LENA™
FOUNDATION

The LENA™ Language Environment Analysis System:
The Interpreted Time Segments (ITS) File

Dongxin Xu, Umit Yapanel, Sharmi Gray, & Charles T. Baer
LENA Foundation, Boulder, CO

LTR-04-2

September 2008

Software Version: V3.1.0

ABSTRACT

The LENA Digital Language Processor (DLP) records real-time audio data that is transferred to a computer for processing by LENA language environment software V3.1.0. Ultimately an Interpreted Time Segments (ITS) file is created that summarizes the processing results for the data contained within the original audio file. LENA Foundation data analyses have focused primarily on estimating Adult Word Counts (AWC), Child Vocalizations (CV), and Conversational Turns (CT) between adult and key child. However, the data processing reveals numerous other factors that have not yet been studied by researchers at the LENA Foundation. In this paper, we reveal the general content of the ITS file. We also provide a detailed description of each component specific to the ITS file and implications with regard to potential research and analyses.

Keywords

Adult word count, audio-transcription, conversation, conversational turns, XML.

1.0 INTRODUCTION

The Interpreted Time Segments (ITS) file is the final output step of the LENA software analysis. One ITS file is created for each recording file that is processed. In essence, the ITS file is an algorithmic transcription of the audio data extracted from the recording file. ITS files are written in standard XML format to facilitate data mining and other research goals.[1] A portion of the data written to the ITS file may be visualized using the LENA software.

2.0 CREATING THE ITS FILE

The ITS file is the culmination of an iterative series of algorithmic procedures and analyses (Figure 1). Briefly, a recording file (typically consisting of 12 or more hours of natural home environment audio) is transferred from the LENA DLP by the LENA software. In a three-step process the audio stream is partitioned into segments which are categorized based on the "speaker" or type of sound each segment represents.

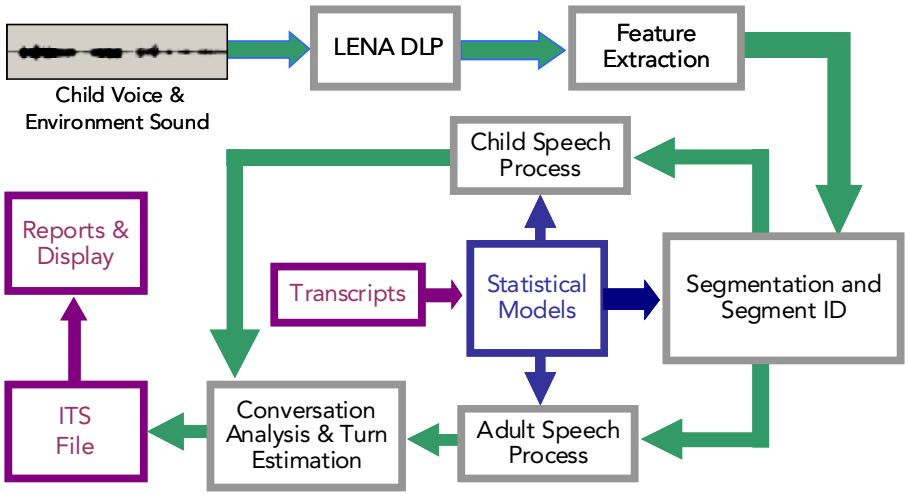

Figure 1. LENA Language Environmental Analysis Audio Processing System.

1 XML code can be converted to other formats or exported to other database files, including Excel. Any standard XML parser can parse the ITS file, and a specially designed, comprehensive ITS file parser (ADEX) is available for users of the LENA Pro software.

In step one, feature extraction and segmentation, certain acoustic features are extracted from the recording and used to partition the audio stream into short segments through a series of iterative modeling algorithms. Recordings made in a natural environment may contain many types of sounds (e.g., people, noise, TV/radio) and periods of silence. The segmentation step can be thought of as marking the time boundaries of different types of sound segments prior to classifying the segment type.

In step two, preliminary classification, the sound segments are identified by a sound category. Segment identifications comprise eight primary categories (Key Child, Other Child, Adult Male, Adult Female, Overlapping sound, TV and other electronic sound, Noise, and Silence), each one denoting a specific statistical model derived from acoustic training data. The acoustic features of a given segment are compared to those of the eight primary models, and the segment is assigned a preliminary classification based on the model representing the closest statistical fit.

In step three, segment identification, audio segment categorization is completed. The fit of each segment to the preliminary classification model is compared to the fit of the segment to the Silence model to determine the likelihood (or certainty) of the preliminary classification. The initial segment identification is retained unless the segment is statistically "close" to the Silence model (i.e., the likelihood of the chosen model falls below a specified threshold). In that case, the segment is reclassified from its primary category into a corresponding secondary (or "Faint") category.[2] For practical analysis purposes, all Faint categories may be combined. Table 1 details the fifteen possible resulting segment identification codes used in the ITS file.

