

**Reliability of the LENA™ Language Environment Analysis System
in Young Children's Natural Home Environment**

Umit Yapanel, Sharmistha Sarkar Gray, & Dongxin Xu
Infoture, Inc., Boulder, CO

ITR-05-1

October 2007

LENA™ Hardware Model: LR-0120 Software Version: V2.3.0

Copyright © 2007, Infoture, Inc. All Rights Reserved

The LENA™ System

The LENA language environment analysis system is a language monitoring and feedback system designed to provide information about the language environment of infants and toddlers to parents, clinicians, and researchers. The LENA System includes the LENA digital language processor (DLP) that children ages 2 through 36 months wear in the pocket of custom-made clothing. It records everything the child says and hears over a continuous 16-hour day. The audio data is transferred to a computer and analyzed by the LENA language environment analysis software. Parents can access automatically generated feedback reports to view objective information about their child's language environment. The Adult Word Count (AWC) report provides estimates of the total number of adult words the child hears, and the Conversational Turns (CT) report provides estimates of the total number of conversational interactions the child engages in with an adult. These reports permit AWC and CT estimates to be viewed as hourly, daily, or monthly totals. Daily AWC and CT percentile ranking estimates based on a normative database are reported in the LENA software.

The LENA System is intended: 1) to provide a measurement tool to help researchers gain insight into the natural language environment of children; 2) to aid professionals in the early detection of language delay; 3) to support home intervention programs directed at improving the language environment of language-delayed or disadvantaged children; and 4) to educate and provide feedback to parents regarding how much they talk to and interact with their children in order to aid them in maintaining and improving their children's language environments.

Abstract

The LENA™ language environment analysis system was designed to provide information about the language environment of infants and toddlers. In this technical report, we describe the reliability of the LENA System in terms of segmentation, adult word counts, and child vocalizations. We also describe unique sources of variability associated with data collection in the natural home environment.

Keywords

Accuracy, adult word count, child vocalizations, LENA language environment analysis system, reliability, segmentation, transcription, variability.

1. Introduction

The LENA™ language environment analysis software V2.3.0 was developed to process and selectively filter audio and interference signals resulting from a natural data collection environment. Interference signals derive primarily from sound segments that do not contribute meaningfully to the child's language environment, such as transient noise (e.g. bumps, rattles) electronic noise (e.g. television and radio), distant speech, and overlapping speech. The primary goals of the audio data processing are to estimate Adult Word Counts (AWC) and Conversational Turns (CT) between the adult and key child. Here, we establish the reliability and accuracy of the audio processing system as well as possible sources of variability that exist in the natural language environment in which data were collected.

2. Accuracy of the LENA System

The spontaneous speech environment in which data for this study were collected is by design natural and without restriction. Traditionally, speaker recognition software is intended for controlled environments; external sounds and events are generally less severe. By contrast, speech produced by people using the LENA System is spontaneous, real, unrehearsed, and representative of each child's typical daily language environment.

The LENA software V2.3.0 selectively segments sounds into meaningful speech and non-speech sounds, then filters out interfering signals resulting from the naturalistic environment. These exclusions derive primarily from sound segments that do not contribute meaningfully to the child's language environment, such as electronic sounds (e.g. television and radio), noise, distant speech, and overlapping speech segments. Table 1 provides a summary of sound categories in a naturalistic environment.

Table 1: Categories of Natural Environment Audio Data

Live Human Sounds	Background Sounds
Adult Male (Near and Distant)	Overlapping Speech
Adult Female (Near and Distant)	Electronic Sounds (e.g. TV/Radio)
Key Child	Noise
Other Child (Near and Distant)	Silence

Segmentation accuracy may be displayed using confusion matrices, in which human transcription and machine-generated classifications are compared visually (Table 2). Data evaluation in matrix format highlights percentage of instances of false positive and false negative classifications. The goal here was to lower the incidence of false positive classifications that inflate the final AWC and CT estimates. False negative classifications were a less serious error, as these were simply excluded from the final estimates, and the quantity of data collected mitigates the impact of such exclusions. In the following sections, we provide a general description of the accuracy of the LENA System audio processing in terms of the segmentation, adult word count and child vocalization identification.

Table 2: Interpretation of data from a 2x2 confusion matrix.

