LENA®

The LENA^{TM} Automatic Vocalization Assessment

Jeffrey A. Richards, Jill Gilkerson, Terrance Paul, Dongxin Xu
LENA Foundation, Boulder, CO

LTR-08-1

September 2008

Software Version: V3.1.0

ABSTRACT

This report describes the development of the LENA Foundation's automatic vocalization assessment (AVA^{TM}) software. AVA software is designed to provide both parents and professionals with automatically generated information about the expressive language development of children ages 2 months to 48 months. Expressive language estimates are produced based on 12- to 16-hour audio recordings collected in the natural home environment using the LENA language environment analysis system. AVA software uses automatic speech recognition technology to categorize and quantify the sounds in child vocalizations (e.g., protophones and phonemes). These quantitative acoustic information data (expressed as "phone" and "biphone" frequencies) are reduced to principal components which are applied as input for age-based multiple linear regression models. The AVA software utilizes these regression models to generate information about expressive language development as standard scores, developmental age estimates, and estimated mean length of utterance (EMLU). AVA expressive language estimates demonstrate statistical reliability and validity comparable to standard expressive language assessments commonly administered by speech language pathologists.

Keywords

Automatic vocalization assessment, AVA, reliability, validity, developmental age, MLU, expressive language

1.0 INTRODUCTION

The purpose of this paper is to present a general overview of the development and functionality of the LENA Foundation's Automatic Vocalization Assessment (AVA™). The AVA software is designed to provide both parents and professionals with information about a child's expressive language development from ages 2 months to 48 months based on the automatic analysis of audio recordings made in the natural home environment.[1] AVA expressive language estimates are reported as standard scores, percentile ranking, developmental age, and estimated mean length of utterance (EMLU). AVA estimates have demonstrated reliability and validity as predictors of potential language delay.

1.1 Background and Motivation for AVA Development

The assessment of language development in young children poses numerous challenges that stem in part from the difficulty of collecting a sufficient quantity of representative language data. Speech language professionals employ a variety of instruments to evaluate a child's language development.[2] These standard assessments incorporate both parent report and clinical observation components to varying degrees, are typically administered by a professional in a clinical setting, and can require from 30 to 90 minutes to complete and score. Consequently, the reliability and validity of these instruments can be affected by such factors as an unfamiliar clinical setting, reliance on parental input, and limited observation time.

There is a clear need for lower cost, less time-intensive methods which can be used in the natural language environment to evaluate child language development. Our goal for AVA was to provide parents and professionals with an automated tool that could be used to screen children for language delay and to generate an objective language development estimate as part of an overall evaluation, diagnostic, and treatment process. In particular, we sought to minimize the effects of confounds inherent to evaluation in a clinical setting by collecting language data in the natural home environment over an entire day in as unobtrusive a manner as possible.

[1] AVA is a component of the LENA Foundations's newly developed language environment analysis (LENA) software and is generated in the Parent, Professional, and Research versions.

[2] Commonly used standard assessments include: the Preschool Language Scale, 4th Edition (PLS-4; Zimmerman, Lee, Steiner, & Pond, 2002); the Receptive Expressive Emergent Language Scale, 3rd Edition (REEL-3; Bzoch, League, & Brown, 2003); the Child Development Inventory (CDI; Ireton, 1992); and the Clinical Linguistic Auditory Milestones / Cognitive Adaptive Test (CAT/CLAMS; Accardo & Caput, 2005).

Two technological advances were necessary precursors to the development of the AVA software. First, the introduction of the LENA System made it possible to distinguish target child vocalizations from others' vocalizations and from other sounds in the natural home environment with accuracy. Second, automatic speech recognition (ASR) systems have evolved to the degree that it is possible to identify approximate phoneme components of speech sounds. The success of the LENA system coupled with sophisticated ASR software made it reasonable to consider designing an automatic expressive language assessment that could successfully operate in a child's natural language environment.

2.0 THE DEVELOPMENT OF AN AUTOMATIC ASSESSMENT OF EXPRESSIVE LANGUAGE ABILITY

Our purpose was to produce an automated assessment of expressive language ability that matched standard assessments. Toward that end we set three development goals: 1) AVA must be consistent with both theories and observations of language development; 2) AVA must correlate well with chronological age; and 3) AVA must correlate well with SLP-administered expressive language assessment scores.

