embedded interaction

Institut für Informatik

Ludwig Maximilians Universität München

LMU

Evaluating the Usage of Perception Technologies in Automated Capture and Access Applications

Sebastian Boring

March 2005

Abstract:

In my work, I introduce an automated capture application that provides access to details, such as scenes in a recorded video, of discrete trial training. Since manual recording and reviewing of material already exists, it is an interesting case study for automated capture technology. Acceptance is dependent on minimal intrusion in human activities during therapy and a well-indexed video, thus I leverage several perception technologies that are as close as possible to common therapy practice. I also critically explore the contribution these technologies have on the overall utility of the system.

Keywords:

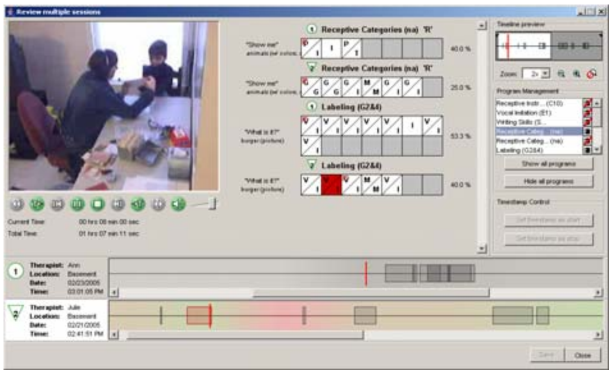
Automated Capture, Perception Technologies

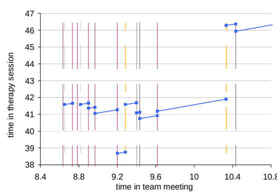
Publications:

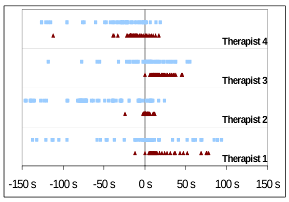
Proceedings of the Seventh International Conference on Ubiquitous Computing, Ubicomp 2005, Tokyo

Contact:

boring@cip.ifi.lmu.de
albrecht@hcilab.org







Supervisor: Albrecht Schmidt
Research Group Embedded Interaction
http://www.hcilab.org

Technical Report No. 004

Table of Contents

1 Introduction

Despite the increasing popularity of capturing activities of everyday life, there are still very few published examinations of real use, intensifying views that automated capture may not be an obligatory capability. Part of the reason for this phenomenon is finding a domain for which frequent and direct access of captured activity is likely as well as finding easily applicable perception technologies to automate capture and annotation of captured media. Thus, identifying a high-need access situation and creating a reliable capture system, exploration of research questions related to capture and access are possible. In this work, I present an example of a domain with one such high-need access requirement-evidence-based behavioral and academic interventions for developmentally disabled children. I will address how automated capture and access to rich media impacts users in the selected domain, specifically, a team of collaborating therapists. Additionally I will determine whether it is worthwhile to employ various perception technologies to annotate captured activity, or whether simpler heuristics for indexing suffice.

An evidence-based approach to intervention therapy attempts to use experiential data of past performance to convey future decisions. For example, importance of determining progress and guiding treatment decisions (particularly evidence that covers a patient's health outside the doctor's or professional's office) is widely acknowledged in the field of medicine. However, there are many situations in education, medicine, and other fields where gathering such evidence is cumbersome, if not even impossible. In previous work, researchers explored the potential of automated capture for the specific case of treating children with autism (CWA) [11]. One example suggested was the support for a team of therapists conducting discrete trial training (DTT), an application of Applied Behavior Analysis methods. An initial capture and access prototype was developed as a technology probe to determine whether practitioners of DTT saw any promise in the resultant prototype. Based on that feedback and the researchers' extensive experiences over the past year with DTT, a small team of researchers (including myself) developed Abaris[1], a complete capture and access system to support home-based DTT therapy. This team's intent is to evaluate the impact of some specific technology decisions on this popular intervention therapy.

This thesis presents the design, initial deployment and results, and evaluation of the Abaris system by paying particular attention to the use of two perception technologies during capture, Anoto's digital pen and paper technology [2] and Nexidia's phonetic-based speech detection [19]. These technologies allow creating a system that is a minimal departure from the existing DTT practice of the considered team of therapists, increasing the chances for adoption. The choice of each technology is strongly motivated by an understanding of the structure and practice of DTT. Preliminary results show that Abaris has been very well received by users, and I will discuss why I think it has been successful. Additionally I will explain, whether the chosen perception technologies played a necessary role in that success. The digital pen and paper were particularly critical to the system. On the other hand, although the speech recognition system did provide a useful indexing service, I imagine that the usage of simpler heuristics would have a similar effect.

The structure of this thesis is as follows. First I will give a concise overview of the specifics of DTT and the structure and communication needs of a therapy team for one child. I will then summarize the goals of the initial study to evaluate an automated capture solution to support DTT. The design and development of the Abaris system with all its relevant features will be described, and then I will present the results of a one-month deployment study,

[1] Abaris was a figure in Greek mythology who was the priest of Apollo and who possessed a golden arrow that, among other things, helped to cure diseases.

focusing on the use of Abaris during routine collaborative meetings of the therapists. Next, I analyze the results of the deployment, with a particular focus on hypotheses about the success of Abaris. Finally, I will discuss the usefulness and costs of the perception technologies as applied to this domain.

2 Related Work

In Ubiquitous Computing, Abaris falls into the category of automated capture and access applications. Several automated capture and access systems from research have helped explore this area, including applications for the classroom such as eClass/Classroom 2000 [5] and for meeting spaces such as Teamspace [21] and Tivoli [20]. Abaris provides users the ability to access information they already access manually (*e.g.*, notes in the classroom and meeting settings) supplemented with additional information to which they would not normally have access. These applications also focus on low need access situations, whereas Abaris is high need. Furthermore, unlike these and other examples, Abaris represents one of the first uses of asynchronous capture of videos for a collaborative access system. NoteLook [7], NotePals [8], and StuPad [24] all allow asynchronous annotation of videos in a collaborative setting, but are not designed for accessing multiple experiences as a collaborative activity. Abaris also differs from other capture and access systems, such as MyLifeBits [9], the Personal Audio Loop [12], Audio Notebook [22], and MIT's personal memory aid [25] in that these are designed for personal use of unstructured live experiences, rather than group access to a structured activity.

Other examples of technology that support the care of developmentally disabled children are not necessarily considered automated capture and access. Many of these technologies are ones with which the child interacts directly, such as Simone Says for teaching speaking language skills [15] and the Discrete Trial Trainer [23], a commercial product aimed at simulated DTT. A commercial product known as mTrials [18], a PDA based data collection system for DTT has enjoyed some success but lacks the additional audio and video inputs and collaborative access interface that Abaris provides.

I also relate this work to other ubiquitous computing projects that have focused on the design, and authentic use, and/or summative evaluation of full-scale systems. PlantCare [14], Labscape [3], mixed reality games [4], tour guides [6], Tivoli, and the eClass system were all designed to be end-to-end solutions to specific domain problems and were deployed for authentic use. All but the first two have published summative evaluation studies. In designing, deploying and evaluating Abaris, I hope to show how acceptance and usability are impacted by specific perception technologies, similar to how Benford *et al.* assess the value of self-reported positioning in the Uncle Roy All Around You game [4].

3 Discrete Trial Training – A short description

As mentioned earlier, the case study for this work is based on a popular intervention therapy for CWA, known as Discrete Trial Training (DTT). It was developed in the 1970's by O. Ivar Lovaas [16] and has evolved as a specific method from the field of Applied Behavior Analysis (ABA) [1]. DTT is currently a best-practice method for teaching basic skills to CWA and other developmental disabilities [13], although it is slightly different from Lovaas' original conception. The basic concept of DTT therapy is that teams of trained therapists have one-on-one sessions with a child to teach basic skills in a structured setting. I will outline the basics of DTT and the therapy team below. Although there are several variations of different DTT practices, the description below is representative of standard practice, as implemented in a home setting.

