The USC CARE Corpus:
Child-Psychologist Interactions of Children with Autism Spectrum Disorders

Matthew P. Black[1,†], *Daniel Bone*[1,†], *Marian E. Williams*[2,†],
Phillip Gorrindo[3], *Pat Levitt*[4,†], *Shrikanth S. Narayanan*[1,†]

[1]Signal Analysis and Interpretation Laboratory (SAIL), Viterbi School of Engineering
[2]University Center for Excellence in Developmental Disabilities, Keck School of Medicine
[3]Medical Scientist Training Program, Vanderbilt University, Nashville, TN, USA
[4]Zilkha Neurogenetic Institute, Keck School of Medicine
[†]University of Southern California (USC), Los Angeles, CA, USA

http://sail.usc.edu

Abstract

We introduce the USC CARE Corpus, comprised of sponta-neous and standardized child-psychologist interactions of chil-dren with a diagnosis of an autism spectrum disorder (ASD). The audio-video data is collected in the context of the Autism Diagnostic Observation Schedule (ADOS), which is a tool used by psychologists for a research-level diagnosis of ASD for chil-dren. The interaction consists of developmentally appropri-ate semi-structured social activities, providing the psychologist with a sample of behavior used to rate the child on a series of autism-relevant symptoms. Our goal with this multimodal cor-pus is to investigate how analytical technology (e.g., speech and language processing) can enhance this observational rating task and provide greater insight into social behavior and communi-cation. We provide demographic statistics on the recruited chil-dren (60 to date), describe the multimodal recording set-up, and discuss current and future work for this novel corpus.

Index Terms: autism, autism spectrum disorders, children's speech, corpus, multimodal recordings, prosody

1. Introduction

Autism spectrum disorders (ASD) are highly heritable neurode-velopmental disorders characterized by a triad of core deficits, including impaired social behaviors, communication, and re-stricted/repetitive behaviors [1]. ASD is considered a "spec-trum" disorder because symptomatology severity in each of the core domains can vary greatly. There have been increased re-search efforts in ASD, as recent prevalence studies indicate that as many as 1 in 110 children are diagnosed with ASD [2]. Stud-ies have shown that early diagnosis and intensive early inter-vention can lead to improved social and communication skills in autistic children [3].

Psychologists in both research and in practice rely heavily upon observational methods for the assessment of social and communicative abilities. The Autism Diagnostic Observation Schedule (ADOS) is one of the most widely used clinical re-search instruments for the assessment and diagnosis of ASD and is appropriate for individuals with varying ages and verbal abil-ities [4, 5]. The semi-structured 30-60 minute interaction pro-vides a trained psychologist with behavioral evidence that can be evaluated along dimensions important in diagnosing autism. Because of the qualitative descriptions of the assessments and the lack of continuous quantitative measures, one challenge in using the ADOS (and with observational methods in general) is the subjective nature inherent to the rating system.

Technology can assist with this process in a number of ways. Audio-video sensors can record the child-clinician in-teraction in a consistent fashion, and state-of-the-art signal pro-cessing methods can facilitate quantitative analyses and mod-eling using the audio-video data. Computational methods may be better suited than human observers in quantitatively tracking certain human behavioral cues (e.g., speech prosody, hand ges-tures). In recent years, there have been new emerging fields (e.g., social signal processing [6], behavioral signal process-ing [7]) concentrating on robustly measuring high-level human behaviors during realistic interactions using audio-video data. These data-driven signal cues could provide researchers and clinicians with a quantitatively dynamic source of information. We wish to emphasize that the clinical acumen of experienced psychologists is invaluable in evaluating and diagnosing chil-dren with ASD, and that we anticipate our work with this corpus will augment, rather than supplant, an expert clinician.