2.1 Training Datasets

In order to develop an automatic assessment of any type, it was necessary first to collect a substantial quantity of audio recordings as well as standard assessment data. These datasets are described briefly below.[3]

Audio Recording Dataset

The audio recording dataset used for developing and testing AVA includes 2,978 unique, 12- to 16-hour recordings contributed by 360 children between 2-48 months of age. From the over 41,400 hours of naturalistic home audio recordings in this dataset we culled approximately 3,033 hours of detected child vocalization data.

[3] For additional information regarding the 6-month normative and 2-year longitudinal data collection procedures from which these data are drawn, please see LENA Technical Report LTR-02-2: "The LENA Natural Language Study."

Standard Assessment Dataset

Participants visited the LENA Foundation Child Language Research Center at least once to be evaluated by a certified speech language pathologist (SLP).[4] Children typically completed three different standard assessments during the hour-long evaluation period, and the choice of assessments for a particular evaluation session depended on the child's age and attention span. For the purpose of developing the AVA expressive language estimate described in this paper we restricted our criterion assessment set to the PLS-4 and REEL-3 expressive language scores.[5] The current dataset includes 1,589 SLP-administered PLS-4 and REEL-3 assessments for these participants. For each child, expressive language scores from both the REEL-3 and PLS-4 and from all available SLP assessment sessions were averaged together and transformed to standard Z-scores to produce a single expressive language index (EL_Z) per child.

2.2 Language Development and ASR Technology

It is not necessarily obvious that current ASR technology could be an appropriate tool for analyzing child vocalizations. First, ASR software is modeled on adult speech. Second, ASR technology may be inappropriate for very young ages as children typically do not produce words or distinctive phonemes until the second year of life (Werker & Pegg, 1992). Indeed, professional transcribers can have great difficulty accurately classifying protophones or phonemes produced by young children in the early stages of language development (Ramsdell, Oller, & Ethington, 2007).

However, young children attempt to communicate from the moment of birth, and the end result (i.e., the goal) of language development can be considered to be the approximation of the ambient adult language. In the second half of the first year of life, infants begin to explore their vocal tract by producing prelinguistic sounds, referred to as "protophones", or precursors to speech (Oller, 2000). Many of these vocalizations resemble phonemes but are not yet well-formed and could be viewed as approximations of the adult target. As children physically develop, acoustic features in their vocal output begin to resemble those of the adult (Fitch & Giedd, 1999). Parents reinforce more adult-like articulations

[4] Approximately 36% (N=126) of participants were evaluated only once. Other participants were evaluated 2-3 times over a 6 month period. Longitudinal participants were evaluated approximately once every 6 months for 2 years.

[5] Twenty-four of our participants had no usable PLS-4 or REEL-3 scores; thus, our sample size for analyses that require these values was N = 336.

by repeating them or providing more attention when they are produced (Reeve, Reeve, Brown, Brown & Poulson, 1992).

In summary, children hear sounds from influential people in their environment and attempt to produce those sounds. Vocalizations resembling sounds in the ambient language are more likely to be reinforced and retained; sounds that are not a part of the phonemic inventory of the ambient language are more likely to be suppressed and be produced relatively infrequently. Accordingly, with age some features of a child's vocal production should grow closer to that of the adult, and an adult-based ASR phone decoder, applied to child speech, could then potentially yield useful information about expressive language development.

The LENA/Sphinx ASR Phone Decoder

To estimate the number of words adults speak to the key child (i.e., the child wearing the LENA DLP), LENA incorporates modified components of the open source Sphinx ASR software. Like most ASR software, Sphinx ASR software has three main stages or components: a feature extraction stage in which the acoustic properties of a speech signal are reduced to statistical features; a phone decoder; and a language model which converts the phone sequence into words, phrases, and sentences. It should be noted that "phones" in ASR technology are approximations of phonemes; they are acoustically similar to phonemes but are more broadly defined sound categories.

As with virtually all ASR software, the Sphinx phone decoder and language model are "trained" using adult speech. ASR error rates for word recognition in an uncontrolled or natural home environment are so high as to make the language model stage unusable for this purpose. However, the LENA software is able to make fairly accurate adult word count estimates in the home environment using the phone decoder to estimate phone counts, from which word counts are then derived. Modified versions of the feature extraction and phone decoder components (but not the language model) of the Sphinx ASR are currently utilized by the LENA software.

We applied the Sphinx phone decoder to child vocalization data collected from our audio recording dataset to obtain phone counts for the 39 phone and 7 filler-phone categories (e.g., breathing, cries) the decoder recognizes. We computed a phone category frequency distribution (or probability density function – PDF) for each recording and compared it to the average phone frequency distribution for adults in our dataset using the Kullbach-Leibler (K-L) Distance Method for estimating the difference between two PDFs. The K-L distance between the average adult and individual child recording phone frequency distributions as a function of chronological age are shown in Figure 1.