The discrete trial: Advocates of DTT believe that even children with severe developmental disabilities can learn correct behaviors through controlled and conditioned training. The discrete trial is one example of this learning model. The idea is to first gain the attention of the child and then make a direct verbal request to the child that requires a well-defined and correct response. For a correct response the child is immediately rewarded with a reinforcing stimulus, such as a piece of candy, a favorite toy, or verbal praise. For incorrect responses, the therapist guides the child in a way to ensure a correct response. In this case the trial is immediately repeated, with the therapist providing whatever prompt is needed to guarantee a correct response. The results of the trial are recorded (*I* for an independent or correct response, otherwise, some mark that indicates the used prompting). If a “correction” trial occurs, the therapist will record both, the initial response and the response of the correction trial.

A DTT program: The therapy system for DTT consists of a collection (10 to 20) of programs for which data is collected. Each program consists of the following specific characteristics. It has a basic skill (*e.g.* Vocal Imitation), a target (*e.g.* the word “snake”), a note for further explaining the task (*e.g.* emphasize “sn” in “snake”), and a specific command (*e.g.* “Say <target_name>”) known as S_D.

A therapy session: Each program/target combination is performed a number of times, ideally distributed randomly throughout the 1-2 hour session. All data is recorded on a score sheet and the overall percentage of independent trials is graphed on a separate sheet after the session reflecting that day’s data. Figure 1 shows both, the score sheet and the graph sheet.

Advancing the program: A given program/target combination can be considered as “mastered” if it fulfills certain pre-defined criteria (*e.g.* 80% correct responses on a given) over some interval of time (*e.g.* three consecutive days). As soon as the target is mastered, a new target will be introduced in the next session. If a sufficient variety of targets is mastered for a program, the program is considered mastered overall and will be practiced throughout a therapy session (without data collection).

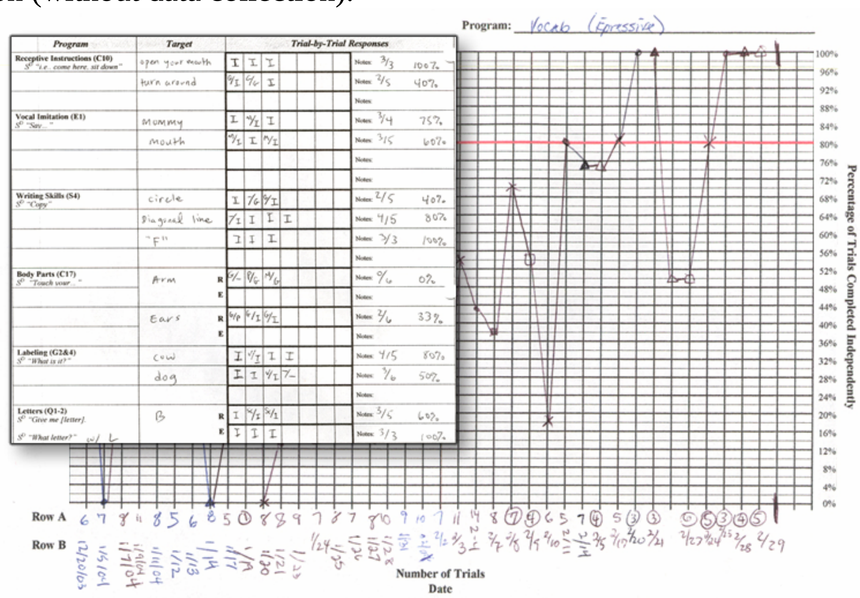


Figure 1: Examples of paper-based forms used by therapists. The left shows the therapy data sheet completed during therapy, and the right shows a graph of the child’s progress for a particular skill.

Before a session, a therapist reviews the child's therapy materials (*e.g.* all the child's past session data sheets, program progress graphs, mastered skills, and narrative notes from other therapist's on the child's progress) which are contained in a notebook. Next, the therapist will prepare the session materials, such as pictures, objects, and writing utensils, and then begin the session by playing and interacting with the child. Finally both, the therapist and the child, are moving to the table to rehearse mastered skills and work on current target skills.

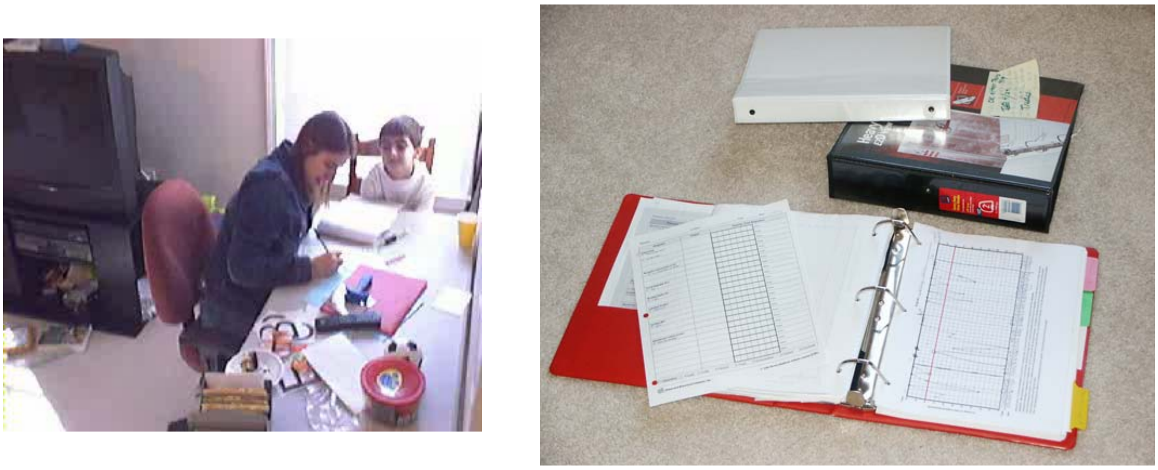


Figure 2: Left shows a therapist engaged in therapy with a child. Right gives an example of a notebook used to store all of the paper-based data.

Evaluating Progress: The team involved in testing Abaris consisted of a parent (trained in DTT but not practicing), three regular therapists, one lead therapist, and a consultant, all providing therapy to a seven year-old, low-functioning child diagnosed with Autistic Disorder (mild to moderate), using the DTT procedure described above.[2] The lead therapist has additional tasks of administrative paperwork, such as determination of which program/target combinations are mastered and scheduling new targets and programs for future therapy sessions, while the consultant is an expert in behavior analysis (but does no direct therapy with the child). To discuss therapy, analyze data, and make necessary adjustments to improve the child's learning ability, the team usually meets every other week. The consultant (who leads the meetings) looks at the manually recorded data, and then asks the therapists for details on how the child is progressing on each skill. If a certain target skill shows no progress and is in place for a long time the team may decide to remove the target and replace it by another one. They could also discuss the reasons for poor performance. In these cases, therapists try to remember what occurred during their session and make hypotheses about what caused the poor performance. After all, the consultant will make suggestions with the team, which are implemented during the next two weeks of therapy. After making these changes, the team reviews the progress again at the next team meeting.

4 Research Goals

With the understanding of this domain, I can now frame the research goals into the design and evaluation of Abaris. I begin with two observations:

[2] Two members of the team were from the research team. All other team members and the child are protected as human subjects under an approved IRB protocol.

- Therapy sessions (though fast-paced and enormously flexible) have a well-defined structure that can be leveraged naturally by perception technologies, potentially providing a suitably indexed recording for later access.
- Team meetings represent a high-need example of access, in which the users who are both capturing and accessing the data absolutely require it to perform their jobs. Furthermore, the meetings consist of a lot of self-reported reflections on past experience between therapist and child, which is a clear opportunity for improvement with real evidence of what occurred during a therapy session.

Even though DTT therapy is one of the most structured and successful treatments available for CWA, there are elements in the process that may lead to inaccuracies in the interpretation of the data, which leads to less efficiency of the overall therapy. These include miscalculation of percentages of trials completed independently as well as inaccurate graphs. Therefore, the goals for Abaris were to be able to address some of these issues and make therapy a better and more useful experience for the therapists and the child.

