Acoustic Topology of Infant Proto-Syllables.

James Robert Piroli

Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/5141
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Acoustic topology of infant proto-syllables

Piroli, James Robert, Ph.D.
The Louisiana State University and Agricultural and Mechanical Col., 1991

Copyright ©1992 by Piroli, James Robert. All rights reserved.
ACOUSTIC TYPOLOGY OF INFANT PROTO-SYLLABLES

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

The Department of Communication Disorders

by

James Robert Piroli
B.S., Eastern Michigan University, 1975
M.A., Eastern Michigan University, 1978
May 1991
ACKNOWLEDGMENTS

"THE LIST"
(In no particular order)

Raymond Manci, MBA
Amelia I. Hudson, Ph.D.
Raymond G. Daniloff, Ph.D.
A Committee of Six
Monica
Joseph Gonzalez, Ph.D.
Laura Hudgins
Michael Bojko
A.C., Ken, Billy, and Gaylen
Marsha Zlatin-Laufer, Ph.D.
R. Prakash Dixit, Ph.D.
Russ Giammanco
Miss Enterprise
Ned C. Stevens
Gary and Gary
Creighton J. Miller, Ph.D.

Betty Piroli
Orlando A. Piroli, Ph.D.
Pam Bloom
Norma Travis, Ph.D.
David, Thomas, and Richard
Marge Chamberlaine
Calvin Howell
HTP
Norma, Lola, and Racquel
Lloyd Sensat
Cunningfox's Cross-Cut
Joan Besing, Ph.D.
David LaVoy
The Aunts
Jimmy Coleman
Cassie
# Table of Contents

**Acknowledgments:** "The List"................................. ii

**List of Tables**.......................................................... v

**List of Figures**.......................................................... vi

**Abstract**........................................................................ vii

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION....................................................... 1</td>
<td></td>
</tr>
<tr>
<td>Infant Vocal Tract Anatomy............................ 2</td>
<td></td>
</tr>
<tr>
<td>Effect of Vocal Tract Shape on Phonological Development... 6</td>
<td></td>
</tr>
<tr>
<td>Estimating Articulatory-Acoustic Control from Infant Vocalizations... 8</td>
<td></td>
</tr>
<tr>
<td>Carrier-Modulation Model of Motor Speech.................. 13</td>
<td></td>
</tr>
<tr>
<td>Idealized Pattern of the CV Syllable...................... 18</td>
<td></td>
</tr>
<tr>
<td>Temporal Structure of the PS............................. 19</td>
<td></td>
</tr>
<tr>
<td>Statement of Problem....................................... 22</td>
<td></td>
</tr>
<tr>
<td>II.</td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTAL DESIGN........................................... 24</td>
<td></td>
</tr>
<tr>
<td>General Research Plan................................. 24</td>
<td></td>
</tr>
<tr>
<td>Subject and Stimuli Selection....................... 26</td>
<td></td>
</tr>
<tr>
<td>Stimuli Analysis........................................... 29</td>
<td></td>
</tr>
<tr>
<td>Spectral Measures........................................ 30</td>
<td></td>
</tr>
<tr>
<td>Measures of Phonatory Instability................... 32</td>
<td></td>
</tr>
</tbody>
</table>

iii
III. RESULTS .................................... 35
  Preliminary .................................. 35
  Temporal Measures ............................ 35
  Burst ......................................... 39
  Measures of Phonatory Instability .......... 42
  Spectral Structure ............................ 45
  Developmental Course of Infant Vocalizations .... 49

IV. DISCUSSION .................................. 68
  General Aspects of Syllable Morphology ... 69
  Temporal Patterning ............................ 71
  Spectral Patterning ............................ 74
  Theoretical Implications ...................... 80
  Developmental Issues .......................... 85
  Clinical Implications and Need for Further Research ... 87

  Summary ...................................... 90

V. BIBLIOGRAPHY ................................. 93

VI. VITA .......................................... 103
LIST OF TABLES

Table                  Page
1. Freidman's Two-Way ANOVA: Ranks on Age...................... 38
2. Correlations by Spearman Rho and Associated Student t Values.... 40
3. Mean Values for Adult and Infants..... 44
4. Transition Aberrancies............... 47
5. Vowel Phoneme Percentages of Current Infants by Age Compared with Irwin (1948)................. 54
6. Stop-Consonant Percentages by Age Level...................... 57
7. Means and Standard Deviations for all Subjects.................... 59
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figures</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Adult and Infant Syllable Duration</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Frequency Histogram</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>F1/F2 Scattergrams for All Subjects</td>
<td>50</td>
</tr>
<tr>
<td>3.</td>
<td>Comparison of F1/F2 Mean Values for Current Subjects With Those</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Reported by Lieberman (1980)</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Infant and Adult Transition Duration</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Frequency Histogram</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Infant and Adult Shimmer Values</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Frequency Histogram</td>
<td></td>
</tr>
</tbody>
</table>
ABSTRACT

