visiBabble for Reinforcement of Early Vocalization

Harriet Fell
College of Computer and Information Science
Northeastern University
Boston, MA, 02115, USA
+1 617 373 2198
fell@ccs.neu.edu

Joel MacAuslan
Speech Technology and Applied Research
54 Middlesex Turnpike
Bedford, MA, 01730, USA
+1 781 276-4540
JoelM@S-T-A-R-corp.com

Cynthia Cress
Communication Disorders Department
University of Nebraska
Lincoln, NE, 68583-0732, USA
+1 402 472 4431
ccress1@unl.edu

Linda Ferrier
Speech-Language Pathology and Audiology
Northeastern University
Boston, MA, 02115, USA
+1 617 373 2892
l.ferrier@neu.edu

ABSTRACT
The visiBabble system processes infant vocalizations in real-time. It responds to the infant's syllable-like productions with brightly colored animations and records the acoustic-phonetic analysis. The system reinforces the production of syllabic utterances that are associated with later language and cognitive development. We report here on the development of the visiBabble prototype and field-testing of the system .

Categories and Subject Descriptors
J.3 [**Computer Applications**]: Life and Medical Sciences

General Terms
Experimentation, Human Factors.

Keywords
acoustic analysis, babbles, early intervention, acoustic landmarks.

1. INTRODUCTION
Communication skills are vital to educational and vocational success. Cerebral palsy, developmental apraxia (DAS), neurological insult/injury (e.g. head injury, encephalitis, meningitis), oral/motor dysfunction, cognitive impairments, tracheotomy, and deafness can all impair or even preclude speech. A child having any of these or other syndromes may not be able to produce a sound when he or she wants to; may produce a limited range of sounds, often vowels and 1-2 consonants; or may not have learned to associate his or her sounds with meaningful referents [2]. During an intervention to promote speech-like vocalizations, non-speaking children tended to have difficulty initiating sounds and participating in vocal imitation. They produced atypical sounds such as elongated vowels, distorted consonants, and non-speech sounds [2].

ASSETS'04, October 18–20, 2004, Atlanta, Georgia, USA.

Because of the atypical sound production of infants in this population [11], traditional intervention strategies to prompt or respond to infant vocalizations may not be sufficient to promote change. Children at risk for speech impairments often produce a higher percentage of vowel-like sounds and a lower percentage of consonant-like sounds during later development than would be expected for typically developing children. The purpose of the visiBabble system is to increase the vocalizations of such infants, and, to increase the sophistication and variety of such vocalizations by differentially reinforcing syllabic versus non-syllabic productions.

The visiBabble system processes vocalizations in real-time. It responds to the child's syllable-like productions with brightly colored animations and records the acoustic-phonetic analysis. The system reinforces the production of syllabic utterances that are associated with later language and cognitive development. As a child interacts with visiBabble, the program collects and analyzes the infant's utterances so that it can be used by a child as a toy/trainer or by a practitioner as a clinical or research implement. The techniques involved with VisiBabble directly reflect best practice in prespeech intervention demonstrated in treatment efficacy studies VisiBabble uses a responsive strategy based on a commonly accepted "play" mode of speech intervention, in which activities are child-initiated, child-focused, and do not include specific training events [16].

The remainder of this paper is organized as follows. Section 2 cites some of the clinical studies that stress the importance of early speech intervention and points out the connection between landmarks that we detect and the infant's control of his or her articulators. Finally, it discusses our prior work validating our landmark analysis and using this analysis to derive a measure that correctly distinguishes pre-speech vocalizations of typically developing children from those of children who are independently known to be developmentally delayed. Section 3 describes the visiBabble system, how we detect landmarks and syllables in real time, our graphic feedback, automatic data collection, and field-testing. Section 4 discusses the results of our field-testing, and section 5 presents our plans for futute work.

2. BACKGROUND AND PREVIOUS WORK

2.1 Babbling - Predictor of Language Abilities

Infant vocalizations are effective predictors of later articulation and language abilities [10, 11, 15]. These studies have been carried out on normally developing children and on children with a variety of early diagnosed problems. These research studies emphasize the importance of early speech intervention for children at risk for being non-speaking. They also point out the difficulty of providing sufficient speech practice and feedback for children with such atypical speech patterns through traditional forms of intervention and interaction.

