Automatic Childhood Autism Detection by Vocalization Decomposition with Phone-like Units

Dongxin Xu, Jeffrey A. Richards, Jill Gilkerson, Umit Yapanel, Sharmistha Gray, John Hansen*
LENA Foundation, 5525 Central Avenue #100, Boulder, CO 80301, USA
dongxinxu,jeffrichards,jillgilkerson@lenafoundation.org
*Center of Robust Speech Systems, University of Texas at Dallas, Richardson, TX 75080, USA
John.Hanseng@utdallas.edu

ABSTRACT

Autism is a major child development disorder with a prevalence of 1/150 in the US [22]. Although early identification is crucial to early intervention, there currently are few efficient screening tools in clinical use. This study reports a fully automatic mechanism for child autism detection/screening using the LENA^{TM} (Language ENvironment Analysis) System, which utilizes speech signal processing technology to analyze and monitor a child's natural language environment and the vocalizations/speech of the child. We previously reported preliminary results in [19] using child vocalization composition information generated automatically by the LENA System employing an adult phone model. In this paper, some extensions have been made, including enlargement of the dataset, introduction of a new child vocalization decomposition with the k-means clusters derived directly from the child vocalizations, and its combination with the previous decomposition. The experiment and comparison consistently shows that the child vocalization composition contains rich discriminant information for autism detection. It also shows that the child vocalization composition features generated with the adult phone-model and the child clusters perform similarly when individually used, and complement each other when combined. The combined feature set significantly reduces the error rate. The relative error reduction is 21.7% at the recording-level and 16.8% at the child-level, achieving detection accuracies of 87.4% for recordings and 90.6% for children at the equal-error-rate points.

Categories and Subject Descriptors

J.3 [**Computer Applications**]: Life and Medical Science – *Health.* J.4 [**Computer Applications**]: Social & Behavioral Science – *Psychology.* I.5.4 [**Computer Methodology**]: Pattern Recognition – *Applications.*

General Terms

Algorithms, Measurement, Performance, Reliability, Languages.

ICMI-MLMI'09 Workshop on Child, Computer and Interaction
November 5, 2009, Cambridge, MA, USA
Copyright 2009 ACM 978-1-60558-690-8/09/11 ...$10.00

Keywords

Autism, Autism Spectrum Disorder (ASD), Speech Signal Processing, Child Speech Signal Processing, Autism Detection, Autism Identification, Child Development, Pattern Recognition.

1. INTRODUCTION

Autism Spectrum Disorder (ASD) has gained more and more attentions in recent years [1]. Significant increases in research grants have been reported [2]. Because of the importance of the early diagnosis for young children with ASD to access effective early intervention services, research and clinical practice have focused more and more on early diagnosis [3,4,7,23]. The American Academy of Pediatrics recommends ASD screening for all children at the 18- and 24-month checkups [5]. However, a survey completed in 2004 indicated that only 8% of primary care pediatricians routinely screen for ASD [6]. For parents with concerns, it typically takes at least 6 months to obtain a clinical diagnosis [3] due to the laborious nature of the existing screening/diagnostic procedures and an insufficient number of trained personnel relative to the large number of children in need of evaluation. Efficient and/or automatic tools for ASD detection can significantly facilitate the evaluation process. This study reports a fully automatic mechanism for early ASD detection using the LENA System, which utilizes speech signal processing technology to analyze and monitor a child's natural language environment and the vocalizations/speech of the child. Preliminary results had been reported in [19]. This paper reports on recent progress in both data collection and detection methods.

ASD is characterized by: (i) qualitative impairments in social interaction shown by abnormalities in such behaviors as eye gaze, body posture, sharing interests and emotions; (ii) qualitative impairments in communication shown by language development issues such as delayed status, problems initiating and sustaining conversations, repetitive patterns; (iii) a restricted repertoire of interests, behaviors and activities shown by an adherence to certain topics, routines, rituals, motor manners, parts of objects and sensory abnormalities [7]. In recent years, increased research efforts have been made towards earlier identification of ASD. For example, [8] reported on the discovery of early attention differences that may lead to earlier identification and new therapies for ASD; [9] reported unusual use of toys in infancy as an indicator of later ASD; [10] reported vocal differences and abnormalities in high risk infants at 12 months; [11] reported decreased responsiveness to their names in 12-month-old high-risk children; [12, 13] focused on specific abnormalities in the prosody of children with ASD. These findings were based primarily on subjective observations and rarely related to automatic or machine-generated objective measures. [14] showed the potential of an automatic measure for prosodic quality rating applied in a laboratory setting. In addition to the efforts associated with detection, there are reports on intervention employing a computer or robot. [15] and [16] described a computer-animated tutor for vocabulary and language learning and a robotic prosody therapist, respectively.