2 For example, a segment is initially classified as "Adult Female" because that model fits the segment best, but the likelihood of this classification compared to that of the Silence model is low. In that case the segment is reclassified as "Adult Female - Faint."

Table 1: Segment Identification Codes Assigned During Segmentation.

Segment ID Code	Segment Description
MAN / MAF	Male Adult / Male Adult - Faint
FAN / FAF	Female Adult / Female Adult - Faint
CHN / CHF	Key Child / Key Child - Faint
CXN / CXF	Other Child / Other Child - Faint
NON / NOF	Noise / Noise - Faint
OLN / OLF	Overlap / Overlap - Faint
TVN / TVF	Electronic / Electronic - Faint
SIL	Silence

After the final identification of segments has been determined, some primary category segments are processed further. Key child segments are analyzed to distinguish and quantify key child vocalizations (including words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries) from non-speech (including fixed signals and vegetative sounds). Male and female adult segments are analyzed to produce an estimate of the number of adult-spoken words in the child's immediate environment. Conversational Turns counts are then estimated based on the previous results. Please refer to technical report LTR-05-2 for additional details on performance. The result of the analyses is the ITS file, a compilation of every facet of data recorded and analyzed. The ITS file can be accessed directly by the LENA software or any XML parser to generate reports and visual representations of the data.

3.0 GENERAL CONTENT OF THE ITS FILE

The DLP unit is equipped with a real-time clock (RTC) for time-stamping audio data. All dates and times are in Greenwich Mean Time (GMT) and are formatted based on ISO 8601 Date and Time Formats Conventions. Please refer to the ISO 8601 documentation for further details on the time-stamp formatting (http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/#isoformats).

Each ITS file contains information on the XML format (e.g., version and encoding style) and file information (e.g., name, version, and time created). An audio header section specifies DLP-related identifying information. An algorithm section details the various LENA software algorithms and models used to analyze the audio data.

Key child and segmentation information follow. The key child information provides key child demographics such as age and gender. The Segmentation section includes each segment of speech spoken by an adult, key child, or other child, as well as other types of non-speech sounds. Please refer to technical report LTR-05-2 for more information on segmentation processes. Table 2 summarizes the components of an ITS File.

Table 2: Components of the ITS File.

Component	Description
Audio Header	Recorder information
Algorithm Version Section	LENA language environment analysis software models and algorithms
Key Child Information	Demographics and mean length of vocalization
Segmentation Information	Machine audio-processing result

Within the ITS file, speech and non-speech sound segments are classified by inclusion in either a pause or a conversation. Pauses may contain a variety of sounds including silence, noise, overlapping speech, distant or faint speech, child or adult vegetative sounds, child or adult fixed signals, or electronic noise (e.g. television or radio).[3] The duration of a pause is by definition greater than or equal to five seconds when it occurs between two consecutive conversations. However, a pause inside a conversation may be less than five seconds in duration.

3 Conversations may contain these same sound categories when the segment duration of the sound is less than five seconds.

4.0 RESEARCH IMPLICATIONS

Starting in 2005, scientists at the LENA Foundation conducted a large-scale study to assess the language environment of children ages 2 months to 36 months from families of varying socioeconomic status (SES) backgrounds. The result of this effort was a massive database of quantifiable adult-child speech phenomena in a natural home environment, known officially as the LENA Natural Language Corpus. The amount of data within this database, and specifically within the ITS files, is vast. Thus, considerable potential for data mining exists.

Segmentation information is conveniently summarized at the end of each ITS file. Thus, the user may easily access segmentation information at the hourly, 5-minute, or even higher resolution levels. From the segmentation summary, the user may acquire estimated values of adult word counts, conversational turns, and key child vocalizations and duration. Some additional information that could be acquired includes an estimate of the number of overlapping segments, the number of interactions between the key child and other children, the length of a television segment, and the amount of child vocalizations during segments containing television, conversation initiator identifications, and the number of instances of floor holding (i.e., monologues), among others.

There are currently four versions of the LENA system, LENA Pro, LENA Pro Graduate Student Version, LENA Language Assessment and LENA Home. The advanced LENA Pro version is specifically designed for pediatricians and speech language professionals and researchers. [The LENA Pro version provides access to the ITS file under special license, and is restricted to use for qualified research purposes.]

5.0 CONCLUSION

The ITS file is a user-friendly, exportable XML-formatted compilation of all algorithmic analyses of the original audio data file. The file is composed of an audio header, a descriptive algorithmic region, segmentation information, and key child information. The file may be mined intelligently by the user to answer new research questions, or may be used simply to report and display data that has been summarized. The content contained within the file has the potential to provide valuable information that has not been easily accessible until now.