A corpus of CV-like proto-syllables was obtained from the archival recordings of three first-born infants, which were made in conjunction with earlier research on the acoustic and interactive aspects of infant language acquisition. Tokens were limited to proto-syllables with perceptual evidence of stop-like or glide-like qualities and an identifiable vowel. Following A/D sampling, measures of developmental change were made including F1/F2 steady states, transition durations, CV durations, maximum F2 velocities, F2 onset, fundamental frequency, shimmer, time-to-peak amplitude velocity, and time-to-peak F2 velocity within the transition. Contrary to expectations, the orderliness of the infant proto-syllables was remarkable with few acoustic events indicative of loss or reduction of articulatory control.

With the exception of transition duration, all intrinsic temporal measures were positively correlated with overall syllable duration. The modal adult and infant syllable durations overlapped at approximately 240 msec with infants exhibiting a strong skew towards
longer durations. F1/F2 values fell in the expected location in the F1/F2 vowel space and were almost perfectly congruent with those reported by previous researchers. Shimmer measures failed to differentiate adults from children, despite trendline correction, suggesting that the glottal pulse for CV frames may be quasi adult-like as early as the fifth month of life. The uniformity of format trajectory suggests that the rather adult-like transitions emerged very early and changed hardly at all across the 5-14 month epic. In terms of distribution of vowels and frequency of occurrence of place of articulation for stops, the infants of the current study compared well with published data.

Theoretical implications are discussed including the presence of a coordinative-structure schemata for vocal motor control of transitions, the possibility of a biologically mediated predisposition to automatically affect rapid transitions and/or a recruitment of the components of reflexive gestures for phonatory purposes, and the fact that monosyllabic frames, once developed, may be stable temporal frameworks for speech output well into the second year of life.
CHAPTER I
INTRODUCTION

During the first twelve to fourteen months, the human infant experiences rapid changes in vocal tract anatomy and physiology which accompany a progression from the essentially reflexive behaviors of deglutition and distress cries, through vocal play and babbling, to the concatenation of sounds which, ultimately, are heralded by parents as the infant's first word. Traditional theories of speech development (Jakobson, 1941/1968; Jakobson and Waugh, 1979; Lenneberg, 1966) view this process as a discontinuous one in which babbling merely reflects the entire range of possible human speech sounds and bears no relationship to the child's later meaningful speech. Recent research, however, has found this process to be a gradual one, the early stages of which are governed by limitations imposed by the shape of the infant's vocal tract, as well as its level of neuromuscular development (Enstrom, 1982; Holmgren, et al., 1986; Kent and Bauer, 1985; Lieberman, 1980; Lieberman, et al., 1971, 1972;
interest in the examination of infant speech and language acquisition during the first year of life, then is the rapid change in the infant’s vocal tract structure and physiology.

**Infant Vocal Tract Anatomy:**

According to Kent and Bauer (1985), it is inappropriate to think of the infant vocal tract as simply a scaled-down version of the adult structure. It is, in fact, more like that of non-human primates than that of an adult human (Lieberman, et al., 1971; Fletcher, 1973; Kent and Murray, 1982; Lieberman, et al., 1972). When compared to the adult, the infant exhibits a shorter vocal tract, a high laryngeal position, direct approximation of the velum and epiglottis, a more anterior tongue mass that rests entirely within the oral cavity, a shorter pharyngeal cavity, and a modest flexion of the oropharyngeal channel (Kent and Murray, 1982; Kent and Bauer, 1985). It has been argued (Lieberman, 1968, 1969, 1980; Lieberman, et al., 1969, 1971, 1972) that these anatomical factors prevent non-human primates and newborn humans from producing the range of sounds that characterize human speech.
As in lower primates (Lieberman, et al., 1971; Fletcher, 1973; Negus, 1962), and the reconstructed vocal tract of the fossil La Chapelle-aux-Saints Neanderthal man (Lieberman, et al., 1972), the infant's laryngeal position is high (Bauer and Kent, 1987; Lieberman, et al., 1971) with the epiglottis at the level of the first cervical vertebre (Noback, 1923) and the cricoid at the level of the fourth cervical vertebre, (versus the third and sixth cervical vertebrae in the adult human, respectively) (Eckenhoff, 1951; Negus 1962). This tract configuration in the newborn infant places the open velum and epiglottis in direct approximation making the infant an obligate nose-breather (Bauer and Kent, 1987; Kent, 1981; Kent and Murray, 1982; Lieberman, et al., 1972; Sasaki, et al, 1977) and, therefore, an obligate nasal vocalizer (Kent and Murray, 1982; Oller, 1978). This velar-epiglottal linkage persists until the neonate acquires what Du Brul (1977) refers to as the distinctively human craneovertebral angle at approximately 4-6 months of age (Sasaki, et al., 1977). It is only after this disengagement that the infant produces routinely non-nasal vowel sounds (Oller, 1978) and reliably impounds intraoral
pressure requisite for the production of plosives and fricatives (Kent, 1981).

