The LENA™ Language Environment Analysis System:

Audio Specifications

Michael Ford, Charles T. Baer,

Dongxin Xu, Umit Yapanel, & Sharmistha Sarkar Gray

Infoture, Inc., Boulder, CO

ITR-03-1

October 2007

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

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^{TM} language environment analysis system was designed to estimate adult and key child interactions in natural home environments. Contrary to controlled linguistic research environments, the speech used by the participants in this study was real, unrehearsed, and representative of each child's typical daily language environment. In this paper, we describe the Audio Processing System in terms of information flow, feature extraction, and segmentation identification. We also reveal the audio specifications that were either met or exceeded during the development and design of the LENA digital language processor (DLP).

Keywords

Audio specifications, feature extraction, segmentation, transcription, Digital Language Processor.

1. Introduction

The LENA[TM] language environment analysis software V2.3.0 was developed to process and selectively filter audio and interference signals resulting from a natural child language environment. Interference signals were derived primarily from sound segments that did not contribute meaningfully to the child's language environment, such as transient noise (e.g. bumps, rattles), media 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.

2. LENA Processing Flow-Chart

The LENA Audio Processing System is comprised of four distinct components: information flow, information processing, algorithmic processing models, and professional human transcriptions (Figure 1).

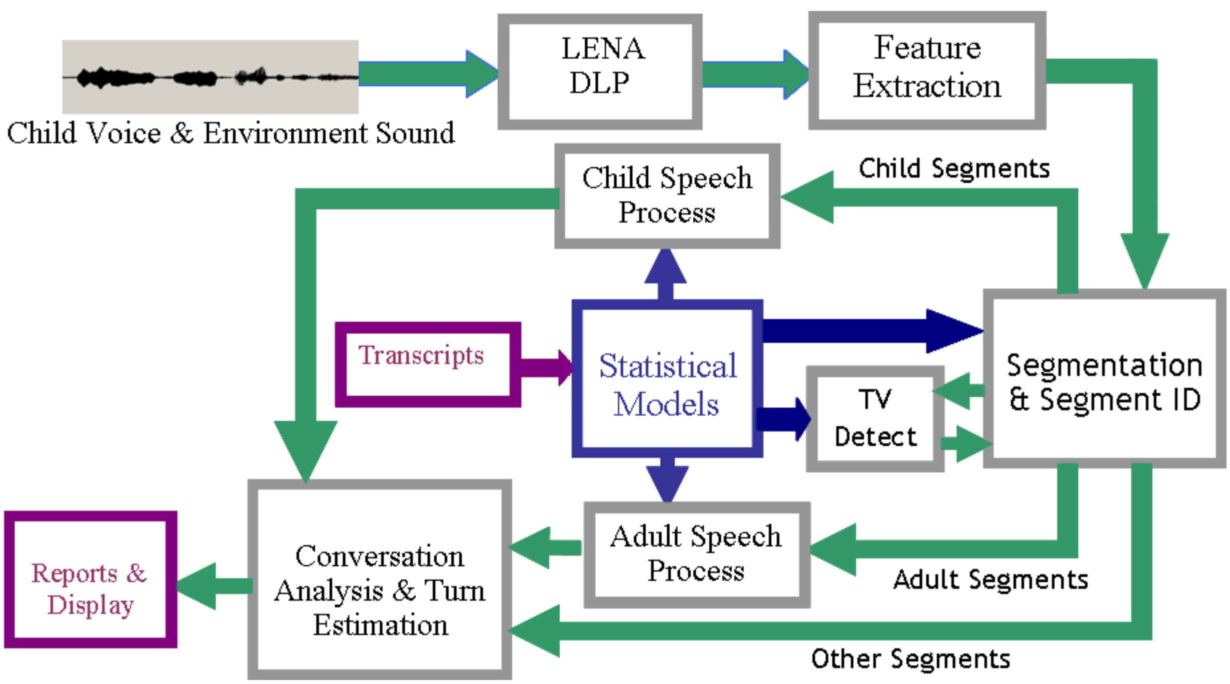

Figure 1. LENA Language Environmental Analysis Audio Processing System.

