The LENA™ Language Environment Analysis System:

Audio Specifications of the DLP-0121

Michael Ford, Charles T. Baer, Dongxin Xu, Umit Yapanel, Sharmi Gray

LENA Foundation, Boulder, CO

LTR-03-2

September 2008

Software Version: V3.1.0

ABSTRACT

The LENA language environment analysis system was designed to estimate adult and key child interactions in natural home environments. Contrary to controlled clinical 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

1.0 INTRODUCTION

The LENA language environment analysis software V3.1.0 was developed to process and selectively filter audio and interference signals resulting from a natural data collection environment. The primary goals of the audio data processing are to estimate Adult Word Counts (AWC), Child Vocalizations (CV), and Conversational Turns (CT) between the adult and key child. Here, we describe the Audio Processing System in terms of information flow, feature extraction, and segmentation identification and detail the audio specifications that were either met or exceeded during the development and design of the LENA digital language processor (DLP).

2.0 LENA PROCESSING FLOW-CHART

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

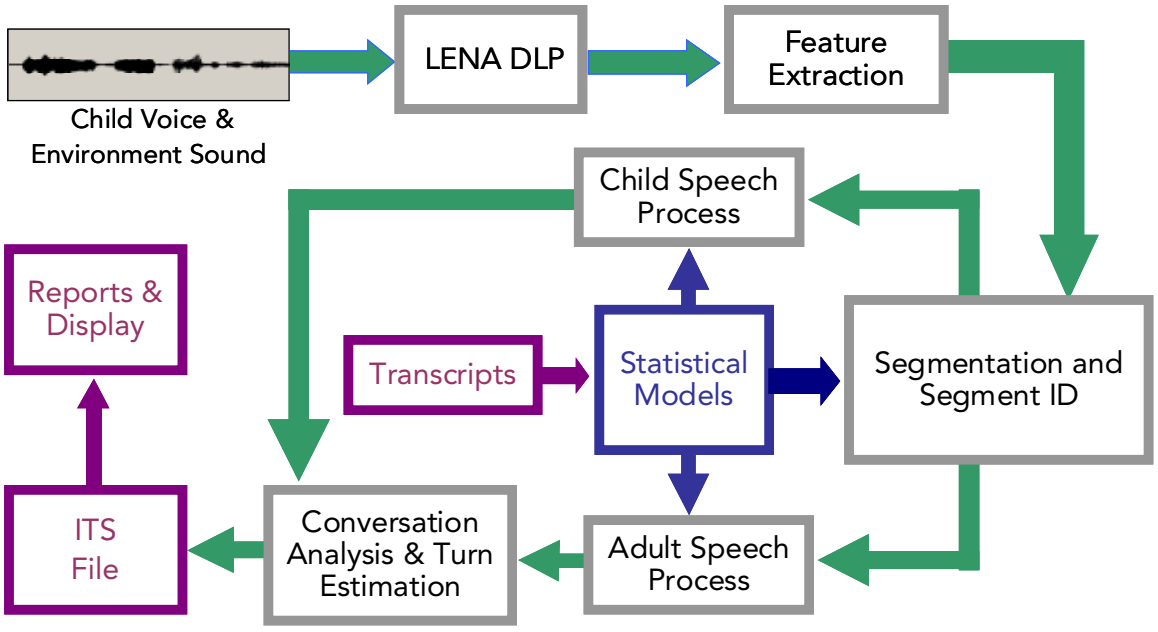

Figure 1. LENA Language Environmental Analysis Audio Processing System.

Initially, an audio file containing recording data from a child's natural home language environment is stored in the DLP. The data are first processed in the DLP 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 primarily used for distinguishing speech signal from non-speech signal; others are used for child speech processing to distinguish child vocalization from other child sounds such as cries, 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 eight categories that identify the source of the audio signal: the key child (wearing the LENA DLP); other child; adult male and adult female; overlapping sounds (at least one human); noise; electronic (e.g. television/radio) sounds; and silence. Based on the statistical fit of each segment to the selected model, the seven categories other than silence are further dichotomized into clear (i.e., high likelihood) and unclear or quiet/distant (i.e., low likelihood) sub-categories.

Professional audio transcriptions were used to train the audio processing models, and the algorithms utilized the models to identify a variety of segments from the audio signals accurately and reliably. 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 V3.1.0 is described in LENA Foundation Technical Report LTR-05-2 and the transcription process in LTR-06-2.

After individual segments are identified, further processing generates key LENA data. Key child sound segments are analyzed through 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) and to estimate the number and duration of vocalizations produced by the child. Adult sound segments are processed to estimate the number of adult words a child hears. Non-speech

sound such as coughing, vegetative sounds, etc., are filtered out and statistical models are used to estimate the number of words spoken in each adult segment. Refer to LENA Foundation Technical Reports LTR-04-2 and LTR-05-2 for information on the segmentation process and speech/non-speech classifications. Statistical modeling is further used to detect Conversational Turns (CT), or back and forth alternation between the key child and an adult. For this purpose a conversation was defined as a contiguous region containing live human speech separated from the next conversation by a pause region of at least five seconds duration which contains only non-live-human speech audio signals. CTs cannot cross conversation boundaries. Results from the audio processing described above are written to the Interpreted Time Segments or ITS file, an XML-coded plain text compilation of every facet of data recorded and analyzed by the LENA software.Please see Technical Report LTR-04-2 for further information on the ITS file.