Because discrete trial training is a structured activity, it is particularly well suited to the use of automatic capture technologies. Additionally, therapists are already trained to be cooperative in the process of manually recording data and are highly motivated to use anything that will save time on laborious paperwork. The paperwork (*e.g.* recording, calculating and graphing all of the data with pen and paper) is very time consuming (often requires one third of the session) and has a high likelihood that the data may be inaccurate due to simple human error. Doing the paperwork is also a source of annoyance as expressed by caregivers and therapists. Due to its tendency to take the therapist's attention away from the child, many caregivers choose not to record data at times, losing valuable evidence and diagnostic information. By designing a system that automates much of this hand analysis and calculation, both the amount of time spent in paperwork and human errors can be reduced.

Abaris was designed as a highly flexible application, because a significant amount of improvisation is required by DTT. Because the system should as close as possible to current therapy practice, Abaris provides pen and paper techniques for collecting data. As noted by Mackay *et al.* in their study of air traffic controllers [17], keeping pen and paper will increase the likelihood of acceptance by the therapists. Thus, the challenge is to design the capture interface in a way that maximizes the inherent structure of the sessions without violating the process. Additionally, an access interface needs to be designed that allows and encourages exploration of evidence without requiring too much time, effort, and distraction during team meetings.

At current team meetings, therapists sometimes speculate about different aspects of therapeutic sessions, such as how a child is responding and why, levels of attention exhibited by the child, or individual characteristics of each therapist's style and work while conducting each trial. These speculations usually appear when there are discrepancies in the child's performance across sessions and across therapists. In addition to trends in the actual performance of child and therapist, discrepancies in the grading of the child by multiple therapists also tend to interfere with measures of progress. These incongruities can result in a mismatch in skills taught and the child's abilities that can then lead to frustration for both, the child and the therapist. Therefore, capture of rich data (*e.g.* video) allows therapists to *see* what the therapists are discussing without being present during the actual session. Thus, the overall goal for the access interface is to provide a means to facilitate those discussions amongst groups of therapists about trends in data using easy access to both, empirical and rich data, to enable data-based decisions for long-term use.

5 The Abaris System

Before I started my work on this particular project, a senior design team created a first prototype of Abaris as shown in Figure 3. During this implementation phase, several perception technologies were been tested to get timestamps of each trial. I will describe below how time stamping and grading a trial were implemented and whether or not they are now deemed to be useful enough to counter the development effort required.

One technique envisioned was to use a tablet PC for entering grades. Instead of using pen and paper, therapists were able to enter grades through this interface where those were time stamped automatically. The interface was connected through a wireless connection, as shown in Figure 4, and therefore synchronized with the server's system clock. However, this technique turned out as not being useful since it was too far apart from the therapist's current methods.

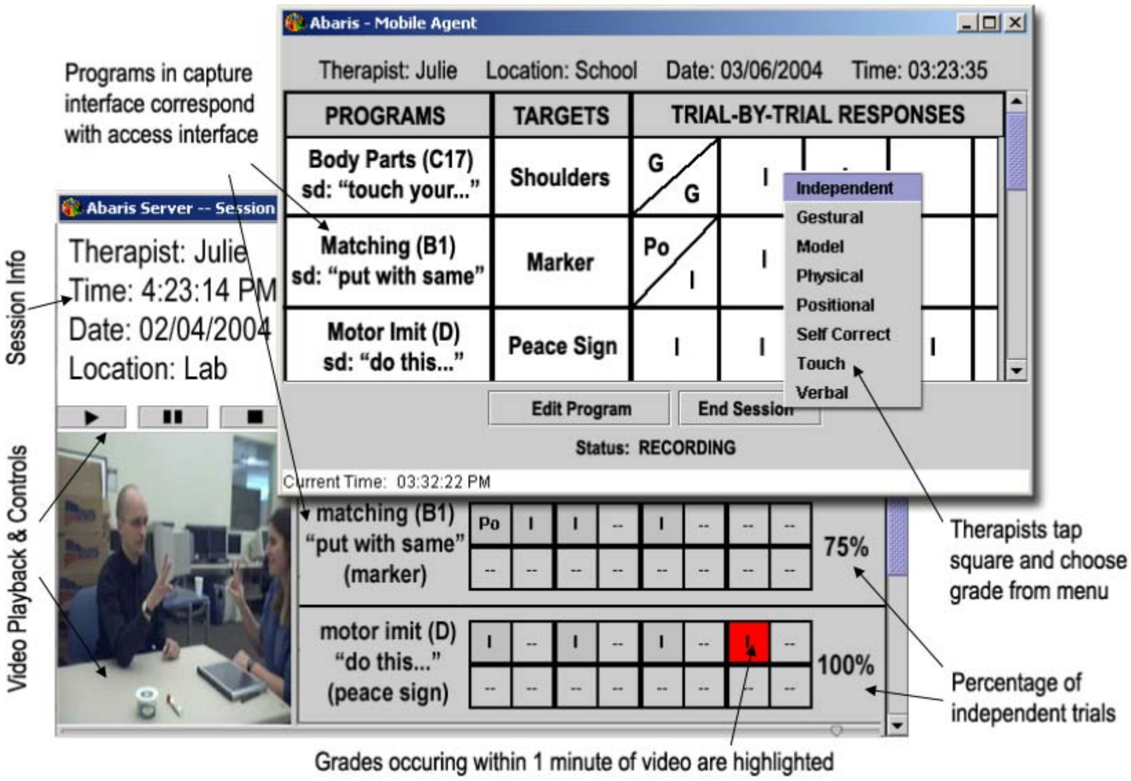


Figure 3: Users score performance data by choosing a value for each trial. They can replay the entire video of a session or go to salient points using discrete data. (In 'Hayes, G. R., Kientz, J. A., Truong, K. N., White, D. R., Abowd, G. D. and Pering T., "*Designing Capture Applications to Support the Education of Children with Autism*"' [11])

Figure 4: Basic system setup for the first implementation of Abaris using a webcam and a tablet PC.

Another consideration was a vision-based solution in which therapists used a simple two-finger gesture on a score sheet to indicate the beginning of a graded trial and the actual grades before and after the trial itself. Though the approach was simple to teach and the vision problem was feasible, researchers found that therapists could not remember to do the gestures at the correct times, resulting in a loss of grading information. Furthermore, the therapists noted that the child was distracted by this extra motion, and the increased cognitive load on the therapists during work was frustrating. Thus, this solution was not usable for Abaris' purpose.

5.1 Design Decisions

As noted with the failure of the first vision technique, I had to consider a variety of factors other than technical feasibility in choosing which perception technologies to use. Thus, I first analyzed the characteristics of a trial to find technologies that are able to figure out the exact timestamps. They should also be not very deviant from the current practices of DTT therapy. Below I will start to describe the system design and the used technologies.

5.1.1 System Design

When I examined the specific characteristics of the trial structure, I concluded that using Nexidia's voice recognition technology (off-the-shelf, phoneme-based speech system) for the "beginning of trial" timestamp is appropriate. Each trial begins with a particular command (*e.g.* "Do this"), which is known before the actual session begins. In using Nexidia, the only change therapists would be forced to endure is to wear a wireless microphone with which the system detects the commands and thus provides the timestamp.

Furthermore, trials should always be graded, with the grade noted on a paper form, close to the end of the trial. By replacing the standard paper with Anoto's digital pen and paper technology, the therapy practices are not changed at all, but the digital pen affords collection of positions and timestamps of every single stroke on the paper. Using a paper-based but enhanced solution preserves the flexibility inherent to writing.

As shown in Figure 5, the system is supplemented by additional devices, such as a high quality duplex printer for augmented datasheets and a web cam for capturing video and audio data.