2.2 Acoustic-Phonetic Landmark Model

Central to the acoustic-phonetic processing are *landmarks*, points in an utterance around which listeners extract information about the underlying distinctive features. They mark perceptual foci and articulatory targets [12, 17]. .

As babbling develops, the infant begins to coordinate control of the vocal folds and the velopharyngeal opening with control of the tongue blade and the lips, and the true consonants appear. In the landmark model, the larynx and the velum are considered secondary articulators, and they are "bound" to control by the primary articulators. That is, implementation of the laryngeal and nasal features depends, in some ways, on the implementation of the primary articulator.

2.3 Closants, Vocants, and Syllables

Consonant-like sounds or *closants* can be detected in the sound waveform from acoustic evidence of abrupt changes in the spectrum of sound. These have been called landmarks by some researchers of adult speech. Landmarks that result from the creation or release of a narrow constriction or closure along the vocal tract are also found in pre-linguistic vocalizations. We can hypothesize that the development of the ability to produce sounds exhibiting landmarks is a necessary skill underlying the production of syllables.

Vowel-like sounds or *vocants* appear early in the vocalizations of infants and are characterized by slowly time-varying spectral patterns. These sounds result from movements of the tongue body, the jaw, and the lips, and are usually produced with the vocal folds positioned to vibrate. A variety of vowel-like sounds appear as the infant learns to control the positioning of these articulators. [1].

A syllable consists of a vowel (or vocant), usually preceded, and/or followed by a consonant (or closant).

2.4 Our Applications of the Landmark Model

We have applied the landmark model in various applications concerning adult speech, for example in detecting emotional stress in speech [7]. We also applied this method to detect landmarks in infant vocalizations and performed validity studies that show that our computer results compare favorably with analysis by trained phoneticians [3, 4, 5]. We then used the detected landmark patterns to describe and quantify the syllable structures produced. This analysis has, in turn, been used to formulate a "vocalization age" that clinically distinguishes between typically developing infants and infants at risk for later speech difficulties [6, 7]. A vocalization age is a normative age-equivalence estimate of the range of speech sounds (inventory of landmark sequences) expected for typically developing children.

This encouraged us to apply the technology more broadly. By revising the algorithms for real-time use, we have been able to create the visiBabble prototype to respond to, and not merely measure, a child's vocalizations. The goal of our first development and field tests was to reinforce syllable-like sounds that are clear precursors to speech without reinforcing cries, coughs, and other non-speech productions. As shown below, visiBabble met this goal.

3. METHODOLOGY

3.1 The visiBabble System

The visiBabble prototype system includes a modern notebook computer (Dell Inspiron, 2.4 GHz Pentium 4 running Windows XP), a microphone, a 15" flat-panel display (see Figure 1), and software, which carries out the following functions:

- Landmark detection – detects landmarks in a child's vocalizations in real-time.
- Graphic feedback -- provides real-time visual response to sound input;
- Data collection – records each session and saves the result as a wav file, collects data on the types and duration of vocalizations produced;
- Experimental formats -- allows the system to run and data to be collected in single-case study formats.

3.2 Finding Landmarks

Our landmark detector is based on Stevens' acoustic model of speech production [17]. The program detects three types of landmarks:

- **glottis:** marks the time when the vocal folds start (+g) and stop (-g) vibrating;
- **sonorant:** marks sonorant consonantal releases (+s) and closures (-s), (e.g., voiced closants);

- **burst:** designates stop/affricate bursts (+b) and points where aspiration/frication ends (-b) due to stop closure.

The visiBabble system can track simple aspects of the acoustic signal in real time, based on a low-resolution spectrogram. In particular, the signal is sampled at 16 kHz and analyzed into a small number, nominally 64, of separate, frequency intervals of ~256 Hz each. A 16 kHz rate provides information up to 8 kHz, sufficiently high to include at least 3-4 formants for an infant and to show the distinction between voicing and other speech sounds: fricatives, stop releases, bursts, etc. (These parameters are suitable for using the FFT and impose no delay of their own beyond 4 ms, i.e., 1/256-th of one second.) The visiBabble system uses only one-half of these frequency intervals because the others differ only in phase.