The collection of realistic corpora is a critical step in many data-driven engineering pattern analysis and recognition realms. Example domains include automatic speech recogni-tion [8, 9], affect/emotion recognition [10], and automatic lit-eracy assessment [11]. In recent years, there has been signifi-cant engineering-related work on analyzing the speech and lan-guage of children with ASD; the experiments have occurred in a variety of social contexts, ranging from isolated speaking tasks [12, 13] and structured clinical assessments [14] to uncon-strained home environments [15]. However, since ASD affects vocal, linguistic, and gestural social behavioral patterns, there is a need for multimodal data of children with ASD.

Towards this end, we introduce the *USC CARE Corpus*, comprised of real, spontaneous child-psychologist interactions, recorded in a controlled clinical environment in the context of the ADOS. The collection of this corpus is the necessary first step in analyzing complex social interactions between expert psychologists and children with ASD. We plan to use the audio-video data for a number of multimodal signal processing re-search projects, ranging from improved modeling of children's spontaneous speech to the analysis of atypical communication patterns and the study of dialogs in clinical settings.

In addition to offering a new problem domain for engineer-ing, this unique corpus has important potential contributions to the ASD community. Ultimately, this research could help sup-

28 – 31 August 2011, Florence, Italy

Module	Subject	Interaction subtasks
1	< phrase speech	free play, response to name, response to joint attention, bubble play, anticipation of routine with objects, responsive social smile, anticipation of social routine, functional and symbolic imitation, birthday party, snack
2	phrase speech	construction task, response to name, make-believe play, joint interactive play, conversation, response to joint attention, demonstration task, description of picture, telling story from book, free play, birthday party, snack, anticipation of routine with objects, bubble play
3	fluent children	construction task, make-believe play, joint interactive play, demonstration task, description of picture, telling story from book, cartoons, emotions, social difficulties/annoyance, friends and marriage, loneliness, creating a story
4	fluent teens/adults	construction task, description of picture, telling story from book, cartoons, emotions, social difficulties/annoyance, friends and marriage, loneliness, creating a story, daily living, current work/school, plans and hopes

Table 1: *A list of the interaction subtasks for each ADOS module [4], along with the intended language level and/or age of the subject.*

port ASD diagnoses with quantifiable and adaptable metrics, provide more accurate stratification of subgroups for targeted interventions, and automatically track children's progress during the treatment. Section 2 discusses the ADOS interaction paradigm in more detail. We describe the USC CARE Corpus (an ongoing data collection) in Section 3, and we explain our initial analyses and future intended work in Section 4. We conclude in Section 5.

2. ADOS Interaction Paradigm

The ADOS is a "gold-standard" research tool for the assessment of the triad of behaviors that together are diagnostic for ASD [4, 5]. There are four modules; the psychologist determines which one to administer, depending on the subject's expressive language level and chronological age (see Table 1).

During the ADOS, the subject interacts spontaneously with a psychologist (and a parent for modules 1 and 2) for approximately 30 to 60 minutes. To ensure that the interaction is standardized, the psychologist follows a predetermined semistructured set of subtasks. Table 1 lists the interaction subtasks for each ADOS module. This table shows that there is significant subtask overlap between the four modules, with more conversational and interview-style subtasks for the fluent-speaking subjects in modules 3 and 4. Modules 3 and 4 are typically administered at a table, while modules 1 and 2 require the child (and parent) to move around the room.

The ADOS was designed for psychologists to make assessments on the subject's proficiency for a number of communication and social interaction skills. Observations are noted by the psychologist in real-time during the interaction, and the psychologist rates the subject's behavior immediately after the session according to the module-specific ADOS coding manual. Each coding manual consists of approximately 28 codes, which are broken down into five main groupings (e.g., communication, reciprocal social interaction, play/imagination/creativity, stereotyped behaviors/restricted interests, and other abnormal behaviors). The codes assess speech (e.g., speech abnormalities/atypical prosody), language (e.g., stereotyped/idiosyncratic use of words/phrases), nonverbal communication (e.g., directed facial expressions, eye contact, use of gestures), and other behaviors (e.g., imagination/creativity, overall quality of rapport).

