# Automatic Babble Recognition for Early Detection of Speech Related Disorders

Harriet J. Fell
College of Computer Science
Northeastern University
Boston, Massachusetts 02115, USA
Tel: +1-617-373-2198
fell@ccs.neu.edu

Joel MacAuslan
Karen Chenausky
Speech Technology and Applied Research
Lexington, Massachusetts 02173, USA
Tel: +1-781-863-0310
starcorp@ix.netcom.com

Linda J. Ferrier
Department of Speech Language Pathology and Audiology
Northeastern University
Boston, Massachusetts 02115, USA
Tel: +1-617-373-5754
l.ferrier@nunet.neu.edu

## 1. ABSTRACT
**We have developed a program, the Early Vocalization Analyzer (EVA), that automatically analyzes digitized recordings of infant vocalizations. The purpose of such a system is to automatically and reliably screen infants who may be at risk for later communication problems. Applying the landmark detection theory of Stevens et al., for the recognition of features in adult speech, EVA detects syllables in vocalizations produced by typically developing six to thirteen month old infants. We discuss the differences between adult-specific code and code written to analyze infant vocalizations and present the results of validity-testing.**

### 1.1 Keywords
infants - pre-speech vocalization - acoustic analysis - early intervention.

## 2. BACKGROUND
There is considerable research to support the position that infant vocalizations are effective predictors of later articulation and language abilities [13, 15, 19, 8]. Intervention to encourage babbling activity in at-risk infants is frequently recommended. However, research and clinical diagnosis of delayed or reduced babbling have so far relied on time-consuming and often unreliable perceptual analyses of tape-recorded infant sounds. While acoustic analysis of infant sounds has provided important information on the early characteristics of infant vocalizations and cry [1, 21] this information has not yet been used to carry out automatic analysis. We are developing a program, EVA, that automatically analyzes digitized recordings of infant vocalizations.

to appear in ASSETS 98 Proceedings

## 3. STAGES OF BABBLING
In order to study the babbling or prelinguistic non-cry utterances of typically developing infants, it is necessary to chart an infant's progress and compare it to a developmental framework. Oller [17], and Oller and Lynch [18] describe a suitable framework, comprised of five stages of babbling:

1. **The Phonation Stage** (0 to 2 months) quasi-resonant or quasi-vocalic sounds
2. **The Primitive Articulation Stage** (1 to 4 months) appearance of primitive syllables combined with quasi-vocalic sounds
3. **The Expansion Stage** (3 to 8 months) open vowels, squeals and growls, yells and whispers, raspberries
4. **Canonical Syllable Stage** (5 to 10 months) well formed syllables and reduplicated sequences of such syllables (We have found that 6 to 12 months is a more accurate range.)
5. **The Integrative or Variegated Stage** (9 to 18 months) meaningful speech, mixed babbling and speech

Infants' well formed syllables are formed of *closants* (consonant-like phonemes produced by oral cavity constrictions) and *vocants* (vowel-like phonemes, i.e. voiced, unconstricted segments).

## 4. EVA PROTOTYPE
Our first version of EVA was developed on a Macintosh computer using SoundScope by GWInstruments. It proved to be a potentially valid tool for counting the number of infant prelinguistic utterances at age 4 and 6 months, categorizing them into three levels of fundamental frequency, and three lengths of duration [7]. This simple analysis was appropriate for classifying babbles in Oller's "Expansion Stage" (3 to 8 months): open vowels, squeals and growls, yells and whispers, raspberries.

Our human judge and EVA showed better agreement (93%) in counting the number of utterances than the trained phoneticians in the Oller and Lynch study [18]. EVA and our human judge agreed on 80% of their categorizations according to duration, and 87% of their categorizations according to frequency, of the 411 commonly recognized utterances.

## 5. SYLLABLE RECOGNITION IN INFANT VOCALIZATIONS

Infant pre-speech vocalizations are included in many general infant assessment scales [2, 5] and have been suggested as important indices of early development [20, 14]. Mastery of syllabic utterances with a consonantal boundary appears to be an important developmental step that is a powerful predictor of later communication skills [9, 16]. Since this skill has its roots in cognitive, motor, social, and linguistic domains as well as the sensory area of hearing, it is a sensitive measure of the infant's development.