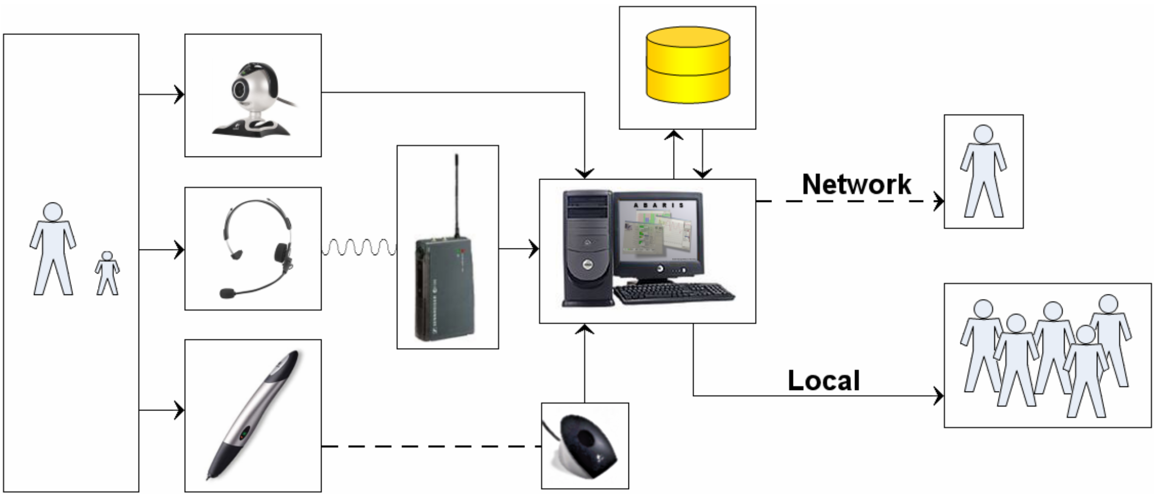


Figure 5: This shows the basic system setup to run the Abaris capture interface.

5.1.2 User Interface Design

After deciding what input methods to use, I designed the access interface to allow quick and easy access to the captured rich data. This application needed to be as close as possible to the therapist's meeting practices because of high likelihood of adoption. The common practice (as described above) uses the graphs first and will go into depth after identifying problems and abnormalities in such a graph. Therefore, the application is designed in the same way. First, an overview of all the data is given and a particular scene can be selected by click afterwards.

Initially, sketches were created to show the ideas and layouts to the therapists. Additionally, core functionalities were discussed to figure out, if the application will meet the therapists' needs. In this phase, therapists mentioned the requirement of network accessibility to reconfigure and/or change data as well as accessing video before a meeting to get a better overview of the meeting's agenda. After two iterations of this process, I was able to start implementing the access side as well as the capture side in November 2004 to permit deployment in February 2005.

5.2 Implementation Details

After defining the layout and functionality of the new Abaris system, I needed to determine which programming languages to use. Because the original system was written in Java and the Java Media Framework (JMF), I decided to use those as well. At the same time, Nexidia's voice recognition technology in the capture interface required to use C/C++ since it was written in these languages. Below I will describe the implementation for each single interface more detailed.

5.2.1 Capture Interface

To give therapists the ability to access videos and certain scenes quickly, indexing and segmentation needs to be done during capturing (see Figure 6 for basic parts of the capture interface). First I will describe every technology separately and then I will illustrate how the different components are working together to meet the requirements.

Anoto's digital pen and paper technology: As described above, Anoto's digital pen technology provides the absolute end timestamp of each single trial. These times need to be synchronized with the video (*e.g.* seconds after the start of the video) after recording is done. Additionally, every single stroke on the paper is provided and therefore, the system knows which cell includes strokes and written grades respectively. Unfortunately, this technique holds several challenges, such as therapists tapping on the paper while thinking about the grade or lines being very close to a cell's border or even a stroke within two cells. The mentioned problems will lead to errors if they are not handled.

My solution is an algorithm that does not value every stroke it gets from the plain-text XML file generated by the digital pen's interaction with the paper. A stroke, by definition, contains at least 6 pixels and more than half of its points inside the 31x20 pixel cell the system is analyzing.

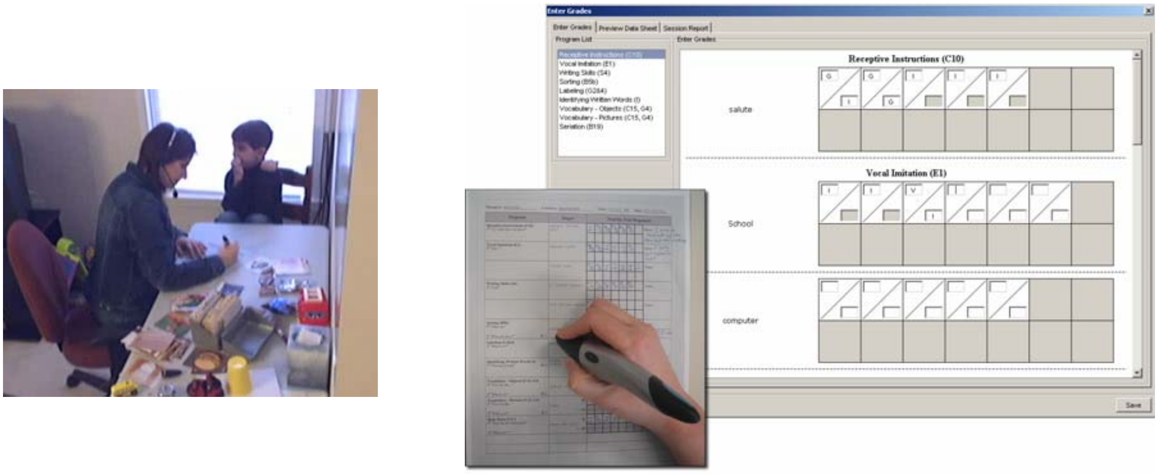


Figure 6: Left shows a scene from a therapy session using the Abaris system. Center shows the digital pen and its specially designed paper used for entering trial grades. Right shows the current system does not use any character recognition, instead providing an interface to enter grades.

Nexidia's voice recognition technology: After evaluating the end timestamps, the system is able to seek out the beginning timestamp within a vector of "guesses" for each trial generated by Nexidia's voice recognition technology. These guesses contain a confidence level (percentage) and a time offset after the start of a session (hundredths of a second). Because several different programs can have the same voice command (*e.g.* expressive tasks for objects and pictures are considered as two different programs, but have the same S_D *"What is it?"*), the system needs the end timestamp to be able to narrow the search for the beginning of a trial.

Nexidia allows two types of identifying voice commands, "searching" and "indexing". The difference between these techniques is the time of actual identifying the commands. "Searching" constructs the pattern file after recording the wave file, while "indexing" is producing this file during recording. Indexing can only be used if the patterns are known prior recording and performing the session which is granted in the application. Additionally, "searching" needs about 20 minutes to evaluate a 1.5 hour wave file and therefore is time consuming and not necessary for Abaris.

A "guess" is considered to be "useful" if its confidence level is above 50 percent and its timestamp is within a time window of two minutes prior the end timestamp. Usually a trial does not last for 2 minutes, but also the grade is not written down immediately after the trial. Additionally, the system must ensure that two trials don't have the same beginning timestamp. Finally, there is the fact that Nexidia is written in C/C++ which needed a link through Java's Native Interface (JNI).

Combination of technologies: To further explain the interactions of all the components, I will give a short description of the process from the computer's view. After starting the application, all relevant data will be queried from the MySQL database, which is connected through Java's Database Connectivity API (JDBC). Next, the therapist selects the type of session ("Regular" or "Maintenance") and has the ability to look at previous sessions' therapy notes and to print the datasheet for the actual session. During viewing of the datasheet, the paper's cells will be stored with their associated programs, targets and trials. Once the printing is done, the session can be started with a single click where all needed speech patterns for Nexidia will be set. Additionally, each codec is being applied to its respective source. The following codecs are being used for capturing the session:

- Video codec (webcam): H263, 352 x 288 pixel, YUV encoding, 15 fps
- Audio codec (webcam): GSM_MS, 16.000 Hz, 8 bit, mono
- Audio codec (Nexidia): WAVE uncompressed, 44.100 Hz, 16 bit, mono

After finishing the session and its recording, the video and audio streams are stored immediately, allowing for the possibility of recreating a session in case of errors or exceptions. After, Anoto's results are used to create the trials and provide them with their associated end timestamps. Then, Nexidia's results are read and, starting with the last trial, filtered for their program and target specific speech patterns. All "guesses" are saved in the database as alternatives to allow later corrections and evaluation of the technology while the best possible "guess" is used to set the start timestamp. Finally, the therapist manually enters every trial's grades and writes the therapy notes, which are narrative accounts of behavior and other factors.

