The Potential of the LENA Audio-based Autism Screen: Comparison among Four Data Sets

Dongxin Xu[1,2], Paul Yoder[3], Brian A. Boyd[4], Yiwen Zhang[5], Fan Jiang[5], Jeffrey A. Richards[1], Jill Gilkerson[1,2], D. Kimbrough Oller[6]

LENA Research Foundation, [2] University of Colorado at Boulder, [3] Vanderbilt University, [4] University of North Carolina at Chapel Hill, [5] Shanghai Children's Medical Center, [6] University of Memphis

LENA RESEARCH FOUNDATION

Abstract

Over the last couple of years, the LENA Research Foundation has studied the possibility of using LENA audio recordings for a childhood autism screen. The studies were based on a specific LENA data set including 106 typically developing children (TD), 49 children with language delay (LD) and 71 children with autism (ASD) whose ages are mainly 15-48 months. There were basically three methods studied: 1) top-down approach based on vocal "islands" and their characteristics of voicing, syllabicity, spectrum tilt, formant bandwidth, pitch control, duration and so on; 2) bottom-up approach based on the phone-level composition and sound-cluster composition of child produced sounds; 3) extension of the bottom-up approach to cover the features about how children may interact with their environment, including the composition of the overlapped sounds and the statistics of what types of sounds may be more or less likely to precede or follow child vocalizations. These studies showed promising results with equal sensitivity and specificity ranging from about 80% to 94% using "leave one child out cross-validation". Now, there comes a unique opportunity for us to further look into the LENA audio-based autism screen studies, comparing third-party data from Vanderbilt University (VU) and the University of North Carolina (UNC) with LENA recordings of children with autism; meanwhile Shanghai Children's Medical Center (SCMC) collaborated with the LENA Research Foundation studying the home language environment of typically developing children in Shanghai, China, of which the recordings can be used to test the LENA audio-based autism screen when applied to the TD children of different languages and cultures. There are 60 daylong recordings of children 25-48 months old in the VU data, 125 preschool environment recordings (1-4 hours) of children 36-68 months old in the UNC data and 432 daylong recordings of children 4-37 months old in the SCMC data. The LENA audio-based autism screen models trained on LENA data can give autism detection sensitivity/specificity from around 80% to more than 90% on these new data sets. These unique third-party data give us not only a chance of validation but also an opportunity to investigate the relationship between the LENA autism probability scores and children's ages; to study recording effectiveness in terms of child vocal information collection versus recording length in different environments such as in a natural home environment and in a more intensive preschool classroom; to study the possibility of applying this method to different languages and cultures; to study other detailed issues such as audio-based effective autism detection features of different ages, different environments and different languages and cultures. Data and results will be illustrated with figures and tables along with the discussion of potential issues.

Review of Methods

Method 1:
Features: characteristics of child vocal "islands", voicing, syllabicity, spectrum tilt, formant bandwidth, pitch control, duration, etc.
Model: Linear Discriminant Analysis (LDA)

Method 2:
Features: composition or frequency of occurrence of phone-level units and sound clusters in child vocalization.
Model: Linear Discriminant Analysis (LDA)