The current standard diagnostic tools in clinical practice include the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule-Generic (ADOS-G) [3]. Some of the existing screens for early identification of autism include the CHAT (the Checklist for Autism in Toddlers), the quantitative CHAT, the Modified CHAT, STAT (the Screening Test for Autism in Toddlers), PDDST-II (the Pervasive Developmental Disorders Screening Test-II), ESA (the Early Screening for Autism questionnaire), ABC (the Autism Behavior Checklist), ASQ (the Autism Screening Questionnaire) and The Developmental Behavior Checklist [3,23]. Because these instruments require parent participation and/or direct observation, rating, and scoring by a trained practitioner, they are labor-intensive and necessarily include some degree of subjectivity. Evaluation in an unfamiliar clinical setting may also influence child behavior and potentially influence the evaluation.

This study introduces an objective, unobtrusive, relatively easy tool for ASD identification based on audio recordings from the natural home environment. An overview of the system, the data, the methods and the most recent progress are provided in the following sections.

2. SYSTEM OVERVIEW

Although the LENA System has been described in detail previously [18,19], a system overview is briefly provided in this section for convenience. As shown in Figure 1, the LENA System begins with a small digital recorder (DLP – digital language processor) worn by the child in the pocket of specially designed clothing [17]. All sounds in a child's environment, including his/her own voice, are recorded continuously in an unobtrusive way for an entire day. The audio data is uploaded to a computer and is analyzed, producing information about the natural language environment and the language development status of the child. Currently, the LENA System provides estimates for the number of adult words spoken near the child (adult word-count), the adult-child interaction (turn-count), the number of distinct child vocalizations (child vocalization count), the amount of audible TV/electronic media, environment noise, overlapped speech, etc. within the child's language environment, as well as information about the child's development, including an automatic expressive language assessment or Automatic Vocalization Analysis score (AVA) and the automatic LENA Autism Screen (LAS) score. This hardware and software combination allows caregivers and professionals to obtain prompt information about a child's environment/development and monitor improvement over time, providing the opportunity to intervene when necessary at an early age [18,19].

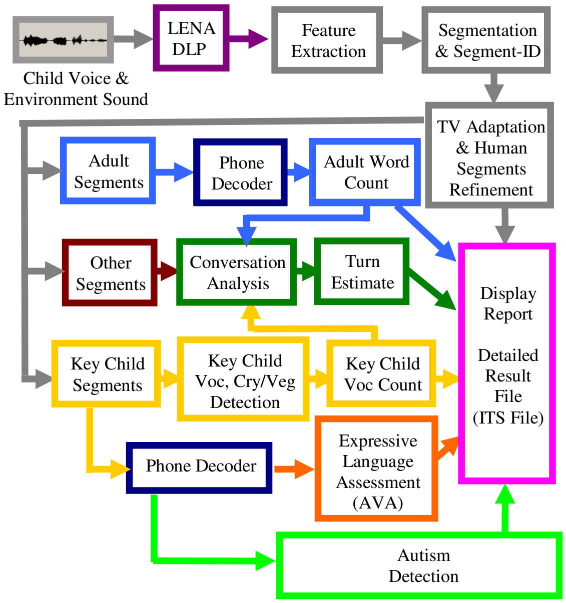

Figure 1: Diagram of the LENA System

As described in [18,19], sounds in the natural environmental recordings are categorized into one of 8 classes: key child, adult male, adult female, other child, TV (including radio and other electronic media sound), noise, silence and overlap (that includes human vocal activity). All non-silence classes are further categorized into clear/faint sub-classes (related to near/far field). Overall, there are 15 sub-classes. After this segmentation and segment-ID process is performed, clear-adult-segments are further processed to produce an adult-word-count estimate. Key-child segments are further processed to delineate normal vocalizations from cries and other fixed signals as well as vegetative sounds. Clear key-child segments are also used to generate child vocalization composition features by using either the open-source Sphinx adult phone-model or the vocalization clusters derived directly from the child vocalization data in this study. The composition features are used to estimate the AVA score and the automatic autism screening score. As a practical consideration, it is required that the full processing time be within 0.5 real-time.