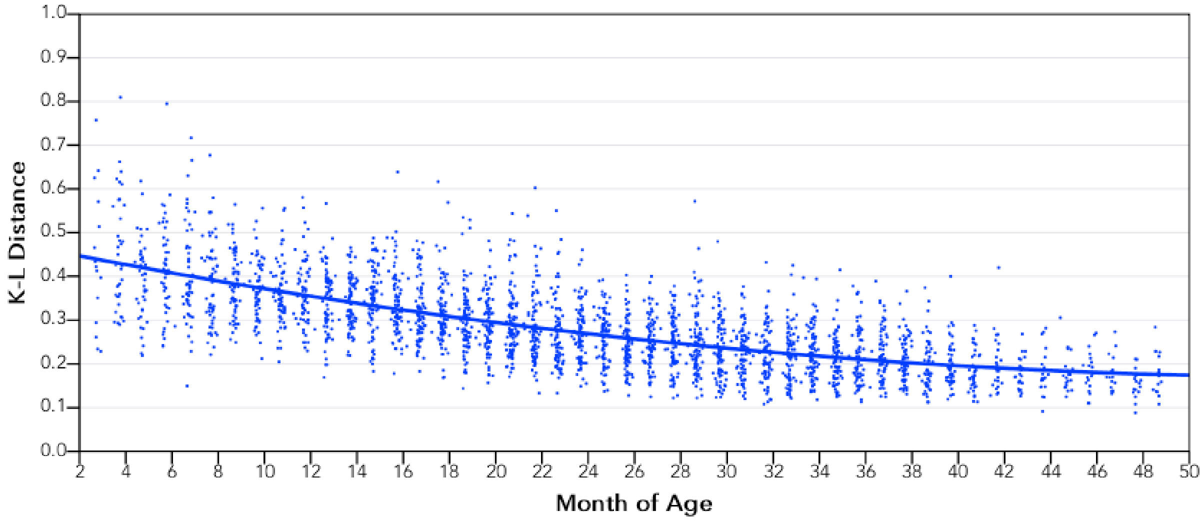

Figure 1. Kullbach-Leibler distance between child and adult phone frequency distributions as a function of chronological age.

As expected, at the youngest ages (i.e., in the early stages of language acquisition) a child's phone distribution differs maximally from the adult phone distribution. But as expressive language develops with age, the child's phone frequency distribution gradually approaches that of the adult. That is, as a child progresses through the stages of language development, his or her phone category distribution gradually converges on that of the adult, supporting the argument that an ASR phone decoder trained on adult speech has the potential to be used to measure expressive language development in young children and so meeting our first goal.

2.3 Predicting Chronological Age from Phone Frequency Distributions

As shown, the phone PDF derived from child vocalizations changes over time to more closely resemble that of the adult, so it is reasonable to expect a relationship between the child's phone PDF and chronological age. Toward that end we built several linear regression models in a two-stage approach in which all phone categories (expressed as proportions of the total number of phones) were included as predictors. In the first stage we applied a global model with a wide age band (over 30 months) to generate a preliminary age estimate. In the second stage we applied a local model with a narrow age band appropriate to the initial age estimate to generate the final age estimate.[6]

Figure 2 shows child phone PDF-based age estimates plotted against chronological age for each recording (r=.90, p<.01). As can be seen, the phone PDF-based model correlates well with chronological age (matching the level of correlation reported for standard expressive language assessments) and so satisfies our second goal.