5.2.2 Access Interface

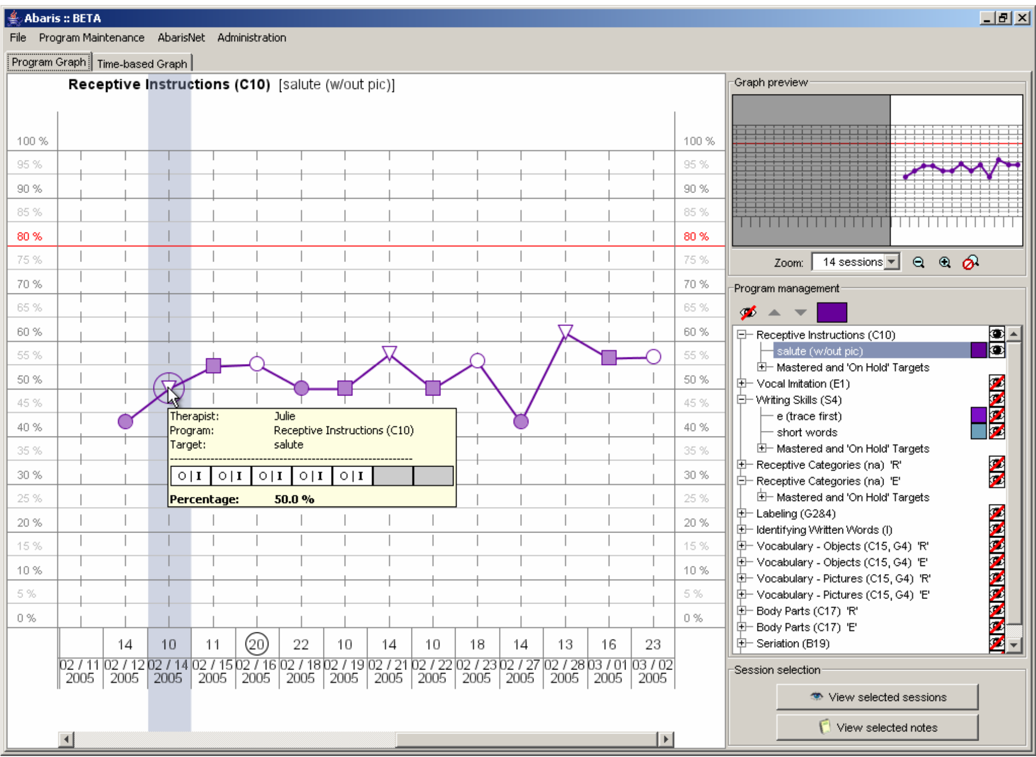


Figure 7: This shows the main access interface displaying a single selected graph on the left, with tool tip indicating information for a specific session. The right shows a view of the entire graph and the list of selectable programs.

The Abaris access interface allows users to review saved and indexed data as well as correct grades and timestamps for places where technical or human error created incorrect data. Therapists need to perform those tasks both locally, at the site of therapy, and remotely, from their homes or offices in preparation for team meetings and therapy sessions. Furthermore, they must be able to access Abaris both individually and in a group setting during team meetings. Of course, Abaris must provide at least the same level of functionality as the traditional pen and paper process, including graphing of empirical data and review interfaces for therapist datasheets.

In addition, several components that make use of information visualization techniques have been added, such as overlaying multiple graphs to detect similarities and/or varieties as well as tool tips to differentiate lines and gain quick access to session information (*e.g.* performance within one program). Within this representation, selecting (multiple) sessions is possible. Figure 7 shows the main part of the access interface.

Program modification: Another interface has been built to modify programs and/or targets. Programs can be separated in “active” or “inactive” states, whereas trials can have three different states: “regular”, “maintenance” or “future”. Finally, programs and targets can be moved to logically group them on the datasheet.

Video review: One major section of the access application is the video access interface (as shown in Figure 8), which allows reviewing a single or multiple sessions. Because reviewing a single session is just a special case of reviewing multiple sessions, I will only describe the first case.

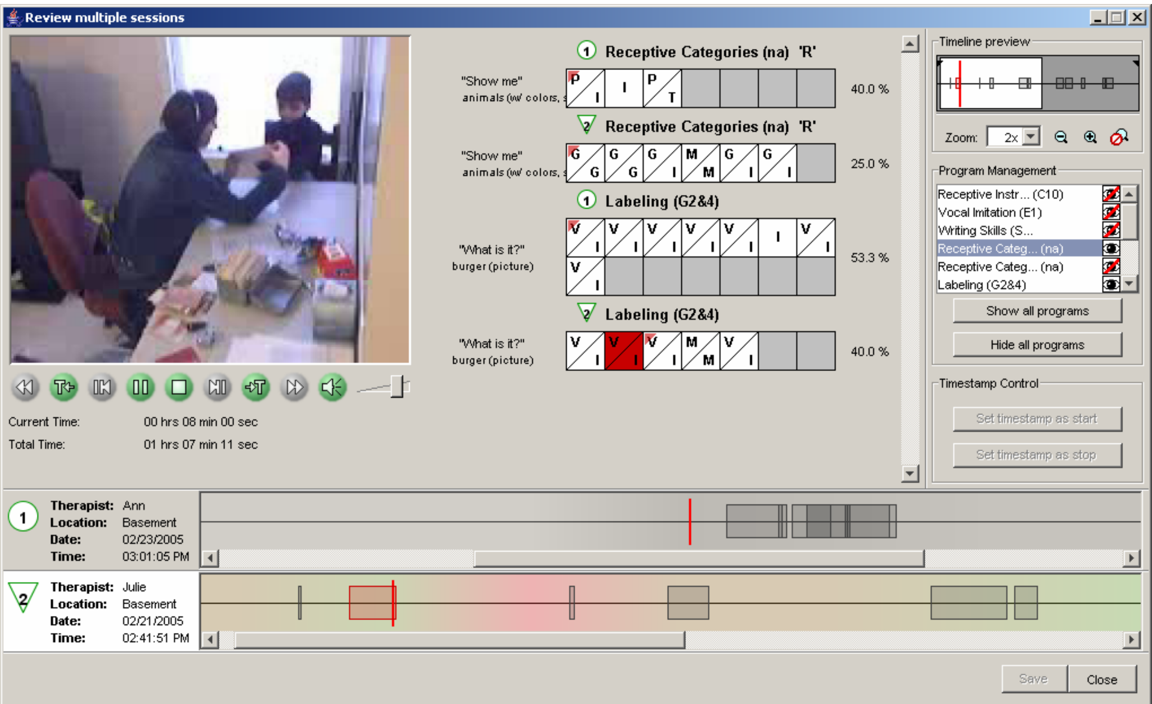


Figure 8: This shows a session browser set up to view two different therapy sessions. The bottom shows the two different timelines, and the left shows the grades for the different trials.

In addition to the current video, this interface visualizes several other data, such as the timeline for the current active video, the selection of programs and the visualization of every single program with its targets as known from the analog datasheets.

The interface provides a timeline for every loaded video to allow switching quickly between videos by using a single click. Every timeline itself has a performance indicator showing the percentage of independently completed trials within a given time interval.

Additionally, all trials of every currently visible program are displayed as rectangles to determine their start and end timestamps. The timeline also allows to correct timestamps of trials, which is useful for later evaluation of the used technologies.