	LENA-Based Target	LENA-Based Non-Target
Transcriber Target	Agreement	False Negative
Transcriber Non-Target	False Positive	Agreement

2.1 Data Set

Among many files in the Infoture Natural Language Corpus, 70 independent 12-hour long audio files were selected using a block randomization scheme

to ensure a test sample representative of the entire data set. This dataset is referred to here as the *test set*. They were selected on the basis of age (2-36 months) as well as mother's socioeconomic status (SES). Two children were selected per age group, one from a relatively higher SES bracket and the other from a comparatively lower SES bracket. Please refer to Technical Report ITR-06-1 for the demographic distribution of the 70 test set files. An algorithm developed by software engineers from the Infoture Research and Development team was used to select six ten-minute segments from each of the 70 audio files. The algorithm automatically detected high levels of speech activity. Each of the six audio sections was concatenated to form a one-hour-long audio file. Thus, 70 hours of data were transcribed from the 70 test set files. Transcriber-determined segmentation information, AWC estimates, and child speech analysis were acquired from these data.

2.2 General Procedures

First, the LENA software V2.3.0 segments the audio file into adult male (near and distant), adult female (near and distant), key child, other child (near and distant), transient noise (e.g. bumps, rattles), overlapping speech, electronic noise (e.g. television/radio), and silence categories of speech. The segmentation process selectively eliminates noise, distant and unclear speech, overlapping speech, and electronic sounds. Next, a language-dependent statistical model is used to estimate the number of words spoken in each near adult segment without recognizing either the content or meaning of the speech. Finally, a statistical analysis is performed on the key child sound segments to detect vocalization by filtering out laughing, crying, and other vegetative and fixed signals. The following sections will discuss the reliability of the segmentation process, AWC estimation, and child vocalization detection, respectively.

2.3 Segmentation

To assess the daily language environment of the child, the processing software must accurately and reliably distinguish between the adult speaker and the key child speaker, as well as distinguish the key child from other children. It is also essential that factors interfering with AWC estimates (e.g., overlapping or unclear speech, distant voices, transient noise such as bumps, and electronic noise such as TV or radio) are identified and selectively eliminated during the audio processing. To assess the accuracy of LENA-based segmentation, segments identified by professional human transcribers were compared to segments identified by the LENA software. For visualization purposes, the accuracy of the LENA System classification is displayed as a confusion matrix (Table 3). Percentage agreement between human-transcribed segmentation and LENA segmentation are shown along the diagonal. Deviations in agreement are shown in the off-diagonal regions.

Table 3: Segmentation agreement between human and LENA language environment analysis software V2.3.0

	LENA System Adult-Near	LENA System Child-Near	LENA System TV-Near	LENA System Other
Transcriber Adult-Near	78	3	4	15
Transcriber Child-Near	7	75	0	18
Transcriber TV-Near	6	0	73	20
Transcriber Other	12	4	6	78

The data displayed in Table 3 shows a high degree of agreement between human and LENA-based segmentations. The LENA System and human-transcribed near-field adult segments were identified with agreement 78% of the time. Similarly, LENA and human-transcribed near-field child and television were identified with agreement 75% and 73% of the time, respectively. Differences in agreement among these categories were very minor, ranging from 0.1% to 7%. In general, the false negative misclassifications outweighed the false positive misclassifications. Larger variations in identification agreement in the category "Other" are primarily due to the presence of overlapping speech segments.

The LENA software V2.3.0 segmentation algorithm was designed specifically to minimize categorical misclassification. As a result, segments that contained overlapping speech were not classified as either adult or child speech and the LENA processing software excluded these regions. The human auditory system has an innate ability to identify speakers in overlapping speech environments; thus, when professional human transcribers segmented these regions, they could distinguish and appropriately categorize the primary speaker. However, even though an astute professional human transcriber may be able to identify the speaker and process the speech in these overlapping segments, it is not known whether an infant or toddler would be able to distinguish similarly noisy language input. In fact, research indicates that an environment in which overlapping speech is present is not as beneficial as quieter environments, and such environments may even be detrimental to the child's cognitive development (Poag, Goodnight, & Cohen, 1985; Wachs, 1982). Thus, the exclusion of these segments from categorization could provide a more accurate representation of the child's meaningful language environment.