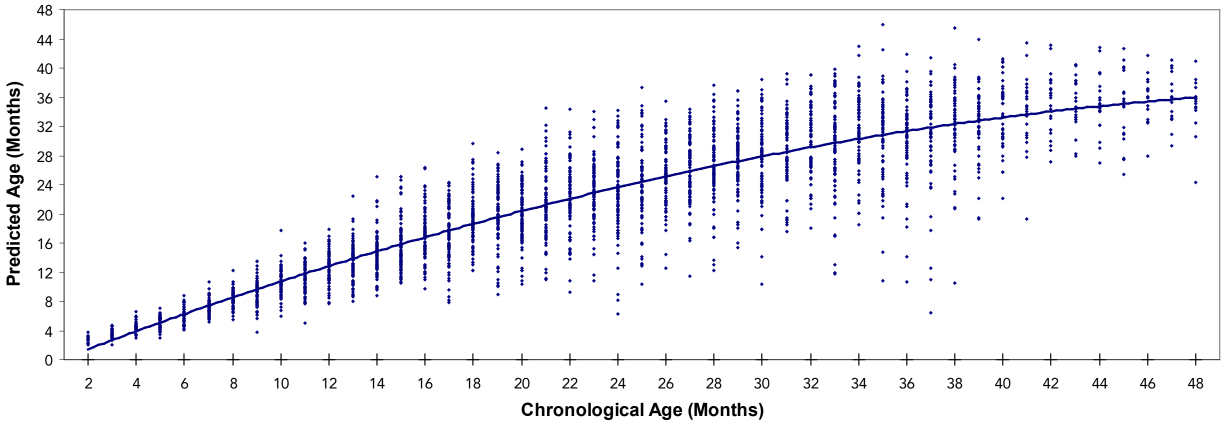

Figure 2. Phone-based age predictions versus chronological age.

[6] An important consideration in linear regression modeling is cross-validation, the goal of which is to ensure that the final model is not overfitted to the training data so that it may be validly generalized to new "unseen" data. One relatively efficient way to protect against overfitting is the "Leave-One-Out Cross-Validation" (LOOCV) method. In LOOCV, the target estimate for each child is generated based on a model including data from all children except the target child (who is "left out."). Each child in turn is "left out" and each target estimate computed from a model built using data from the others. Thus, all available data are utilized, and a given child's target estimate is computed from a model which never "saw" that child's data. This method helps to confirm the generalizability of the resulting models. Results for all models presented here were generated using LOOCV.

2.4 Predicting Expressive Language Ability from Phone Frequency Distributions

A "Uniphone" Approach

The statistical approach described above used to generate estimates of chronological age can easily be extended to generate estimates of expressive language ability. As previously mentioned, PLS-4 and REEL-3 expressive language scores from one or more assessment sessions were averaged together to produce a single expressive language index, EL_Z for each child. A series of age-based linear regression models (similar to those used to predict chronological age) were constructed to predict EL_Z from the 46-category child phone PDFs.[7] The 46 phone and filler-phone categories output by the Sphinx ASR system can be thought of as "uniphones." That is, each category is meant to approximate a single phoneme.

Expressive language estimates AVA_Z generated using these optimized "uniphone" models (and LOOCV) correlated well to SLP-determined scores ($r=.72$, $p<.01$). To produce final age models for AVA estimates we included all participants in the appropriate age ranges (i.e., no LOOCV process was employed). Note too that the expressive language ability estimates AVA_Z are computed in Z-score form. To enhance interpretability, for the recorded AVA estimate we transformed these Z-scores to a standard score format (AVA_{SS}) more commonly used in language assessments.[8]

A "Biphone" Approach

Given the viability of the "uniphone" approach, it is reasonable to investigate whether additional information regarding expressive language complexity might be derived by extending from single category "uniphones" to sequential pairs of categories, i.e., "biphones." For example, the decoded phone sequence "P A T" contains the phone pairs "P-A" and "A-T". Frequencies for each "uniphone" pair may be calculated and a "biphone" PDF constructed. Note that "uniphones" are included as single phones paired with an utterance start or stop marker.

[7] We fine-tuned these models by optimizing the age bands for each age. Rather than using a fixed, symmetric age band at each age, we applied a Dynamic Programming (DP) approach and basic smoothing constraints to each model. Ultimately, 47 separate expressive language models were generated, one for each age group (2-48 months) under consideration.

[8] Our standard scores assume a Gaussian distribution with Mean=100 and Standard Deviation=15.

Conceptually, these "biphone" frequencies could then be used as predictors for linear regression models similar to those built for the "uniphone" case. However, there is a considerable potential problem created by moving from "uniphones" to "biphones." Given 46 phone categories plus the utterance start and end markers, the total number of possible pairs is 48*48 = 2,304. Including so many predictors in a linear regression model could easily lead to the model overfitting the training data, resulting in poor generalization to novel samples.

To resolve this issue, a principal components analysis (PCA) reduced the dimensions of the "biphone" space from over 2,300 to under 100.[9] Exploratory analyses suggested that reducing the "biphone" space to 50 dimensions provided optimal results. That is, we reduced the over 2300 "biphone" combinations to 50 principal components to use as predictors in multiple linear regression models to estimate expressive language scores, exactly as described above in the "uniphone" case. AVA estimates generated from the "biphone" approach (and LOOCV) correlated well with SLP-based expressive language composite scores ($r = 0.75$, $p<.01$), satisfying our third and final goal.