The small size of infants and their higher laryngeal positioning result in an appreciably shorter vocal tract, 7 to 9 cm from the glottis to the lips, compared to 17 cm for an adult male (Bauer and Kent, 1987; Kent and Bauer, 1985; Kent and Murray, 1982). The acoustic theory of speech production (Chiba and Kajiyama, 1958; Fant, 1960; Stevens and House, 1955), demonstrates a relationship between vocal tract configuration and formant structure that depends directly on the vocal tract area function (i.e., the cross-sectional area of the vocal tract as a function of the distance from the vocal folds to the lips). One would predict that the infant's shorter vocal tract would be associated with significantly higher formant frequencies than would the adult vocal tract (Bauer and Kent, 1987). Researchers have indeed shown that the younger and smaller the child, the higher are the formant frequencies for specific vowels (Kent and Murray, 1982; Lieberman, 1980). Because of the length and configuration of their supralaryngeal vocal tracts, infants cannot produce the adult formant patterns
that they hear (Kent, 1981; Lieberman, 1980). Although earlier researchers suggested that infants can produce formant ratios analogous to but not exactly at adult values (Lieberman, 1980; Lieberman, et al., 1971; Lieberman, et al., 1972), recent work by Crelin (1987) suggests that the child's vocal tract prevents adult-like vowel formant ratios until 2.5 to 3 years of age unless the child engages in "hyper-extended" oro-mandibular movements for vowel phonation.

In addition to a shorter supralaryngeal vocal tract, the infant's tongue lies completely within the oral cavity, while in adults the posterior third of the tongue is vertically positioned forming the anterior wall of the oral-pharyngeal cavity. (Buhr, 1980; Crelin, 1972; Irwin, 1947; Lieberman, et al., 1972). In the infant, then, there is practically no supralaryngeal portion of the pharynx present in the direct airway out of the larynx when the velum shuts off the nasal cavity (Kent, 1981; Lieberman, et al., 1972). This lack of a supralaryngeal-pharyngeal region prevents the infant from independently manipulating the cross-sectional area of the oral and pharyngeal cavities via a midpoint constriction, the
presence of which is requisite for the production of the quantal or point vowels [a], [i], and [u] (Lieberman, et al., 1972; Stevens, 1969).

Effect of Vocal Tract Shape on Phonological Development:

The infant's phonetic inventory is limited by a compressed formant space due to his immature vocal tract shape and inability to smoothly implement inter-articulator coordination (Browman and Goldstein, 1989; Lieberman, et al., 1971). Since the newborn infant possesses a supralaryngeal vocal tract configuration that approximates a uniform cross-section or slightly flared tube (Lieberman, et al., 1971), it is not surprising that the vowel repertoire of the zero to four month old infant is limited primarily to [æ], [I], [E], and [ʌ] (Irwin, 1949; Lieberman, 1980). Relative to consonant production, the infant initially produces only the glottal fricative [h], the glottal stop [ʔ], and perhaps [l] as well (Irwin, 1949). While infants appear to have the anatomical structures necessary for the production of both labial and dental consonants (Buhr, 1980; Lieberman and Crelin, 1971; Lieberman, et al., 1972), the fact that these sounds do not occur initially (Irwin, 1957) suggests a
general inability to produce rapid articulatory maneuvers (Lieberman, et al., 1972).