As reported in [18]: the segmentation/segment-ID accuracy varies from 70.5 to 82.0%; the adult word-count performance in terms of the Relative Root Mean Square Error varies from 42% for 1 minute measuring length to below 7-8% after 5 hours of measuring time; and the AVA scores correlate at r=0.75 with comparable scores assessed by human speech language pathologists using standard language assessments. The remainder of this paper focuses on the automatic ASD detection system utilizing child vocalization composition.

3. CHILD VOC-DECOMPOSITION

As mentioned above, childhood ASD is characterized by abnormalities in social interaction, communication, language development and repetitive stereotyped behavior. It is reasonable to assume that certain characteristics of these abnormalities could be exhibited and detected within a day-long audio recording. Specific abnormalities of vocalization and prosody in children with ASD in fact have been reported before [10,12,13]. For a fully automatic ASD detection/screening tool, one of the major considerations is to find a robust discriminant feature for ASD detection which can be automatically generated. It has been found in this study that the child vocalization composition contains rich information to distinguish the children with ASD from other children. The composition analysis is commonly used in Chemistry and other scientific areas to distinguish different materials. Moreover, the child vocalization composition is also robust to noise, interference and recognition error. As long as the majority of the child vocalizations are decomposed correctly, a moderate amount of interference and recognition error will not greatly change the overall composition statistics or destroy the discriminant information in the composition feature. The decent performance in this study somehow proved this to a certain extent.

To the best of our knowledge, the prior research in early child vocal composition has not yet been approached in a similar detailed quantitative way. We believe that this is in part due to the lack of data, as estimation of child vocalization composition requires a large number of audio samples. The fully automatic LENA System now provides a relatively easy way to make the composition analysis possible by obtaining a large number of audio samples.

The child vocalization decomposition with the open-source Sphinx adult phone-model is reported in [19]. One concern about this method is the appropriateness of applying an adult model to child vocalization/speech. It should be noted that the final purpose here is not to recognize phones produced by a child using adult phones as criteria. The phone decoder with an adult model here serves as a decomposer. As long as it works objectively and consistently, and the resulting composition contains rich discriminant information for ASD detection, it is not critical whether a particular vocalization is recognized as [a] or [i] or other categories using adult speech as criterion. Of course, one may argue that more accurate phone decomposition with adult speech as criterion may generate better discriminant information. This is an empirical question we leave for future research; we do not believe it is critical for the final purpose of ASD detection at the current stage. Based on similar reasoning, child vocalizations could be decomposed with self-organized clusters derived directly from the child vocalization data. This approach may have obvious advantages since there is no data-model-mismatch which is the major concern of using an adult phone-model. No matter what the reasoning, the current experiment shows that both the decomposition based on the adult phone-model and the one based on child-clusters perform similarly when individually applied, and complement with each other when combined, resulting in significant improvement in performance.

As mentioned in [19], the clear key-child segments in a day-long recording are used for vocalization decomposition regardless of which decomposer is used. For the adult phone-based decomposition, the open-source Sphinx system is used, which contains 39 regular adult English phone models such as [t], [a] and 7 filler models to absorb pause, breath, hesitation, possibly crying and other categories in clear key-child segments. There are a total of 46 categories collectively referred to as uni-phones in the study. The frequency of a uni-phone is defined by the count of that uni-phone normalized by the total count of all uni-phones in a recording. All such frequencies constitute the uni-phone probability distribution (or probability density function – pdf if we regard a discrete distribution as a special case of the continuous distribution with the Dirac delta function as the bridge between discrete and continuous cases). The composition of a child's vocalization can be quantified by this pdf function. In addition to the uni-phones, to make use of the dynamic information contained in phone-sequences, uni-phone pairs (called bi-phones) are also tested. Since the bi-phone pdf function has high dimensionality (roughly 46x46 = 2116), Principal Component Analysis (PCA) is used to reduce the dimensionality to 50 (called bi-phone-50 in the study) [18,19]. Similarly, tri-phone and longer phone-sequences could potentially be utilized.

To achieve the cluster-based decomposition, an unsupervised k-means clustering method was applied to child vocalization data to generate phone-like clusters. Potentially, the advantage of this method is that the natural modeling units self-organized from child vocalization data itself can better and more naturally handle the issues of ill-defined child pronunciation and its large variation, avoiding the data-model-mismatch that can occur when an adult model is applied to child data. Currently, 64 clusters were generated on the acoustic feature of mel-frequency cepstrum (mfc) with an order of 13 and its first and second order derivatives, constituting the feature with 39 dimensions, which is also the feature used by the Sphinx adult phone-model. These clusters were generated during the previous AVA study before our ASD detection efforts. The data involved in the clustering were clear key child segments automatically obtained from 2979 day-long recordings with the LENA segmentation subsystem. These recordings were from typical and delayed children, no children with ASD were available at that time. With 64 self-organized phone-like clusters, child segments in a natural day-long recording can be decoded to produce cluster sequences and the pdf of clusters can be generated as the composition information of child vocalizations. This approach is called cluster-64.