Deriving Developmental Age from AVA

Developmental age estimates from language assessments are typically computed by determining the median age for which an assessment's raw (i.e., unstandardized) score is obtained. Then, a child who receives that raw score is assigned a developmental age equal to that median age. Because our approach is based not on raw counts of observed behaviors but on the probability distribution of phone categories, we have taken a somewhat different approach. In brief, we estimate developmental age AVA_{DA} by making an adjustment to a child's chronological age that reflects his or her AVA estimate as described more specifically below.[10]

Because we are utilizing a linear regression-based method to calculate AVA, it is possible to obtain extreme values that could result in nonsensical developmental age estimates. Thus, we first applied boundary constraints to AVA_Z such that estimates greater than 2.33 standard deviations from the mean (approximately equivalent to the 1st and 99th percentiles) were reset to these boundaries. Next, we computed a smoothed estimate of AVA_Z variability for each age month, SD(Age), using values derived from our training data

[9] Principal component analysis is a standard statistical tool commonly used for data reduction.

[10] This approach is conceptually related to the potentially more familiar calculation of developmental quotients or percent delay.

set. Finally, we computed the developmental age estimate AVA_{DA} = Chronological Age + b_1*SD(Age)*AVA_Z, for which coefficient b_1 was derived from our training data set.

Estimated Mean Length of Utterance (EMLU)

We produce one additional AVA-transformed measure: an Estimated Mean Length of Utterance (EMLU). Speech and language professionals have traditionally used "mean length of utterance" (MLU) as an indicator of child language complexity. This measure, formalized by Brown (1973), assumes that since the length of child utterances increases with age, one can derive a reasonable estimate of a child's linguistic sophistication by knowing the average length of his/her utterances, or sentences.[11]

LENA Foundation transcribers computed MLU for 60 children 15 months-48 months of age, roughly two children for each age month. Transcribers followed transcription and morpheme-counting guidelines described in Miller and Chapman (1981), which were in turn based on Brown's (1973) original rules. They identified 50 key child utterances in each file and counted the number of morphemes in each utterance. The MLU was calculated by dividing the total number of morphemes in each transcribed file by 50. The AVA-based MLU estimate EMLU is computed as a transformation of the standardized estimate AVA_{SS} based on a simple linear regression model using the developmental age estimates AVA_{DA} to predict transcriber-determined MLU.

[11] Brown and others have associated utterance length with various developmental milestones (e.g., productive use of inflectional morphology), reporting consistent stages of language development associated with the average length of a child's sentences. Utterance length is considered to be a reliable indicator of child language complexity up to an MLU of approximately 4.5 morphemes (Brown, 1973).

3.0 RELIABILITY AND VALIDITY OF AVA ESTIMATES

3.1 AVA Test-Retest Reliability

Test-retest reliability for AVA is shown in Table 1 which details comparisons between AVA scores from recordings two months apart for participants ages 2 months-48 months. For reference, test-retest reliability values for our sample's PLS-4 and REEL-3 expressive language scores from administrations two months apart are shown.[12]

Table 1: Observed Expressive Language Standard Score and Developmental Age Test-Retest Reliability over Two Months: Ages 2-48 Months.

		Correlation[a]		Mean Difference	
Measure	N	Standard Score	Developmental Age	SS Mean	SS SD
AVA	**318**	**0.65**	**0.98**	**0.34**	**7.9**
PLS-4	218	0.62	0.95	2.00	13.0
REEL-3	188	0.73	0.92	0.85	4.3

[a]All p< .01

As can be seen from the Correlation columns, AVA test-retest reliability is very similar to that of the standard assessments. The Mean Difference column shows that on average there were no significant differences between AVA scores from recordings made two months apart. AVA scores from audio recordings collected one month apart correlated similarly well (r=.76, p<.01). Developmental age estimates were highly correlated (r=.99, p<.01). Estimates of ages one month apart as would be expected differed on average by approximately one month (M=1.1, SD=1.7).

[12] All reported reliability and validity correlations are based on LOOCV AVA estimates.

Using a subset of 155 participants with six months of consecutive recordings, test-retest reliability for the AVA standard score was examined in two ways: first for single monthly scores, and then by averaging AVA scores from multiple recordings to produce a single estimate per child (similar to the derivation of the criterion score).[13] Test-retest values for monthly AVA scores correlated consistently well (r = 0.74 – 0.80, all p<.01). Averaged AVA scores from the first and last three successive months (1,2,3 vs. 4,5,6) correlated as well (r = 0.85, p<.01). Finally, the average of AVA scores from every other month (1,3,5 vs. 2,4,6) yielded the highest correlations (r = 0.91, p<.01).

