A^3: HCI Coding Guideline for Research Using Video Annotation to Assess Behavior of Nonverbal Subjects with Computer-Based Intervention

JOSHUA HAILPERN, KARRIE KARAHALIOS, JAMES HALLE,
LAURA DETHORNE, and MARY-KELSEY COLETTO
University of Illinois

HCI studies assessing nonverbal individuals (especially those who do not communicate through traditional linguistic means: spoken, written, or sign) are a daunting undertaking. Without the use of directed tasks, interviews, questionnaires, or question-answer sessions, researchers must rely fully upon observation of behavior, and the categorization and quantification of the participant's actions. This problem is compounded further by the lack of metrics to quantify the behavior of nonverbal subjects in computer-based intervention contexts. We present a set of dependent variables called A3 (pronounced A-Cubed) or Annotation for ASD Analysis, to assess the behavior of this demographic of users, specifically focusing on engagement and vocalization. This paper demonstrates how theory from multiple disciplines can be brought together to create a set of dependent variables, as well as demonstration of these variables, in an experimental context. Through an examination of the existing literature, and a detailed analysis of the current state of computer vision and speech detection, we present how computer automation may be integrated with the A3 guidelines to reduce coding time and potentially increase accuracy. We conclude by presenting how and where these variables can be used in multiple research areas and with varied target populations.

Categories and Subject Descriptors: H.5.1 [**Multimedia Information Systems**]: Evaluation/methodology; K4.2 [**Social Issues**]: Assistive technologies for persons with disabilities

General Terms: Measurement, Reliability, Experimentation, Human Factors

This research was supported by the National Science Foundation (grant NSF-0643502). The views expressed in this article are our own and do not reflect those of the National Science Foundation.
This is an extended version of the article "A3: A Coding Guideline for HCI+Autism Research Using Video Annotation," presented at ACM SIGACCESS-ASSETS 2008. Halifax, Canada.
Authors' addresses: J. Hailpern, K. Karahalios, Department of Computer Science, University of Illinois, 61802; email: {jhailpe2, kkarahal}@cs.uiuc.edu; J. Halle, Department of Special Education, University of Illinois, 61802; L. DeThorne, M. Coletto, Department of Speech and Hearing, Science, University of Illinois, 61820; email: {halle, lauras, mcoletto}@Illinois.edu.

Additional Key Words and Phrases: Autism, ASD, nonverbal, intervention, coding, guideline, video, annotation, reliability, point-by-point agreement, Kappa, audio feedback, visualization

ACM Reference Format:
Hailpern J., Karahalios K., Halle J., Dethorne L., and Coletto M. 2009. A3: HCI coding guideline for research using video annotation to assess behavior of nonverbal subjects with computer-based intervention. ACM Trans. Access. Comput. 2, 2, Article 8 (June 2009), 29 pages.
DOI = 10.1145/1530064.1530066. http://doi.acm.org/10.1145/1530064.1530066.

1. INTRODUCTION

When working with nonverbal subjects (those who do not use language to communicate) interacting with a computer-based intervention system, one wonders how to assess their behavior in relation to the computer system (e.g., "how does a computer system increase speechlike vocalizations or turn taking?" or "are subjects engaged with the visual and/or auditory aspects of a computer system?"). Unlike traditional HCI experiments, this subject pool cannot engage in talk/think-aloud protocol, answer a questionnaire, or have a meaningful discussion on the merits of the technology. Traditional psychometric assessments and standard HCI task-based studies do not easily apply to this group of subjects. As a result, researchers are forced to rely upon analysis of the subjects' behavior to assess their interaction, engagement, and reaction to computer-based interventions. Since no formal parent report or normative tools are currently available to directly assess a child's vocal response during computer interaction, we focused our methods on direct observation. However, this raises additional questions such as what behaviors should be considered, how to sample, and thereby, how to assess behavior.

This article illustrates how theory from multiple disciplines can be brought together to create a set of dependent variables for use in video annotation, which can be used to assess the interaction between nonverbal subjects and computer-based interventions. This coding guideline called A3 (pronounced A-Cubed), or Annotation for ASD Analysis is the primary contribution of this research. A3 examines the interactions of nonverbal subjects with computer feedback in terms of engagement and attention, specifically focusing on speechlike vocal behavior. This article presents the full coding guideline and justification grounded in existing literature. By basing each variable on preexisting work, we leverage the large body of literature that already establishes the usefulness and application of each variable. Further, A3 is demonstrated in an experimental context through agreement analysis of over 1200 minutes of experimental video footage.

In addition to the creation of A3 and reliability demonstration in an experimental context, we examine the existing body of literature and the current state of computer vision and voice detection. We apply state-of-the-art capabilities of computer technology to the A3 guideline/annotation process and detail how, and to what degree, the process may be automated. This analysis is grounded in the technology capabilities of today, rather than what computer automation may be capable of in the future. These solutions may be integrated into the annotation process, reducing the burden on coders and researchers.

Because A3 is based on diverse literature from a wide span of disciplines, the applications to nonverbal subjects interacting with computer-based systems are far reaching. This article concludes by discussing how, and where, the A3 guideline may be used.

Though this work is situated in an experimental context studying Autistic Spectrum Disorder (ASD), the application reaches beyond this population. A3 can be utilized in diverse methodologies (e.g., single subject and group design, short term and longitudinal studies) examining a wide range of populations including individuals with limited speech output (e.g., verbal apraxia and selective mutism) and across multiple domains (e.g., HCI, behavioral science, and clinical practice). As researchers, it is incumbent upon us not only to construct tools, but also to reliably test and measure the performance of our solutions. It should be noted that A3 was not designed to, or has been tested to, be used as a tool for diagnosing autism, or performing general behavioral assessment. A3 provides an operationalized system, grounded in existing literature from multiple disciplines, for quantitatively assessing the behavior of nonverbal subjects interacting with computer-based interventions. Investigators and clinicians are encouraged to adopt its use to their own purproses as guided by their own knowledge of best practice. Because A3 focuses on vocalization and interaction of nonverbal participants with computers, we believe that the coding guidelines outlined in this article can be applied to many populations, experimental designs, and contexts.

2. EXPERIMENTAL CONTEXT

This article is situated, and demonstrated, in an experiment that lasted over a 12-month period [Hailpern et al. 2009], which focused on children with Autistic Spectrum Disorder (ASD). ASD is a developmental disorder exemplified by delays in empathy, basic social interaction, communication, and language use/acquisition. As the name connotes, it is not a monotonic diagnosis, but rather a spectrum with ranges in severity and symptom presentation. The focus of this research was to examine the effect of computer-generated feedback on the behaviors of five nonverbal children with ASD. The feedback generated (audio and visual) was directly based on the individual participant's vocalizations. The interaction of each participant was assessed in terms of his or her engagement with the computer/feedback, attention, and vocal behavior.

The five participants were exposed to a variety of computer-based feedback systems over six sessions that were spaced approximately one week apart. Sessions lasted about 40 minutes, and consisted of eight trials, each about two minutes in length. With parental consent, sessions were video taped, totaling over 1,200 minutes of video footage for data collection. The experimental context focused on nonverbal low-functioning individuals with ASD who varied in age (3–8 years), but could all be characterized at the prelinguistic stage in terms of intentional communication.

3. EXISTING LITERATURE

At the outset of the original experiment, researchers reviewed existing literature across multiple disciplines in HCI and Behavioral Sciences. Researchers

concluded that no one source provided a set of dependent variables for observation of engagement and vocalization of nonverbal subjects in computer-based intervention contexts. Although the literature has established methods for coding the behaviors of nonverbal subjects, no system has been designed for explicitly examining the interactions between nonverbal subjects and computer-based intervention. We present a brief summary of existing literature pertaining to children with ASD, and nonverbal subjects. Specific related work is cited in Section 3.3 to provide justification for each variable and Section 6 in order to assess the current ability to automate video/audio annotation.

3.1 General Overview of Direct Observation of Behavior

For many researchers, and many domains, direct observation has been the primary mode of data collection. Before the advent of video systems and digital technology, direct observation was conduced through hand written ethnographies, descriptions and hand sketches of observed behaviors by psychologists who explored man's individual and social behavior, and reports by anthropologists examining the behavior of societies [Bijou et al. 1968; Clifford et al. 1986]. As technology has progressed, applications of these same principles have been applied to new domains (Human Factors, Computer Science, CSCW). In addition, the process of gathering observational data has become more operationalized. With celluloid[1] and now digital video, rewatchable sessions can now be annotated and linked with specific behaviors. This allows researchers to quickly refer back to the actual events, rather than rely exclusively on notes and memory [Lee 1974; Retherford et al. 1993; Rosenblum et al. 2004]. To reflect the broad spectrum of disciplines that use video annotation and the study of behavior, many guides and sets of variables have been created. We briefly discuss here the approach of behavioral scientists and researchers in human computer interaction.

3.2 Behavioral Sciences

Researchers in the behavioral sciences have studied behavior from a broad range of perspectives: including both diagnostic and thereputic. Even within different applications, the focus on behavioral observation is quite varied. Some researchers has examined aspects of speech and/or sound production [Koegel et al. 1998; Wetherby et al. 1998; Woods and Wetherby 2003; Owens 2007], while others have focused on interaction (with and without researcher/clinician being physically present) [Baskett 1996; Wetherby et al. 1998]. Diagnosis via observation [Lord et al. 2000] and communication skill acquisition [Prizant et al. 1997; Sheinkopf et al. 2000] have also been examined.

There is an interesting parallel between subjects in infant research and older nonverbal subjects in that both populations are nonverbal. In some respects, they present similar levels of verbal competence, and consequently similar coding challenges. Although our work has a different purpose and is

[1]As used in film reels.

examined in a different context, it shares many of the same critical aspects of behavioral assessment as that of infant research [Segal et al. 1995; Hayne et al. 2000; Luo and Baillargeon 2005].