In the current phase of our project, we are developing tools to classify babbles in Oller's "Canonical Syllable Stage" (6 to 11 months): well formed syllables and reduplicated sequences of such syllables. There are several aspects of these vocalizations that we plan to analyze automatically:

1. Detecting and Classifying syllables (CV, VC, CVC, V)
2. Identifying reduplicated sequences of such syllables
3. Classifying closants by manner of articulation (e.g., stop, fricative, nasal, or liquid)
4. Classifying vocants as central versus peripheral. [3]

In this paper we report on our work on the first part of this analysis, detecting syllables. We present our methods and the results of validity testing.

## 6. THE COMPUTER SYSTEM AND ALGORITHMS

Our syllable-detection software is built on a program written by Liu [12] to detect landmarks (à la Stevens [22]) in adult speech. It runs on a SUN Sparc workstation equipped with the Entropic Signal Processing System (ESPS) software library environment with Waves [6].

The Liu-Stevens Landmark Detection Program was developed as a part of an adult-speech recognition system for analysis of continuous speech [11]; that system is founded on Stevens' acoustic model of speech production [23]. Central to this theory are *landmarks*, points in an utterance around which listeners extract information about the underlying distinctive features. They mark perceptual foci and articulatory targets. The most common landmarks are acoustically abrupt and are associated with consonantal segments, e.g., stop closures and releases. Such consonantal closures are mastered by infants when they produce syllabic utterances. Liu's program was tested on a low-noise (Signal to Noise ratio = 30 decibels) database, LAFF, of four adult speakers speaking 20 syntactically correct sentences. Error rates were determined for three types of landmarks -- Glottis, Sonorant and Burst. The overall error rate was 15%, with the best performance for glottis landmarks (5%).

The Liu-Stevens program first sends speech input through a general processing stage in which a spectrogram is computed and divided into six frequency bands. Then coarse- and fine-processing passes are executed. In each pass, an energy waveform is constructed in each of the six bands, the time derivative of the energy is computed, and peaks in the derivative are detected. Localized peaks in time are found by matching peaks from the coarse- and fine-processing passes. These peaks represent times of abrupt spectral change in the six bands.

In type-specific processing, the localized peaks direct processing to find three types of landmarks. These three types are:

1. g(lottis), which marks the time when the vocal folds transition from freely vibrating to not freely vibrating or vice-versa.
2. s(onorant), which marks sonorant consonantal closures and releases, such as nasals.
3. b(urst), which designates stop or affricate bursts and points where aspiration or frication ends due to a stop closure.

We have modified this program to accommodate the particular acoustic characteristics of infant speech and the signal-to-noise ratio in our recordings.

The first change was to adjust the boundaries of the six frequency bands to better capture abrupt changes in F0, F2, and F3 (F1 is unused). Using ranges for formants in infant vocalizations cited in the literature [10, 4, 9], we set the bounds on the six frequency bands as shown in the table below. Additionally, we created a seventh band, composed of the union of bands three through six, for future work in detecting burst landmarks. (Bursts are not detected in these infant recordings as reliably as they had been in Liu's adult recordings.) See Table 1.