The common visualization of programs and their trials gives more functionality, such as indicating the current running (if applicable) and the next upcoming trial within a program. In here it is also possible to change grades if inconsistencies and/or erroneous grades occur.

5.2.3 Network Interface

Because therapists need to access the child's data remotely from their homes or offices, a network interface (AbarisNet) has been built within the access interface. It provides almost all the same functionality as the local interface. To access data remotely, I needed to create a server inside the local implementation of the access interface. The server's attributes, such as users and their access rights, can be configured within the access interface. It will also be used to create sets of ports for streaming of video and audio.

Data stored in the MySQL database can be accessed directly (although user verification will be done by both, server and database). Additionally, the user can select the period of time he/she wants to review, to avoid long download times. For this reason, I decided to use streaming of video and audio data (file sizes of about 300 MB per video). Streaming also provides security as well as avoiding redundant data on several machines. Since the whole application is written in Java, I used the Real-Time Transfer Protocol (RTP) as defined by JMF's RTP API. In addition, I reduced the file size by recompressing the video files to a smaller resolution (in this case 176 x 144 pixels) which causes a file size of one fourth of the original file size.

RTP defines ports in a certain order. The server must define and create separate streaming sessions for video with an audio track. Every stream needs two ports, one for the actual media data (even numbers), and one for communication and synchronization between those (odd numbers). Additionally, the server defines two more ports for communication between itself and the client, such as changing video position or videos, starting, pausing and stopping. Because reviewing multiple videos needs to be possible within the remote interface as well, the following equation demonstrates how to determine the number of ports p for n videos:

$$p = 4 \cdot n + 2;$$

The client has to request the port set (defined by first and last port) from the server. The server itself will assign the ports to the client and will block them from being used for other clients' requests until the client has ended its session. The additional communication between client and server is implemented in stream control messages that include the track number, the requested function as well as a value (*e.g.* the new time to be set) if applicable. The server needs to execute those functions, because the client only has virtual control over the streams. Figure 9 shows the communication protocol between the applications.

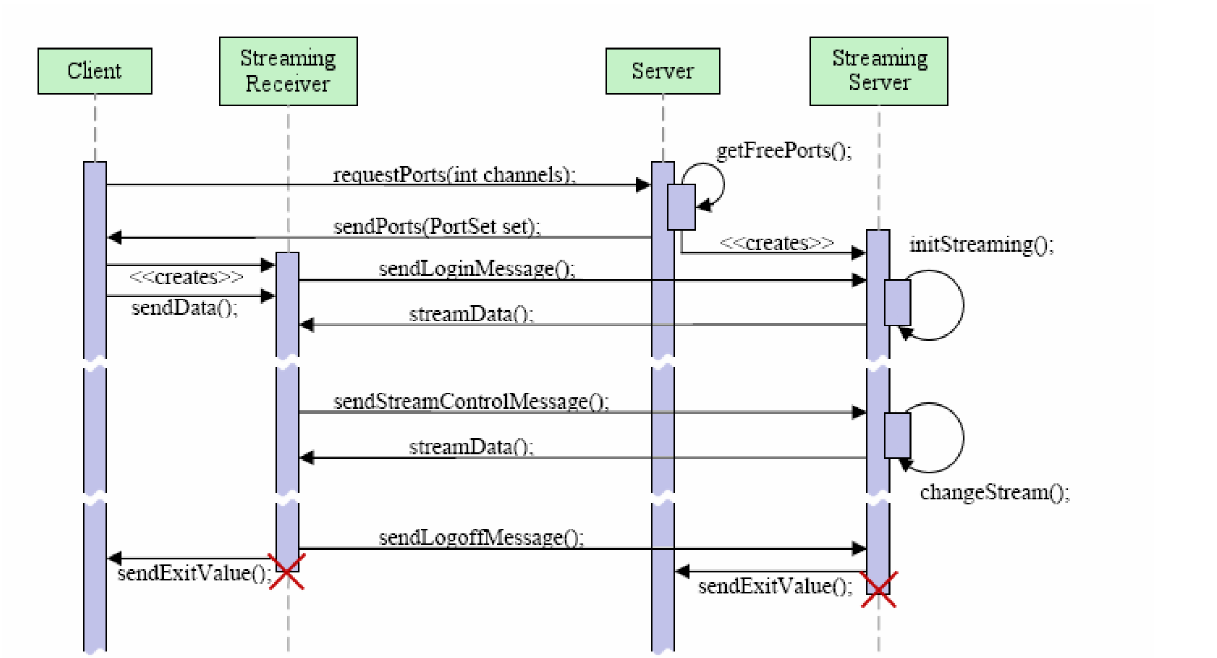


Figure 9: Shows a sample communication between client and server to establish and run a streaming session. As indicated, both, server and client create a sub process as described in 4.2.4.

5.2.4 Improving the System

During implementation of the access interface, I discovered several problems with the original design. For example, every dialog that has been opened by the application allocated memory (about 50 MB) which has not been released after closing the window. Thus, after opening four consecutive video windows, no main memory was available for any new windows and the application ran out of memory. To avoid this issue, I re-implemented the video windows as stand-alone processes which communicate via Java's Remote Method Invocation (RMI) with the main application (as shown in Figure 9).

I also changed algorithms for detecting current and upcoming trials by using more efficient sorting algorithms to be able to use as many resources as possible for video playback. Additionally, repainting algorithms (for both, timeline and graphing window) have been improved to only repaint necessary regions.

6 Results of User Studies

The Abaris system described above has been deployed in one child's home for his team of therapists to use for one month. Meetings to discuss the progress of the child occurred during the second and third week of deployment, and a researcher was on hand during their first sessions in case they had any questions. Before the first meeting, the lead therapist and the consultant were trained on the access interface. During the meetings, researchers were also present to answer questions. In the first meeting, one of the researchers controlled the access interface according to the therapists' requests. In the second one, the lead therapist was comfortable enough to take control.

6.1 Use of Capture System

So far, the team has captured 25 sessions, consisting of 1615 trials and 21.4 hours of recorded data, including every session that has taken place during the study. The capture interface appeared to be easy for the therapists to learn, because the digital pen allowed them to perform their work in the exact same way they had done it before. Although the interface appeared to be easy for therapists to use, they initially demonstrated skeptical attitudes about its use. Despite the skepticism, participants used Abaris in all of their sessions for which it was available. The only benefit of use at this stage was removing the need for users to "hand graph". This consistent use is remarkable given that at first all users were contributing to this groupware system while receiving little benefit [10]. At the first meeting that made use of Abaris, participants were then able to experience the benefits of access. The team of researchers believes this willingness of therapists to contribute on the capture side before receiving benefits was due in large part to the conscious effort during design to maintain nearly identical work practices that reduce or maintain the same level of effort.

Therapists reported allocation of session time both before and during the deployment. Overall, work time for these hourly employees decreased slightly, but time saved not graphing data was redistributed to time spent teaching or playing with the child (see Figure 10). With Abaris, therapists can devote a greater percentage of their paid time to interaction with the child, either in the form of therapy or play. Shorter sessions overall also have the benefit of cost savings for the parents.

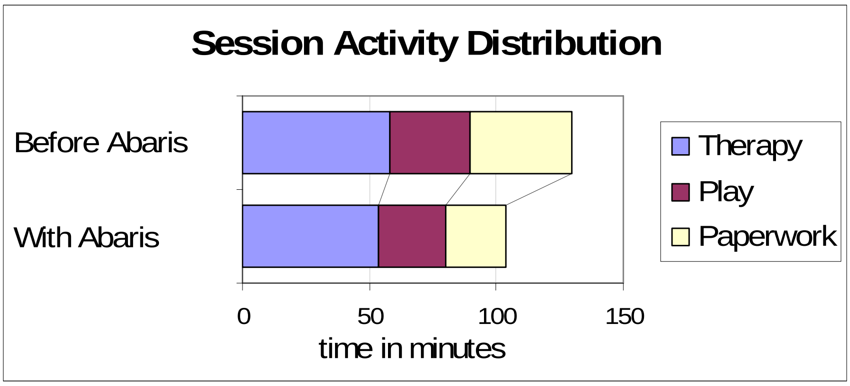


Figure 10: Comparison of the activity makeup of each session before and after deployment.

