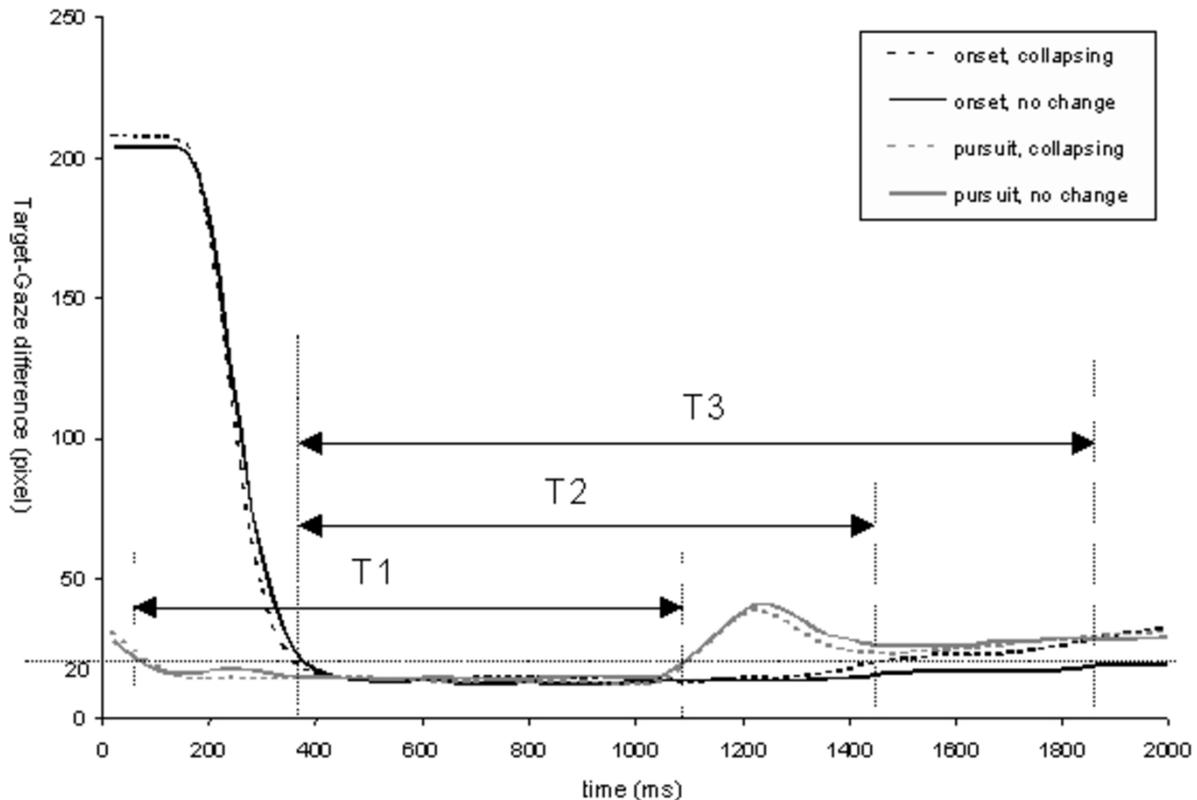


Figure 8 Target-Gaze differences as time unfolds with calibration windows T1, T2, T3