<table>
<tr><th>Band</th><th>Bounds<br>for an Adult Male</th><th colspan="2">Purpose</th><th>Bounds<br>for an Infant</th></tr>
<tr><td>1</td><td>0-400Hz<br>$F_0 \sim 150$</td><td colspan="2">To capture $F_0$</td><td>150-600Hz<br>$F_0 \sim 400$Hz</td></tr>
<tr><td></td><td>$F_1 \sim 500$Hz</td><td colspan="2">Ignore $F_1$</td><td>$F_1 \sim 1000$Hz</td></tr>
<tr><td>2</td><td>800-1500Hz</td><td rowspan="2">For intervocalic consonantal segments<br>a zero is introduced in this range:<br><br>"Bands 2 and 3 overlap in the hope that one of these bands will capture a spectral prominence.[12]</td><td rowspan="4">At a sonorant consonantal closure, spectral prominences above $F_1$ show a marked abrupt decrease in energy.<br><br>Onsets and offsets of aspiration and frication will lie in at least one of these four bands.</td><td>1200-2500Hz</td></tr>
<tr><td>3</td><td>1200-2000Hz<br>$F_2 \sim 1500$Hz</td><td>1800-3000Hz<br>$F_2 \sim 3000$Hz</td></tr>
<tr><td>4</td><td>2000-3500Hz<br>$F_3 \sim 2500$Hz</td><td></td><td>3000-4000Hz</td></tr>
<tr><td>5</td><td>3500-5000Hz</td><td></td><td>4000-6000Hz<br>$F_3 \sim 5000$Hz</td></tr>
<tr><td>6</td><td>5000-8000Hz</td><td colspan="2">Spans the remaining frequency up to 8000Hz.</td><td>6000-8000Hz</td></tr>
<tr><td>7</td><td></td><td></td><td>A threshold might be used on this band to detect +b/ -b landmarks</td><td>1800-8000Hz</td></tr>
</table>

**Table 1: Spectral Bands Used for Landmark Detection (adult vs. Infant)**

We use a high-pass filter (cutoff: 150Hz) on our source files before applying the landmark program. This lowers the interference from ambient noise.

We were not satisfied with the initial marking of voicing (+g/ -g) done by Liu's program on our digitized samples. This algorithm works by looking only at the energy in Band 1. It assumes that high energy in this band indicates the presence of voicing. Our infant vocalizations usually exhibited lower energy than the adult male samples in the LAFF and TIMIT databases used by Liu. The ESPS get_f0 function [6] (which measures periodicity to calculate F0 in voiced regions) appeared to be more reliable at finding the voiced parts of the infant signals. We integrated this with Liu's program by multiplying Liu's coarse-pass Band 1 energy by the "probability of voicing" returned by get_f0 (a 0/1 value).

Experimentation with get_f0 resulted in settings of Min F0 = 150Hz and Max F0 = 1200Hz. (An earlier study [7], of four infants found an average F0 between 290Hz and 320Hz.) Utterances with fundamental frequency in the range were handled appropriately by the landmark software. We did not attempt to handle squeals, i.e. vocalizations with F0 > 1200. The lower threshold of 150Hz was sufficient to filter out noise. This left sounds that might be classified as growls but that were nonetheless suitable for analysis by the landmark program.

The program applies a variety of rules to check and possibly modify the initial +g/ -g settings. Adult rules and infant rules differ for two reasons. In data samples of an adult male reading single sentences, no pauses were expected. So the original code inserted a +g/-g into any -g/ +g interval of duration greater than 150ms. In a ten-second sequence of infant babbles, there are likely to be pauses that are at least 150ms long, so we adopted 350ms for this insertion rule. We adjusted thresholds related to vocalic energy levels to accommodate the faint babbles uttered by some of our infants.

## 7. VALIDITY TESTING

We measured the extent to which the landmarks found by EVA to mark syllable boundaries corresponded with distinctions made by trained human listeners (the judges).

We started with an inter-judge reliability study to assure the consistency of independent hand-marking of spectrograms by the judges. We then conducted a small study comparing the results of EVA to the landmarks agreed on by the judges.

## 7.1 Subjects

Five subjects were enlisted into the study, four typically developing and one with hydrocephaly, surgically rectified immediately after birth but with some slight gross motor delay. The typically developing subjects comprised three male and one female; one male was African American. All infants had English-speaking parents. (In addition, we collected one sample from each of three typically developing children at ages 12, 13 and 14 months to allow working on the software and testing it on a small number of syllabic utterances.) See Table 2.

| Subject | Sex | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|
| TW | boy | x |  | x | x |  | x |  |
| EK | boy | x |  | x | x |  | x | x |
| NZ | boy |  | x | x |  | x | x |  |
| JD | boy |  |  |  | x | x | x | x |
| LS | girl |  | x | x |  | x | x |  |

x - recorded

**Table 2: Subjects and Months Recorded**

## 7.2 Data collection procedures

Parents of the normal infants interested in participating in the study were initially contacted by phone and biographical data was collected. Infants with early medical problems, including histories of middle ear infections, were eliminated.