By the fourth to sixth month of life the infant's vocal tract grows into a configuration more adequate for speech-like vocalization and develops increased neurological maturity which undergirds the acquisition of speech motor skills (Kent, 1981). The most pronounced anatomic change to occur is the descent of the infant's larynx within the neck. This results in elongation of the pharynx, emergence of perpendicularly intersecting oral and pharyngeal tubes (the typically human craneovertebral angle), and separation of the velum and epiglottis (DuBrul, 1977; Sasaki, et al., 1977). The concurrent growth of the facial skeleton downward and forward increases the length and depth of the oral cavity. The resulting changes in vocal tract dimensions give rise to an enhanced range of fully non-nasal vowels sounds (Crelin, 1987; Oller, 1978, 1986), as well as palatal, velar, and pharyngeal constrictions which produce vocal tract shape discontinuities necessary for the quantal vowels [i], [u], and [a] (George, 1978; Kent, 1981; Lieberman, et al., 1971; Oller, 1978; Wood, 1982). Along with these structural
changes, neurological maturation enhances the infant's ability to impound intraoral air pressure and to rapidly modulate vocal tract width at physiologically optimal locations, resulting in enhanced production of supralaryngeal consonants (Holmgren, et al., 1986; Kent, 1981; Wood, 1982).

By six months of age the infant has also developed the motor skills necessary to achieve reduced variability, increased anticipation, and improved economy in the execution of serially ordered motor acts (Bruner, 1973). When applied to the production of speech, these developments result in a reduced variability of segment durations and spectra (Kent, 1976; Kent and Forner, 1980) and an increase in anticipatory coarticulation (Thompson and Hixon, 1979). At roughly the same time, the infant demonstrates a peak in the occurrence of repetitive motor behaviors (Thelen, 1981), including the use of the single or reduplicated articulatory movements associated with babbling (Koopsman-van Beinum and van der Stelt, 1986).

**Estimating Articulatory-Acoustic Control From Infant Vocalizations:**

Historically, research examining the infant's
developmental patterns of articulatory and language skills has concentrated on the age of emergence of phones and the relationship, if any, between emergence of specific phonetic and linguistic behavior and other identifiable motor milestones. This research was based upon phonetic transcription of infant productions, diaries, and anecdotal evidence (Irwin, 1947, 1948, 1949; Irwin and Chen, 1946 a, 1946 b). These early researchers have been criticized for their classification of infant speech sounds with respect to traditional phonetic transcription techniques which are based upon an adult phonetic system: imposition of adult phonetic inventories inevitably biases the analysis towards the native language of the adults in the child's language community (Bauer and Kent, 1987; Holmgren, et al., 1986) and forces pre-speech vocalizations into language-specific molds by imposing these adult phonetic categories on non-meaningful utterances (Holmgren, et al., 1986; Kent and Bauer, 1985). Since the infant's vocal tract is not nearly a linear transform of the adult in terms of size and shape (Fant, 1960), it seems likely that infants may in fact vocalize in a manner quite different from adult
linguistic categories (Bauer and Kent, 1987; Crelin, 1972, 1987; Kent and Bauer, 1985). Despite these difficulties, (perhaps because of them), phonetic transcriptions of infant sound productions between researchers are in substantial accord (cf., George, 1978; Irwin, 1948). Although such qualitative analysis remains an essential first step in a study of infant speech sounds, more recent work has attempted to quantify the vocal tract dynamics underlying infant sound production via non-invasive acoustic analysis (Bauer and Kent, 1987; Enstrom, 1982; Lieberman, et al., 1971, 1972; Zlatin and Koenigsknecht, 1976).

Based upon the results of recent acoustic analyses of infant vocalizations (Bauer and Kent, 1987; Buhr, 1980; Kent and Bauer, 1985; Kent and Murray, 1982; Laufer and Horii, 1977), the following generalizations can be drawn concerning the acoustic features of infant vocalizations. (1) The formant frequencies of infant vowels are approximately double those of adult males (Fant, 1960), with the approximate center of the three-dimensional vowel space having formant frequency values of $F1 = 1000$ Hz, $F2 = 3000$ Hz, and $F3 = 5000$ Hz (Bauer and Kent,
In view of the fact that estimates of the infant's vocal tract length are approximately half those of adult males, these formant values are not surprising and can be predicted by the acoustic theory of speech production (Chiba and Kajiyama, 1958; Fant, 1960). Additionally, the initial vocalic formant frequencies are appropriately modeled by a vocal tract configuration which approximates a single, right-circular tube, consistent with the infants vocal tract anatomy. (2) The majority of infant comfort-state vocalizations have a duration of 400 ms or less with mean duration increasing over the first year of life (Bauer and Kent, 1987; Kent and Murray, 1982). (3) The mean fundamental frequency (f0) typically falls within the range of approximately 300 to 600 Hz but with enormous variability (from less than 100 to 1000 Hz or higher).