2.4 Adult Word Count (AWC)

Adult word count (AWC) is an estimate of the number of adult words a child hears per hour or per day. The research and development teams at Infoture have developed novel instrumentation to estimate AWC in natural, spontaneous speech environments. Data for the reliability assessment were selected from the Infoture Natural Language Corpus. To assess the accuracy of the AWC estimates, we compared LENA System-detected AWC to the AWC reported by the human transcribers. Results are shown in Figure 1.

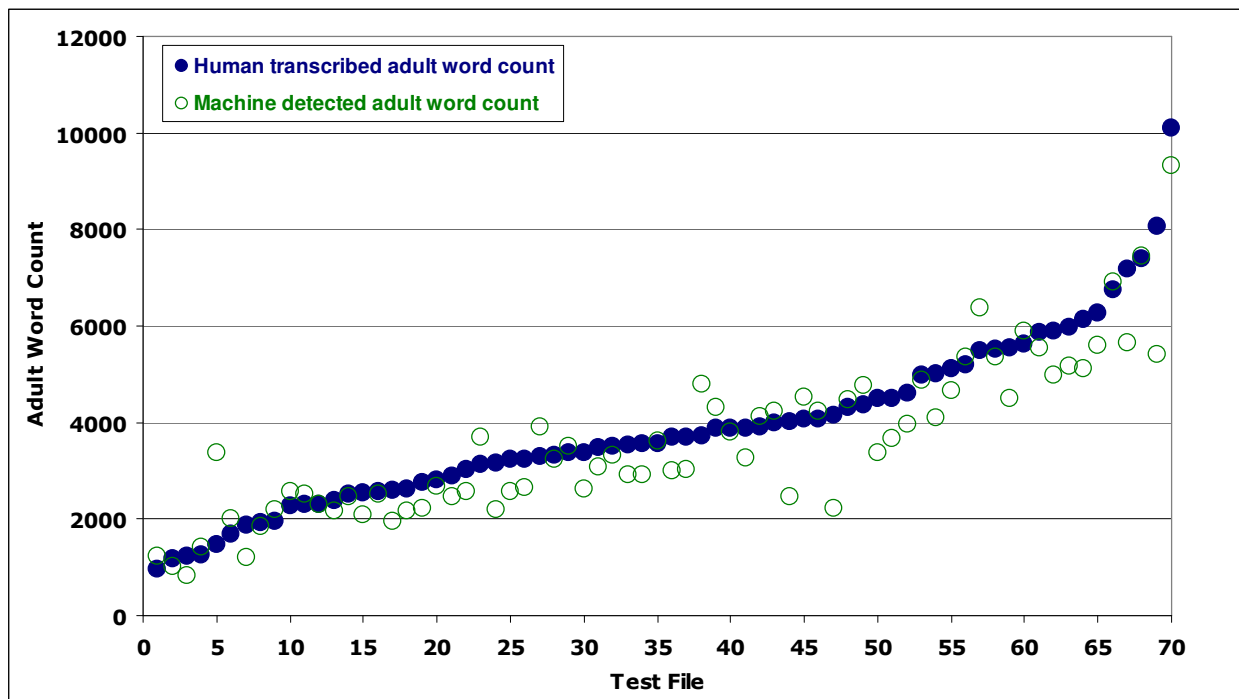


Figure 1. Human and LENA-based AWC estimates for 70 test files

The data in Figure 1 demonstrate a significant correlation between human and LENA-based AWC estimates. The Pearson's Product Moment correlation (r) revealed a 92% linear relationship between the two variables. The LENA mean word count was approximately 7% lower than the word count reported by the transcribers, primarily a result of underestimation (i.e., false negative misclassification). LENA-based AWC estimates are generally lower because LENA System segmentation processes eliminate overlapping speech, possibly

providing a more accurate representation of a child's language environment. In addition, the estimated reliability of LENA analyses is likely to be higher than reported due to inherent error present within the human-based transcriptions. Please refer to Technical Report ITR-06-1 for further information on transcription reliability.

To further assess the accuracy of the LENA audio processing software V2.3.0, LENA-based and transcribed AWC estimates were compared through complete transcriptional analysis of two twelve-hour sessions randomly selected from the Infoture Natural Language Corpus. In one file, "File 1 – Typical Quiet Day," the language environment was generally quieter than the language environment of the second file, "File 2 – Typical Active Day." The two 12-hour files were analyzed in their entirety to assess the accuracy of the LENA-based AWCs relative to human-based AWC as influenced by environmental intensity.