Parents filled out an Infant State Form to testify that their infant had followed a normal schedule and was in an appropriate mood for recording. Infants whose parents reported their infants were in an atypical state were requested to reschedule. The infants were then recorded interacting with their parents in a sound-proof booth. Digital recordings were made using a high quality lavaliere microphone with a wireless transmitter and receiver. The following equipment was used for recording:
Shure Brothers, Inc. Lavalier Unidirectional Microphone
Shure Brothers, Inc. Transmitter SCI-CL
Shure Brothers, Inc. Wireless Receiver SC4-CL
Shure Brothers, Inc. Mike-to-Line Amplifier 11
Panasonic SV-3700. Professional Audio Digital Tape-Deck.

Parents were instructed to play with their infants as they normally would and to elicit vocalization by showing the infants toys and talking quietly to them. They were instructed to stop talking when their infants vocalized. At the end of the recording, parents filled out a second part of the infant state form describing the infant's state and vocalization behaviors during the recording period.

## 7.3 Data Analysis by Humans

The recordings were divided into two sets, a development set and a test set. Recordings were first examined by a graduate student in Speech-Language Pathology to identify where syllabic babbles occurred. They were then digitized into the Waves program and edited to remove mothers' utterances, non-speech utterances (e.g., coughs, sneezes, and crying) and background noise. Syllabic utterances were hand-marked using a broad phonetic transcription.

In order to remove parents' utterances, non-speech utterances, and background noise from the tapes of the babies we recorded, the tapes were screened to mark the locations and phonetic content of the reduplicative babbles. We digitized only those sections containing clear babbling, omitting sections obscured by extraneous noise or voices. The resulting files were high-pass filtered to remove noise below 150Hz. All files contained recordings of syllables that consisted of, at minimum, a vowel nucleus (a vocant), and optionally, an oral or glottal constriction (a closant).

## 7.4 Hand-marking conventions

Two phoneticians (LF, KC) trained together on a set of recordings that contained a representative variety of closants and vocants, working out a system of hand-marking and establishing the following conventions:

(1) Mark closants at the beginning of F0 or at the burst if there was one (i.e., at oral release) with the symbol 'h#'. However, if intervocalic glottal closures are perceptually long enough to function as a syllable boundary between vocants, count them as closants and mark them accordingly with 'h#'. Otherwise, include them with the vocant and do not specifically mark them. /h/ and glottal stops are generally not considered closants. For stops with unclear, fricative-like releases, mark them as closely as possible to the beginning of the release of the constriction, not at the end of frication. For true fricatives, mark both the beginning of frication and the oral release with the symbol 'sh' to delineate the extent of frication.

(2) Use the symbol 'em' to delineate nasals, marking both oral closure and oral release.

(3) Use the symbol 'hv' to delineate vocants, marking the beginning and ending of audible F0 energy. Long vowels and diphthongs, even if they contain a change in pitch or amplitude, are considered as one vocant.

(4) Ignore squeals, utterances with no F0 and bilabial trills (raspberries). To indicate regions in files that should be ignored, delineate them with the symbol 'ax-h'. Also bracket with this symbol nonlinguistic noises and

utterances obscured by noise that was not removed by filtering.