Two therapists reported that the clip-on microphone was a bit too heavy for some of their typical clothing and could be uncomfortable. Most preferred to use a head mounted boom microphone. A few incidents occurred in which the child became fascinated by the microphone and would reach out and play with it, a behavior that typically occurs when therapists wear jewelry the child finds interesting. Although this behavior can be common for some CWA, it may not happen in all cases.

Simple usage errors sometimes had large impact. One of the therapists forgot to press the record button at the beginning of her session, resulting in no video for the session. In one incident, placement of the Anoto paper in the printer backwards resulted in incorrect detection of the timestamps. These errors can be prevented, but because of its improvisational nature, I could not predict all of the exceptions to the therapy. For example, the lead therapist wanted to change the success criteria for one type of program, but she had no way of doing this with the current interface. Basing Abaris on pen and paper input allowed for a lot of improvisation, but it was very difficult to plan and address all cases.

6.2 Use of Access System

Therapists used the access interface for discussion during two meetings, which lasted 2.5 hours and 1.5 hours, respectively. Each meeting was video recorded and observed, and afterwards the therapists were debriefed on the experience with the system, in which discussion was similar to that of a focus group. Between the first and second meetings, I instrumented the access interface so that we could produce logs of its use providing some empirical evidence of access behaviors. In the second meeting, the team used the access interface to view the video six times, and video viewing took up 20.4 percent of the meeting time. Visualizations of interesting data in these logs are present in Figure 11. The top graph is a typical example of comparing a program across two therapists viewed by the lead therapist before the meeting. The middle graph shows various artifacts in the interface - the timeline and the trial grades - were used to navigate to the desired portion of video. The bottom graph is a detailed version of a portion of the middle graph. That this kind of browsing occurred six times during the meeting is an indication that the team found the value of viewing video outweighed the cost of finding the appropriate session. For 18 months prior to these two meetings with Abaris, the team had access to digital video recordings of the sessions at the site of the meeting, and not once was a segment viewed during a meeting, reportedly because it took too long to find a relevant clip.

Due to the complexity of the data for this type of therapy, the therapists reported the access interface to be complicated at first. They received two hours of training before they expressed enough comfort to use it on their own. Although the ease of use was not as good as I would have liked, the therapists reported they thought the benefits to the system were worth the time it took to learn the access interface. Additionally, the access interface is intended for expert users (*e.g.* the lead therapist and the consultant), allowing them to use the system with all of their clients once they are past this initial learning curve.

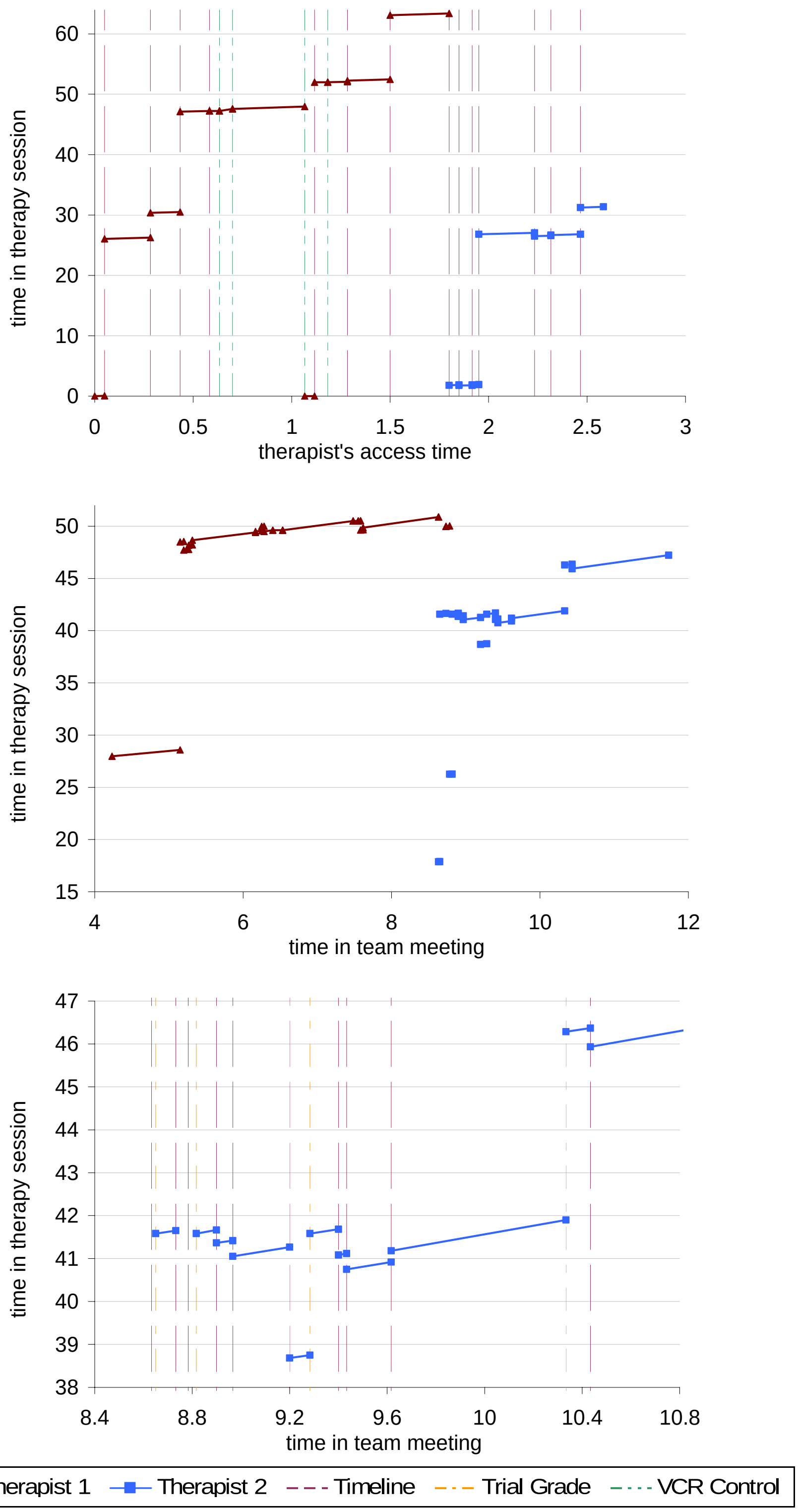


Figure 11: Visualizations from the logs of the second meeting of the team. Top shows access from the lead therapist before the meeting while the center graph shows access to videos during the meeting. The red (with triangle data points) plot show access to video of one therapist, while the blue (with square data points) plot shows viewing a different therapist. In the bottom graph, I expand one segment of the top graph (between 8.4-10.4 minutes into the meeting) and show how different artifacts are used to facilitate navigation.

7 Discussion

The fact that Abaris is considered a useful system by its target user group is encouraging, but as researcher and designer, I wanted to better understand which features contribute to its usefulness and which do not.

7.1 Do perception technologies make a difference?

The integration of trial time predictions and the recorded video are a reasonable first guess at the success of Abaris. As seen in Figure 11, skimming to an appropriate portion of the video was quick enough to encourage use. End times of trials were equated with the time the grade for that trial was written on the Anoto paper. Beginning trials were estimated based on suggested locations of the appropriate verbal command. Four separate therapy sessions have been selected, one for each therapist, and Abaris has been used to create "ground truth" timestamps for the beginning and ending of each trial by manually noting when trials began. Figure 12 shows the error distribution of prediction versus ground truth. A negative error indicates a time prediction earlier than ground truth, and a positive error indicates a prediction after ground truth.