The child in File 1 is a 10-month old male who was assessed at an average pre-language skill level by Infoture's Speech Language Pathologist. As revealed in Figure 2, LENA-based AWC estimates were very similar to human-based counts. The LENA mean word count estimate was approximately 2% lower than the transcriber-reported word count. Here, the red squares reveal boundary points indicative of changes that occurred in the level of activity or environment. Blue and green circles represent human and LENA-based AWC estimates, respectively. Notably, the human and LENA-based values were consistently similar for the entire duration of the analysis, even in the presence of an alternative dialect of English near the start of the timeline.

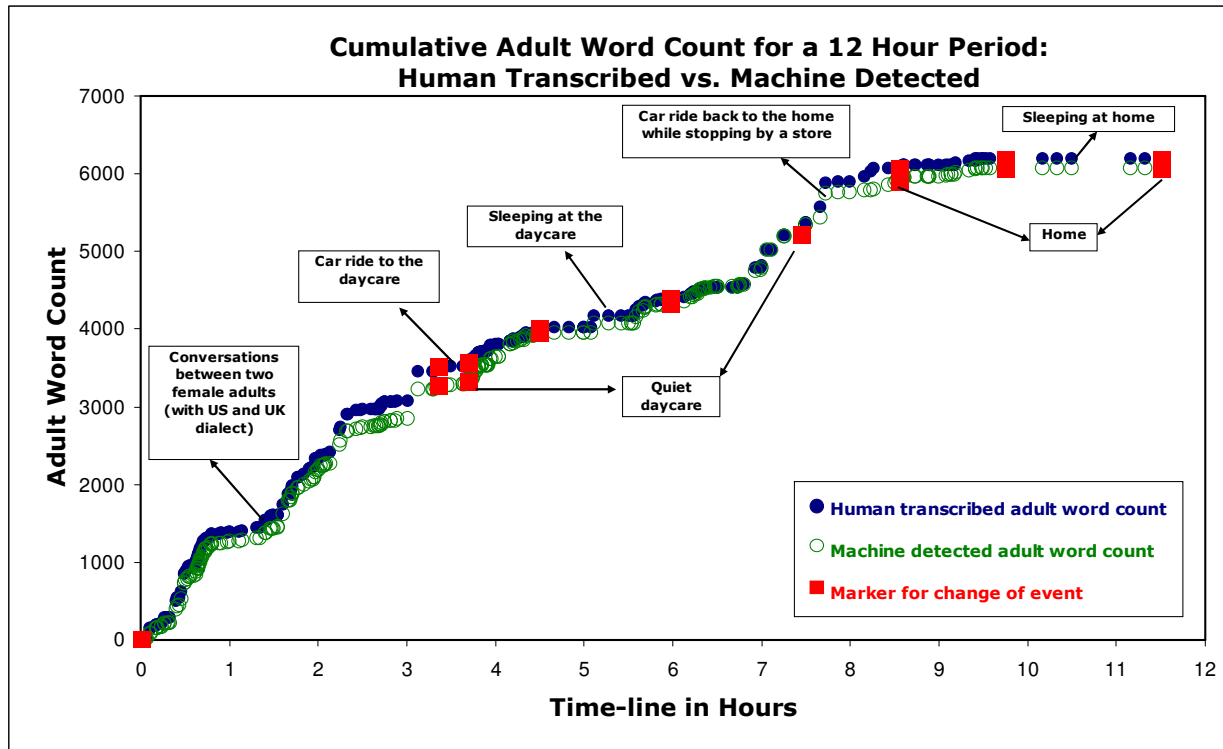


Figure 2. LENA and human-identified AWC spanning a continuous 12-hour recording session in a predominantly quiet environment (File 1 – Typical Quiet Day).

The key child in File 2 was a 31-month old female who was rated with an *outstanding* language skill level by the Infoture Speech Language Pathologist. The AWC estimate for File 2 revealed a very interesting phenomenon in the LENA-based modelling. As shown in Figure 3, the correlation between human-transcribed and LENA-based AWC estimates was lower during the child’s participation in outdoor and other less-quiet activities. As a result, human and LENA-based AWC estimates deviated during these time periods. However, the AWC began to follow human-based estimates once quiet activity was again resumed. This deviation was a direct result of the segmentation process. During more active periods such as those that occur outdoors, variations in human and LENA-derived AWC were due in large part to the presence of multiple overlapping speech

segments. Ultimately, LENA-based AWC estimates for the 12-hour period deviated from the human AWC estimates by 39%. Recall that the LENA System was designed to assess the language environment a child is exposed to. As already indicated, however, high intensity or boisterous environments are not beneficial to the cognitive development of the child, thus the LENA appropriately penalizes the situation (Poag, *et.al.*, 1985; Wachs, 1982). As a result, the lower LENA AWC estimates in active environments may be more reflective of what the child is actually absorbing.