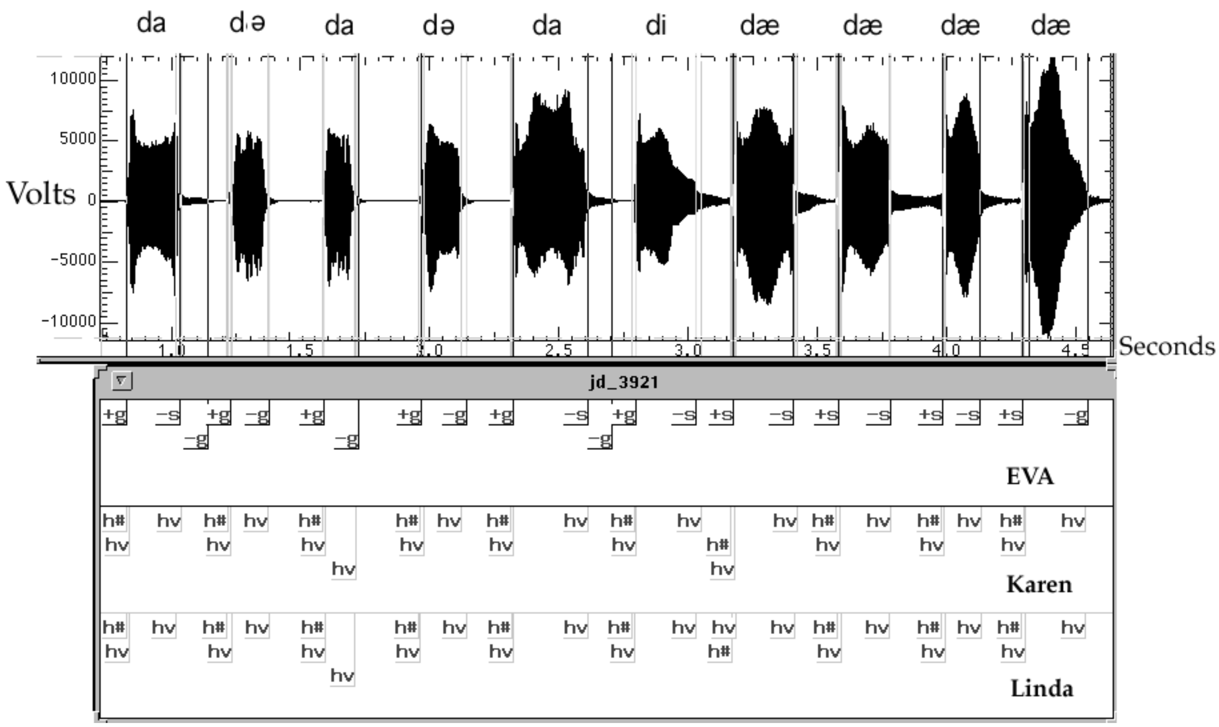


**Figure 1: Sample Marking of Landmarks by Human Judges and EVA**

## 7.5 Computer Marking of Landmarks

As described above, Liu's program marks three kinds of landmark, g(lottis), s(onorant), and b(urst) We presently make use only of the glottis and sonorant detectors. Our sonorant-consonant detection algorithm appears to work well on infant vocalization. (Interestingly, Liu's adult version showed high error rates for sonorant-consonants but not for bursts; but we find the reverse.) Between the glottis and sonorant features, we are able to mark voiced syllables. Preliminary results are discussed below while experimentation and analysis continue.

## 7.6 Experimental Results and Analysis

We performed an inter-judge reliability study and compared EVA to human judgment.

### *7.6.1 Inter-Judge Reliability*

The criterion for inter-judge agreement on perceptual hand-marking of landmarks was set at 95%. We defined inter-judge reliability as

$$\frac{\text{number landmarks agreed upon by both}}{\text{number agreements + number disagreements}}.$$

Inter-judge reliability on the existence, type and placement of landmarks was carried out by having the two trained phoneticians independently mark 15 arbitrarily selected, digitized samples not in the training set. These samples ranged in length from 3 to 10 seconds. Each judge used headphones to listen to the infant utterances, and viewed both the time waveform and wide-band spectrogram on-screen. The judges worked independently, entering their judgments into separate files. At no point did they refer to analyses carried out by the EVA program.

Since the longest file was 10 seconds long, the worst screen resolution was approximately 10 ms per pixel. (Waveforms were typically scaled to the screen width.) The accuracy was limited primarily by screen resolution and each judge's ability to manipulate the computer's mouse.

Between them the two judges marked a total of 204 landmarks (see Figure 1 for an example). The judges marked 198 landmarks that were close to each other. One judge marked 4 landmarks that the second judge did not, and the second judge marked two that the first did not. The two judges agreed on presence and position, to within 50ms, on 185 of the 204 landmarks (91%) and on position, to within 100ms, on 194 of the 204 landmarks (95%). We felt that this agreement met our original training criterion of 95% inter-judge reliability on perceptual hand-marking of landmarks. All utterances marked after the completion of this test were then analyzed by one judge. Problematic utterances were marked and agreement jointly negotiated with the other judge.