Regardless of the specific population, studying the behavior of nonverbal subjects requires a reliable coding system that often is not published. Consequently, investigators in such areas are often forced to "reinvent the wheel."

3.3 HCI Research

Computer Scientists, particularly in HCI, and Social Scientists have developed a broad set of coding guidelines for a large array of tasks [Whyte 1980; Suchman 1987]. However, few of these focus on the evaluation of subjects who are noncommunicative. Even fewer address those subjects with ASD. Guidelines exist that have dealt with higher functioning subjects [Piper et al. 2006; Cassell et al. 2007; Gillette et al. 2007; Tartaro and Cassell 2008] and most, gathered data through subjective (qualitative) observation [Kerr et al. 2002; Michaud and Théberge-Turmel 2002; Parés et al. 2005].

Although the literature in the HCI and Social Science disciplines is comprehensive and examples of coding guidelines are robust, an established quantitative coding system that addresses the behavior of nonverbal subjects and interventions using computer systems that provide auditory and/or visual feedback does not exist. This article addresses this gap by detailing the construction of A3 coding guideline and the reliability of the variables in an experimental research setting.

4. A3 CODING GUIDELINES

From existing areas of research, we drew the most relevant and salient dependent variables in relation to vocalization and engagement to help quantify the interaction of nonverbal subjects and computer-based intervention systems. By drawing theory and experimental findings from 19 sources spanning five disciplines (Special Education, Speech and Hearing Science, Infant Research, Diagnostic Observation, and HCI), we built upon the research of others and grounded our guidelines in the bodies of literature across multiple areas of science. This resulted in a set of 17 momentary metrics and four durational measures, which totaled 21 dependent variables for analyzing the behavior of nonverbal subjects, which we termed A3. The name A3 references the original purpose of studying ASD behavior with computer intervention, though the application of this system is far broader.

To facilitate the refinement and reliability of the variable definitions, four coders were employed to annotate video from the experimental context. All coders were from the Department of Speech and Hearing Science at the University of Illinois at Urbana-Champaign. The first two coders were undergraduate seniors, while the second two were graduates of the undergraduate program (one was pursuing a master's degree in Speech and Hearing Science). All video coders had class experience in phonetic transcription and three of the four had worked as coders on relevant research projects.

4.1 Process of Refinement

Beginning in October 2007, researchers began a seven-week iterative design cycle in which the definitions of the 21 metrics for examining behavior were refined. Each week, two coders reviewed the same video from the experimental context and annotated all 21 variables based on the current definitions using the video annotation system VCode [Hagedorn et al. 2008]. At the end of each week, coders met with researchers and an examination of agreement was performed using VData [Hagedorn et al. 2008]. Discussion of discrepancies resulted in modifications to the variables's descriptions. Each successive week, data were coded in light of the new definitions.

At the conclusion of the seven weeks, two additional coders were recruited due to a change in the availability of the original coders. Though this required training to begin again, the addition of two fresh perspectives ensured that any assumptions about variable definitions made by the first pair of coders (this potential confound is referred to as observer drift [Kazdin 1977]) were revealed and explicitly noted in the guidelines through several weeks of training/guideline-refinement.

This process continued until a point-by-point inter-rater agreement level [Kazdin 1982] of 85% was reached across all variables in one session. Although an agreement level of 80% is considered acceptable [Kazdin 1982], we wanted to ensure that the construction of variable definitions was "above" standard. We used a point-by-point agreement calculation with a tolerance of one second (i.e., two events were said to agree if the secondary coder's mark was within 0.5 seconds on either side of the primary coder's mark) for calculation of agreement. See Section 5 for a more detailed discussion of agreement calculations employed in this research.

4.2 Video Annotation Software

To facilitate the video annotation process, many digital tools have been developed [Burr 2006; Kipp 2007; Noldus Information Technology 2007; SALT Software 2007], all of which can be used in conjunction with A3. With the onset of these digital systems, software designers can enhance them to aid in the coding process. For our own research, we created a tool called VCode (Figure 1), published in 2008 [Hagedorn et al. 2008]. VCode was designed to ease the burden on coders, specifically related to coding schemes such as A3. Although VCode is not a contribution of this article, we provide a brief description of VCode features related directly to, and aid the understanding of A3.

The VCode environment is focused on providing a unified interface for coders to allow them to make the most informed annotations possible. Coders may view multiple simultaneous video streams at the same time, with one full size. Coders can switch between camera angles/streams with one click, allowing them to get the best angle or perspective. Each variable is listed along the right hand side of the screen represented by a unique color also used in the timeline (bottom center) of the screen, for each annotation. Events can be categorized as either momentary (occurring once) or durational (having a start and end time). A single diamond represents momentary events, while

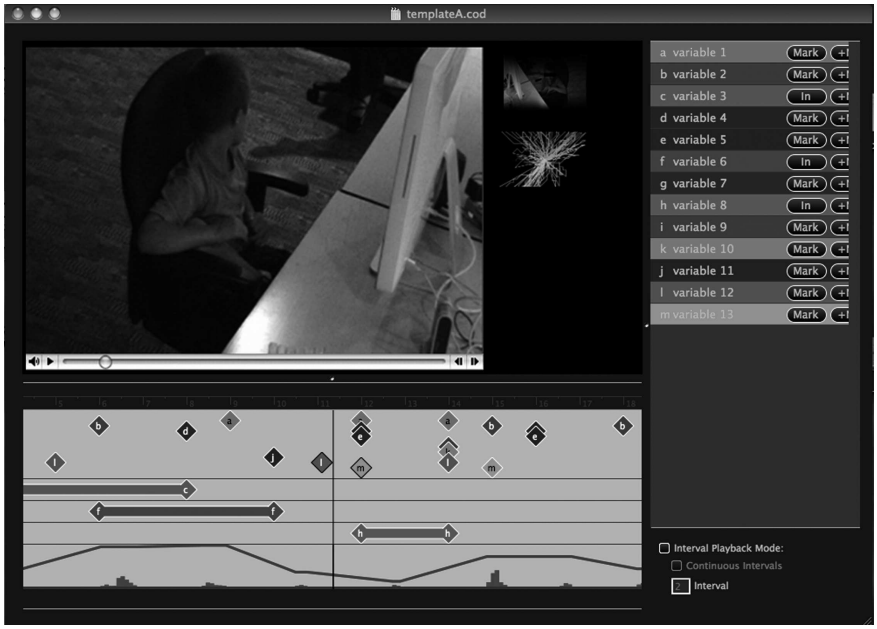


Fig. 1. The VCode video annotation environment. Multiple video streams are displayed along the top left. Variables to are shown along the right. The bottom portion of the window contains the timeline, representing momentary events (single diamonds), ranged events (diamonds with a tail), and extra data such as video volume (along the bottom of the timeline).

ranged events are displayed as two diamonds connected by a tail. Because all annotations are viewable on the same timeline, coders can see and code based on the relationship between different variables, a key aspect of A3. Along the bottom of the timeline are secondary data (e.g., computer logs or video volume) that provides coders a more robust set of information from which to code, and increases annotation quality [Quek et al. 2003].

4.2.1 *Playback Modes.* VCode also facilitates multiple modes of video playback. Due to the nuances of different variables, traditional Standard Playback (from start to finish with the viewer being able to pause) is not always the most practical or relevant. Often, particular actions have ambiguous start and end times, thus making it difficult to pinpoint, or annotate, the initiation of a behavior. For these "difficult" variables, a video playback technique called Interval Playback is commonly used. Interval Playback divides a video into discrete segments of equal length, which are viewed individually, and a coder marks whether or not a specific event occurred during that time period. For example, coders may be asked to watch a 3-second segment of video, and then mark whether or not the child smiled during that interval. Traditional VCR techniques ask coders to play a video for N seconds and then pause the tape as close as they can to N seconds. VCode eliminates the burden on coders by having three modes of playback.

—Standard Playback. Watch a video from start to finish. Coders have the ability to pause and start video as well as rewind, fast-forward, and view frame-by-frame. In addition, coders can also jump through video by annotation.
—Interval Playback. Video is divided into discrete units of length N (specified at annotation time). The video jumps from frame to frame, N seconds apart. This captures behavior at discrete moments, eliminating start and end times of events, and reduces the annotation to a binary operation of an event occurring or not occurring.
—Continuous Interval Playback. Video is divided into discrete units of length N (specified at annotation time). The video plays for N seconds then pauses. When coders press play again, the video will play the next interval allowing behavior to be captured in durational spans eliminating the need to identify event start/end times.

The two modes of interval playback allow for two levels of video analysis; coarse (Interval Playback) and fine (Continuous Interval Playback). A3 does not use Interval Playback. Both of these modes have been described in the behavioral observation literature [Alberto and Troutman 2005].

4.3 A3 Variables

The following section is a detailed description of the dependent variables examined in the A3 coding guideline. These descriptions focus on the rationale for each variable and the major choices made when constructing the variable definitions. The actual guide (with the specific topographical, physical, or behavioral features for annotation) is presented in Appendix A. In general, we can divide our dependent variables into measures of (a) engagement and (b) vocal behavior of the subject. We begin by highlighting the engagement variables, because engagement is often viewed as a prerequisite to learning and communicative exchange. We follow with a review of the variables based on child vocalization.

Before reviewing the individual variables, it should be noted that one variable, Phonetic Transcription (or marking the phonemes uttered by subjects), was dropped early in the analysis process due to coder feedback. Coders stated that because of poor audio quality and extremely poor articulation by subjects, accurate and reliable phonetic transcription would not be possible. It should also be noted that here we present only justification for each variable, and refer the reader to Appendix A for variable descriptions.

4.3.1 *Engagement Variables.* With the exception of the metric Time In Chair (Section 4.3.1.5), which was gathered with standard playback, all these metrics were gathered with the Continuous Interval Playback Mode set to three seconds. This mode is ideal for variables that are not discrete (i.e., difficult to specify their exact starting or ending point, or exact duration).