The spectral intervals are grouped into six broad bands. 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. These peaks represent times of abrupt spectral change in the six bands. Energy in bands 2 (1200 - 2500 Hz.) and 3 (1800 - 3500 Hz), e.g., provides evidence of voicing or, in some cases, of bursts. The distinction between these is readily made in the time domain (voicing persists much longer than bursts) as well as by appeal to information in the other spectral bands: voicing provides a power spectrum that decays with frequency approximately as 1/frequency2, whereas most other speech sounds have flatter spectra.

For the poorly formed or unstable closants and vocants typical of infants, wide frequency bands are well suited to recognition: Higher frequency resolution would require averaging over bands anyway. It would require spending more time computing and – worse – more time sampling the signal for the initially higher resolution.

3.3 Graphic Feedback

Figure 2: Characters used in the visiBabble Prototype

The visiBabble prototype responds to the child's utterances with five different brightly colored animations that cycle to avoid habituation: (a train, a bird, a frog and two cartoon creatures that move across the screen, see Figure 2). It responds to the start of each syllable it detects by advancing the current animation one step.

The system determines that a syllable has started either by noting the onset of voicing (a +g landmark) or by a voiced closant that occurs at least 100 ms after the start of the previous syllable. Admittedly, a syllable might start with a burst before the voicing onset but, to avoid responding to noise, visiBabble waits for the onset of voicing. The system responds in no more than 0.25 second of the corresponding acoustic event.

3.4 Experimental Formats

Single case study designs [14] are particularly suited to our preliminary tests of visiBabble since they provide the freedom to conduct a study on a small heterogeneous group of subjects. The prototype program can be run in a variety of "formats":

- Baseline (recording, no graphic display);
- Response (graphic display is always present, while recording);
- A-B-A (no display, display on, no display). The length of A or B phases can be changed.

3.5 Data Collection

As visiBabble runs, it makes a digital recording of the session which it saves in wav format. It also saves a record of the times and types of landmarks it found during the session. Data is collected during all phases of all formats to allow a comparison of behavior during the baseline and active phases.

3.6 Syllable Analysis

After a visiBabble session, a post-processor uses the landmark types and times recorded by visiBabble to analyze the child's session for clinical or research purposes by:

- Removing landmarks from areas of the recording that have been corrupted by noise, as well as landmarks produced as artifacts of the process
- Grouping landmarks into "syllables" of certain specific types such as "+g/-s/-g", using information about their order and spacing. For each recording session, creates a profile using the number of syllables, number of syllables types, and duration of each syllable type.
- Grouping syllables into "utterances"--series of syllables occurring closely together--based on timing considerations.
- Tabulating the average number of syllables per utterance, as well as the number of utterances comprised of 1, 2, 3, and more syllables, for each subject and session.

The analyses of landmarks and syllables are conducted and recorded separately for the B phase and two A phases. (See Table 2.)

3.7 Field Testing

As part of the software development, a prototype of the system was beta-tested by a typically-developing 11-month-old and was evaluated in trials with five children with severe expressive impairments, whose starting ages ranged from 28 months to 7.5 years, and one premature but typically developing infant whose starting age, corrected for prematurity, was 11 months. It is currently being tested with two additional typically developing but premature infants. The system will be iteratively modified in response to the results of this field-testing.

Preliminary questions on the use of the visiBabble include:

1) What features of infant vocalization can the system respond to in real time?

2) What graphic feedback do infants find appealing?

3) What changes have to be made in the graphic feedback to avoid habituation?

4) Do the infants show increased babbling, e.g. increased number of syllables, during the treatment (B) phases?

5) Do infants adjust the amplitude of their utterances in response to the visual reinforcement?

6) Do infants adjust the pitch of their utterances in response to visual reinforcement?

7) Do infants increase the variety of syllable types and complexity of their utterances?

8) What changes appear in variety of syllable types and complexity of utterances as an infant matures?

9) Do parents perceive changes in their infants' vocalizations in response to the visiBabble program?

The ABA design allows direct comparisons of the child's productions (items 4 to 8) with and without the system's visual feedback. Both the rate and the variety of syllables may be tested for the stimulating effect of the system by several techniques.

3.8 Human Issues - Iterative Design

Our beta-testing with a typically developing 11-month-old old showed that our system was responsive to a child of that age. On days when he wasn't cranky, as reported by his parents, he showed an interest in the visual response screens. These sessions were run by the child's parents in a particularly noisy environment. Noise from the heating system, a vacuum cleaner, parents talking, and the computer itself were often louder than the child and clearly affected the output. The child was also very interested in the buttons on the display.

