## A Longitudinal Study of Infant Speech Production Parameters: **A Case Study**

Hynek Bořil<sup>1</sup>, John H.L. Hansen<sup>1</sup>, Dongxin Xu<sup>2</sup>, Gill Gilkerson<sup>2</sup>, Jeff Richards<sup>2</sup>

{hynek, John.Hansen}@utdallas.edu



<sup>1</sup>Center for Robust Speech Systems (CRSS) Erik Jonsson School of Engineering & Computer Science University of Texas at Dallas

LENA

<sup>2</sup>LENA Foundation Boulder, Colorado, USA







LENA Users Conference 2011: April 17-19, 2011

Denver, Colorado, USA

## Introduction

- Infant/children's speech processing plays important role in detection of language delay and early communication disorders, automated reading tutoring, or emotional state assessment.
- In order to design and improve performance of such applications, good understanding of children's speech structure and its development over

- . Study of infant speech production within the interval of 11 to 35 months of age, sampled with a step of 4 months.
- Analyzed parameters: fundamental frequency, formants, vocal tract length, average spectrum, spectral slope, duration of voiced segments, number of voiced segments in conversational turn, and recently proposed

gitudinal data from a healthy female infant acquired within the interval of 11 to 35 months of age, with a sampling step of 4 months. Analyzed data set comprises 7 recordings acquired at the age of 11, 15, 19, ..., 35 months.

For each age sample, segments containing child's utterances were select yielding 5 minutes of speech per age.

Recordings are stored in 16 kHz/16 bit format.

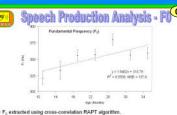
Corpus

ch database acquired by the LENA Foundation.

Subjects are recorded starting from 2 months through 36 months of age

Each recording is conducted during an 'ordinary' day in the child's natural

- Age 0-5 months: Fo increases in both cry and non-cry utterances Age 0-12 months: F. increases in hunger cries.
- Age 3, 6, 9 months: F<sub>0</sub> slightly increases at 6 months, followed by slight
- Age 8-26 months: not significant changes of F. in monosyllables and bi-
- Age 8 months = 3.5 years: difficult to track any significant longitudinal trends in
- Age 5-17 years: males significant F<sub>0</sub> drop starting from 11 to 15 years, no
- significant change later; females significant drop at 7-12 years, no significant change later.
- Age 8.5-11.5 years: F., in females decreases





Age 0-5 months: F. increases in both cry and non-cry utterances

# Design of a simple automatic age classifie



VTL - Literature Overview

Overview of LENA Corpus

More than 65,000 hours of recorded data.

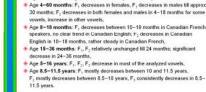
**Data Subset Utilized in Present Study** 

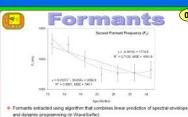
- Location of higher formant frequencies is believed to be more correlated with VTL compared to lowest formant frequencies, which are strongly dependent on
- technique used for normalizing inter-speaker vocal tract differences to improve automatic speech recognition), the warping parameter proportional to VTL differences is estimated from the inverse of higher formant frequencies
- In our study, inverse of F., used for estimation of VTL changes
- ♦ VTL varies with the production of distinct → averaging F, over longer speech
- segments (utterances) to estimate average, phone independent VTL



- months remains almost steady
- observed trend seems to correlate well with the accelerated VTL growth







- and dynamic programming (in WaveSurfer).

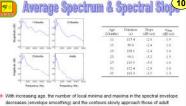
  Average F<sub>2</sub> displays descending trend with age. Similar was observed also for F<sub>3</sub> and F<sub>3</sub>.
- $(F_i: slope | a_{p_1} = -2.21 \text{ Hz/month, correlation coefficient } R^2 = 0.22; F_2: a_{p_2} = -8.56 \text{ Hz/month, } R^2 = 0.71; F_2: a_{p_3} = -6.90 \text{ Hz/month, } R^2 = 0.48).$ rrelation coefficients of the F., and F., trends are considerably higher than in F



- Despite the obvious accuracy limitation of this technique compared to MRI, the



- Average spectra of the infant subject are compared with the average spectrum extracted from the spontaneous speech of 23 US native adult female subjects from the UTScope database (average age  $\mu$  = 22.2 years, standard deviation  $\sigma = 3.6 \text{ years}$
- Average spectrum is extracted from a 25ms window shifted with a skip rate of
- Complementary average spectral slopes are extracted on the frame basis by fitting a straight line into the short term amplitude spectra in log frequency-log amplitude domain by means of linear regression.



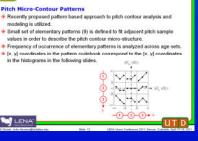
- Spectral slope considerably flatter than the one usually seen in adult speakers (typically

## Voiced Segment Duratio 11

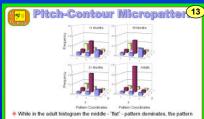
- . In children (typically 5 years and older), the duration of vowels, consonants syllables, and sentences tend to reduce with age and approach the one in adults In our study, the duration of voiced speech segments is analyzed. Voiced
- segments are formed by a non-interrupted sequence of voiced frames.
- . Average number of voiced segments per conversational turn is also counted

Average	Age (Months)						
	11	15	19.	23	27	31	35
Segment Duration (s	0.31	0.22	0.32	0.33	0.32	0.43	0.37
* Segments Turn	2.3	2.5	2.0	1.8	1.8	2.4	2.1

. It can be seen that unlike reported for older children, the durations tend to increase for our subject, while the number of voiced segments per conve



Pitch-Contour Micropatter 12



distribution in early infant speech is way more uniform. With increasing age, the 'flat' pattern becomes gradually more and more pronounced and the whole distribution approaches the one seen in adults.

## Automatic Age Classification 14

- The observed production variations can be expected to have a direct impact on coding commonly used in speech systems.
- To evaluate this hypothesis, we train separate acoustic models for each age sample set, yielding 7 models representing 11-months speech, 15-months speech. . . . 35-months speech.
- If the speech coding is age-sensitive, the models will capture age-dependent speech characteristics and can be used for speakerdependent automatic age classification.
- Gaussian mixture models (GMM) are utilized to model probability density functions (pdf's )of speech parameters
- Task pick a pair of most likely adjacent age models out of 6 possible neighbor
- A GMM maximum a posteriori classifier utilizing modified perceptual linea predictive cepstral coefficients (PLP) provided classification accuracy of 70 %





- This study has presented an initial analysis of speech production development in a healthy female infant subject conducted on the longitudinal data spanning 11-35 months of age.
- Trends in the reduction of formant frequencies and extending vocal tract length with age confirm our intuition and observations presented in the literature.
- On the other hand, the observed gradual increase of F<sub>S</sub> after the 12 months of age together with the increase of average voiced segment duration and reduction of voiced segments in conversational turns are somewhat surprising.
- Novel approach to children pitch contour analysis and modeling utilizing a codebook of pitch micro-contour patterns has been presented and shown to capture additional aspects of speech production maturing.
- inally, age dependency of common speech coding strategies has been exploited in the design of a speaker-dependent automatic age classifie
- Many of the observed differences between the child's and adult speech production can be directly utilized in improving current speech processing techniques.