Initially, an audio file containing recording data from a child's natural home language environment is stored in the DLP. The data are initially processed to minimize disk space and battery power consumption. The audio data on the DLP are transferred through a USB port onto a computer where the data are further processed and acoustic features are extracted.

Various acoustic features are extracted for different purposes. Some features are good for distinguishing speech signal from non-speech signal; Some are good for child speech processing to detect child vocalization from other child sounds such as cry, vegetative sounds and fixed signals.

At the heart of the LENA system is the capability for the algorithmic models to segment and appropriately identify sounds of varying amplitude and intensity. Features extracted from the audio data were segmented through iterative modelling processes into sub-categorical components consisting of key child, other child (near and distant), adult male (near and distant) and adult female (near and distant), noise, silence, overlapping sounds, and electronic (e.g. television/radio) audio categories. Professional audio transcriptions were used to train the audio processing models and the algorithms worked with the models to accurately and reliably identify and detect a variety of segments from the audio signals. For example, it was necessary for the speech processing algorithms to differentiate adult speech from child speech, and to differentiate the speech of the key child from the speech of other children or non-speech sounds (e.g. cries or vegetative sounds). Thus, algorithmic models were built and optimized using the professionally transcribed segmentations as a basis for accuracy. The accuracy and reliability of the LENA software V2.3.0 is described in Technical Report ITR-05-1, and the transcription process in ITR-06-1.

Key child sound segments were further processed through additional iterative processing to distinguish segments containing key child speech (including words, babbles, and pre-speech communicative sounds such as squeals, growls, or raspberries) from non-speech (including fixed signals and vegetative sounds). Refer to Technical Report ITR-06-1 for information on segmentation processes and child speech/non-speech classifications.

Adult sound segments were processed to estimate the number of adult words a child hears. It can be summarized as hourly, daily, and monthly AWC. Adult speech was segmented as either near-field (close and audible) or far-field (distant and quiet) using a statistical modelling approach. A language-dependent statistical model was used to estimate the number of words spoken in each adult segment, including filtering out adult non-speech sound such as coughing, vegetative sound, etc

Statistical modeling was further used to detect Conversational Turns, or back and forth alternation between the key child and an adult. For the purposes of statistical modelling, a conversation was defined as a contiguous region containing live human speech. Two conversations should be separated by a pause region of at least 5-second which contains non-live-human signals or background sound, such as noise. Conversational turns (CT) were defined as alternation counts between key child vocalization and adult effective speech (with word count) within each conversation and each alternation can not go across conversation boundaries..

The result of the analyses described above is the Infoture Time Segment (ITS) file, a compilation of every facet of data recorded and analyzed. Please see Technical Report ITR-04-1 for further information on the ITS file. LENA software engineers continue to work ambitiously to improve the algorithmic-based feature extraction and segmentation analyses. We intend to release upgraded versions of the software annually.

3. Lena System Audio Specification

The LENA System consists of a novel Digital Language Processor (DLP) that was developed by hardware and software engineers at Infoture. Here, we describe the performance goals associated with the unit, as well as hardware and operational performance.

3.1 Performance Goals

The LENA DLP is used for full-day recording sessions, for a maximum of 16 consecutive hours. Thus, LENA DLPs must be stable and maintain high levels of inter-unit reliability; performance goals center on these two aspects of the design. Infoture hardware engineers observed that signal level directly affected AWC. For example, if the signal variation was +/- 1 dB, a maximum of 4% variance was observed. However, if the signal variation was +/- 2 dB, the maximum variance observed was 18%. In the example below, showing the signal variation of the current model LR-0120, six DLP units were chosen at random to determine how well they recorded between two different passes. As revealed in Figure 2, the signal variation between passes was quite marginal for all DLP units tested.