As a result of these sessions, we now ask that the computer be placed behind the microphone and that observers, if they must speak, do so as quietly as possible and also behind the microphone. We have also placed black tape over the display buttons.

Our field-testing with developmentally delayed children revealed another attractive nuisance, the microphone (see Figure 1). We placed the microphone between the flat-panel display and a neck pillow that also served to cover the space below the display and hide the cables. Rubbing, breathing, and chewing on the microphone can interfere with visiBabble's operation. The last of these is probably not very good for the child either. In future versions of visiBabble, the microphone will have a wind screen and be hidden inside the pillow.

4. RESULTS

We describe here the results from twenty-one sessions, with three children with severe expressive impairments and one premature child (see Table 1 and Figure 3). Three completed sessions, one with each of three children, were not used in this study. In a session with K1, the external monitor did not work when attached to the computer. In a session with K2, there were adults talking, both about and even *to* the infant, throughout the session. Therefore, the data was deemed not valid for this study. In one K3 session, all the vocalizations were laughter.

One developmentally delayed child, K5, was clearly not yet a candidate for visiBabble and the details of sessions with this child are not included in this analysis. This child vocalized fewer than ten times in six complete sessions. The visiBabble system was not effective in increasing her vocalizations. It isn't designed to elicit vocalizations from children who can't voluntarily produce speech sounds. If there is nothing to respond to, the system cannot reinforce and motivate further speech sounds.

In the tests reported here, the system rarely responded to noise and whispering that can be heard in the background. The exception to this occurred when such sounds overlapped with the child's utterances. For the twenty-one sessions analyzed, landmarks not caused by the child's voluntary vocalizations were discarded before the analysis.

The results of a sample session are shown in Table 2. The subject of this session, K2, is a twenty-eight-month-old female child with developmental disabilities. She is an intentional communicator with a few vowel and consonant sounds.

Each session consisted of a 10-minute visiBabble run in A-B-A format. The no-feedback, A phases (A1 and A2) each ran for 2.5 minutes. The B phase, with feedback, ran for 5 minutes. Data was collected throughout the 10-minute session and we compared the results during the B phase with the combined results of the A1 and A2 phases. Our results, therefore, are a comparison of vocalizations during five minutes with feedback and five minutes without.

In the twenty-one sessions of children with severe expressive impairments, number of syllable and utterance productions was significantly greater in the visiBabble test (B) segments compared to the control sessions (sum of A1 and A2). The average vocal production varied across test and control conditions: mean number of syllables: 676 test vs. 442 control, and mean number of utterances 522 test vs. 355 control. The quantity of vocalizing was significantly different at $p < .01$ between test and control conditions, using nonparametric Wilcoxon signed-rank comparisons to account for the non-Normal population distribution: number of syllables $S(21) = 208$, $p < .001$, and number of utterances, $S(21) = 189$, $p < .006$. [9]). See figures 3a for syllable production across subject sessions and 3b for number of utterances across subject sessions.

This difference was demonstrated not only in the total sessions, but also for eighteen of the twenty-one sessions individually. In fact, the total syllable production rose in the test (B) condition for every child, by amounts ranging from 26% to 203% (Student's $t(5)$ from 15.7 to 29.9, $p < .05$ to $p<.005$). The number of utterances rose similarly.

More importantly, the number of syllable *types*, a measure of the variety of syllables produced, also rose substantially and significantly, from a mean of 9.0 in the control (A) condition to 11.9 in the test (B): $S(21) = 195$, $p < .003$, signed-ranks test. See figure 3c for the number of distinct syllable types across subject sessions.

There were no statistically significant differences in the mean number of syllables per utterance (1.30 in A vs. 1.25 in B: most utterances were monosyllabic for these infants), mean syllable duration (247 vs. 241 ms), or mean number of landmarks per syllable (2.81 vs. 2.89).

These results are very encouraging. We observed an approximately 50% increase in syllable and utterance production and a 33% increase in the variety of syllable types when visiBabble was active.

5. FURTHER DEVELOPMENT

There are several features we plan to add to the visiBabble system. We have observed that some young infants are not always interested in our visual feedback. They may not be focused on the part of the screen where the bird is flying or the frog is hopping. We will add feedback that occupies more of the screen, e.g. fireworks or large faces that wink or smile. We may add sound or tactile feedback to the responses.