3.2 AVA Validity

The validity of AVA estimates was examined by correlating AVA standard scores to the PLS-4 and REEL-3 derived averaged expressive language index EL_Z. Using the subset of 155 participants with six months of complete data described previously, AVA standard scores were examined in two ways: first as single monthly scores, and then by averaging AVA scores from multiple recordings to produce a single estimate per child. Individual AVA scores for each month correlated well with EL_Z (r = 0.69 – 0.76, all p<.01). Averaged AVA scores from three successive months (1,2,3; 2,3,4; etc.) correlated as well or better with EL_Z (r = 0.77 – 0.80, all p<.01). Finally, the average of AVA scores from every other month provided the highest correlations with EL_Z (months 1,3,5: r = 0.88; months 2,4,6: r = 0.86, both p<.01).

We also examined the validity of AVA developmental age estimates by comparing them with chronological age and with PLS-4/REEL-3 estimates assessed within 6 weeks of the recording from which AVA was derived. AVA developmental age correlated well with chronological age (r= 0.97, p<.01) and with age estimates from the criterion measures (r= 0.93, p<.01).

Alternately, the validity of the AVA estimate can be assessed by the accuracy for identifying potential expressive language delay. AVA scores from all available recordings were averaged to produce a single composite AVA score per child. As Figure 3 illustrates, using a threshold score of 77.5 (i.e., 1.5 standard deviations below the mean) AVA correctly identifies 11/19 = 58% of participants similarly identified by criterion measures. Eight (42%) of the SLP-identified children with possible delay have below average AVA estimates but do not meet

[13] The LENA Parent software reports the average of three AVA scores from recordings made within a 120-day period. LENA Pro and LENA Research software report these three-recording average AVA scores plus AVA scores for individual recordings.

the threshold criteria. The false positive rate is low; 3/317 = 1% of typically developing children in the sample fall below the AVA threshold, suggesting a possible expressive language delay.

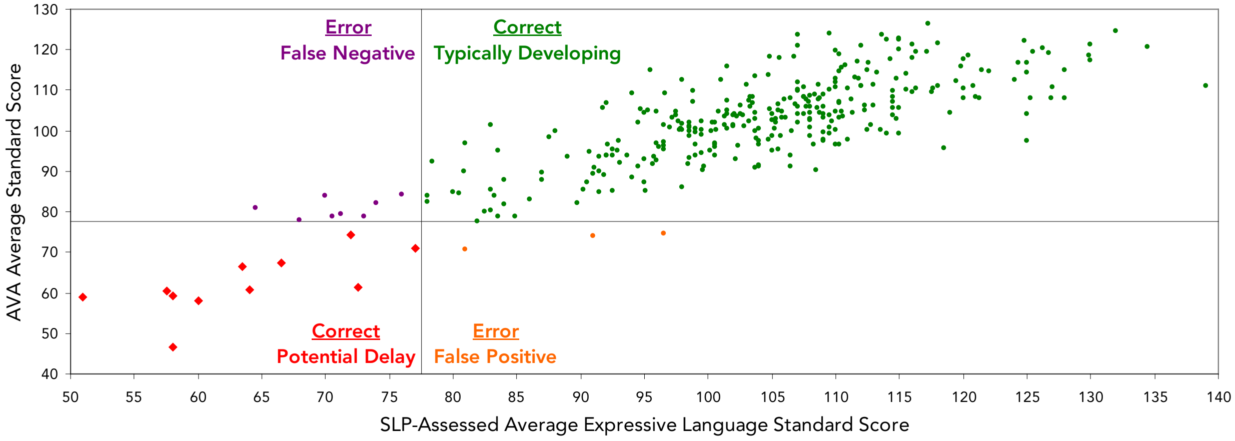

Figure 3. AVA Detection of SLP-Assessed Expressive Language Delay

3.3 EMLU Validity

The validity of the EMLU is demonstrated by its (on average) close approximation of actual MLU (Mean Difference = .15, SD = 0.74) and by the correlation between EMLU and actual MLU scores (r=.78, p<.01).