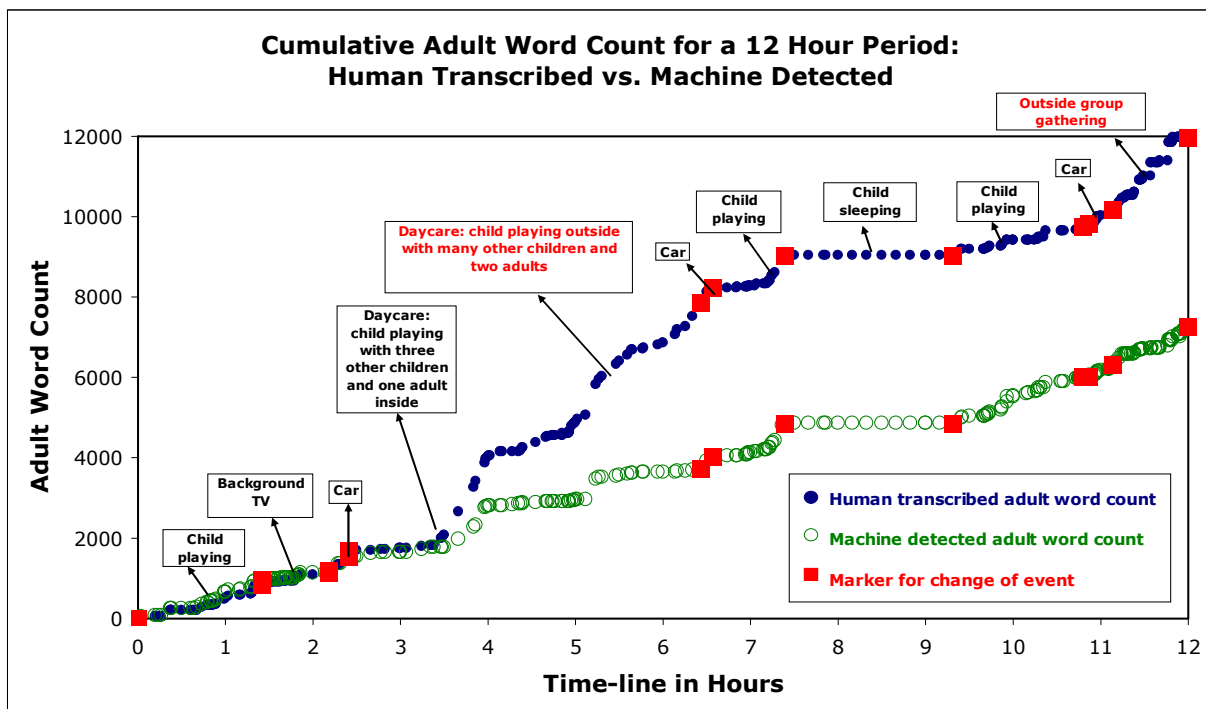


Figure 3. LENA and human-identified AWC spanning a continuous 12-hour recording session in a predominantly noisy environment (File 2 – Typical Active Day).

2.5 Child Vocalizations

Accurate and reliable processing of the audio data is critical to the accurate assessment of the language environment a child is exposed to in a natural state. Of equal importance is the ability of LENA-based algorithms to distinguish between key child speech and key child non-speech. The challenge was to define what should be considered speech versus non-speech. Sounds that were considered speech included words, babbles, and pre-speech communicative sounds or “protophones” such as squeals, growls, or raspberries (see Oller 2004 for more detail on protophones). Sounds considered non-speech were further subdivided into either fixed signals or vegetative sounds. Fixed signals contain sounds that are instinctive emotional reactions to the environment (e.g. crying, screaming, laughing). Vegetative sounds are those sounds resulting from respiration (e.g. breathing) or digestion (e.g. burping). Please refer to Technical Paper ITR-06-1 for further information on the child vocalization classifications.