Though our prototype system just responds to the detected start of syllables, it is also capable of responding to other aspects of the child's vocalizations, e.g. variation in pitch or energy, the duration of syllables or utterances, or the complexity of syllables in terms of landmark structure. We plan further tests with infants and children on these aspects of the system. We envision a system where a speech pathologist, for example, might choose to work with a child on producing longer utterances and set the visiBabble system accordingly.

For research purposes, we plan to add to the information saved by the visiBabble system. We currently save a digital audio recording of each session and the landmark analysis as it was computed in real time. From this, we are able to identify the syllables that visiBabble found and hence responded too. In future systems, we will likewise record which response was displayed so that we might determine whether certain responses are particularly effective. We will also save the pitch information as it was computed during the session. Our summary program will then be augmented to classify syllables according to pitch contours as well as landmark content.

We hope to see visiBabble become a product that is useful as a clinical and research tool for work with at-risk infants or older non-speaking children. We also intend to produce a version that can be used as a training toy for these infants and children.

Table 1: Syllable Count and Duration Summary

Summary Developmentally Delayed Children									
	All		A1		B		A2		A1 & A2
Syllables	Number	Mean Duration	Number	Mean Duration	Number	Mean Duration	Number	Mean Duration	Number
K1_1	47	0.252	4	0.289	26	0.247	17	0.251	21
K1_2	26	0.238	4	0.283	21	0.234	1	0.152	5
K1_3	28	0.263	8	0.291	18	0.256	2	0.218	10
K1_4	14	0.226	4	0.198	7	0.268	3	0.166	7
K1_5	115	0.386	43	0.381	60	0.395	12	0.363	55
K2_1	20	0.187	0		18	0.191	2	0.153	2
K2_2	62	0.214	15	0.171	30	0.13	17	0.401	32
K2_3	73	0.241	24	0.175	20	0.275	29	0.274	53
K2_4	29	0.181	8	0.117	18	0.165	3	0.444	11
K2_5	94	0.291	20	0.45	69	0.251	5	0.215	25
K3_1	77	0.264	17	0.26	50	0.262	10	0.282	27
K3_2	114	0.289	21	0.315	76	0.3	17	0.208	38
K3_3	92	0.267	30	0.28	45	0.261	17	0.261	47
K3_4	102	0.302	14	0.349	71	0.305	17	0.25	31
K3_5	108	0.245	33	0.291	59	0.229	16	0.212	49
K4_1	21	0.195	1	0.209	19	0.201	1	0.064	2
K4_2	8	0.243	0		8	0.243	0		0
K4_3	6	0.165	0		5	0.19	1	0.04	1
K4_4	27	0.201	6	0.159	17	0.218	4	0.191	10
K4_5	31	0.258	6	0.268	25	0.255	0		6
K4_6	24	0.231	7	0.341	14	0.177	3	0.233	10
All wtd means	**1118**	**0.269**	**265**	**0.293**	**676**	**0.261**	**177**	**0.267**	**442**