Each code has a written description, and the psychologist chooses the value that best describes the subject's behavior; while in some cases codes are based on single subtasks, most codes consider overall behavior throughout the evaluation. The summary ADOS algorithm includes those codes that were shown in the standardization research to best predict an autism diagnosis. The algorithm codes are summed to attain communication and social interaction subtotals (each with predetermined autism and ASD cut-offs), and a total score is computed for a final ADOS classification. In addition, the clinician administering the test is asked to give an overall diagnosis of autism/ASD,

based on the ADOS scores as well as other information that may influence the validity or interpretation of the scores [4].

Psychologists are trained to administer and code the ADOS using a stringent training protocol that includes reaching agreement on multiple ADOS protocols with an expert evaluator. This training process is time-consuming and challenging. One of the main challenges is due to the qualitative nature for some of the codes in the ADOS coding manuals. For example, in modules 2-4, speech abnormalities (i.e., atypical prosody) are coded based on descriptions such as, "Little variation in pitch and tone; rather flat or exaggerated intonation, but not obviously peculiar, or slightly unusual volume, and/or speech that tends to be somewhat unusually slow, fast, or jerky." From this qualitative description, it is clear that coding for prosodic abnormalities requires knowledge of normative prosody and the range of atypicalities associated with ASD. Thus, it takes a great deal of specialized training to learn to reliably score the ADOS. Moreover, the scoring is categorical and does not provide continuous measures that can be utilized for population stratification.

Our goal is to contribute to overcoming some of these limitations through the development of engineering algorithms and tools that are based on the analysis of a large corpus of ADOS sessions. Incorporating engineering methods that quantitatively assess communication/social interaction behaviors could help support scalability and the analysis and decision capabilities of psychologists. In addition, it has the potential to contribute to research aimed at better understanding variations in the communication and social patterns of children with ASD.

3. The USC CARE Corpus

3.1. Background

The Center for Autism Research in Engineering (CARE) was established in 2009 at the University of Southern California (USC), with the goal to better incorporate engineering and computer science methodologies into autism research through interdisciplinary collaborations. Early and ongoing work between the Signal Analysis and Interpretation Laboratory (SAIL) at USC and the USC University Center for Excellence in Developmental Disabilities at Children's Hospital Los Angeles (CHLA) experimented with child-computer interaction applications [16, 17].

More recently, we have teamed up with researchers at the Zilkha Neurogenetic Institute and the Boone Fetter Clinic (BFC) at CHLA to record ADOS evaluations for an ongoing prospective clinical and genetics study on the relationship between ASD and gastrointestinal dysfunction (GID). The subsequent ADOS sessions from the Los Angeles-based families who agreed to be recorded for the study make up the *USC CARE Corpus*.

Category	Count/Statistic
Age (years)	*mean*: 9.3, *std. dev.*: 3.1, *range*: 5.2-17.0
Gender	*male*: 49, *female*: 11
Native language	*Spanish*: 38, *English*: 22
Ethnicity	*Hispanic/Latino*: 34, *White*: 9, *Other*: 8, *unk*: 9
ADOS module	*#1*: 17, *#2*: 13, *#3*: 28, *#4*: 2
ADOS diagnosis	*autism*: 37, *ASD*: 9, *no ASD*: 8, *unk*: 6

Table 2: *Demographic statistics of the 60 recorded subjects administered the ADOS (as of March, 2011). The unknown ("unk") entries have not been uploaded to the database yet.*

3.2. Recruited Subjects

All recruited subjects with a prior clinical diagnosis of ASD were administered the ADOS (to verify the diagnosis). The subjects were required to be between 5 and 18 years of age to participate in the study. In addition, all participating families were required to be native speakers of either English or Spanish. As part of the study, the parents filled out a number of standardized questionnaires (e.g., on their child's verbal abilities and social functioning), which are included as part of the corpus.