Method 3:
Features: extension of method-2 features to include "interactive features" such as composition of child co-vocalizations with environmental sounds and probabilities of environmental sounds preceding or following child vocalizations.
Model: AdaBoost (approximation to logistic regression), good for high dimensional data
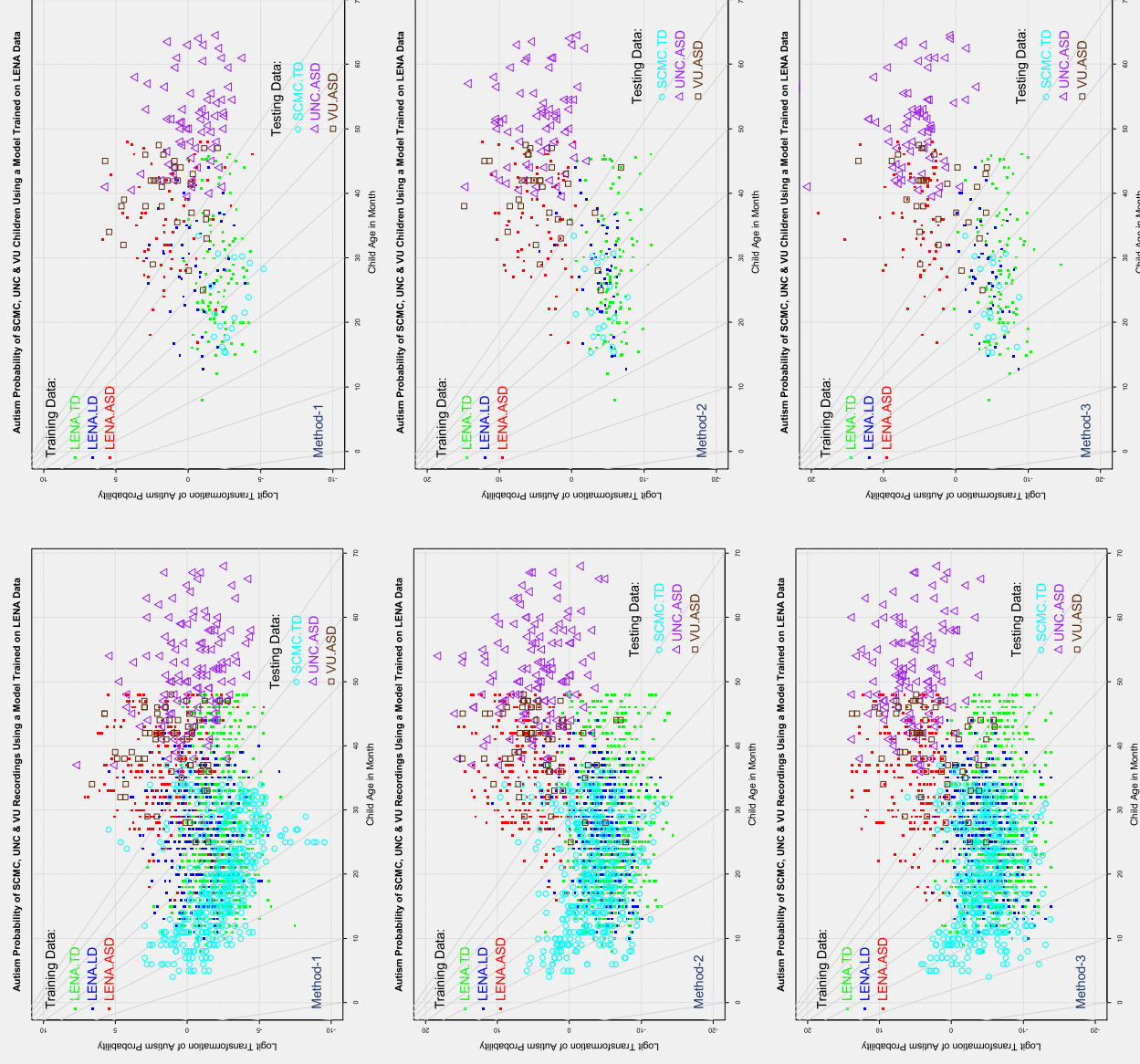
Data

LENA DATA (LENA.TD; LENA.LD; LENA.ASD)
Shanghai Children's Medical Center –TD (SCMC.TD)
University of North Carolina—ASD (UNC.ASD)
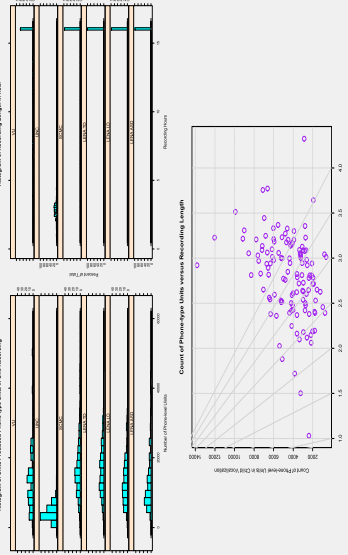Vanderbilt University—ASD (VU.ASD)

	Child Group	Number of Children	Number of Recordings	Children's Age (month)	Recording Length (hour)
Training	LENA.TD	106	802	8 – 48	mainly 16h
	LENA.LD	49	333	10 – 44	mainly 16h
	LENA.ASD	71	228	16 – 48	mainly 16h
	SCMC.TD	22	432	4 – 37	mainly 16h
Testing	UNC.ASD	67	125	36 – 68	1 – 4h
	VU.ASD	31	59	25 – 48	mainly 16h

Objective

Testing three methods using new data sets with models trained on old LENA data; validating the fundamental methodology; identifying potential issues; seeking further improvement in the future.

Sensitivity/Specificity & Discussion

Equal Sensitivity/Specificity is used to show how ASD data are distinguished from non-ASD data. Both Recordings and Children in the Training and Testing Data Sets are tested

%	Method-1	Method-2	Method-3
Recordings in Training Data Set	79.8	95.6	98.2
Children in Training Data Set	81.3	96.8	100
Recordings in Testing Data Set	78.8	87.1	87.5
Children in Testing Data Set	82.7	91.8	92.7

- The LENA audio-based autism screen works fine on the third-party data with the models trained on the LENA data.
- The relatively short UNC recordings (1-4 hours) in intensive class-room environments seem fine with good performance and provide quite a few phone-type units of child vocalizations.
- We are not sure if other ASD children in UNC environment were confused with key ASD children, which could intensify the effect and make UNC data quite "ASD-like".
- Method-1 models basic characteristics of child vocalization which may not work very well for elder children in UNC set, but works fine for young children in SCMC set of different language.
- Method-2 relies more on categories of child sound. This may explain why it degrades when applied to different language of SCMC set. However, Method-3 with extra interactive features works fine on SCMC set of different language.
- Compared with Method-2, Method-3 seems degraded on VU data set. This needs further investigation.
- It might be necessary to investigate in detail each feature when applied to the data of different language and elder ages.
- It is promising to extend this screen to ages above 48 and below 15 months in the future.

Are recordings of 1-4 hours good enough?