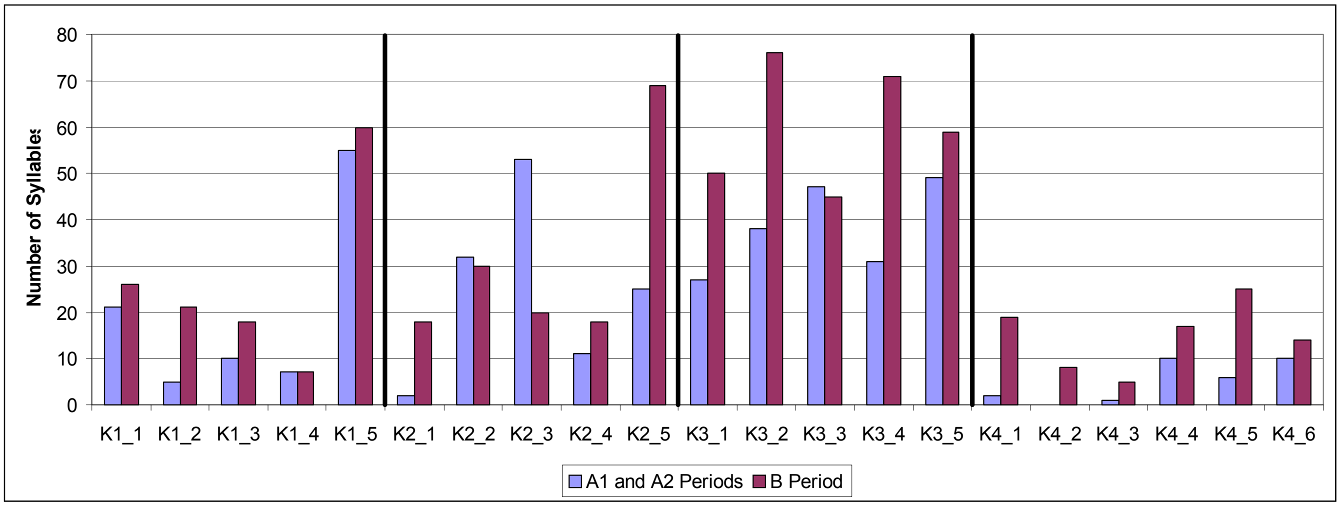

Figure 3a: Comparison of Number of Syllables in Twenty-one Sessions with Developmentally Delayed Children

Notice that the number increased during the visual-feedback or "B" phase in nearly every session.

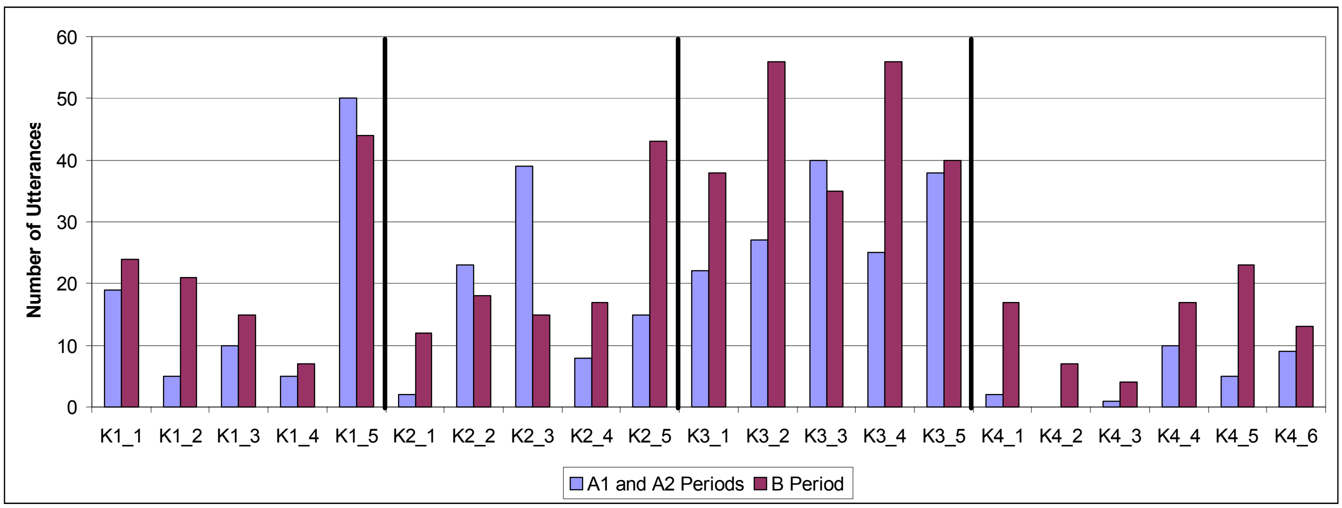

Figure 3b: Comparison of Number of Utterances in Twenty-one Sessions with Developmentally Delayed Children