LENA software engineers continue to improve the algorithmic-based feature extraction and segmentation analyses. We intend to release upgrade versions of the software annually.

3.0 LENA SYSTEM AUDIO SPECIFICATION

The LENA System includes a Digital Language Processor (DLP) that was developed by hardware and software engineers at the LENA Foundation. Here, we describe the performance goals associated with the DLP, 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, the unit must be stable and maintain high levels of inter-recorder reliability, and the performance goals center on these two aspects of the design. LENA Foundation 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 DLP-0121, 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.

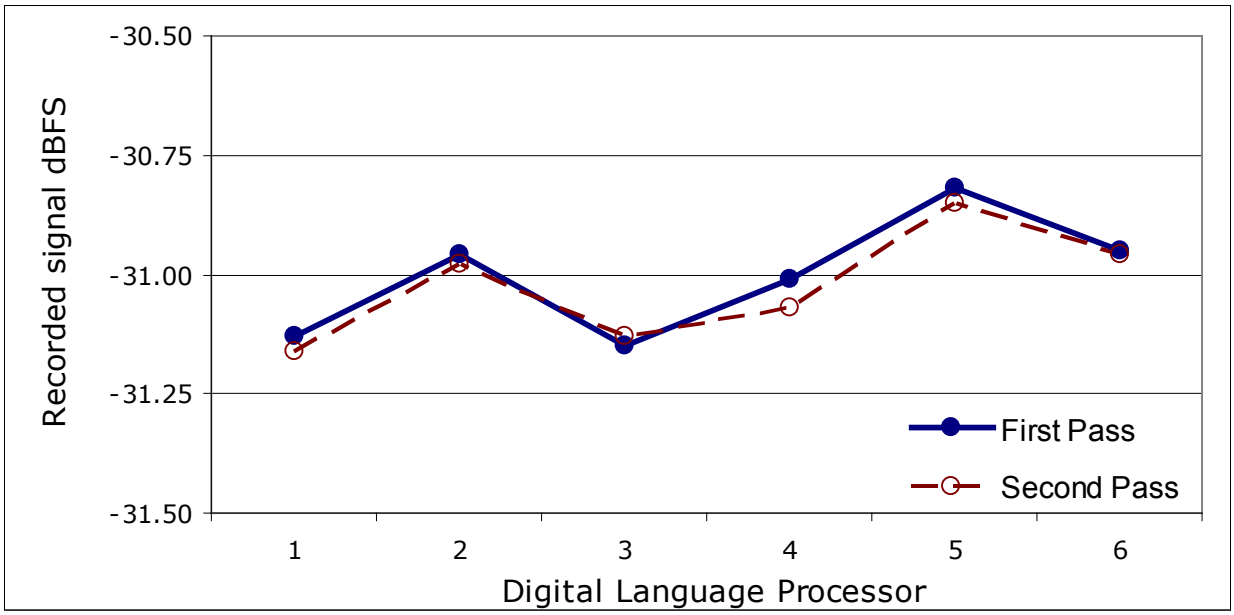

Figure 2. Signal reliability using six DLP units chosen at random.

LENA Foundation 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 minimize variation (67 dBC SPL to -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. The current DLP model DLP-0121 meets or exceeds US/Canada compliance standards. Standards for compliance that were either met or exceeded are shown in Table 1.

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

Standards for Compliance	Description	DLP-0121
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	✔

[a]UL: Underwriters Laboratories Inc; EU: European Union

3.2 Hardware

Audio data were collected using an omnidirectional microphone with a flat 20 Hz-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 512 MB flash memory using a 4:1 Adaptive Differential Pulse Code Modulation compression scheme (DVI-4 ADPCM). The flash 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 USB 2.0 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 DLP-0121 unit peak operating power is 50 mW. A primary 450 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 time-stamping 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 approximately 5 years.

3.3 Simple Operation

The LENA DLP was designed for usability. It is equipped with a power switch and a test button (Figure 3). A visual feedback mechanism 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 compact (3-3/8" x 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 (DLP-0121), actual size.

4.0 CONCLUSION

We have described the four components of 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 categorical components including male and female adult, key child, other child, overlapping speech, noise, electronic noise, and silence. Key child segments are further segmented into speech/non-speech, and adult segments are processed to generate AWC estimates. Adult child alternations are processed into CT estimates. The LENA DLP is simple to operate and the inter-unit signal variation is low, as assessed by test-retest reliability. The DLP model DLP-0121 has either met or exceeded US/Canada compliance standards.