4.3.1.1 *Smiling.* The variable Smiling was chosen because it is typically associated with pleasure or enjoyment [Field et al. 2001]. Although the source of the smile could not always be determined, we hypothesized that a higher rate

of smiles would reflect conditions that were generally more enjoyable to the subject.

4.3.1.2 *No Face.* The No Face variable was used to identify three-second intervals when the child's face could not be seen, and no coding determination could be made as to whether or not a smile occurred. This variable was identified because of a concern that surfaced during the coding process: coders found that when they summarized the data, they had difficulty discerning intervals when no smiles had occurred from those they were unable to code. Although its accuracy is reported, this variable was not directly used in the analysis. Rather its agreement was useful for demonstrating that coders were "on the same page," and allowed agreement calculations for Smiling, which is dependent upon being able to see the face. The absence of both Smiling and No Face marks are an indication that the child was not smiling.

4.3.1.3 *Oriented at Screen.* In order to assess visual attention to content, we created an "orientation arc" for the evaluation of the child's gaze. See Baskett [1996] and Field et al. [2001] for others who opertaionalized this procedure. If gaze was directed within this arc within the 3-second interval, the subject was considered to be Oriented at Screen. The arc's width ($\sim 90^\circ$) was used to accommodate the behavior in which children with ASD use peripheral vision as primary visual input [Howlin 1986]. See Appendix A for illustration/diagram of the arc.

4.3.1.4 *Auditory Focus.* Much like Oriented to Screen, the Auditory Focus variable was used to assess auditory attention. Unlike visual attention, which has a more observable physical indicator, auditory attention must be observed indirectly. Auditory Focus was observed via changes in subject proximity to (moving closer) and physical contact with the speaker which is a possible surrogate for inferring interest in computer audio. In addition, Auditory Focus was also annotated if the subject oriented to the screen/speaker after a new sound was made by the computer [Clifton et al. 1999; McCalla and Clifton 1999].

4.3.1.5 *Time in Chair.* To assess the willingness to attend to computer stimuli, we coded the duration a child would spend in his/her chair [Sajwaj et al. 1972; Lovaas 2003]. We hypothesized that increased time sitting was a proxy for engagement with the computer.

4.3.2 *Verbal Variables.* Verbal metrics were collected to examine vocalizations during the experimental and control conditions. Coding of vocalizations was facilitated through use of a decision tree, which is incorporated in the full coding guide (Appendix A). Figure 2 presents the child sound decision tree in isolation as a hierarchical set. From this perspective, researchers can better understand the relationships between different variables. With the exception of Turn Taking (Section 4.3.2.7), the following sections are presented hierarchically at decision points in the tree rather than one for each variable. These variables were assessed with Standard Playback. It should be noted that we

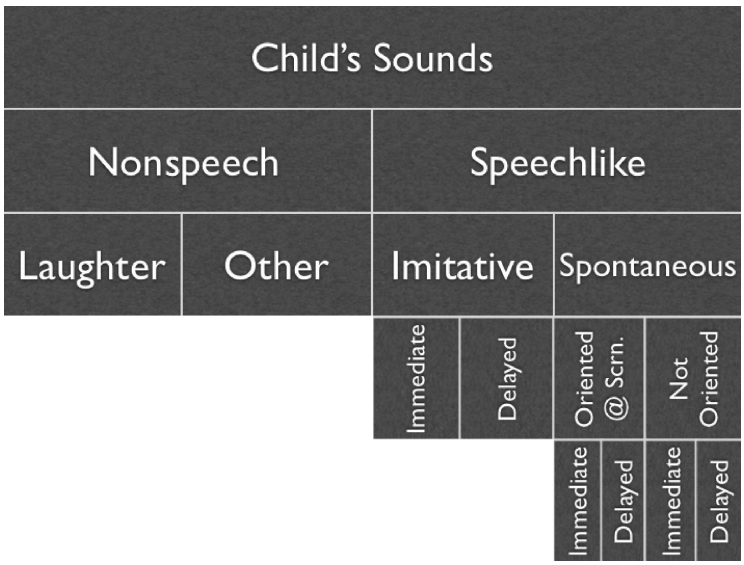


Fig. 2. Child Sound Decision Hierarchy.

present only justification for each variable, and refer the reader to Appendix A for variable descriptions.

4.3.2.1 *Child's Sound (Speech vs. Nonspeech)*. The most basic question coders had to address was whether or not a sound was considered "speechlike." Specifically, we define a Speechlike sound as one that could be phonetically transcribed. This decision point attempts to screen sounds that have the potential to lead to conventional speech and those that may be related to ticks, breathing, self-stimulatory behavior, or other forms of expression that are not used in speech production (laughing, screaming, etc) [Woods and Wetherby 2003].

4.3.2.2 *Nonspeech Sounds (Laugher vs. other)*. Though the range of Nonspeech Vocalizations is large, we asked coders to distinguish Laughter as another means to examine children's pleasure and engagement during the activity [Gena et al. 1996]. (We realize that this requires an inference that may not be accurate for many children with ASD.) In addition, we wanted to differentiate Laughter from vocal self-stimulatory behaviors. Compared to other children with developmental delays, those with ASD tend to produce more nonspeech sounds [Sheinkopf et al. 2000]. By annotating Nonspeech Vocalizations, we also hoped to examine the impact of the external stimuli on their nonspeech vocalization and in comparison to their Speech-Like Vocalization.

4.3.2.3 *Speechlike Sounds (Imitative vs. Spontaneous)*. A critical distinction made in studying the communicative behavior of children with special needs is between sound production that is Imitative (repeating a sound previously heard) or Spontaneous (without an immediate model) [Halle 1987]. Using this distinction, we hoped to explore the nature of speechlike sounds

produced and if there is a direct relationship between what is prompted (human/computer) and what is vocalized.

4.3.2.4 *Imitative Sounds (Immediate vs Delayed).* To explore the imitative sounds produced by subjects, we divided them into those that occur immediately after a source (within five seconds) and those that occur after a more prolonged time [Prizant et al. 1997]. This distinction is particularly relevant for children with autism due to echolalic tendencies [Rapin and Dunn 1997]. Theory suggests that words/sounds in delayed imitation are stored outside of the subject's short-term or working memory [Gathercole and Baddeley 1993; Bradford-Heit and Dodd 1998].

4.3.2.5 *Spontaneous Sounds (Orientation to Screen).* While imitative sounds are, by definition, based on audio stimuli, we wanted to delve deeper into spontaneous sounds, and their relationship to screen orientation. Eye gaze is indicative of engagement. When paired with vocalization, it is a key communicative development [Baskett 1996; Wetherby et al. 1998]. For each spontaneous sound produced, we explored whether or not that sound was made while oriented to the screen. This allowed us to examine the direct relationship between spontaneous sound production and orientation.

4.3.2.6 *Spontaneous Sounds (Immediate vs. Delayed).* Much like imitative sounds, we wanted to understand if there was any correlation between spontaneous sound production and auditory stimuli. To explore this relationship, we asked coders to mark spontaneous sounds that were made within five seconds (immediate) of a source sound, and those made after a longer period of time (delayed).

4.3.2.7 *Turn Taking.* An important skill in oral communication is that of turn taking, or waiting for others to finish their utterance before vocalizing [Wetherby et al. 1998; Owens 2007]. With all speechlike sounds, we asked coders to determine if the subject waited for the source (be it a researcher or computer-generated sound) to "finish" their sound production. In other words, did the child wait for his/her turn to talk (or not interrupt).

4.3.3 *Other Metrics.* Previous sections dealt with variables related to Engagement and Verbal Behavior. Two nonverbal/engagement variables useful in analyzing subject behavior are presented here.

4.3.3.1 *BIGmack Switch.* The BIGmack™ Switch[2] is an assistive technology device that plays a pre-recorded sound for individuals with speech and language delays [Reichle et al. 2002]. With one subject, who was suspected of having limited motor control of his vocalizations, this device was used to simplify the task of producing sounds.

4.3.3.2 *Nonchild Audio.* These data points served two purposes. Primarily, they were collected to help clean data logged on the computer by marking

[2]http://www.ablenetinc.com/item_detail.aspx?ItemCode-1000201

sounds (other than those made by the child) in the video that interfered with automatic data gathering. A second purpose was to familiarize coders with the video they were about to watch without forcing them to annotate a very complex variable. Because Nonchild Audio was coded as a first pass, by itself, coders were required to watch the entire video once, before examining more specific details.

4.4 A3's Four Pass System

The actual annotation process was divided into four passes, each of which asked coders to focus on a specific category of dependent variables while they watched a video in its entirety. Since many of the different variables required different view modes, this pass breakdown not only aided the examination of the data, but also was optimal for the annotation process. The specific variables for each pass can be seen in Appendix A.

4.4.1 *Pass 1—Data Cleaning.* In addition to providing visual and audio information to subjects, computers have the ability to perform real-time analysis of audio (from a microphone) and visual (from an attached camera) data. To draw corollaries between changes in the state of the computer system and the behavior of subjects, HCI systems often log a robust set of data. Though these data can be rich and useful in behavior and interaction analysis, this data set can often be muddied by nonvalid audio data (audio input that came from non-subject sources). In other words, sound from a researcher, an object falling, or the subject hitting the computer/table can appear in the logged data. The first pass focuses on reducing the "noise" from the audio data to help filter the logged information.

4.4.2 *Pass 2—Attentiveness and Engagement.* The second pass is a collection of variables that can be used to assess the engagment, response, and interaction of the subject diuring presentation of the computer's audio and visual feedback. The data are collected using a Continuous Interval Playback.