*7.6.2 EVA to human judge comparison:*

To test EVA-human reliability, the following rules were used to match EVA's landmark vocabulary with that of the human judges. Landmarks that were further than 200ms apart were assumed to represent different events.

For the purposes of the first test, we considered as valid only those landmarks hand-marked and agreed upon to within 100ms by both human judges. As shown in the table, we did not count h# (glottal closure) marks at the start of voiced regions because these are not (yet) in EVA's repertoire.

| Human Judges | EVA |
|---|---|
| h# (at voicing start)<br>hv<br>hv | (do not match/count)<br>+g<br>-g |
| | or |
| h# (within voicing)<br>hv<br>hv | +s<br>(no correspondence)<br>-g or -s (if both marked, use -g, do not count -s) |
| hv (no preceding h#)<br>hv | +g<br>-g |
| em (only one instance)<br>em<br>hv<br>hv | +g<br>(no correspondence)<br>-s<br>-g |
| em<br>em (no hv/hv following) | +g<br>-g |
| ay (initial glide) | (ignore region at this point) |
| sh (only one instance)<br>hv<br>hv | +s<br>(no correspondence)<br>-g |

**Table 3: Correspondence of EVA and Human Markings**

In this test, we considered the human consensus as the standard. H denotes the total number of validly hand-marked landmarks. We counted three kinds of errors for EVA: A <u>deletion</u> is a missed landmark where EVA detected no landmark in the vicinity of the hand-labeled landmark. The deletion rate is given by:

Deletion rate = (D/H) x 100%
where D is the number of deletions.

An <u>insertion</u> is a false landmark, i.e., one which was detected by EVA but not by the judges. The insertion rate is given by:

Insertion rate = (I/H) x 100%
where I is the number of deletions.

A <u>shift</u> is a landmark, found by EVA and matched by our correspondence rules, that is more than 100ms from the corresponding hand-labeled landmark. The shift rate is given by:

Shift rate = (S/H) x 100%
where S is the number of shifts

Then the total error rate is defined as:

E = Deletion rate + Insertion rate + Shift rate.

*7.6.3 First Comparison Experiment*

As a preliminary test of EVA's agreement with human landmark detection, we ran EVA on 15 digitized samples that were marked by both judges. Together, the 15 digitized samples had 128 validly marked landmarks (agreed on by both judges). Comparatively, EVA had 9 insertions and 3 deletions. Of the 125 landmarks found by EVA that corresponded to the valid human landmarks, only 1 was more than 100ms from either judge. The error rates were:

| | | | | |
|---|---|---|---|---|
| Deletion rate | = | (3/128) x 100% | = | 2% |
| Insertion rate | = | (9/128) x 100% | = | 7% |
| Shift rate | = | (1/128) x 100% | = | 1% |
| Total Error rate | = | 2.3 + 7.0 + 0.8 | = | 10% |

The insertions were a +s/ -s pair 12 ms apart, three +g/ -g pairs 32, 35, and 37ms apart, and an unmatched -g. The deletions were all at the end of one digitized sample and just after a region that was not counted due to disagreement between the judges.

The table below shows the statistics on the 124 landmarks, that were agreed on by both judges and corresponded to EVA landmarks (within 100ms):

| Distance between Landmarks | EVA to judge 1 | EVA to judge 2 | inter-judge |
|---|---|---|---|
| Average | 15.4ms | 13.7ms | 15.0ms |
| Standard Deviation | 16.1ms | 16.3ms | 16.3ms |
| Maximum | 93 | 93 | 90 |
| number > 50ms | 5 | 6 | 6 |
| percent > 50ms | 4.0% | 4.8% | 4.8% |
| number > 70ms | 4 | 4 | 2 |
| percent > 70ms | 3.2% | 3.2% | 1.6% |

**Table 4: EVA vs. Human Comparison 1**

EVA counted 67 syllables in these samples while the judges found 64. In places where one syllable boundary was missing, 0.5 syllable was counted. EVA deleted 1.5 syllables found by the judges and inserted 4 syllables, the +s/-s and +g/-g pairs described above.