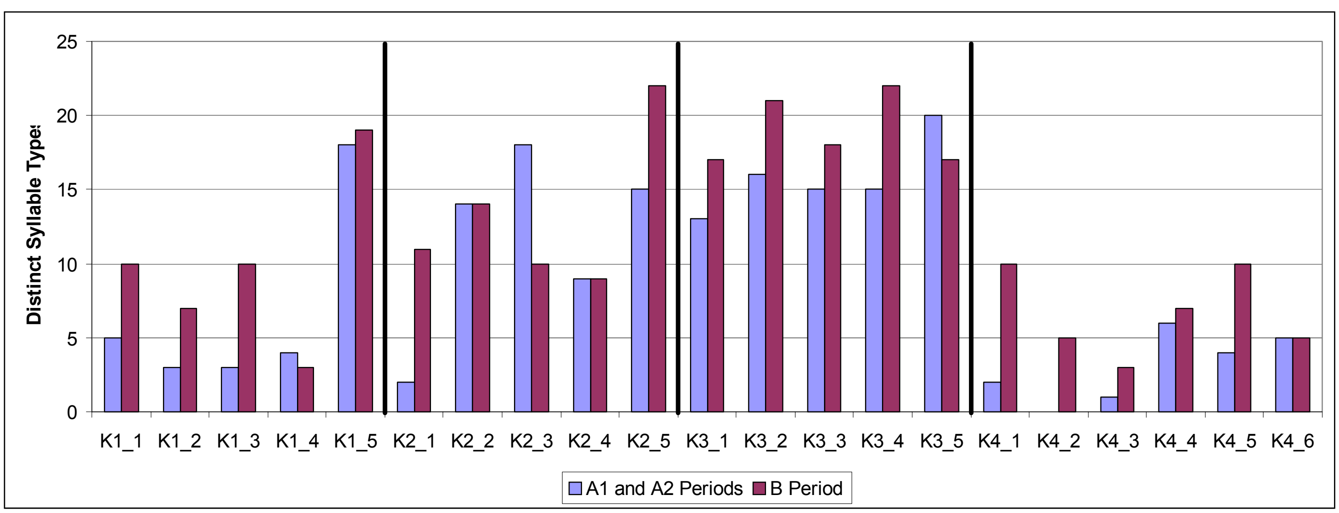

Figure 3c: Comparison of Number of Syllables Types in Twenty-one Sessions with Developmentally Delayed Children

Table 2: Sample Summary of the Data Collected During a 10-minute A-B-A visiBabble Session.

A1 2.5 minutes - no display; B 5 minutes with responsive display; A2 2.5 minutes - no display.
This particular child essentially stopped vocalizing when the feedback ended as shown in column A2.

K2_20030723_1911ed.lm Created using visiSyl_1_2 and v_collectSyllables

SYLLABLE	**ALL**		**A1**		**B**		**A2**	
TYPE	**COUNT**	**MEAN DURATION**	**COUNT**	**MEAN DURATION**	**COUNT**	**MEAN DURATION**	**COUNT**	**MEAN DURATION**
+g-g	12	0.242	1	0.394	11	0.228		
+g+s	1	0.49					1	0.49
+g-s	1	0.064			1	0.064		
+s-g	3	0.144			3	0.144		
+s-s	2	0.008			2	0.008		
+b+g-g	2	0.081			1	0.113	1	0.048
+b+g+s	1	0.144			1	0.144		
+b+g-s	1	0.176	1	0.176				
+g-g-b	4	0.268			3	0.249	1	0.326
+g+s-g	18	0.337	5	0.644	13	0.219		
+g+s-s	4	0.139			4	0.139		
+g-s-g	7	0.22	2	0.236	5	0.213		
+s-g-b	1	0.104					1	0.104
+s-s-g	6	0.239	3	0.405	3	0.072		
+b+g-g-b	3	0.407			3	0.407		
+b+g-s-g	1	0.538			1	0.538		
+g+s-g-b	3	0.696	3	0.696				
+g+s-s-g	7	0.689			7	0.689		
+s-s-g-b	1	0.33			1	0.33		
+g+s-s-g-b	1	0.257	1	0.257				
+s+s	1	0.105					1	0.105
+g+s+s	5	0.219	2	0.265	3	0.188		
+g+s+s-g	1	0.16			1	0.16		
+g-s-s-g	2	0.406	1	0.418	1	0.394		
+s+s-s-g	1	0.12			1	0.12		
+s-s+s+s	1	0.136			1	0.136		
+s-s-s-g	2	0.112			2	0.112		
+g+s-s+s-s-g	2	0.165	1	0.225	1	0.105		
ALL TYPES	**94**	**0.291**	**20**	**0.45**	**69**	**0.251**	**5**	**0.215**
Number of Landmarks	**Number of Syllables with the Number of Landmarks Shown at the Left**							
2	20		1		17		2	
3	49		13		33		3	
4	22		4		18		0	
5	1		1		0		0	
6	2		1		1		0	
7	0		0		0		0	
Mean Landmarks/Syllable	3.106		3.4		3.058		2.6	
Different Syllable Types	28		10		22		5	
Number and Duration of Utterances	58	0.518	14	0.68	43	0.448	1	1.25
Syllables per Utterance:	**Number of Utterances with the Number of Syllables Shown at the Left**							
1	39		10		29		0	
2	11		2		9		0	
3	5		2		3		0	
4	0		0		0		0	
5	1		0		0		1	
6	2		0		2		0	
Mean Syllables/ Utterance	1.603		1.429		1.581		5	