4.0 DISCUSSION

The original development goal for AVA was to provide both parents and professionals with an automated expressive language assessment for use in a natural home environment that demonstrated levels of reliability and validity similar to that of existing standard clinical assessments. As described above, we were able to meet this goal successfully in our test sample. Using the PLS-4 and REEL-3 as criterion references, AVA estimates demonstrate a comparable degree of variability and test-retest reliability and correlate well with these measures in our sample data. But to maximize the utility of AVA as a tool to screen for potential language delay, it is important to understand why AVA works and to clarify what it does and does not do.

4.1 Why AVA Works[14]

Although the rationale for AVA as described above is framed in terms of “phones,” it is not clear whether AVA in fact assesses “phonemic” growth in infants. A more conservative interpretation is that the AVA analysis estimates the overall degree of acoustic organization of infant vocalizations with respect to certain features that provide infrastructure for speech – some possibilities of the kinds of infrastructural features that are assessed by AVA include instantaneous spectral patterns, burst characteristics, rapid amplitude and spectral changes, and so forth.

Speech sounds are not fixed objects; rather, they are dynamic entities that show extensive acoustic variation depending on circumstance (e.g., gender, voice quality, rate of speech, dialect, etc.) and phonetic context (i.e., the same phoneme has different acoustic characteristics depending on which other phoneme precedes it and which follows it). As a result, in order for an ASR system such as Sphinx to identify words and sentences accurately, it must be adapted to each individual speaker, and once the adaptation is complete and the ASR system is being used for practical purposes, the speaker must be careful to pronounce words at a constant rate and volume and to use only the limited set of words that the system is trained for.

Of course, the 46 averaged phone and filler-phone Sphinx models used by AVA cannot be used effectively to identify words or sentences from individual speakers in normal conversation, nor are they intended to here. The goal for AVA was more general: to incorporate the aggregate characterization of the whole acoustic space of adult English, based on the 46 Sphinx models, and thus to provide a basis for comparison of the adult acoustic space to that of infants and toddlers at various points in development. The degree to which an infant’s acoustic space resembles the adult average can be interpreted as a measure of the infant’s vocal linguistic development.

[14] We acknowledge here the substantial contribution of D. Kimbrough Oller to this section.

This argument is not meant to imply that AVA measures infant "phonemes" in any direct way. The reliability for identification of a sound that might be interpreted as any particular phoneme in a vocalization is low for both adult and infant sounds. However, as described earlier a direct comparison of average adult and child phone distributions (quantified by the Kullbach-Leibler distance) strongly suggests that whatever acoustic features of the child's vocal output are being captured using the Sphinx ASR system, as a child ages his or her profile grows to more closely resemble that of the adult's.

Another way to characterize AVA is that it estimates the degree of organization of an infant's vocalizations as speech-like sounds. Newborn's vocalizations are very little like speech; they show a certain level of diffuse organization, but that organization is quite distant from the organization of mature speech for which there are well-formed phonemic, syllabic, and phrasal categories. Such well-formed categories are not present in early infant vocalizations. However, across the first few years of life infants progress through stages in which new categories of sound are introduced, refined, and elaborated. Ultimately, the sound quality of their vocalizations matches that of mature speech to a significant degree, and they come to possess the categories of adult speech. The goal of AVA then is to assess this growth in mature categorical organization at a general level. AVA does not at present characterize individual sound categories of any infant but rather estimates the general conformity of the degree of categorical organization in the infant with that of mature speech.

4.2 Development Challenges

Although our approach was developed toward the goal of minimizing uncertainty, with any assessment tool some degree of measurement error is unavoidable. In this case, deriving a vocalization-based measure in the context of a recorded natural home environment presents considerable challenge. Major sources of variability for the LENA system include environmental factors (e.g., ambient noise, overlapping speech, channel acoustics such as echo or wind), speaker factors (e.g., age, gender, pitch, etc.), and hardware and operating system factors.[15] Given the ASR phone-based approach utilized in the calculation of AVA, confusion of the vocalization activity of the target child with that of other children and/or adults, especially the mother, is an important source of estimation error. Over time, LENA Foundation engineers expect to be able to reduce the confusion between key child vocalizations and adult speech and other noise, which should improve the reliability of AVA.

Another source of error results from the limitations imposed by our training sample. Despite the LENA corpus being the largest of its type, there is a relative sparsity of data from very young and older children. Across the age range there would ideally be substantially more children per age month, balanced geographically and by SES. A more extensive dataset would facilitate exploring additional sophisticated statistical techniques such as unsupervised mixed mode modeling, classification and clustering approaches, or higher order entropy measures. A further expansion including more complex variables (such as diphones, triphones, and syllables) could be feasible. Also, extending the sample to children with specific developmental disorders might offer the possibility of an automatic identification of disorders for which there are specific acoustic markers or perhaps parent-child interaction markers.