4. DETECTION & ANALYSIS METHOD

Unlike most ASD detection research in which only a few variables are examined (e.g. attention [8], pitch range [13]), the uni-phone pdf, bi-phone-50 and cluster-64 pdf approaches utilize high dimensional features. Although each individual component in the uni-phone, bi-phone-50 or cluster-64 may not contain significant discriminant information, the combination of them can be powerful enough to achieve much better performance. A data-driven approach is used to find the optimal transform to convert high-dimensional data into low- or 1-dimensional space. Specifically, Linear Discriminant Analysis (LDA) [20,24] is utilized to obtain the linear projection with optimal Fisher-Ratio. We started with a relatively simple method such as LDA and have not yet extended to more sophisticated methods such as support vector machine based on two considerations. The first one is the easy implementation of cross-validation with a simple method; the second is the natural choice of starting with a simple smooth-enough model to study the basic characteristics of the task. With the LDA and certain assumptions, the posterior probability of a child's recording belonging to the ASD-class can be estimated. A formal description of the method could be given as follows.

For a day-long recording of a child, the uni-phone pdf, or bi-phone-50, or cluster-64 pdf is calculated, annotated as $X_i = (x_{i1}, x_{i2},..., x_{id})^T$ where i is the recording index and d is the dimension of a pdf function. The child class-ID is coded 1 for ASD and 0 otherwise, annotated as c_i. A linear projection $W = (w_1,..., w_d)^T$ with optimal Fisher Ratio

$f = W^{T} S_{B} W / W^{T} S_{W} W$ is searched, where S_B and S_W are between-class and within-class scatter matrices respectively. With the optimal W, the multi-dimensional input data X can be converted into a one-dimensional value: $y = W^{T} X$. Under the assumption that y is Gaussian distributed for both ASD-class and non-ASD-class with equal variance (this is actually the underlying assumption of LDA [20,24]), the means m_1 for ASD and m_0 for others and the variance σ^2 can be estimated. With the *a priori* probabilities p_1 for ASD and p_0 for others, the posterior probability of a recording being ASD-class given the input X could be calculated as:

$$P(c = 1 \mid X) = \frac{p_1 G(y - m_1, \sigma^2)}{p_1 G(y - m_1, \sigma^2) + p_0 G(y - m_0, \sigma^2)}$$

where $G(y - m, \sigma^2)$ is the Gaussian function with mean m and variance σ^2.

With a decision threshold t, any recording with a posterior probability above t could be considered belonging to the ASD-class. By varying t from 0 to 1, the performance ROC curve can be obtained and the equal-error-rate (EER) point on the ROC can be determined, i.e., the point with the miss-detection-rate equal to the false-alarm-rate. EER is used as the performance measure for comparison of different cases. It should be noted that the choice of the prior p_1 or p_0 does not affect the ROC and the corresponding EER.

One important consideration for data-driven approaches is the generalization or the potential of models over-fitting to the training data. To obtain realistic performance estimation, cross-validation is needed. To make full use of the data available, the leave-one-out-cross-validation (LOOCV) scheme [21] is utilized. As reported in [19], various levels of targets are left out for cross-validation, including recording, child and recorder. In the recording-left-out test, the posterior probability (pp) of a recording is calculated with the LDA and Gaussian models described above trained using all recordings but the targeted one itself. The left-out designation is circulated through all recordings to obtain the pp for all of them. Similarly, in the child-left-out test, in addition to the target recording, all other recordings from the same child are left-out for the model training. In the child-and-recorder-left-out test, all recordings from the same child or the same recorder as the target recording are left out for its model training. By performing various levels of left-out-cross-validations, we are attempting to ensure that it is an acoustic signature of ASD captured by the models and reflected in the performance report, not the confounding signatures of a particular child or recorder.

Because young children develop rapidly, the key ASD-related acoustic characteristics at different month-ages could be significantly different. To further improve the performance and test month-age effects, age-normalization is tested. For each month-age a, the mean and variance are estimated for each input parameter x_j using the recordings from typically developing children of age range: $[a - band, a + band]$, where $band$ is called age-band to control the size of age-range for the smoothness of age-normalization. The normalization results in transformed parameters with 0-mean and unit-variance for each month-age: $\bar{x}_j = (x_j - age\_mean_j) / age\_std_j$. This transformation can be regarded as part of the modeling process and is tested under the leave-one-out-cross-validation scheme.