4.4.3 *Pass 3—Vocalization Analysis.* The heart of the A3 coding guidelines is a detailed analysis of the subject's vocalizations, specifically focusing on those that are considered to be "speech like," or have the potential for contributing to meaningful speech.

4.4.4 *Pass 4—Time in Chair.* The fourth pass examines the time a subject spends in the chair. Though not all subjects with developmental disorders are willing to sit for an extended amount of time, analysis of this behavior can be used to examine participation with the computer system.

5. A^3 RELIABILITY & EFFICACY[3]

By situating the creation of A3 in an experimental context [Hailpern et al. 2009], researchers were able to utilize the collected video data to test reliability

[3]The numbers reported here are based on a larger dataset then those reported on in Hailpern et al. [2008].

Table I. Point-by Point Percent Agreement

Variable	% Agreement
Nonchild audio	75.96
Duration: Non child audio	82.39
Smiling	90.71
No face	86.01
Oriented at screen	98.09
Auditory focus	87.37
Laugh	63.89
Duration: Laugh	91.30
Nonspeech vocalization	83.86
Speechlike vocalization	90.50
Duration: Speech like vocalization	92.84
Turn taking	81.60
Immediate + Screen + Spontaneous	79.30
Delayed + Screen + Spontaneous	79.45
Immediate + Spontaneous	78.31
Delayed + Spontaneous	84.62
Delayed imitation	*None Recorded*
Immediate imitation	85.71
Time in chair	88.64
Duration: Time in chair	82.05
BIGMack Switch	87.50

and efficacy of the A3 system. During data collection using VCode, coders were asked to annotate all 21 variables. Of the 268 trials, 11% were checked for reliability (∼132 minutes of video footage). Not all trials were equal in length, nor did all variables occur equally in all sessions. As a result, we examined agreement values across randomly sampled sessions. We present reliability calculations, coder's feedback, and a demonstration of efficacy of A3 in the experimental setting.

5.1 Point-by-Point Agreement

Using point-by-point agreement method [Kazdin 1982], overall agreement between coders across all variables was 88%. Agreement was defined using a conservative tolerance of one second (two events were said to agree if secondary coder's mark was within 0.5 seconds on either side of primary coder's mark). Percent agreements are presented in Table I. Five variables had an agreement above 90%, 11 above 85%, and 15 were above 80%. However, five variables fell below the 80% benchmark for reliable data collection. The sections under of 5.1.2 address these discrepancies.

5.1.1 *Robustness of Agreement Calculations.* To observe how agreement would change with an increase in the timing tolerance, we increased timing tolerance from 1.0 to 5.0 seconds at 0.1 second intervals. Surprisingly, the number of matched points changed only when the tolerance was increased to 1.5 seconds (observers' marks were said to agree if the secondary mark was within 0.75 seconds on either side of the primary coder's mark). Of the few variables whose reliability increased, the change resulted in a gain of at most 0.5%. This suggests that the coders were likely accurate in the placement of

Table II. Combined Data from the Four Spontaneous Speechlike Vocalization Variables

Variable	% Agreement
Immediate Spontaneous	83.19
Delayed Spontaneous	80.57
All Spontaneous	84.62

their marks. Moreover, we can surmise that if there were a lack of agreement in the data, it was likely due to a disagreement of what was coded and not due to ambiguity regarding whether an event had occurred.

5.1.2 *Discussion of Variables with Low Agreement Scores.*

5.1.2.1 *Laughter.* The variable with lower inter-rater agreement values is Laughter, with an agreement of 63.89%. However, upon further inspection, we noticed that it also had a low frequency of occurrence. Lower levels of agreement are often achieved on low-rate behaviors because there are fewer opportunities to observe the target variable [Birkimer and Brown 1979].

Coders also mentioned difficulty in distinguishing Laughter from Speechlike Vocalizations and Nonspeech Vocalizations. Often coders found that vocalizations may have been laughterlike, but matching positive affect with the vocalization was difficult. Perhaps this difficulty is exacerbated by the differences in affect expression that characterizes ASD [Wetherby et al. 1998]. Despite this characteristic, reliability of Laughter may be improved through increased microphone quality. An explicit plan to strengthen the reliability should be considered.

5.1.2.2 *Spontaneous Speechlike Vocalizations.* The subdivisions within Spontaneous Speechlike Vocalizations (with and without orientation at screen and delayed vs. immediate) also resulted in low agreement between coders (78.31% to 84.62% with a mean of 80.42%). To explore the effect of subdivision on reliability, we combined data across variables (Table II). To combine two variables, we treated all marks for both variables the same and recalculated agreement. This analysis can suggest at what level of granularity (distinction between variations of a variable) these variables can be reliably coded. When we eliminated the distinction between sounds made while the subject looked at the screen versus those made when looking away, reliability improved to 80%. It appears that such small distinctions may have been too fine-grained to code accurately unless multiple camera angels are available. Otherwise, we recommend differentiating only Immediate and Delayed Spontaneous Speech.

Upon further examination of the Spontaneous Speechlike data, we discovered the potential for double-counting disagreements. Every Speechlike Vocalization was coded as either Spontaneous or Imitative. However, every disagreement in Speechlike Vocalization is a guaranteed disagreement for the Spontaneous/Imitative distinction. Thus, our lower agreement values may have been a direct result of "double counting."

From this analysis, we believe that the best approach for implementation of A3 is to code all levels of Spontaneous Speechlike Vocalization that are of

Table III. Kappa Statistic for Variables Coded Using Interval Playback (set to 3 seconds)

Variable	Kappa
Auditory Focus	0.64
Oriented at Screen	0.84
No-Face	0.77
Smiles vs. No Smiles	0.63

theoretical interest, and consider the possibility that distinctions between Immediate and Delayed Imitations may need to be collapsed to achieve reliability. To analyze the effects of screen attention and vocalization, we suggest performing a post-hoc analysis between Oriented at Screen and Speechlike Vocalization. Although this may not link each specific vocalization to the visual display, a trend should be apparent and extrapolation should be possible.

5.1.2.3 *Nonchild Audio.* The second variable that had less than 80% agreement was Nonchild Audio. One plausible explanation for the poor agreement in this variable may be due to the poor quality of the audio recording. Some coders were better able to hear quieter sounds and, thus, would mark them, while other coders could not hear these sounds and they would escape notation. As a result, there was a discrepancy between the coders.

We propose an audio "threshold" be set to differentiate between sounds to code and sounds not to code through the use of a low-pass filter. With the addition of a low-pass filter (either to be employed by the computer or presented as a visualization inline with the coding timeline), we believe we can surpass the 80% agreement level as recommended by Kazdin [1982].

5.2 Kappa Statistics

Most variables were coded on an infinite timeline. In other words, marks could be placed at almost any location during the duration of the video. As a result, the probability of a chance agreement was extremely low, and thus reduced the need and applicability of Cohen's Kappa [Kazdin 1982]. However, for the variables collected using the Continuous Interval Playback, we calculated Kappa analysis because the video was annotated in discrete, binary segments (set to 3 seconds) of required observations, and thus subject to chance observation.

The Kappa statistics calculated from the data suggest a high level of agreement (Table III). Kappas ranged from 0.64 (Good) to 0.84 (Very Good). Our interpretation of agreement follows from that of Altman and Byrt [Altman 1991; Byrt 1996]. These findings add further support to the findings of the point-by-point agreement analysis.

For Smiling and No Face calculations, we used a 2-tier evaluation metric similar to that used by Reid et al. [1985]. The first Kappa accounts for the level of observer agreement on whether they could make a judgment about a subject's smiles (could they see the subject's face—No Face). We then eliminated all of the intervals in which either coder marked No Face implying that they could not assess whether the subject was or was not smiling because the assessment of agreement on Smiling depended on both observers being able to see the subject's face. The intervals, in which both observers coded the subject

as either smiling or not, were examined for agreement. In other words, coding of smiles depended upon both observers agreeing that they could see the subject's face.

5.3 Efficacy of A3 in Experimental Context

Hailpern et al. [2009] performed analysis of the experimental context using the A3 coding guidelines. This article examined the impact of different forms and modalities of computer generated feedback on the Spontaneous Speechlike Vocalizations of five nonverbal participants with ASD. Results demonstrated statistically significant effects of different feedback systems on the subject's vocalization and highlighted their individualized preferences. Most notably, analyses using the A3 coding guidelines revealed statistically significant differences in vocalization rate based on the modality of the feedback with three children producing more Spontaneous Speechlike Vocalizations in the presence of audio feedback, and one child responding more consistently to visual feedback. One child's Spontaneous Speechlike Vocalization rate was unaffected by the type of feedback systems. One of the three children who responded significantly in the presence of audio feedback also demonstrated a statistically significant response to conditions with both audio and visual feedback. These results illustrate the capacity of the A3 coding guidelines to quantitatively assess vocalization, but they also support qualitative observations made by researchers.

During the experimental period, researchers made qualitative observations of participants' behavior, their engagement, and which forms/modalities of feedback produced better responses. Upon analysis of the variables collected using A3, statistical findings mirrored most qualitative observations made by researchers. This suggest that the A3 coding guidelines can quantitatively capture qualitatively observed behaviors.

This investigation supported the effectiveness of A3 for quantitatively assessing vocal behaviors in nonverbal participants [Hailpern et al. 2009]. Currently, we are expanding our data analysis in the experimental context to further assess the impact of computer-based feedback systems on measures of engagement, as well as the differential impact on speechlike versus nonspeech vocalizations in the same participants.

5.4 Coder Feedback

By the last video, coders spent approximately 20 minutes to annotate one minute of video footage. This represents a significant decrease from the initial 40 minutes per one minute of video footage (self reported by coders). Coders felt that the majority of their time was spent on the third pass of the annotation system, due to the complexity and scrutiny required to differentiate among different types of vocalization. This difficulty was exacerbated due to the poor articulation of the the population used in the experimental context and the poor quality of the video camera's microphone. Improvement of microphone quality could help improve accurate labeling of the vocalizations.