### *7.6.4 Second Comparison Experiment*

Second Comparison Experiment: In this test, the digitized samples were hand-labeled by only one phonetician and compared to EVA's performance. (Based on the previous two-judge comparison, perhaps 2-3% of the differences should be attributed to the judges, not to EVA.) The 11 samples were all of subject LS, five recorded at 7 months and six recorded at 8 months. The phonetician picked these samples to test EVA because they contained relatively few phonemes not yet in EVA's repertoire, e.g. glides. The 7-month samples, however, were known to contain substantial amounts of vocal fry in the utterances, which might interfere with EVA's evaluation.

In the 11 digitized samples, 150 landmarks were marked by the human judge. EVA detected 2 landmarks that the judge did not and omitted 17 that the judge noted. Of the remaining 135 landmarks found by EVA that paired with the valid human landmarks, only 2 were more than 100ms from their corresponding marks. The rates of disagreement (which may or may not be EVA errors) were:

| | | | |
|---|---|---|---|
| Omission rate | = (19/150) x 100% | = | 11.3 % |
| Extra-detection rate | = (2/150) x 100% | = | 1.3 % |
| Shift rate | = (2/150) x 100% | = | 1.3 % |
| Total Error rate | = 11.3 + 1.3 + 1.3 | = | 13.9 % |

The table below shows the statistics on the 133 landmarks that corresponded to EVA landmarks to within 100ms (i.e. all but the two shift errors).

| **Distance between Landmarks** | **EVA-to-Judge** |
|---|---|
| Average | 19.2ms |
| Standard Deviation | 19.7ms |
| Median | 13.1 |
| Maximum | 86.5 |

**Table 5: EVA vs. Human Comparison 2**

EVA detected one +s/-s pair not found by the phonetician (judge); a second phonetician deemed that this pair was unclear. EVA omitted 19 landmarks labeled by the human judge. Of these, 12 landmarks (6 syllables) came from the 7-month recordings and represented syllables with substantial vocal fry; 4 landmarks (2 syllables) were found in a single digitized sample in a region of low vocal energy; the others may have been erroneous deletions by EVA. In all cases, the second phonetician concluded that the disagreements were within the bounds of typical inter-judge (human) differences.

## 8. CONCLUSIONS

We are optimistic, at this point, that EVA can validly and reliably find landmarks associated with infant syllable boundaries. This is the first step in establishing a method of automatically determining whether infant utterances are demonstrating the syllabic complexity to be expected at the end of the first year of life.

## 9. CURRENT AND FUTURE WORK

We are now using EVA to count the number and duration of syllables in all recordings of the five infants in this study. We are also noting the syllable patterns that EVA finds (e.g., +g/-g, +s/-s, +g/-s/-g) to see whether the distribution of these patterns changes over the five recordings for each infant.

We plan to expand EVA to include categorization of syllables according to initial sounds (whether there is an initial consonant and if so, whether it is a stop, fricative, nasal, or liquid) and use this to develop an inventory of phones in infants' repertoires at each age level. The data will be used to establish an initial normative data base on the pre-speech development of typically developing infants.

## 10. ACKNOWLEDGMENTS

We would like to thank Sharlene Liu for making her code available to us, Ursula Muroff for recording the infants in this study, and Corine Bickley for her suggestions and her help as a third phonetician.

This work was sponsored in part by NIH grant #R41-HD34686.

## 11. REFERENCES

[1] Bauer, H.R., 1988. Vowel Intrinsic Pitch in Infants, *Folia Phoniatrica 1988*, **40**, 138-146.

[2] Bayley, N. 1969. *Manual for the Bayley scales of infant development*. New York: The Psychological Corporation.

[3] Bickley, C. 1989. Acoustic evidence for the development of speech, *RLE Technical Report No. 548*, 111-124.

[4] de Boysson-Bardies, B., Halle, P., Sagart, L., and Durand, C., 1989. A crosslinguistic investigation of vowel formants in babbling. *J. Child Lang.,* **16**, 1-17.