As indicated above, one child may contribute multiple recordings. The posterior probability of an individual child being classified to the ASD-class could be estimated from the product of the individual recording probabilities, assuming the independence of different recordings: $pp = \sqrt[n]{\prod_{i=1}^{n} pp_i}$ where pp_i is the posterior probability of i-th recording and n is the number of recordings for the child.

The details and comparisons of different leave-out-cross-validation and age-normalization were reported in [19], where a conclusion can be made that for a realistic performance estimation and comparison, it is necessary and good to use leave-one-child-out-cross-validation with age-normalization inside the cross-validation. In the remainder of this paper, only this scheme is used for performance analysis.

5. DATA & EXPERIMENT RESULT

The current study includes 2-stage data: Set-1 and Set-2. Set-1 contains the samples of typically developing and language-delayed children (i.e., children who do not have ASD, but have been diagnosed with a language delay). It also contains the sample of children with ASD, called "ASD-sample-1". Set-2 currently contains only the ASD sample, called "ASD-sample-2". The collection of additional typical and delayed samples for Set-2 is currently ongoing and not available yet for the current study.

The Typically Developing Sample includes 76 typically developing children with 712 day-long recordings [25]. Expressive language development status for these children was confirmed by an evaluation by a certified speech-language pathologist. The PLS-4 standard scores (Expressive language standard scores for the Preschool Language Scale, Fourth Edition [28]) were averaged at 104.8 (SD=12.1). The month-age of this sample ranges from 8 to 48.

The Language-Delayed Sample includes 30 delayed children with 290 day-long recordings, confirmed by an evaluation by a certified speech-language pathologist [25], with PLS-4 [28] standard scores averaged at 86.2 (SD=13.3). The month-age range is from 10 to 40.

ASD-sample-1 includes children between the ages of 16 months to 48 months who had been formally diagnosed with ASD. They were recruited from January to June 2008 nationwide; parents were required to provide documentation of an ASD diagnosis by one or more trained professionals and typically elected to mail photocopies of their comprehensive evaluations. All documentation was reviewed to confirm the validity of parental reports of a diagnosis of ASD. Parent-report assessments of symptom severity and language development were obtained during the course of the study. A total of 34 children (28 male, 6 female) with 225 day-long recordings and average age of 33 months (SD=7.9) were included in this sample. Average parent self-report symptom ratings for the Modified Checklist for Autism in Toddlers [26] were 9.7 (SD=4.9; Range 1-18) and for the

Social Communication Questionnaire [27] were 19.8 (SD=6.1; Range 9-31).

ASD-sample-2 includes children diagnosed with ASD, recruited from January to July 2009, following the same procedure as for ASD-sample-1. Parent-report assessments of symptom severity and language development were similarly obtained during the course of the study. A total of 35 children with 105 day-long recordings, ranging in age from 27-48 months (M=38.5, SD=6.6) were included in this sample. Parent self-report average symptom ratings for the M-CHAT [26] were 9.8 (SD=4.7; Range 0-19) and for the SCQ [27] were 18.2 (SD=5.7; Range 7-32).

There are in total 140 children (1227 recordings) in Set-1 and 175 children (1332 recordings) in Set-1 and Set-2 combined. Figure 2 shows the recording distribution over age (Note that a child may have multiple recordings at different month-ages.)