In further discussion, coders mentioned some confusion over the No Face variable, specifically in the boundary condition when the child has part of his or her face covered. Feedback from coders included specific requests for a more explicit definition of the features that must be seen to justify annotating No Face. For future experiments, we propose specifying features of the face (e.g., lips, cheeks) that most clearly provide access to determination of subject: re affect.

6. AUTOMATION AND A3

Although coding time was reduced from a self-reported 40 minutes/1 minute of video to 20 minutes/minute of video, coding remains time consuming. For researchers, this impacts the time required to gather data, as well as delays data analysis and the ability to progress in the research itself. Likewise, clinicians need to reduce time demands in order to provide clinical utility [Perlman 2008]. As VCode was developed to assist in the coding process, further computerized systems can be developed to reduce the burden on coders and the time required to annotate a video. Specifically, through the use of computer automation, the task of coders may be substantially reduced if not eliminated. The addition of computer-based automation has the potential to not only increase the speed at which video-annotation can be accomplished, but also to improve the quality of the data gathered for the experimental evaluation [Burr 2006].

This section details how automation, applied to A3 could be used to enhance the collection of data. In order to uncover ways the technology could be used to automate the coding process, we examined the existing body of literature and the current state of computer vision and voice recognition software. By examining what technology is capable of today, we are able to assess the degree to which computers could be used by coders and researchers to augment the video annotation process. While research is continuing to make strides in computer vision and speech recognition, we focus our discussion on the state-of-the-art and what can be implemented today, rather than on where technology may be in the future.

6.1 Degrees of Automation

With such a complex set of variables and behaviors incorporated in the A3 guideline, the varying forms of assistance technology can provide are prodigious. We therefore break down each form of technological assistance into one of four categories:

REPLACEMENT: When a coder can be replaced entirely by an automated process for collection of a specific variable. This requires the automated system to have a low rate of false positives and low rate of false negatives. By using a Replacement system, coders need not manually annotate a video for a specific behavior. This form of assistance requires technology to achieve nearly the same agreement as that of a coder. Though this is by far the most demanding on the technological support system, it does greatly reduce the burden on the

coder and speeds up time to annotate. For this automation, incorporation with the existing coding practices is not required. Rather, researchers can omit manual annotation of the specific variable.

REVIEW: When a coder is only needed to double check data annotated by an automated process. This requires the automated system to have a low rate of false negatives and a moderate to low rate of false positives. Systems that use the Review model relatively accurately mark when a specific behavior occurs in a video. However, due to a modest rate of false positives, a coder must check over each annotation to assess its validity. This substantially reduces time required to annotate, because coders are not required to watch a video in its entirety. Rather, they jump from mark to mark. Low false negatives are a requirement, in that, coders will not be watching nonmarked footage. Hence, missed behavior will not be marked. For this automation, incorporation with existing practices can be achieved by creating annotations that exist in the coding environment (e.g., as actual marks on a timeline) or as a secondary display of information (e.g., as in the graphical data shown on along side the annotation timeline).

QUICK INDEX: When the automated process can help guide coders where to "look" for a behavior, but not provide actual annotations for use in data collection. This requires the automated system to have a low rate of false negatives and can accommodate a high rate of false positives. Quick Index automated systems point out time spans of interest to coders, allowing them to skip areas of inactivity. By maintaining a low rate of false negatives, coders can be assured that the area they are skipping is of little or no value to the current variable. These Quick Index systems can be used when a computer can detect that "something" is occurring, but are unable to identify or categorize the exact behavior that is being demonstrated. For this automation, incorporation with existing practices can be achieved by creating a video stream whose colors change to alert coders that a certain condition is occurring (e.g., as a secondary video stream) or as a secondary display of information (e.g., as a graphical data shown along side the annotation timeline).

MANUAL: When the automation process has a high rate of false positives and a high rate of false negatives, coders must annotate without any assistance. Manual annotations require time and the full concentration of a coder. While these forms of annotations are time demanding, with proper training and regular agreement checks, the accuracy of these annotations (as demonstrated in this article) can be reliably used for data collection.

6.2 Variable by Variable Automation

This section presents an analysis of each variable and to what degree current technology can facilitate automated annotation. Based on this analysis of existing technology, researchers can design systems to automate their own use of the A3 annotation guidelines, thereby improving accuracy while reducing coding time. Further, researchers can augment existing open source video annotation software, such as VCode [Hagedorn et al. 2008], to incorporate these forms of automatic video annotation. We break down the variables at key

Table IV. Degree of Automation by Variable

Variable	Automatic	Review	Quick Index	Manual
Nonchild audio		○	●	●
Smiling		◑	●	●
No face			◑	●
Oriented at screen		●	●	●
Auditory focus		◑	●	●
Laugh		○	●	●
Nonspeech vocalization		○	●	●
Speechlike vocalization Sounds			◑	●
Turn taking		◑	●	●
Imitative vs. Spontaneous			○	●
Immediate vs. Delayed		◑	●	●
Orientation + Vocalization	○	◑	●	●
Time in chair	●	●	●	●
BIGMack Switch	●	●	●	●

● Possible/Usable ◑ Probable ○ Potential

decision points paralleling section 4.3. A summary of variables and their degree of automation is presented in Table IV.

6.2.1 *Engagement Behavior Variables.* Behavior detection in video is a growing field of research in areas such as scene based automatic annotation [Ponceleon and Srinivasan 2001; Chen et al. 2002], automatic event logging [Banerjee et al. 2004], and object of focus identification [Bertini et al. 2004]. We focus here on systems related to A3 Engagement Variables.

6.2.1.1 *Smiling.* Because positive affect is part of the definition of Smiling, detection of emotion is critical towards automating the process of smile detection. A growing area of HCI research has been focusing on bringing emotion and emotion detection into day-to-day computing [Crane et al. 2007]. Within this subset of work, emotion detection has been a critical component [Zacks et al. 2001; El Kaliouby and Robinson 2004; Wang et al. 2007; Zeng et al. 2007]. In traditional HCI, this detection can be used to adapt computer systems to respond more appropriately to human reaction [Setlur and Gooch 2004]. In addition, a growing set of work in facial feature detection has been focused on disability research [Kaliouby and Teeters 2007; Madsen et al. 2008]. Other researchers have examined detection of emotion/arousal with other nonvideo metrics [Chang and Ma 2002; Jones and Troen 2007]. This work, combined with research on automatic smile detection [Valstar et al. 2006; 2007] can provide a Review-based automation system. With poorer accuracy, these techniques could still provide a time saving gain of a Quick Index system, simply by identifying "happy" emotional points.

6.2.1.2 *No Face.* Research on face detection has long been a major focus of computer vision systems [Zhao et al. 2003]. Given this robust set of data, computer systems could relatively easily note moments when no face is visible. However, the definition of No Face only is marked when the face is occluded enough to prevent analysis of a smile. Work in machine learning [Osadchy et al. 2007] allows for facial orientation analysis, which could

further be used to determine No Face. Working with research on smile detection (Section 6.2.1.1), systems could conceivably be constructed to determine when a smile was not present, however, that does not necessarily mean the face was occluded enough. In conclusion, technology has great potential here, however, current algorithms are not robust enough to facilitate Replacement or Review systems. At best, a Quick Index system could be constructed to alert coders to the presence of a face, and thus require coders to mark when a No Face truly occurred.

6.2.1.3 *Oriented at Screen.* In addition to using face detection techniques [Zhao et al. 2003], eye-tracking technology [Jacob 1991] can be used to assess when subjects are looking at a screen [Wang et al. 2007]. Further, work in machine learning [Osadchy et al. 2007] can also utilize facial pose and orientation to determine looking direction. Because A3's guideline examines facial orientation within an arc, we believe that existing technology can provide a fairly robust Review automation system, for annotating when subjects are oriented. However, it would still be necessary for coders to review these marks to reduce the false positive rates and determine when faces detected are not those of the subject.

6.2.1.4 *Auditory Focus.* Auditory Focus is a complex relational analysis between audio produced by the computer and the child's orientation to the computer. However, these components, when broken down, can be detected relatively accurately by a computer system. Section 6.2.1.3 detailed the detection of subject orientation to screen. That combined with logging of computer-produced sound occurrences, can create a Reviewable automation system, requiring coders only to validate the behavior. For the more mundane physical changes in proximity to an audio speaker or making physical contact with a speaker, using computer vision becomes increasingly difficult. At best, computer vision systems can detect motion [Zhang and Kurtev 2003; Xiao et al. 2008]. Setting a certain degree of motion, systems could be automated to note locations for Quick Index.

6.2.1.5 *Time in Chair.* Time in Chair could be easily determined by a series of pressure sensors, detecting the presence/absence of the weight of the subject in the chair. The pressure sensors would have to be calibrated to detect full sitting, but not those when child is not "sitting" (e.g., one leg on chair and one leg standing), resulting in Automatic automation.

6.2.2 *Verbal Variables.* Voice detection and analysis is a rapidly evolving field of research covering dialogue transcription [Johnson 2007; SALT Software 2007; Studiocode Business Group 2007] and Voice-Computer Interaction [IBM 2005; Nuance Communications 2008]. A large hurdle when identifying and classifying speech comes from speaker identification [Nishida and Ariki 1998; Kwon and Narayanan 2004]]. However, the largest burden still comes from identification of what is being said. Though this is a robust field of research, it does require a large degree of training data for each subject. Much work has focused on identifying what is being said for subjects with good diction

and vocalization. Thus it is unclear how well it would function with nonverbal sound production or vocalization of individuals with speech impairments. As a result, much of this classification may occur largely Manually. However, by at least noting when sound does occur (though volume analysis), coders may be able to use Quick Index from this degree of automation.