Figure 12: On the left, paired error distributions (in seconds) for Anoto-predicted end of trial (red triangles) and Nexidia-predicted beginning of trial (blue squares) for four of the programs used in the deployment. On the right, error distributions are shown for one session of each of the four therapists (lead and three

For each of the programs, the error distribution of the Anoto predictions is much narrower than that for Nexidia. The Anoto predictions occurred temporally after the actual end time, as expected, because trials should be graded after they occur. The distribution of errors for Nexidia is wider. When viewed grouped by therapist, these error distributions have substantial variation in practice between therapists. Therapist 2 had Anoto predictions that were very tightly bunched near the actual end of trials. This therapist followed the practice of writing trial grades right after the trial was performed, as opposed to other therapists who ensured delivery of a reinforcing reward first. This is actually considered good practice for DTT, and Abaris benefits from this practice.

The phoneme-detection of Nexidia, and my accompanying algorithm for assigning assumed beginning of trial times, produced a significant amount of error. Errors are not surprising, given the nature of the therapy, with graded and mastered trials often having the same spoken command and occurring in rapid succession. However, because the interface was still usable, as reported based on use during team meetings and the overwhelming positive reaction of the team of therapists in discussions, this error may not be limiting. If this size of error makes no discernable difference, I hypothesize that speech detection may be unnecessary if I can find an alternative approach that introduces no additional errors.

Unfortunately, voice recognition only provides a best guess for the beginning of a particular trial, because many trials for which grades are not recorded use the same spoken command. For example, a therapist may be grading a child's ability to mimic hand clapping, for which the spoken command is "do this" coupled with the therapist modeling hand clapping. Prior to this trial of interest, a therapist may ask the child to perform any number of other activities with the same command of "do this", and then end with the final request "do this" while hand clapping. Thus, I developed a simple algorithm for determining the most likely beginning of a trial based on a combination of the time that the trial likely ended (from the Anoto data) and the time that different spoken commands were used.

7.2 Use of Alternative Methods

Considering the narrow distribution of the Anoto errors for trial endings in Figure 12, there are several suggestions for potential temporal heuristics that might produce begin trial estimates at least as good as Nexidia. I have anecdotal evidence that for Therapist 2, a fairly reliable heuristic was a function of the program type and whether or not a correction trial was needed. The current results give me confidence that there is an upper bound on the error distribution for estimating the start of a trial, and I will experiment with a variety of algorithms to find one that is both accurate and precise enough without impinging on the therapy itself.

8 Conclusions and Future Work

I presented my work on the design, development, and deployment of the Abaris system for supporting therapists who do discrete trial training therapy for CWA. The results show that initial therapist reaction is largely positive which I attribute to the closeness of the system to the therapists current practice. Practice has been improved by allowing therapists to spend less time in paperwork and more time in the therapy itself. Also, a team of researchers evaluated the usefulness of the recognition techniques employed by Abaris by comparing its accuracy with the ability to find the needed video segments. Time stamping using Anoto digital pen technology was useful for this practice, and while the errors were introduced by using Nexidia voice recognition, the indices to the video still were useful in practice. Thus, I hypothesize that you could do just as well as the voice recognition with a trained heuristic based on therapist and program type.

Although my initial results are promising and have lead to some interesting insight, there is still room for exploration with this technology. We have submitted one paper for publication to Ubicomp this year, and as of the time of writing, Abaris is still being used by the team of therapists. I plan to continue supporting use of the system over the next few months to start to look at some of the long-term uses of the system and evaluate its abilities to affect the dynamics of the team of therapists. Additionally, I am continuing to look for more ways of improving the system and adding features that can make Abaris even more useful to the therapists. Plans for the Abaris system include more visualization of data that otherwise wouldn't be possible with the paper system, using it as a test bed for more recognition, automation, and multimodal interaction techniques and finding ways of sharing a child's therapy information with all those interested in his progress, not just those present at therapy meetings, such as through a web based information portal. Lastly, researchers are hoping to use Abaris to contribute to the field of autism interventions by enabling domain experts to analyze the science behind DTT therapy itself and improve on its methods.

9 Acknowledgements

I would like to thank the entire ubicomp team at the Georgia Institute of Technology. Special thanks are going to Julie A. Kientz, Gregory D. Abowd and Gillian R. Hayes for helping and supporting me during the whole project. Also, I would like to thank Albrecht Schmidt from the University of Munich for giving me the opportunity to work at the Georgia Institute of Technology.

References

1. Alberto, P.A. and A.C. Troutman, *Applied Behavior Analysis for Teachers*. 6th ed. 2003: Prentice Hall.
2. Anoto, Inc., *http://www.anoto.com*. 2005.
3. Arnstein, L., et al., *Labscape: A Smart Environment for the Cell Biology Laboratory*, in *IEEE Pervasive Computing Magazine*. 2002.
4. Benford, S., et al. *The error of our ways: The experience of self-reported positioning in a location-based game.* in *Ubicomp 2004*. 2004. Nottingham, UK: Springer-Verlag.
5. Brotherton, J.A. and G.D. Abowd, *Lessons learned from eClass: Assessing automated capture and access in the classroom.* ACM Transactions on Computer-Human Interaction, 2003.
6. Cheverst, K., et al. *Developing a context-aware electronic tourist guide: some issues and experiences*. in *SIGCHI conference on Human factors in computing systems*. 2000.
7. Chiu, P.I. *NoteLook: Taking Notes in Meetings with Digital Video and Ink*. in *ACM Multimedia*. 1999. Orlando, FL.
8. Davis, R.C., et al. *NotePals: Lightweight Note Sharing by the Group, for the Group.* in *CHI 1999*. 1999. Pittsburgh, PA.
9. Gemmell, J., et al. *MyLifeBits: Fulfilling the Memex Vision*. in *ACM Multimedia '02*. 2002.
10. Grudin, J., *Groupware and Social Dynamics: Eight Challenges for Developers.* Communications of the ACM, 1994.
11. Hayes, G.R., et al. *Designing Capture Applications to Support the Education of Children with Autism*. in *Ubicomp 2004*. 2004. Nottingham, UK.
12. Hayes, G.R., et al. *The Personal Audio Loop: Designing a Ubiquitous Audio-Based Memory Aid*. in *Mobile HCI 2004*. 2004. Glasgow, Scotland.
13. Heflin, L.J. and R.L. Simpson, *Interventions for Children and Youth with Autism: Prudent Choices in a World of Exaggerated Claims and Empty Promises. Part I: Intervention and Treatment Option Review.* Focus on Autism and Other Developmental Disabilities, 1998. **13**(4): p. 194-211.
14. LaMarca, A., et al. *PlantCare: An investigation in practical ubiquitous computing systems*. in *Ubicomp 2002*. 2002. Goteborg, Sweden.

15. Lehman, J.F. *Toward the use of speech and natural language technology in intervention for a language-disordered population*. in *Third International ACM Conference on Assistive Technologies*. 1998.
16. Lovaas, O.I., *Teaching Developmentally Disabled Children: The Me Book*. 1981, Austin, Texas: Pro-Ed.
17. Mackay, W.E., et al. *Reinventing the Familiar: Exploring an Augmented Reality Design Space for Air Traffic Control*. in *CHI 1998*. 1998. Los Angeles, CA.
18. Mobile Thinking, Inc., *mTrials*. 2002: San Diego, CA.
19. Nexidia, Inc., *http://www.nexidia.com*. 2005.
20. Pedersen, E.R., et al. *Tivoli: An Electronic Whiteboard for Informal Workgroup Meetings*. in *ACM INTERCHI 1993*. 1993. Amsterdam, The Netherlands.
21. Richter, H., et al. *Integrating Meeting Capture within a Collaborative Team Environment*. in *ACM Conference on Ubiquitous Computing*. 2001. Atlanta, GA.
22. Stifelman, L.J., *The Audio Notebook*, in *Media Laboratory, MIT*. 1997.
23. The Discrete Trial Trainer, *http://www.dttrainer.com*. 2004: Columbia, SC.
24. Truong, K.N. and G.D. Abowd. *StuPad: integrating student notes with class lectures*. in *CHI 1999 Extended Abstracts*. 1999.
25. Vemuri, S., et al. *An Audio-Based Personal Memory Aid*. in *Ubicomp 2004*. 2004. Nottingham, UK.