We began recording the ADOS evaluations in April, 2010. As of March, 2011, we have collected data from 60 subjects; our goal is to record 100 subjects. Table 2 provides statistics on the participants whose demographics have been uploaded to the database. The majority of the recruited families were native speakers of Spanish because the experiments took place in Los Angeles, California. Note that the gender imbalance in recruited subjects is due to the fact that males are four to five times more likely than females to be diagnosed with ASD [2].

3.3. ADOS Codes

Three research-certified psychologists administered the ADOS evaluations; co-author M. E. Williams was the lead psychologist, and she oversaw the training of the other two psychologists. The administering psychologist coded the subject according to the module-specific ADOS manual (Section 2). The resulting code scores are a critical part of the USC CARE Corpus, since they represent standardized expert coding of ASD-relevant behaviors. We also have the final ADOS diagnosis from the psychologist; most of the subjects met the autism or ASD cut-offs (Table 2).

Some of our intended future work will make use of these code scores and final ADOS classification. This information can be used to cluster children with similar characteristics and/or to facilitate the use of supervised learning techniques to automatically categorize typical from atypical behavior. As part of our future work, we may also collect similar ADOS data from typically developing children, which would allow us to train normative behavioral models.

3.4. Multimodal Data Collection

All ADOS evaluations took place in the BFC at CHLA, with floor dimensions of 3.3m x 2.3m and ceiling height of 2.6m. The BFC is a shared multi-use clinical space, so we used a portable smart-room solution with multiple audio-video sensors to unobtrusively record the interaction. All of the sensors operated in the far-field to ensure that they were not disruptive to the natural flow of the interaction and to maximize the ecological validity of the experiments. In addition, we did not place any sensors directly on the subject or the subject's clothing, to minimize the possibility of exacerbating anxiety states (many individuals with ASD have anxiety and sensory sensitivities [2]).

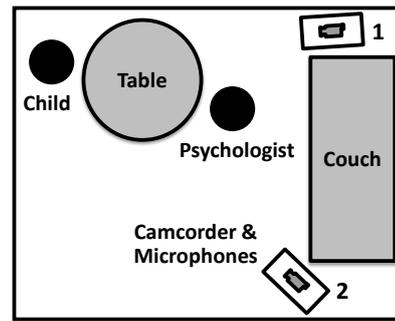Two Sony HDR-SR12 High Definition Handycam Cam-

Figure 1: *The layout of the clinical room, showing the location of the participants and the placement of the camcorders and microphones.*

corders were mounted on tripods approximately 1m off the floor and 2m from the subject. They operated in the corners of the room opposite to the child to capture the child's body and face while seated at the table. See Figure 1 for the layout of the clinical room. We used the highest quality video settings: 1080i resolution with a 16:9 widescreen aspect ratio (1920x1080 pixels, 59.94 interlaced frames per second, H.264/MPEG-4 AVC compression).

We recorded stereo audio signals from each camcorder's internal microphones (48 kHz, 16-bit). In addition, we recorded audio from two high-quality directional shotgun microphones (SCHOEPS CMIT 5 U), which were mounted next to the camcorders. We used the Edirol R-4 Pro recorder to capture the uncompressed audio (48 kHz, 24-bit).

The audio-video equipment takes approximately ten minutes to set up and five minutes to dismantle. All audio-video signals were synchronized to the nearest frame of video by clapping before and after the sessions and manually marking these clap times in each of the audio-video signals. We currently have 50 hours of data (for each of the two channels of video and six channels of audio); the average session duration is 49.5 minutes.

4. Initial Analyses & Future Work

4.1. Manual Transcription & Language Processing

We are in the process of manually transcribing and segmenting the USC CARE Corpus into individual speaker turns; these transcribers are blinded to the code scores and final ADOS classification. The resulting transcriptions will provide us reference landmarks indicating when each person was speaking, the word-level lexical content of the speech, and enriched transcription of: partial-words, stuttering, disfluencies, nonverbal *communication* (e.g., laughs, sighs), nonverbal *vocalizations* (e.g., grunts, babbles), mispronunciations, and neologisms. In addition, each utterance is labeled as a question, fragment, interruption, and/or complete thought. The transcription manual we devised was adapted from the SALT transcription guidelines [18], used in many ASD-related studies.