6.2.2.1 *Child's Sound (Speech vs. Nonspeech vs. Nonchild-Audio).* Identifying the difference between speech and nonspeech vocalizations would require a robust speech detection engine to make this distinction. While much of the existing literature has focused on identifying words and letters in speech [IBM 2005; Nuance Communications 2008], very little has been done to examine those sounds produced that are not speechlike. Even more complex is identification of sounds that are not speech at all, rather other miscellaneous sounds in the environment. Some more recent research in the area of speech separation has begun examining and differentiating different sources of sounds coming from a solo microphone [Bach and Jordan 2006; Olsson and Hansen 2006]. Without further work, automation is limited to identifying when sound occurs in a video. This identification, of Quick Index automation, should decrease annotation time. However, the burden of identification and classification is on the shoulders of coders.

Additional automation can be done to detect the "breaks" between vocalization, and should aid in the breaking up of vocalizations. However, when there are simultaneous sounds (subject vocalizing while something is occurring in the room), meaningful (and separate) vocalizations may be linked together due to background noise. Thus, this breakup can be helpful, but at most, would provide a Reviewable automation.

6.2.2.2 *Nonspeech Sounds (Laugher vs. other).* Research on detection of laugher has mainly focused on audio processing [Melder et al. 2007]. Some new work has added the additional analysis of vision and audio processing to detect laughter [Petridis and Pantic 2008]. This new work demonstrates an 80% precession rating. This vein of research, in concert with presence of audio and emotion detection (Section 6.2.1.1), has the potential to act as a Quick Index or possibly a Review system (with suitable improvement in the technology). Detection of other types of nonspeech sounds will probably remain at a very course level of Quick Index automation, simply based on the presence of audio in a video. Though this will reduce time required for video analysis by allowing coders to skip over nonrelevant portions of video, current speech recognition systems are not able to differentiate between generic nonspeech sounds and those that are Speechlike.

6.2.2.3 *Speechlike Sounds (Imitative vs. Spontaneous).* The distinction between Imitative and Spontaneous can be automated to a small degree though analysis of the phonemes via speech detection [IBM 2005; Nuance Communications 2008] and syllable detection [Howitt 2000]. However, though this distinction may be somewhat effective, gauging appropriate start/stop of vocalizations falls under the same accuracy concerns in Section 6.2.2.1. Further, not all sounds come from a source that can be imitative. Thus, difficulties with

speaker identification [Nishida and Ariki 1998; Kwon and Narayanan 2004] are also problematic. The resulting automation, at best, is Quick Index identification, requiring a heavy degree of analysis, and examination of spaces that are both marked and non-marked. Until technology better supports a host of concerns outlined here, it is most like a Manual annotation system.

6.2.2.4 *Speechlike Sounds (Immediate vs. Delayed).* In general, computers should be able to differentiate between sounds with a short delay and sounds with a long delay. This can be accomplished through simple analysis of the presence of sound and the time between sounds. However, there is a great deal of review necessary due to concerns outlined in the above sections related to relevant sounds and appropriate division of overlapping sounds. Yet this additional automation should greatly help coders classify a majority of the sounds. They will still need to relisten to each vocalization to ensure it is divided appropriately. This level of Review automation is quite plausible and easily implemented.

6.2.2.5 *Vocalization and Orientation to Screen.* Given the state of facial orientation software discussed in Section 4.3.1.3, computer systems could go through a set of annotations made by coders and mark those vocalizations that were made while oriented at the screen given the A3 guideline. Post automation, coders can Review those marks, just to verify the occurrence of the orientation. Given the rapid improvement of computer vision technology, this degree of automation may become a Replacement for manual annotation.

6.2.2.6 *Turn Taking.* Much like the problems from speaker identification [Nishida and Ariki 1998; Kwon and Narayanan 2004], work analyzing turn taking requires identification of two sounds occurring simultaneously. For those experimental set ups that use computer-based audio, a simple logging system would allow for automation between the existence of sound in the video that occur at the same time that the computer logs generating the sound. However, without better source identification, the limitation of the automation is a set of Quick Index automated location detection, in which, coders can review events when there was video sound, and sound in the video.

However, for turn taking conditions with a human subject, automation is dependent upon technological advances in speaker identification systems. The main existing solution requires each subject to have a microphone, and use subtraction to determine overlap. Until the time of accurate and robust systems, coders must Manually annotate multiperson turn taking.

6.2.3 *BIGmack Switch.* By attaching a pressure sensor to a BIGmack Switch, computers can log this behavior, and Replace coders. Pressure sensors can also be applied to other external therapeutic and communicative devices (e.g., GoTalk 20+ [Attainment Company 2008]) to accurately log usage.

6.3 Implications of Automation

Based on the analysis of the existing technology, computers can greatly augment the video annotation process through automation. The creation or

modification of automated systems can improve accuracy and reduce coding time of the A3 annotation guidelines. We provide a summary table listing each variable with the degree of automation available (Table IV). Overall, computer based automation can aid coders by highlighting areas of meaningful activity, reducing the time required to view video segments that have no relevant behavior. For eight variables, automation is accurate enough to only require coders to validate the automated marks, thereby reducing time required and greatly improving accuracy. The resulting use of these techniques may be through individual monitoring systems that log specific behavior (e.g., checking visual orientation) a plug-in to a system like VCode that analyzes the video/audio streams, or a completely new package that provides multiple forms of automation for clinicians and research, aggregating and logging many data points. Technology is currently not advanced enough to fully remove coders from the video annotation process. However, as video and sound analysis techniques continue to improve, so will the burden placed on manual coding.

7. FORMS OF USE

Though demonstration of A3 came in the context of coding variables from video of children with ASD interacting with technology, we strongly believe that the application of A3 extends well beyond this particular situation.

7.1 Application External to ASD

Because this guideline focuses on nonverbal subjects interacting with technology, we believe that A3 can be applied to other areas of HCI research that target nonverbal subjects. This could include infants, non verbal subjects with verbal apraxia, autism, or other severe and profound disabilities. Because our system also focuses on computer feedback and categories of vocalizations, A3 could also be used in situations where subjects use augmentative and alternative communication (AAC) speech generation devices (e.g., Go-Talk [Attainment Company 2008]) [Reichle et al. 2002].

Beyond expanding A3 to multiple disorders, it can also be used for data driven clinical assessments and intervention. During speech and behavior therapy, researchers often incorporate different technological tools. This ranges from paper constructions to AAC devices and computerized speech feedback systems (the type of application used in the experimental context). In the therapeutic context, clinicians also must be able to assess progress made by subjects. A3 allows practitioners to assess changes in behavior, as well as the subjects' engagement with technological devices used in therapy.

7.2 Use of the A3 Guidelines

Though A3 does not cover every possible dependent variable for video analysis, it does provide a robust library of features to annotate. By understanding the interaction between subjects and computers through video annotation and A3, researchers can evaluate the software intervention itself. However, not all features of A3 may be applicable, or worth analyzing for every experiment. Thus, A3 is designed to act as a source for dependent variable selection. When

designing an experiment, researchers can select the most applicable set, definition, and justification of variables from the guideline, and use that sub set in their own research. By utilizing the four-pass system, researchers can extract passes that directly target the type of analysis they wish to perform.

Variables are not limited to simple frequency counts. They can be presented as rates of behavior (e.g., vocalizations/10 seconds) or ratios (e.g., Speechlike Vocalizations: Nonspeech Vocalizations). By analyzing ratios, researchers can assess the influence on multiple behaviors concurrently. A3 presents multiple related perspectives on many types of variables (e.g., Speechlike Vocalizations vs. Nonspeech Vocalizations). We believe that the strength of A3 is its flexibility allowing researchers to examine their own set of variables in the most efficient way, using a set of operationalized dependent variables.

7.3 Coding Time

Though the coding time for all 21 metrics was 20 minutes/1 minute of video, this was the most intensive form of video annotation required by the A3 system. As researchers refine their selection of variables, the time required to annotate video will be greatly reduced. In addition, improvements in microphone quality will also improve speed of analysis in that coders will not need to repeat as many segments of video for clarification. Time accelerators based on computer-based automation can be found in Section 6.

8. CONCLUSION AND FUTURE WORK

Research with subjects who are nonverbal poses challenges for researchers in terms of creating educational solutions and testing computer-based interventions. Without an existing and unified set of metrics for assessing these subjects' interaction with technology, a reliable and standardized method for comparing and contrasting the performance of different systems does not exist. This article presents A3 (Annotation for ASD Analysis), a collection of dependent variables for video annotation that can be used to assess the interaction between nonverbal subjects and computer based interventions. As researchers, it is incumbent upon us to faithfully describe our scheme and its relationship to other tools, and provide data to support its reliability in the field.

In addition to providing the operationalized coding process and detailed justification grounded in existing literature, this article presents an analysis of existing technology and how it can be used to automate the annotation of many of the variables. The degree of automation can reduce the time required for coders to annotate, in addition to increasing accuracy, and reducing costs for researchers. This analysis is grounded in the capabilities of existing technology, based upon a detailed examination of literature. Lastly, we present a detailed description of where and how the A3 coding guidelines can be used. For future experiments, we hope to design tools to further aid in the automation of A3 variables, as well as improve the description of some of the variables based on of the lessons learned. Overall, A3 provides researchers a unified source of

operationalized variables that can be used to assess the interaction between nonverbal subjects and computer-based intervention systems.

APPENDIX A. A^3 CODING GUIDE

The following is the coding guide distributed to coders. Included here is the full description of the behavior that must be observed to mark each variable. The variables are broken down into four passes, and the mode of video playback is also noted.

A^3 CODER GUIDE

*denotes ranged event

PASS 1: *Use Standard Playback Mode*

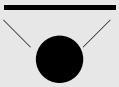
***Non-Child Audio:** Durations when audio/sound made & see on bars (you can hear it) and is NOT coming from the child /overlaps with child with child's sounds OR sound caught on volume bars, but is not identified as from child (unknown source). Include REGULAR heavy breathing. Mark whole segment if "contaminated"

PASS 2: *Use Interval Payback Mode (Continuous Playback)*

Smiling: When the child appears to be smiling during the past 3 seconds, and in the past 3 seconds, we at some point could see his face.