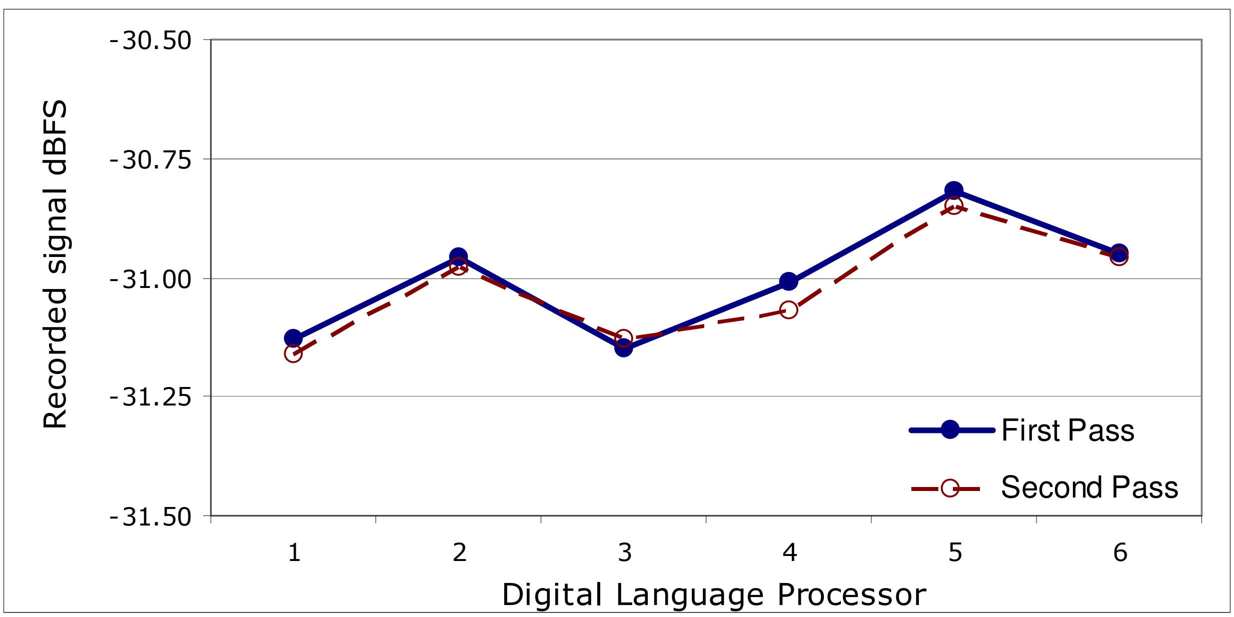

Figure 2. Signal reliability using six DLP units chosen at random.

Infoture hardware engineers sought to produce consistent inter-recorder sensitivity to minimize variation of report output from different DLP units.

The target sensitivity was set to record 67 dBC SPL as -30 dBFS in the audio file. An additional performance goal was to achieve inter-recorder variation of no more than +/- 1 dB. Currently, inter-recorder (between unit) variation is less than +/- 0.5 dB and intra-recorder (within a single unit) variation is less than +/- 0.1 dB. Additional performance goals included a flat frequency response (+/- 1dB 100-4000 Hz), on/off axis linearity of sensitivity and frequency range, and low signal distortion. Finally, the unit was designed such that the recording was unaffected as the battery discharged to a lower operational limit. Standards for compliance there were either met or exceeded are shown in Table 1. The current DLP model LR-0120 meets or exceeds US/Canada compliance standards.

Table 1: Compliance standards met or exceeded by the LENA DLP.

Standards for Compliance	Description	LR-0120
UL 60065	UL Standards for audio, video, and similar electronic apparatus – Safety requirements	√
CAN/CSA-C22.2 No. 60065	Canada – Standard for audio, video, and similar electronic apparatus – Safety requirements	√
UL 696	UL Standard for Safety – Electric Toys	√
EN 55022	EU Standard for Information Technology – Radio disturbance characteristics	√

3.2 Hardware

Audio data were collected using an omnidirectional microphone with a flat 20-20 KHz frequency response. Extreme frequencies were suppressed, as they were unlikely to contain human speech activity. Low frequency data were suppressed through a 70 Hz high-pass filter. Digital data were recorded using a 10 KHz low-pass filter to suppress high-frequency sounds. Frequencies were recorded using a 16 KHz 16-bit sigma-delta analog to digital (ADC) converter with 8x over-sampling digital interpolation.