We plan to use these transcriptions to analyze the children's language use, turn-taking trends, and other surface behaviors, as was done in [14]. Importantly, the transcriptions can also be used to train a number of speech-related signal processing tools (voice activity detector, acoustic models, language models) for the specific environmental conditions and speaker demographics of the corpus.

4.2. Speech Signal Processing

Speech is one modality in which state-of-the-art signal processing can make a profound impact. As highlighted in Section 2,

the assessment of children's speech is one important aspect of the ADOS. There are a variety of atypicalities associated with the prosody of verbal autistic children. Studies have suggested that individuals with ASD have problems with lexical stress and pragmatic prosody [19]. Others have reported durational abnormalities, with the speech either too fast or slow [20]. Consistent with Kanner's description more than 60 years ago [21], current listeners often report a "bizarre" quality to the speech (e.g., monotonous intonation) [19, 22].

One of our future goals with this corpus is to develop automated methods to assess the various dimensions of prosody (e.g., rate, intonation, volume) within the context of the ADOS. While significant work has been done on assessment of prosody during children's constrained speaking tasks [12, 13], the USC CARE Corpus presents new challenges and opportunities due to the spontaneous nature of the speech. As part of this future work, we will also have to account for the rich linguistic diversity of the recruited subjects (Table 2).

We also plan to develop methods to automatically detect nonverbal vocalizations (e.g., grunts, shrieks, babbles) during the ADOS. It was found that these non-lexical vocalizations were a key feature in automatically separating typically developing children from those diagnosed with ASD [15]. This is a challenging learning problem, since these behaviors are idiosyncratic and can be rarely occurring events.

4.3. Multimodal Signal Processing

One of the unique aspects of the USC CARE Corpus is the multiple channels of synchronized audio and video. As discussed in Section 2, the assessments for the ADOS, and the social communication impairments for ASD, are inherently multimodal. In fact, some of the behavioral abnormalities are due to atypical synchrony between expressive modalities. Therefore, multimodal signal processing techniques are needed to fully capture the unique differences in communication of children with ASD.

Our initial plans will concentrate on processing specific subtasks within the ADOS interaction (Table 1). We will first consult with the psychologists to discover the cues most relevant for the subtasks. This will help inform the multimodal feature extraction and an appropriate automatic learning method. We can validate our methods using the psychologists' codes (Section 3.3). In some cases, we will manually code events at a finer-grained temporal scale, which we can use to train (and evaluate) specific behavioral detectors/classifiers. Combining data-driven and expert-inspired knowledge in a multimodal signal processing framework also is an area of future work.

5. Conclusions

In this paper, we introduced the USC CARE Corpus, a large multimodal corpus of ADOS evaluations. This data is important to facilitate the analysis of complex interactions involving children with ASD while in a controlled clinical environment. In addition to describing the unique elements of the ADOS and the multimodal recording set-up, we also provided an outline for future work in speech, language, and multimodal signal processing with this novel corpus.

We believe that the collection of the USC CARE Corpus represents a key step towards better incorporating engineering methodologies into the behavioral sciences and health-care related domains, including neurodevelopmental disorders. This is a primary goal of behavioral signal processing.

6. Acknowledgements

7. References

[1] American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Ed. 4 text revision (DSM-IV-TR), *American Psychiatric Association*, Washington, D.C., 2000.

[2] National Center on Birth Defects and Developmental Disabilities. (2010, March) Autism Spectrum Disorders (ASDs). Centers for Disease Control and Prevention. [Online]. Available: http://www.cdc.gov/ncbddd/autism/

[3] G. Dawson, S. Rogers, J. Munson, M. Smith, J. Winter, J. Greenson, A. Donaldson, and J. Varley, "Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model," *Pediatrics*, vol. 125, no. 1, pp. 17–23, 2010.