To assess the accuracy of LENA-based child speech identification, human and LENA-based segment classifications were compared using the 70 test set files described in Section 2.1. As Table 4 demonstrates, human and LENA-based child speech identification were in agreement 77% of the time. The LENA algorithm misclassified a child speech vocalization as a non-speech vocalization 23% of the time. Similarly, human and LENA-based speech detections both categorized a sound as non-speech 66% of the time. The LENA processing algorithms misclassified key child non-speech as meaningful speech 34% of the time.

The primary source of LENA-based misclassification is the lack of context in the algorithmic-based analyses. Professional human transcribers have a distinct advantage over the algorithms that results from the innate ability of the human auditory system to identify context associated with each sound.

For example, a sound without context may sound like a squeal (speech), but when the context of the situation is noted, the same sound may clearly be a scream (non-speech, fixed signal). Nonetheless, the high degree of classification agreement between the transcription and the LENA System is noteworthy. Infoture engineers continue to work to improve the accuracy of the child speech segmentation algorithms.

Table 4: Transcriber vs LENA-algorithmic based Detection and Classification of Sound as Either Speech or Non-Speech Key Child Vocalizations

	LENA System Child Vocalizations	LENA System Child Cry/Veg/Fixed
Transcriber Child Vocalizations	77	23
Transcriber Child Cry/Veg/Fixed	34	66

3. Reliability Over Time

The data described in Section 2 reveal the high degree to which the LENA audio processing system accurately and reliably assesses the language environment of the child. In this section, we discuss the reliability of the LENA processing over time.

The AWC error data obtained from the same 70 test file in Figure 4 indicate that the accuracy of the LENA-based AWC estimate is a function of recording duration. The line details the percentage of difference between the human transcriber word counts and the LENA Adult Word Count estimates. Initially, human and LENA-based AWC estimates differ by as much as 25%; however, this variation decreases virtually logarithmically as a function of time. The variability starts to plateau after approximately one continuous hour of recording, and ultimately the error stabilizes at a rate of variability of <5%. This time-dependent variability curve is primarily a result of the regression

modeling approach used for the analyses. Specifically, over time, false positives and negatives cancel out, thus minimizing the effect of individual misclassifications. This being the case, it is of great importance that LENA DLP users record for a minimum of one hour to achieve relatively accurate estimation of AWC. It should also be noted that LENA norms was designed for use in a 12-hour long spontaneous speech environment, and it is not intended for use in a controlled environment or for use with short excerpts of read speech.

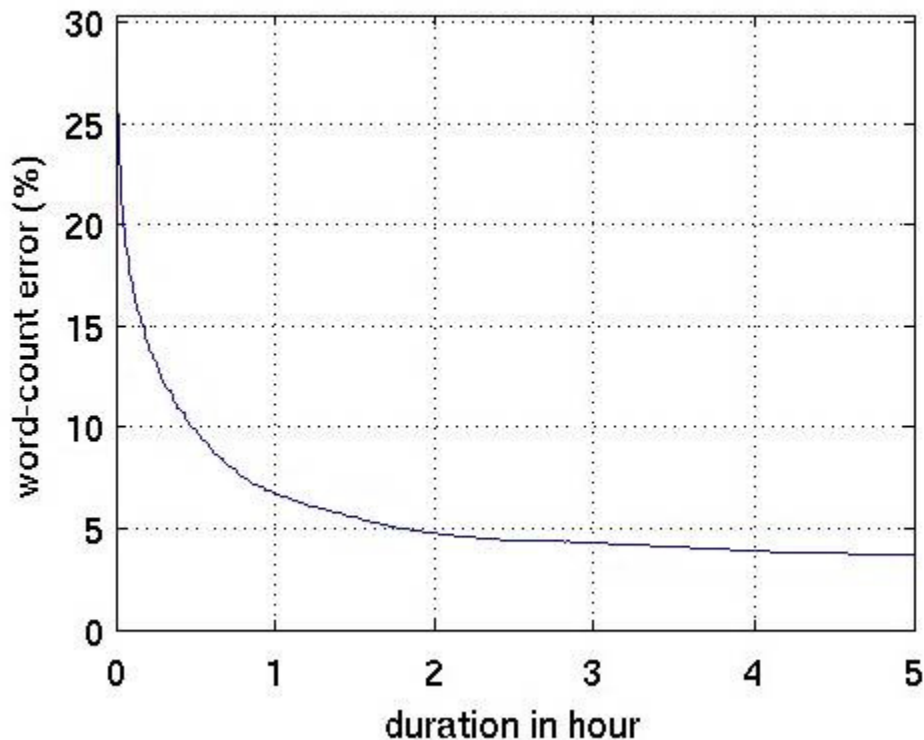


Figure 4. LENA-based AWC error as a function of recording duration

4. Sources of Variability

The LENA System was designed to function as an audio processing unit rather than as a speech recognition device (i.e., the LENA System does not identify individual words, but provides word count estimates based on acoustic information in the audio stream). The algorithms used to process