Initially, audio data were written to a 512 MB flash drive using a 4:1 Adaptive Differential Pulse Code Modulation compression scheme (DVI-4 ADPCM). The memory uses an internal error correcting code (ECC) for data storage and recovery. Complete discharge of the battery will not result in loss of audio data. Data were uploaded to a host computer through a 2.0 USB high-speed port with a sustained audio transfer rate to host of approximately 4 MB/sec (~ 2.5 minutes per 16 hours of audio). Once uploaded, the data were decompressed to the PCM audio format with one 16-bit channel at a 16KHz sample rate.

The LR-0120 unit maximum operating power is 50 mW. A primary 750 mAh battery provides a minimum of 30 hours of recording when new. The recording is safely discontinued when battery power is depleted. The DLP contains a real-time clock (RTC) for timestamping recordings, as well as providing a time base for built-in ADC sample rate calibration. The unit comes equipped with a dedicated real-time clock battery power for life of ~ 10 years.

3.3 Simple Operation

The LENA DLP was designed with the user in mind. It is equipped with a power and record button (Figure 3). An LCD display allows the user to easily identify when the unit is sleeping or recording, as well as the battery status. The unit easily attaches to LENA-designed clothing in a protective pocket that snaps shut. The DLP is very compact (3-3/8” 2-3/16” x 1/2”) and of minimal weight (< 2 oz) in relation to children, thus minimizing the distraction associated with the presence of the recorder.

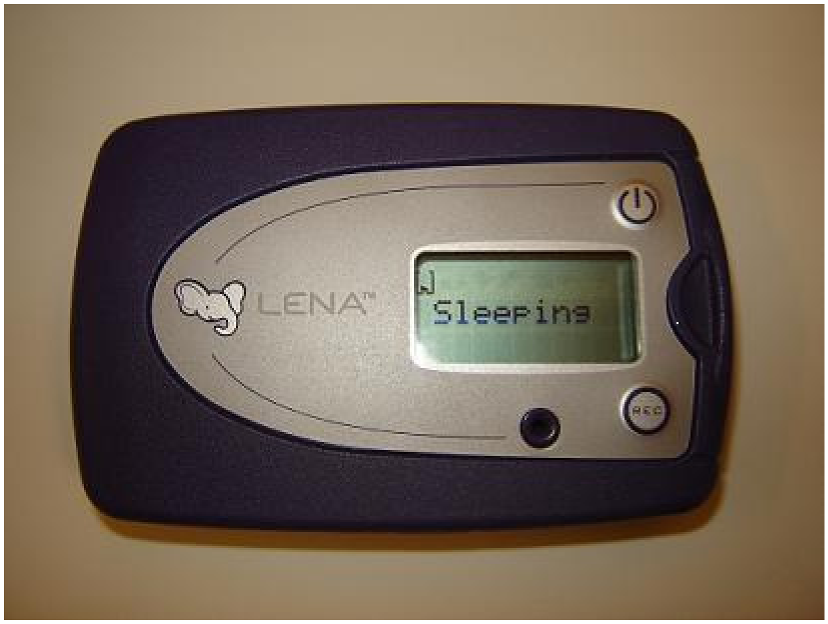

Figure 3. The LENA digital language processor (LR-0120).

4. Conclusion

We have described four components to the audio processing system: information flow, processing, statistical modeling, and transcriptions used for model training. Features extracted from the audio are segmented through iterative modeling processes into sub-categorical components including adult by gender (near and distant), key child, noise, overlapping speech, other child, and electronic noise. Key child segments are further segmented into either speech or non-speech. Adult sound segments are processed into AWC estimates and adult child alternations are processed into CT estimates. The LENA DLP is simple in operation and the inter-unit signal variation is low, as assessed by test-retest reliability. The DLP model LR-0120 has either met or exceeded US/Canada compliance standards.