[4] C. Lord, S. Risi, L. Lambrecht, E. Cook, B. Leventhal, P. DiLavore, A. Pickles, and M. Rutter, "The Autism Diagnostic Observation Schedule–Generic: A standard measure of social and communication deficits associated with the spectrum of autism," *J. of Autism and Developmental Disorders*, vol. 30, pp. 205–223, 2000.

[5] K. Gotham, S. Risi, A. Pickles, and C. Lord, "The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity," *Journal of Autism and Developmental Disorders*, vol. 37, no. 4, pp. 613–627, 2007.

[6] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, pp. 1743–1759, 2009.

[7] M. P. Black, A. Katsamanis, C.-C. Lee, A. C. Lammert, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Automatic classification of married couples' behavior using audio features," in *Proc. Interspeech*, 2010.

[8] J. J. Godfrey, E. C. Holliman, and J. McDaniel, "SWITCHBOARD: Telephone speech corpus for research and development," in *Proc. ICASSP*, 2002.

[9] K. Shobaki, J.-P. Hosom, and R. A. Cole, "The OGI kids' speech corpus and recognizers," in *Proc. ICSLP*, 2000.

[10] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Language Resources and Evaluation*, vol. 42, no. 4, pp. 335–359, 2008.

[11] A. Kazemzadeh, H. You, M. Iseli, B. Jones, X. Cui, M. Heritage, P. Price, E. Anderson, S. S. Narayanan, and A. Alwan, "TBALL data collection: The making of a young children's speech corpus," in *Proc. Interspeech*, 2005.

[12] J. P. H. van Santen, E. T. Prud'hommeaux, and L. M. Black, "Automated assessment of prosody production," *Speech Communication*, vol. 51, no. 11, pp. 1082–1097, 2009.

[13] J. P. H. van Santen, E. T. Prud'hommeaux, L. M. Black, and M. Mitchell, "Computational prosodic markers for autism," *Autism*, vol. 14, no. 3, pp. 215–236, 2010.

[14] P. A. Heeman, R. Lunsford, E. Selfridge, L. Black, and J. van Santen, "Autism and interactional aspects of dialogue," in *SIGdial Meeting on Discourse and Dialogue*, Tokyo, Japan, Sept. 2010.

[15] D. K. Oller, P. Niyogi, S. Gray, J. A. Richards, J. Gilkerson, D. Xu, U. Yapanel, and S. F. Warren, "Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development," *Proceedings of the National Academy of Sciences*, vol. 107, no. 30, pp. 13 354–13 359, 2010.

[16] M. P. Black, E. Flores, E. Mower, S. S. Narayanan, and M. E. Williams, "Comparison of child-human and child-computer interactions for children with ASD," in *IMFAR*, 2010.

[17] E. Mower, M. P. Black, E. Flores, M. E. Williams, and S. S. Narayanan, "Rachel: Design of an emotionally targeted interactive agent for children with autism," in *ICME*, 2011.

[18] J. Miller and A. Iglesias, "Systematic Analysis of Language Transcripts (SALT)," Language Analysis Lab, University of Wisconsin-Madison, 2006, (Version 9) [Computer Software].

[19] J. McCann and S. Peppe, "Prosody in autism spectrum disorders: a critical review," *International Journal of Language & Communication Disorders*, vol. 38, no. 4, pp. 325–350, 2003.

[20] J. J. Diehl, D. Watson, L. Bennetto, J. McDonough, and C. Gunlogson, "An acoustic analysis of prosody in high-functioning autism," *Applied Psycholinguistics*, vol. 30, pp. 385–404, 2009.

[21] L. Kanner, "Autistic disturbances of affective contact," *Nervous Child*, vol. 2, pp. 217–250, 1943.

[22] W. H. Fay and A. L. Schuler, *Emerging Language in Autistic Children*. Baltimore, MD: University Park Press, 1980.