No Face: Can not see face (to determine smile) at all in the past 3 sec.

Oriented @ Screen: When the child, in the past 3 seconds, was facing towards the screen within 90 degrees. Spinning through the 90 degree arc should not be counted. Rather, time where the facing direction is within the arc for at least a "moment.



Auditory Focus: During the time frame, did the child get closer in proximity to, or touch the speaker. OR Child is not in the visual arc, in response to computer sounds, orient to the visual arc.

PASS 3: *Use Standard Playback Mode*

Use a 2 second pause between end and start to delineate between vocalizations
Also mark sounds even if not recorded by computer

***Laugh:** The sound should NOT be able to be transcribed as a speech like vocalization. To qualify as laughter it needs to be paired with a positive affect.

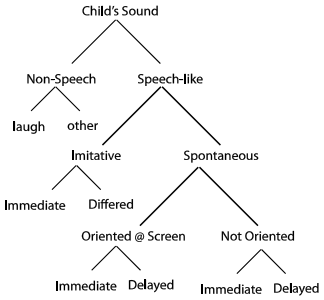
Non-Speech Vocalization: When a vocalization made is non speech, and is not a laugh. Includes gulps, screech, grunt, lip pops, ticks, heavy sighs, etc. Lasts for 2 seconds before code again.

***Speech Like Vocalizations:** When a vocalization is made by the child, the phonetic construction of the sound should be noted as an event. Marked at the start of the sound
Use the annotation hotkey to write the sound made
New Sounds are formed by gap of 2 seconds, OR separated by a non-speech sound, or laugh, or computer sound (while the child is not making a speech-like sound).

Turn taking: the computer makes a sound, and then the child starts sound if nucleus of final syllable (vowel) has been initiated
Speech Like Sounds Only

BIGmack Switch: When a child presses the switch, mark this at the start of the press. Do not mark when switch is pressed by anyone else.

For each speech like vocalization mark one of the following Correspondences.



Spontaneous: creating a non-imitative vocalization

I+S spontaneous (Immediate, Oriented @ Screen, Spontaneous): while oriented towards screen (within in 2 seconds of starting sound), speaker, within 5 seconds of end of source sound

D+S spontaneous (Delayed, Oriented @ Screen, Spontaneous): while oriented towards screen (within in 2 seconds of starting sound), speaker, after 5 seconds of end of source sound

I spontaneous (Immediate Spontaneous): while NOT oriented towards screen (within in 2 seconds of starting sound), speaker, within 5 seconds of end of source sound

D spontaneous (Delayed): while NOT oriented towards screen (within in 2 seconds of starting sound), speaker, after 5 seconds of end of source sound

Imitative: the child attempts to echo or repeat a sound previously heard/made by computer or human Must match 50% of phonemes of ATTEMPTED target OR same number of syllables as whole. Target resets after each new non-child sound. Cannot imitate self OR echoed sound

Differed imitation: time frame for Differed Imitation ends after a new sound is made by speaker, or researcher

Immediate imitation: within 5 seconds from source

PASS 4: *Drag Position for fast skim*

***Time In Chair:** marking times during the video when the child is seated in the chair. This includes having the child's butt on the chair, 2 legs on the chair (ie. sitting cross-legged, sitting on feet) or otherwise in the chair in a manner generally deemed "sitting". Not included is one leg on chair, one leg standing.

ACKNOWLEDGMENTS

We would like to thank all of the participants and their families, our four terrific coders (Ashley Sharer, Christine Holloway, Sammi Goldenberg, Christine Renee Birn), our friends, family, and loved ones.

REFERENCES

ALBERTO, P. AND TROUTMAN, A. 2005. *Applied Behavior Analysis for Teachers*. Prentice Hall, Upper Saddle River, NJ.

ALTMAN, D. 1991. *Practical Statistics for Medical Research*. Chapman and Hall, London.

ATTAINMENT COMPANY 2008. GoTalk. http://www.attainmentcompany.com/xcart/home.php.

BACH, F. R. AND JORDAN, M. I. 2006. Learning spectral clustering, with application to speech separation. *J. Mach. Learn. Res. 7*, 1963–2001.

BANERJEE, S., COHEN, J., ET AL. 2004. Creating multi-modal, user-centric records of meetings with the Carnegie Mellon meeting recorder architecture. In *Proceedings of the ICASSP Meeting Recognition Workshop*.

BASKETT, C. B. 1996. The effect of live interactive video on the communicative behavior in children with autism. M.S. Thesis, University of North Carolina at Chapel Hill.

BERTINI, M., DEL BIMBO, A., CUCCHIARA, R., AND PRATI, A. 2004. Semantic video adaptation based on automatic annotation of sport videos. In *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*. ACM New York. 291–298.

BIJOU, S., PETERSON, R., AND AULT, M. 1968. A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *J. Appl. Behav. Anal. 1*, 2, 175.

BIRKIMER, J. C. AND BROWN, J. H. 1979. A graphical judgmental aid which summarizes obtained and chance reliability data and helps assess the believability of experimental effects. *J. Appl. Behav. Anal. 12*, 4, 523–533.

BRADFORD-HEIT, A. AND DODD, B. 1998. Learning new words using imitation and additional cues: Differences between children with disordered speech. *Child Lang. Teach. Ther. 14*, 2, 159.

BURR, B. 2006. VACA: a tool for qualitative video analysis. *CHI'06 Extended Abstracts on Human Factors in Computing Systems*. ACM Press.

BYRT, T. 1996. How good is that agreement? *Epidemiol. 7*, 5, 561.

CASSELL, J., KOPP, S., TEPPER, P., FERRIMAN, K., AND STRIEGNITZ, K. 2007. *Trading Spaces: How Humans and Humanoids Use Speech and Gesture to Give Directions. Conversational Informatics*. John Wiley & Sons, New York, 133–160.

CHANG, E. AND MA, C. 2002. Implicit speech recognition: making speech a first class object on computers. In *Proceedings of the 2nd International Conference on Human Language Technology Research*. Morgan Kaufmann Publishers Inc.

CHEN, S., SHYU, M., LIAO, W., AND ZHANG, C. 2002. Scene change detection by audio and video clues. In *Proceedings of the IEEE International Conference On Multimedia and Expo (ICME'02)*.

CLIFFORD, J., MARCUS, G. E. AND SCHOOL OF AMERICAN RESEARCH. 1986. *Writing Culture*. University of California Press, Berkeley, CA.

CLIFTON, R. K., PERRISA, E. E., ET AL. 1999. Does reaching in the dark for unseen objects reflect representation in infants? *Infant Behavior and Development 22*, 3, 297–302.

CRANE, E. A., SHAMI, N. S. ET AL. 2007. Let's get emotional: emotion research in human computer interaction. CHI '07 extended abstracts on Human factors in computing systems. San Jose, CA, USA, ACM.

EL KALIOUBY, R. AND ROBINSON, P. 2004. Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures. Computer Society Conference on. Computer Vision and Pattern Recognition. Washington, DC, Springer. 3, 154.

FIELD, T., FIELD, T., ET AL. 2001. Children with Autism Display more Social Behaviors after Repeated Imitation Sessions. *Autism 5*, 3, 317–323.

GATHERCOLE, S. AND BADDELEY, A. 1993. Working Memory and Language, Psychology Pr.

GENA, A., KRANTZ, P., ET AL. 1996. Training and generalization of affective behavior displayed by youth with autism. *J. Appl. Behav. Anal. 29*, 3, 291–304.

GILLETTE, D., HAYES, G., ET AL. 2007. Interactive technologies for autism. CHI 07, San Jose, CA, ACM.

HAGEDORN, J., HAILPERN, J., ET AL. 2008. VCode and VData: Illustrating a New Framework for Supporting the Video Annotation Work? Advanced Visual Interfaces. Napoli, Italy, ACM-PRESS.

HAILPERN, J., KARAHALIOS, K., ET AL. 2009. Creating a Spoken Impact: Encouraging vocalization through audio visual feedback in children with ASD. CHI 2009. Boston, MA, ACM.

HAILPERN, J., KARAHALIOS, K., ET AL. 2008. A3: A Coding Guideline for HCI+Autism Research using Video Annotation. ACM SIGACCESS- ASSETS 2008. Halifax, Canada, ACM-PRESS.

HALLE, J. 1987. Teaching language in the natural environment: An analysis of spontaneity. *J. Assoc. Pers. Sev. Handicaps 12*, 1, 28–37.

HAYNE, H., GROSS, J., ET AL. 2000. Repeated reminders increase the speed of memory retrieval by 3-month-old infants. *Dev. Sci. 3*, 3, 312–318.

HOWITT, A. W. 2000. Automatic syllable detection for vowel landmarks, Massachusetts Institute of Technology 1.

HOWLIN, P. 1986. An Overview of Social Behavior in Autism. Social Behavior in Autism. E. Schopler and G. B. Mesibov. New York, NY, Plenum. 113–115.

IBM. 2005. IBM ViaVoice.

JACOB, R. J. K. (1991). The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. Inf. Syst. 9*, 2, 152–169.

JOHNSON, A. 2007. About vprism video data analysis software. 2007, from http://www.camse.org/andy/VP/vprism.htm.

JONES, C. M. AND TROEN, T. 2007. Biometric valence and arousal recognition. Proceedings of the 2007 conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: Activities, artifacts and environments. Adelaide, Australia, ACM.

KALIOUBY, R. E. AND TEETERS, A. 2007. Eliciting, capturing and tagging spontaneous facial affect in autism spectrum disorder. In *Proceedings of the 9th International Conference on Multimodal Interfaces*. Nagoya, Aichi, Japan, ACM.