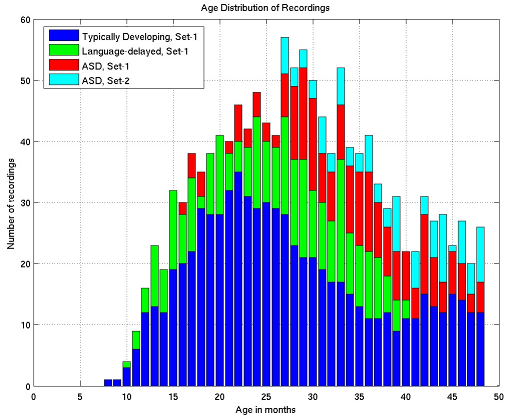

Figure 2: Recording distribution over age

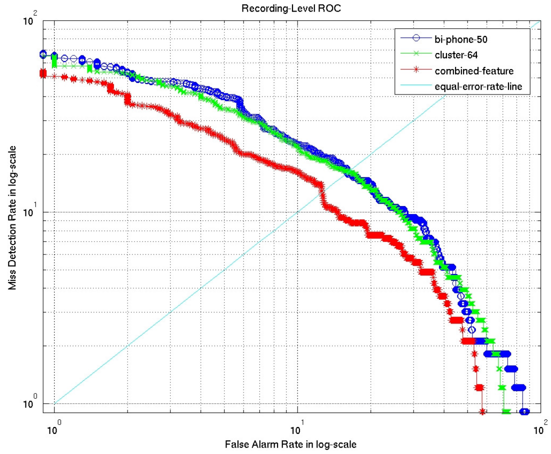

Figure 3: Recording-Level ROC for bi-phone-50 (circle-blue), cluster-64 (x-green) and combined-feature of bi-phone-50 and cluster-64 (star-red) for the task of detecting ASD-sample recordings from typical and delayed sample recordings

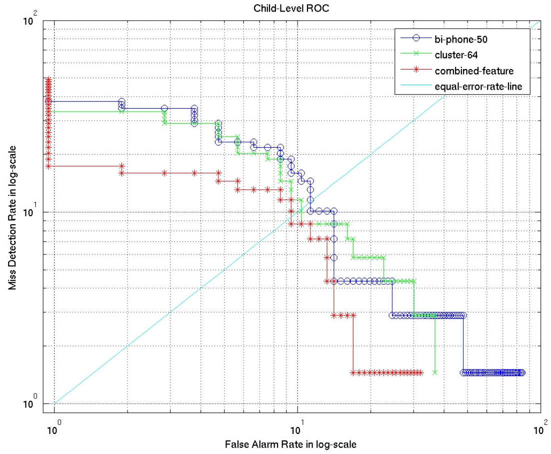

Figure 4: Child-Level ROC for bi-phone-50 (circle-blue), cluter-64 (x-green) and combined-feature (star-red) for the task of detecting children with ASD from typical and delayed children

Table 1: Recording-Level Equal-Error-Rate of Leave-one-child-out-cross-validation on data Set-1 and Set-2 combined

Task	uni-ph	uni-bi-p	bi-ph-50	cl-64	b50-cl64
a vs d	24.1%	20.7%	19.7%	21.2%	15.5%
a vs t	14.5%	14.2%	13.2%	12.8%	11.5%
a vs d+t	17.6%	15.8%	**16.1%**	16.3%	**12.6%**

Table 2: Child-Level Equal-Error-Rate of Leave-one-child-out-cross-validation on data Set-1 and Set-2 combined

Task	uni-ph	uni-bi-p	bi-ph-50	cl-64	b50-cl64
a vs d	11.6%	13.0%	10.1%	16.7%	10.0%
a vs t	11.6%	10.0%	11.6%	11.6%	9.2%
a vs d+t	12.3%	13.0%	**11.3%**	10.4%	**9.4%**

Table 3: ASD-Classification Error-Rate on Set-2 (ASD versus typical+delayed), using the model trained on Set-1 and the threshold of the equal-error-rate-point of the leave-one-child-out-cross-validation-test on Set-1

	uni-ph	uni-bi-p	bi-ph-50	cl-64	b50-cl64
Recording Level	20.0%	16.2%	14.3%	13.3%	14.3%
Child Level	14.3%	14.3%	11.4%	5.7%	5.7%

As mentioned previously, in this paper we focus on the equal-error-rate of the leave-one-child-out-cross-validation on Set-1 and Set-2 combined together. The age-normalization is also cross-validated under the same framework. There are 3 tasks tested. The first one is to detect the children with ASD (or their recording) from the children with language delays not including ASD, annotated as “a vs d”; the second is to detect the children with ASD (or their recording) from the typically developing children, annotated as “a vs t”; the third one is to detect the children with ASD (or their recording) from the typically developing children and children with language delays together, annotated as “a vs d+t". Table-1 shows the recording-level equal-error-rates for 5 different features: Sphinx uni-phone (uni-ph); bi-phone-50 (bi-ph-50); uni-phone and bi-phone-50 combined together (uni-bi-ph); cluster-64 based decomposition feature (cl-64) and the combination of bi-phone-50 and cluster-64 feature (b50-cl64.) As can be seen, the performance is significantly improved by combining bi-phone-50 and cluster-64 features. Figure 3 shows the corresponding recording-level ROC curves for bi-phone-50, cluster-64 and the combination of them. The ROC curves for bi-phone-50 and cluster-64 are similar. However, the ROC of the combined feature is uniformly better than other features in the graph using any decision threshold. At the equal-error-rate point, the relative error reduction of the combined feature over the bi-phone-50 is 21.7% (= (16.1-12.6)/16.1).