[15] Please see LENA Technical Report LTR-05-2: "Reliability of the LENA Language Environment Analysis System in Young Children's Natural Home Environment" for more details.

Additional challenges remain. Although we are utilizing ASR technology to parse the acoustic space in the vocalizations of the child, we do not collect syntactic information, and we do not identify specific words or grammar. We are applying ASR models based on adult speech to children, in some cases pre-verbal children, so the meaningfulness of these vocalizations in relation to language development is yet to be completely understood. There remains as well a conceptual hurdle in describing adequately what exactly it is that AVA measures. We are deriving a quantity based on the distribution of phoneme-like structures ("phones") present in the child's vocalizations, but more work remains to be done to clarify what those structures represent.

4.3 Directions for Future Research

Although AVA does not identify words, one advantage the LENA system offers is the ability to collect a large sample, indeed an entire day, of child vocalization output. It may be that one needs to identify only a small number of words that carry particular power to estimate language development. If this were the case, it would be easier to develop a specific word-based method using existing technology than if there were a large number of words that would need to be identified. It should be noted in this context that applying existing ASR technology to older children is less problematic.

As well, the LENA system makes it feasible to collect multiple samples of child vocalization data. As noted previously, averaging AVA scores derived from three or more separate recording sessions improves the stability of the estimate and increases the correlation between AVA and SLP-based estimates. A multiple sampling approach can reduce the variability associated with any single administration, but work remains to be done to identify the optimal parameters (e.g., hours of recording, range of sampling period) to ensure the highest reliability and validity.

Many expressive language estimates include a parent-report component. The LENA Developmental Snapshot (LDS), an online 52-item self-report measure of expressive and receptive language development has demonstrated very good reliability and validity compared with standard measures and may be completed in 10 minutes or less.[16] Preliminary analyses indicate that combining information from the LDS with AVA estimates may significantly improve the estimation of expressive language ability.

[16] Please see LENA Technical Report LTR-07-2: "The LENA™ Developmental Snapshot" for more detailed information.

5.0 CONCLUSION

This paper has described the development of an automated vocalization assessment (AVA) that generates an estimate of a child's expressive language development reported as a standard score, developmental age, and estimated mean length of utterance (EMLU). The LENA system enables the analysis of full-day audio recordings that are collected in the natural language environment, and the AVA software takes advantage of the full range of data to enhance the stability of the estimate.

We set three development goals for AVA: 1) AVA must be consistent with both theories and observations of language development; 2) AVA must correlate well with chronological age; and 3) AVA must correlate well with SLP-administered expressive language assessment scores. In this paper we have presented a theoretical justification for the AVA approach and shown that AVA estimates not only correlate well with chronological age but also demonstrate reliability and validity comparable to that of standard expressive language assessments (i.e., the PLS-4 and REEL-3) commonly administered by speech language pathologists.

A primary advantage of AVA is that it is an automatic tool that may be utilized by both parents and professionals as a standalone language and development screening tool or as a confirmatory instrument in conjunction with other assessments to identify potential language delay. Because the LENA system allows an essentially unobstructed view into the natural language environment of the child, the AVA score provides a different and potentially more accurate determination of a child's actual ability than the typical clinical setting.

REFERENCES

Brown, R. (1973). A First Language: The Early Stages. Cambridge, MA: Harvard University Press.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. The Journal of the Acoustical Society of America, 1999, 106, 3, 1511-1522.

Miller, J.F. & Chapman, R.S. (1981). The Relation between Age and Mean Length of Utterance in Morphemes. Journal of Speech and Hearing Research, 1981, 24, 154-161.

Oller, D. K. (2000). The Emergence of the Speech Capacity. New Jersey: Lawrence Erlbaum Associates.

Ramsdell, H., Oller, D. K., Ethington, C. A. (2007). Predicting phonetic transcription agreement: Insights from research in infant vocalizations. Clinical Linguistics and Phonetics, 21, 10, 793-831.

Reeve, L., Reeve, K. F., Brown, A. K., Brown, J. L., & Poulson, C. L. (1992). Effects of delayed reinforcement on infant vocalization rate. Journal of the Experimental Analysis of Behavior, 1992, 58, 1, 1-8.

Werker, J.F., & Pegg, J.E. (1992). Infant speech perception and phonological acquisition. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Ed.), Phonological development: Models, research, and implications (pp. 285-311). York Publishing Company.