6. ACKNOWLEDGMENTS

We would like to thank Gwen Sterup and Andrea Heinrich, graduate students in speech pathology at the University of Nebraska - Lincoln, for their enormous help in collecting the data for this study.

This work was sponsored in part by NIH Grant R41 DC005534.

7. REFERENCES

[1] Bickley, C. Acoustic evidence for phonological development of vowels in young children. *MIT Speech Communication Working Papers IV*, (1984), 111-124.

[2] Cress, C.J., and Ball, L. Strategies for promoting vocal development in young children relying on AAC: Three case illustrations, *Proceedings of RESNA '98* (Minneapolis, MN, June 1998). RESNA Press, 44-46.

[3] Fell, H.J., Ferrier, L.J., Sneider, D., and Mooraj, Z. EVA, An Early Vocalization Analyzer: an empirical validity study of computer categorization, *Proceedings of Assets '96* (Vancouver, BC, April 1996), ACM Press, 57-61.

[4] Fell, H.J., MacAuslan, J., Ferrier, L.J., and Chenausky, K., Automatic babble recognition for early detection of speech related disorders, *Proceedings of Assets '98* (Marina del Rey, CA, April 1998), ACM Press, 59-66.

[5] Fell, H.J., Ferrier, L.J., Espy-Wilson, C., Worst, S.G., Craft, E.A., Chenausky, K., MacAuslan, J., and Hennessey, G., Automatic analysis of infant babbling in EVA, the Early Vocalization Analyzer, *Proceedings of ASHA 2000* (Washington, D.C., November 2000), abstract only.

[6] Fell, H.J., MacAuslan, J., Ferrier, L.J., Worst, S.G., and Chenausky, K., Vocalization age as a clinical tool, *Proceedings of ICSLP '02*, (Denver CO, September 2002). 2345-2348.

[7] Fell, H.J. and MacAuslan, J., Automatic Detection of Stress in Speech, *Proc. 3rd Int'l Workshop on Models & Analysis of Vocal Emissions for Biome.. Applications* (MAVEBA 2003, Florence, Italy, December 2003), pp. 9-12.

[8] Fell, H.J., MacAuslan, J., Cress, C.J. and Ferrier, L.J., Using Early Vocalization Analysis for visual feedback, *Proc. 3rd Int'l Workshop on Models & Analysis of Vocal Emissions for Biomed. Applications* (MAVEBA 2003, Florence, Italy, December 2003), pp. 39-42.

[9] deGroot, M.H. 1975. *Probability and Statistics*, 483-486, Addison-Wesley, Reading, MA.

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

[11] Levin, K., Babbling in infants with cerebral palsy, *Clinical Linguistics and Phonetics*, 13, 4 (1999), 249-267.

[12] Liu, S., Landmark detection of distinctive feature-based speech recognition, *Journal of the Acoustic Society of America*, 96, 5, Part 2 (1944), 3227.

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

[14] McReynolds, L.V. and Kearns, K.P. *Single Subject Experimental Designs in Communication Disorders*, University Park Press, Baltimore: MD, 1983.

[15] Menyuk, P., Liebergott, J., Shultz, M., Chesnick, M.. and Ferrier, L.J., Patterns of early language development in premature and full term infants, *Journal of Speech and Hearing Researc*h, 34 (1991), 1.

[16] Shriberg, L. and Kwiatkowski, J. Phonological disorders II: A conceptual framework for management, *Journal of Speech and Hearing Disorders*. 47 (1982), 242-256.

[17] Stevens, K.N., Manuel, S., Shattuck-Hufnegel, S., and Liu, S., Implementation of a model for lexical access based on features, *Proceedings of ICSLP '92*, (Banff, Alberta, 1992), 1, 499-502