Table-2 shows the child-level equal-error-rates for 5 features. The combined feature of bi-phone-50 and cluster-64 is again significantly better than other features, resulting in the relative error reduction of 16.8% (= (11.3-9.4)/11.3) over bi-phone-50. Figure 4 is the corresponding child-level ROC curves for bi-phone-50, cluster-64 and the combination of them. Again, the combined feature is uniformly better than the other two features alone.

Table-3 shows the error-rate of Set-2 using the model trained on Set-1 and the decision threshold at the equal-error-rate point of the leave-one-child-out-cross-validation on Set-1. It consistently shows the performance improvement of the combined feature over bi-phone-50. However, since Set-2 currently contains only ASD data, this result could be biased due to the choice of the decision threshold. Therefore, the result of Table-3 is not as solid as Table-1 and Table-2, and should only be used for reference before the new samples of typically developing and delayed children for Set-2 are available.

6. CONCLUSION & DISCUSSION

This study reports a fully automatic ASD detection method using the LENA System. This paper extends the results previously reported in [19] by incorporating additional data from children with ASD, introducing a new cluster-based child vocalization decomposition approach, and combining the child vocalization decomposition based on the Sphinx adult phone-model with the decomposition based on new clusters derived directly from child data. The following are points for conclusion and discussion.

First, with more ASD data in the cross-validation test, a similar performance to that reported in [19] is obtained, which further confirms that the child vocalization composition contains rich discriminant information for ASD detection. This is one major discovery of the study.

Child vocalization decomposition could be done using either adult phone-model or clusters derived directly from child vocalizations. Performance for the two methods are similar when applied individually. When combined together, the performance is significantly improved. This suggests that the two approaches may capture different discriminant information for ASD detection, and may complement each other when combined together. So far, using all available data, a 9.4% child-level equal-error-rate and 12.6% recording-level equal-error-rate are achieved in the leave-one-child-out-cross-validation test for ASD detection.

Up to now, only smooth linear modeling has been attempted. In the future, non-linear models and large margin classifiers will be explored. It is possible that more and more unlabeled data will be available in the future. The semi-supervised approach may be considered to take advantage of the large amount of unlabeled data to further improve the performance.

One major goal of automatic ASD detection is the early identification of children under 24 months, or even under 18 months. Currently, the majority of our ASD sample data are from children older than 24 months of age. It is not clear how the model trained on our current data can perform for children under 24 or 18 months of age. This might be a challenging task for future research.

Future directions may also include modeling of other types of information in the audio recording, such as social interaction, emotion, etc. Reducing the variation of different recordings for a child is also important in order to reduce the cost of the multiple recordings currently necessary for improved detection performance. Combining the automatic ASD screen with other existing screening instruments may also be important. Additional data and data diversity are always needed for more rigorous tests, especially when considering that ASD is a spectrum disorder and a practical screening tool should ultimately be a universal one.

7. ACKNOWLEDGEMENT

We greatly acknowledge Terry Paul for his conception of the LENA System and for personally funding and directing its development as well as the development of the LENA Foundation Natural Language Corpus.

8. REFERENCES

[1] Archive of Abstracts of International Meeting for Autism Research (IMFAR) http://www.autism-insar.org

[2] J. Singh, J. Illes, L. Lazzeroni, J. Hallmayer, "Trends in U.S. Autism Research", 77, 146.2, 7th Annual International Meeting for Autism Research (IMFAR), May 2008, London

[3] Susan E. Bryson, Sally J Rogers, Eric Fombonne "Autism Spectrum Disorders: Early Detection, Intervention, Education, and Psychopharmacological Management", Canadian Journal of Psychiatry, Vol 48, No 8, Sept. 2003

[4] G. Dawson, J. Osterling, "Early Intervention in Autism", in "The Effectiveness of Early Intervention", M. Guralnick (Ed.), Baltimore: Brookes, 1997

[5] American Academy of Pediatrics, Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee. "Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening." Pediatrics. 2006; 118: 405-20

[6] C. Johnson, S. Myers and the Council on Children with Disabilities, "Identification and Evaluation of Children with Autism Spectrum Disorders", Pediatrics, Vol 120, No 5, Nov. 2007

[7] T. Charman, G. Baird "Practitioner Review: Diagnosis of Autism Spectrum Disorder in 2- and 3-year-old Children",

Journal of Child Psychology and Psychiatry 43:3 (2002), pp 289-305

[8] A. Klin, D. Lin, P. Gorrindo, G. Ramsay, W. Jones "Tow-year-olds with autism orient to non-social contingencies rather than biological motion" Nature advance online publication 29 March 2009

[9] S. Ozonoff, S. Macari, G. Young, S. Goldring, M. Thompson, S. Rogers, "Atypical object exploration at 12 months of age is associated with autism in a prospective sample" Autism, Vol 12, No 5 pp 457-472, 2008.