[5] Capute, A. J., Palmer, F. B., Shapiro, B. K., Wachtel, R. C., & Accardo, P. J. 1981. Early language development: Clinical application of the language and auditory milestone scale. In R.E. Stark (Ed.), *Language Behavior in Infancy and Early Childhood*. New York.

[6] ESPS with Waves, 1993. Entropic Research Laboratory, Inc. AT&T Bell Laboratories

[7] Fell, H.J., Ferrier, L.J., Schneider, D., & Mooraj, Z. 1996, EVA, An early vocalization analyzer: an empirical validity study of computer categorization. *Proceedings of Assets '96*, 57-61.

[8] Jensen, T. S., Boggild-Andersen, B., Schmidt, J., Ankerhus, J., & Hansen, E. 1988. Perinatal risk factors and first-year vocalizations: Influence on preschool language and motor performance. *Developmental Medicine and Child Neurology*, **30**, 153-161.

[9] Kent, R. D. & Murray, A. D. 1991. Acoustic features of infant vocalic utterances at 3, 6 and 9 months. In R. J. Baken & R. G. Daniloff (Eds.), *Readings in Clinical Spectrography of Speech* (pp. 402-414). New Jersey: Singular Publishing Group and Kay Elemetrics.

[10] Kent, R. D. & Read, C. 1992. The Acoustic Analysis of Speech, San Diego: Singular Publishing Group, Inc.

[11] Liu, S. 1994. Landmark detection of distinctive feature-based speech recognition. *Journal of the Acoustical Society of America,* **96**, 5, Part 2, 3227.

[12] Liu, S. 1995. *Landmark Detection for Distinctive Feature-based Speech Recognition*. Ph.D. Thesis. Cambridge, MA: Mass. Inst. Tech.

[13] Locke, J. L. 1989. Babbling and early speech: Continuity and individual differences. *First Language*, **9**, 191-206.

[14] Locke, J. L. 1993. *The child's path to spoken language*. Cambridge, MA: Harvard Univ. Press.

[15] Menyuk, P., Liebergott, J., Shultz, M., Chesnick, M, & Ferrier, L.J. 1991. Patterns of Early Lexical and Cognitive Development in Premature and Full-Term Infants. *J. Speech and Hearing Res.*, **34**, 88-94.

[16] Menyuk, P., Liebergott, J., & Shultz, M. 1995. *Early Patterns of Language Development in Full Term and Premature Infants*. Lawrence Erlbaum Associates: Hillsdale, N.J.

[17] Oller, D.K. 1980. The emergence of sounds of speech in infancy. In G. H. Yeni-Komshian, J.F. Kavanagh, & C.A. Ferguson (eds.), *Child Phonology Volume 1: Production*. New York: Academic Press.

[18] Oller, D.K. & Lynch, M.P. 1992. Infant utterances and innovations in infraphonology: Toward a broader theory of development and disorders. In Ferguson, C.A., Menn, L., & Stoel-Gammon, C. (Eds.), *Phonological Development: Models research implications* (pp. 509-536). Timonium, Maryland: York Press.

[19] Oller, D. K. & Seibert, J. M. 1988. Babbling of prelinguistic mentally retarded children. *American Journal on Mental Retardation*, **92**, 369-375.

[20] Proctor, A. 1989. Stages of normal noncry vocal development in infancy: A protocol for assessment. *Topics in Language Disorders*, **10**(*1*), 26-42

[21] Stark, R.E. 1981. Stages of Development in the first year of life, In *Child Phonology*: *Vol. I*, G. H. Yeni-Komshian, C. A. Ferguson, J. Kavanagh (Eds.), New York: Academic Press.

[22] Stevens, K.N. 1992. Lexical access from features. *Speech Communication Group Working Papers*, Volume VIII, Research Laboratory of Electronics, MIT, 119-144.

[23] Stevens, K.N., Manuel, S., Shattuck-Hufnagel and Liu, S. 1992. Implementation of a model for lexical access based on features. *Proc. Int'l. Conf. Spoken Language Processing*, Banff, Alberta, **1**, 499-502.