KAZDIN, A. 1977. Artifact, bias, and complexity of assessment: the ABCs of reliability. *J. Appl. Behav. Anal. 10*, 1, 141.

KAZDIN, A. E. 1982. Single-Case Research Designs: Methods for Clinical and Applied Setting. USA, Oxford University Press.

KERR, S. J., NEALE, H. R., ET AL. 2002. Virtual environments for social skills training: The importance of scaffolding in practice. In *Proceedings of the fifth international ACM conference on Assistive technologies*. Edinburgh, Scotland, ACM Press.

KIPP, M. 2007. Anvil - the video annotation research tool.

KOEGEL, R. L., CAMARATA, S., ET AL. 1998. Increasing speech intelligibility in children with autism. *J. Autism Dev. Disord. 28*, 3, 241–251.

KWON, S. AND NARAYANAN, S. 2004. Speaker Model Quantization for Unsupervised Speaker Indexing International Conference Spoken Language Processing Jeju, Korea.

LEE, L. 1974. Developmental Sentence Analysis. Evanston, IL, Northwestern University Press.

LORD, C., S. RISI, ET AL. 2000. The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord. 30*, 3, 205–223.

LOVAAS, O. I. 2003. Teaching Individuals with Developmental Delays: Basic Intervention Techniques. Austin, TX, PRO-ED, Inc.

LUO, Y. AND BAILLARGEON, R. 2005. Can a Self-Propelled Box Have a Goal? *Psychol. Sci. 16*, 8, 601–608.

MADSEN, M., KALIOUBY, R. E., ET AL. 2008. Technology for just-in-time in-situ learning of facial affect for persons diagnosed with an autism spectrum disorder. In *Proceedings of the*

10th International ACM SIGACCESS Conference on Computers and Accessibility. Halifax, Nova Scotia, Canada, ACM.

MCCALLA, D. D. AND CLIFTON, R. K. 1999. Infants' means-end search for hidden objects in the absence of visual feedback. *Infant Behav. Dev. 22*, 2, 179–195.

MELDER, W. A., TRUONG, K. P., ET AL. 2007. Affective multimodal mirror: Sensing and eliciting laughter. In *Proceedings of the International Workshop on Human-Centered Multimedia*. Augsburg, Bavaria, Germany, ACM.

MICHAUD, F. AND THÉBERGE-TURMEL, C. 2002. Mobile robotic toys and autism. In *Socially Intelligent Agents - Creating Relationships with Computers and Robots*. K. Dautenhahn Ed., Springer, 125–132.

NISHIDA, M. AND ARIKI, Y. 1998. Real time speaker indexing based on subspace method-application to TV News Articles and Debate. In *Proceedings of the 5th International Conference on Spoken Language Processing*.

NOLDUS INFORMATION TECHNOLOGY. 2007. The observer. www.noldus.com/.

NUANCE COMMUNICATIONS 2008. Dragon NaturallySpeaking. www.nuance.com/naturallyspeaking/.

OLSSON, R. K. AND HANSEN, L. K. 2006. Linear state-space models for blind source separation. *J. Mach. Learn. Res. 7*, 2585–2602.

OSADCHY, M., CUN, Y. L., AND MILLER, M. L. 2007. Synergistic face detection and pose estimation with energy-based models. *J. Mach. Learn. Res. 8*, 1197–1215.

OWENS, R. E. 2007. *Language Development: An Introduction* 7th Ed. Allyn & Bacon, Boston, MA.

PARÉS, N., CARRERAS, A., ET AL. 2005. Promotion of creative activity in children with severe autism through visuals in an interactive multisensory environment. In *Proceedings of the Conference on Interaction Design and Children*. ACM Press.

PERLMAN, A. 2008. Speech and hearing science proseminar. University of Illinois, Champaign, IL.

PETRIDIS, S. AND PANTIC, M. 2008. Fusion of audio and visual cues for laughter detection. In *Proceedings of the International Conference on Content-Based Image and Video Retrieval*. ACM Press.

PIPER, A., O'BRIEN, E., MORRIS, M., AND WINOGRAD, T. 2006. SIDES: a cooperative tabletop computer game for social skills development. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM Press.

PONCELEON, D. AND SRINIVASAN, S. 2001. Automatic discovery of salient segments in imperfect speech transcripts. In *Proceedings of the 10th International Conference on Information and Knowledge Management*. ACM Press, 490–497.

PRIZANT, B. M., SCHULER, A. L., WETHERBY, A. M., AND RYDELL, P. 1997. Enhancing language and communication: Language approaches. In *Handbook of Autism and Pervasive Developmental Disorders 2nd Ed.* D. Cohen and F. Volkmar, Eds. Wiley, New York.

QUEK, F., MCNEILL, D., ROSE, T., AND SHI, Y. 2003. A coding tool for multimodal analysis of meeting video. In *Proceedings of the NIST Meeting Room Workshop*.

RAPIN, I. AND DUNN, M. 1997. Language disorders in children with autism. *Semin. Pediatr. Neurol. 4*, 2, 86–92.

REICHLE, J., BEUKELMAN, D., AND LIGHT, J. 2002. *Implementing an Augmentative Communication System: Exemplary Strategies for Beginning Communicators.* Brookes Publishing Company, Baltimore, MD.

REID, D. H., PARSONS, M. B., MCCARN, J. E., GREEN, C. W., PHILLIPS, J. F., AND SCHEPIS, M. M. 1985. Providing a more appropriate education for severely handicapped persons: increasing and validating functional classroom tasks. *J. Appl. Behav. Anal. 18*, 4, 289–301.

RETHERFORD, K., SEWARDS, D., AND HESS, L. 1993. *Guide to Analysis of Language Transcripts*, Thinking Publications, Eau Claire, WI.

ROSENBLUM, K., ZEANAH, C., MCDONOUGH, S., AND MUZIK, M. 2004. Video-taped coding of working model of the child interviews: a viable and useful alternative to verbatim transcripts? *Infant Behav. Dev. 27*, 4, 544–549.

SAJWAJ, T., TWARDOSZ, S., AND BURKE, M. 1972. Side effects of extinction procedures in a remedial preschool. *J. Appl. Behav. Anal. 5*, 2, 163–175.

SALT SOFTWARE 2007. www.languageanalysislab.com/salt/.

SEGAL, L. B., OSTER, H. COHEN, M., CASPI, B., MYERS, M., AND BROWN, D. 1995. Smiling and fussing in seven-month-old preterm and full-term black infants in the still-face situation. *Child Dev. 66*, 6, 1829–1843.

SETLUR, V. AND GOOCH, B. 2004. Is that a smile? Gaze dependent facial expressions. In *Proceedings of the 3rd International Symposium on Non-Photorealistic Animation and Rendering*. ACM Press.

SHEINKOPF, S. J., MUNDY, P., OLLER, D. K., AND STEFFENS, M. 2000. Vocal atypicalities of preverbal autistic children. *J. Autism Dev. Disord. 30*, 4, 345–354.

STUDIOCODE BUSINESS GROUP. 2007. Studiocode business group - supplier of studiocode and stream video analysis and distribution software. http://www.studiocodegroup.com.

SUCHMAN, L. A. 1987. *Plans and Situated Actions: The problem of Human-Machine Communication*. Cambridge University Press.

TARTARO, A. AND CASSELL, J. 2008. Playing with virtual peers: Bootstrapping contingent discourse in children with autism. In *Proceedings of the International Conference of the Learning Sciences*. ACM Press.

VALSTAR, M. F., GUNES, H., AND PANTIC, M. 2007. How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th International Conference on Multimodal Interfaces*. ACM Press.

VALSTAR, M. F., PANTIC, M., AMBADAR, Z., AND COHN, J. F. 2006. Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. In *Proceedings of the 8th International Conference on Multimodal Interfaces*. ACM Press.

WANG, J., YIN, L., AND MOORE, J. 2007. Using geometric properties of topographic manifold to detect and track eyes for human-computer interaction. *ACM Trans. Multimed. Comput. Commun. Appl. 3*, 4, 1–20.

WETHERBY, A. M., PRIZANT, B. M., AND HUTCHINSON, T. A. 1998. Communicative, social/affective, and symbolic profiles of young children with autism and pervasive developmental disorders. *Am. J. Speech-Lang. Pathol. 7*, 79–91.

WHYTE, W. 1980. *The Social Life of Small Urban Spaces*. Conservation Foundation, Washington, DC.

WOODS, J. J. AND WETHERBY, A. M. 2003. Early identification of and intervention for infants and toddlers who are at risk for autism spectrum disorder. *Lang. Speech Hear. Serv. Sch. 34*, 180–193.

XIAO, Z., POURSOLTANMOHAMMADI, A., AND SORELL, M. 2008. Video motion detection beyond reasonable doubt. In *Proceedings of the 1st International Conference on Forensic Applications and Techniques in Telecommunications, Information, and Multimedia and Workshop*.

ZACKS, J., TVERSKY, B., AND IYER, G. 2001. Perceiving, remembering, and communicating structure in events. *J. Exper. Psych. Gen. 130*, 1, 29–58.

ZENG, Z., PANTIC, M., ROISMAN, G. I., AND HUANG, T. S. 2007. A survey of affect recognition methods: audio, visual and spontaneous expressions. In *Proceedings of the 9th International Conference on Multimodal Interfaces*. ACM Press.

ZHANG, Z. AND KURTEV, S. 2003. Independent motion detection directly from compressed surveillance video. In *Proceedings of the 1st ACM SIGMM International Workshop on Video Surveillance*. ACM Press.

ZHAO, W., CHELLAPPA, R., PHILLIPS, P. J., AND ROSENDELD, A. 2003. Face recognition: A literature survey. *ACM Comput. Surv. 35*, 4, 399–458.

Received November 2008; revised February 2009; accepted February 2009.