[10] L. Zwaigenbaum, S. Bryson, J. Brian, W. Roberts, P. Szatmari, B. Mackinnon, S. Mitchell "Early Language Impairments in High-Risk Infants Subsequently Diagnosed with Autism" S4.9, 4th IMFAR, May 2005, Boston

[11] A. Nadig, S. Ozonoff, G. Young, S. Macari, S. Rogers, M. Sigman, A. Rozga "Response to name in 12-month-old siblings of children with autism or typical development" P3B.1.8 4th IMFAR, May 2005, Boston

[12] Joanne McCann, Sue Peppe, "Prosody in Autism Spectrum Disorders: a Critical Review", International Journal of Language & Communication Disorder, Vol 38, No 4, Oct-Dec, 2003

[13] J. Diehl, D. Watson, J. McDonough, C. Gunlogson, E. Young, L. Bennetto "Acoustic and Perceptual Analysis of Prosody in High-Functioning Autism" P1B.2.8, 4th IMFAR, Boston, May 2005

[14] E. Prud'hommeaux, J. Van Santen, R. Paul, L. Black "Automated measurement of expressive prosody in neurodevelopmental disorders" 31 154.31, 7th IMFAR, May 2008, London

[15] A. Bosseler, D. Massaro, "Development and Evaluation of a Computer-Animated Tutor for Vocabulary and Language Learning in Children with Autism" Journal of Autism and Developmental Disorders, Vol 33, No 6, December, 2003

[16] E. Kim, E. Newland, R. Paul, B. Scassellati "Robotic Therapist for Positive, Affective Prosody in High-Functioning Autistic Children" 6.114.6, 7th IMFAR, May 2008, London

[17] http://www.lenababy.com/LenaSystem/AboutLena.aspx

http://www.lenababy.com/

[18] D. Xu, U. Yapanel, S. Gray, J. Gilkerson, J. Richards, J. Hansen "Signal Processing for Young Child Speech Language Development" 1st Workshop on Child, Computer and Interaction, Oct. 2008, Chania, Crete, Greece. Also available:http://www.lenafoundation.org/DownloadFile.aspx/pdf/SignalProcessing_ChildSpeech

[19] D. Xu, J. Gilkerson, J. Richards, U. Yapanel, S. Gray "Child Vocalization Composition as Discriminant Information for Automatic Autism Detection" 2009 International Conference of the IEEE Engineering in Medicine and Biology Society, Sept. 2-6, 2009, Minneapolis, Minnesota, USA.

[20] R. Duda, P. Hart, "Pattern Recognition and Scene Analysis", A Wiley-Interscience Publication, New York Wiley, 1973

[21] S. Haykin, "Neural Networks, a Comprehensive Foundation", 2nd-Edition, Prentice-Hall Inc. 1999

[22] http://www.cdc.gov/ncbddd/autism/faq_prevalence.htm#whatisprevalence

[23] J. Pinto-Martin, A. Weissman, D. Mandell "Screening and Detection of ASD: The State of the Science in Research and Practice" Chapter VI, Trends in Autism Research, Editor: O.T.Ryaskin, Nova Science Publishers, Inc.

[24] http://en.wikipedia.org/wiki/Linear_discriminant_analysis

[25] J.Gilkerson & J.Richards, The LENA Foundation Natural Language Study (LENA Foundation Technical Report LTR-02-2). http://www.lenafoundation.org/TechReport.aspx/Natural_Language_Study/LTR-02-2

[26] D.Robins, D.Fein & M.Barton, "Modified Checklist for Autism in Toddlers" 1999 http://www2.gsu.edu/~psydlr/Diana_L._Robins,_Ph.D._files/M-CHAT.pdf

[27] M.Rutter, A.Bailey & C.Lord, "The social communication questionnaire" Los Angeles: Western Psychological Services. 2003

[28] LL.Zimmeman, V.Steiner & R.Pond "Preschool Language Scale" Fourth Edition, San Antonio: The Psychological Corporation, 2002