data from the LENA DLP are probabilistic modeling approaches. As a result, any variability in the source signal will affect the system performance, with the effect being dependent on the amount of degradation caused by the interference. Some of the audio processing challenges unique to the natural child language environment are summarized in Table 5, which classifies the sources of variability and elaborates on the effect of these variations on final system performance. The most significant sources of variability in the LENA System include interferences from environmental factors, speaker differences, effects of clothing, and inter-recorder and hardware differences.

Table 5: Sources of variability, and challenges faced during naturalistic, real-time data collection.

Sources of Variability	Natural Environment	Traditional Environment
Environmental	Background noise	No background noise
	External conversations	No external conversations
	Overlapping speech	Overlapping speech controlled
	Channel Acoustics	Relevant
Speaker Variations	Speaking style	Controlled
	Rate	Controlled
	Accent or dialect	Controlled
	Pitch	Controlled
	Sickness	Controlled
Clothing Effects	Thickness	Not Applicable
	Sound absorption rate	Not Applicable
Hardware Effects	Inter-processor variability	Relevant
	Hardware and operating system variability	Relevant

Environmental effects were a major source of variability. Moreover, channel acoustics were a primary source of error in this category. Echo and reverberation effects resulting from room size, flooring type, environmental location, and far-field effects could negatively impact signal integrity.

The AWC estimates are also influenced by variations of speech quality introduced by different speakers in the audio files. These variations include speaking style, rate, accent or dialect, pitch (e.g., pitch variations resulting from parentese) and voice changes resulting from a variety of health conditions.

The child wears the LENA DLP for a continuous span of 12-16 hours. As a result, the signal quality is affected by the child's clothing. This source of variability was greatly minimized through the production of custom-made clothing tailored specifically to optimize the quality of audio files from the DLP. It should be noted that the LENA clothing has been rigorously tested to ensure that variability associated with the clothing is minimized; the LENA clothing should be worn during audio recording sessions for optimal recording quality.

Finally, hardware-related sources also contributed to the overall variability of the estimates. Error may have been introduced by the inherent variability between different DLP instruments as well as variability within the computer hardware and operating systems.

5. Conclusion

We have described the accuracy and reliability of LENA language environmental analysis software V2.3.0 in terms of segmentation, AWC estimates, and child vocalization classification. Confusion matrices showed that LENA and human-based segmentations had a high level of agreement. Misclassifications were primarily false negatives resulting from the elimination of overlapping speech, and thus the LENA estimates were likely

more representative of the meaningful language environment of the child. Similarly, LENA and human-based estimates of AWC were highly correlated ($r = 92\%$) when random concatenated sections were transcribed; this reliability increased as a function of time. Full-day transcriptional analyses of two select files (one representing a typical quiet day; the other representing a typical active day) revealed that LENA System and human-based AWC estimates deviated during segments containing substantial noise and overlapping speech. The lower estimates from the algorithmic models may again be more reflective of what the child is absorbing. In addition, the reliability of LENA analyses is likely to be higher than reported due to inherent error present within the human-based transcriptions. LENA and human-based child vocalization classifications were predominantly in agreement, with a slightly greater tendency for the algorithmic models to misclassify non-speech sounds as speech. Finally, the natural child language environment introduced sources of variability not generally seen in traditional data collection environments, including environmental factors, speaker variations, and clothing and hardware effects.

References

- Poag, C.K., Goodnight, J.A., & Cohen, R. (1985). The Environments of Children: From Home to School. In R. Cohen (Ed.), *The Development of Spatial Cognition* (pp. 71-113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Wachs, T.D. (1982) Relation of home-noise confusion to infant cognitive development. Annual Meeting of the American Psychological Association.
- Oller, D.K. (2000) *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.