While acoustic analysis provides information that is more objective than a perceptual analyses, most recordings of infant vocalizations are unusable because of acoustic characteristics which are peculiar to the infant's vocal tract anatomy (Buhr, 1980). These include vocal tremor (vibrato),
harmonic doubling, abrupt f0 shifts, vocal fry (or roll), and intrusive noise episodes (Kent and Murray, 1982). Particularly problematic is the juxtaposition of the velum and epiglottis resulting in the obligate nasal coupling which persists through approximately six months of age. This nasal coupling obscures vowel formants by introducing a nasal murmur, anti-resonances, and oral formant damping. Nasalization combined with harmonic doubling results in an acoustic pattern with poorly resolved formant structure, making measurements of the second and third formants extremely difficult (Buhr, 1980; Kent and Bauer, 1985).

Difficulties with acoustic analysis of infant vocalizations result not only from anatomy, (which is dynamically changing), but also from the fact that the coordinative structures suspected to be the neural plans for linguistic utterances have not been learned and concatenated. As the syllable is the primary putative coordinative structure for vocalization (Kelso, et al., 1986; Kozhevnikov and Chistovich, 1966; Stetson, 1951), close acoustic examination of the syllable and its infant precursors is a logical next step in delineating
developmental trends in the growth and development of pre/post linguistic infant vocalization.

**Carrier-Modulation Model of Motor Speech:**

In 1940, Dudley enunciated a simple conceptual model of connected speech based upon concepts of radio engineering. Specifically drawing on amplitude and frequency modulation (AM/FM) technology, Dudley proposed that a continuous speech utterance consists of an acoustically powerful carrier wave which slowly drifts from one vocalic spectral pattern to another. This powerful acoustic wave is periodically interrupted by rapid transitional shifts which exemplify amplitude and frequency modulation. Amplitude modulation consists of a fall in amplitude during vocal tract constriction. Frequency modulation consists both of a downward shift of f0 and of rapid shifts of formant frequencies during modulation. In radio terminology, the information-bearing (viz: consonantal) signal is imposed upon the vocalic carrier as a rapid amplitude/frequency modulation, exemplified by the transition into and out of the consonantal constricitions. Research by Lieberman, et al. (1967) and Furui (1986) has shown that the
majority of information for consonant as well as vowel identity resides in the vocalic transition into the vowel and not in the period of maximum consonantal closure, verifying Dudley's intuition that the modulated carrier wave bears the consonantal information (Lieberman, et al., 1971). Thus, in Dudley's conceptualization, labio-lingual motor commands are organized to produce a slowly varying sequence of vowel shapes (Kent and Moll, 1972). This slow vocalic drift of the carrier from one open shape (vowel) to another (vowel) is overlaid but little interrupted by rapid consonantal closure/opening movements (lips, tongue tip, tongue blade) which create the bursts, aspiration, transitions, and other information bearing elements for consonants. Consonant clusters are envisaged as complex (multiple), successive carrier modulations which affect both each other and adjacent vowels.

The acoustic layering (shingling) of information caused by consonantal modulation of vowels was termed coarticulation (Menzerath and de Lacerder, 1933) and later coproduction (Fowler, 1980) and is based on classical experimental studies by Rousselot (1901). This modulation process results in the smearing of
consonant and vowel information across the entire modulation-demodulation syllabic event. Hence Haskins Labs' early demonstration of the perceptual unity of the syllable and the decades later demonstration by Furui (1986) that vocalic transitions, not steady states, are the irreducible minimal information bearing elements for consonants as well as vowels. Pursuing this line of reasoning, Ohman (1968) studied the spectral patterns of VCV utterances to ascertain whether Dudley's model predicted observed consonant-vowel coarticulation. Ohman found that the patterns of formant coarticulation (perturbation) coincided very well with the Dudley vowel modulation model. Intervocalic consonant gestures began slowly (almost coincidentally) with vowel onset and accelerated to appreciable size only near the transition point, during which time both vowels still exerted a strong (coarticulatory) influence on the consonant. This suggests that the slow vowel drift continues in those structures unoccupied by the consonantal gesture (cf. Kent and Moll, 1972). Indeed, vowel influenced vowel across the consonantal episode.

Ohman proposed a model postulating lingual-
labial-mandibular movement which proceeded slowly and steadily in vowel-directed motions at all times, while consonantal modulations were characterized by rapid labial, apical, and dorsal movements which only minimally (or partially) suppressed the steady movement toward the vowel. Kent and Moll (1972) and Ohman (1968) speculated that intrinsic lingual and labial gestures for consonants were effected via distinct neuronal pathways separate from those commanding the slower extrinsic muscle driven movements for vowel production. Thus, speech motor patterning consisted of phasic modulations of steadily ongoing vowel movements executed by distinct neuronal/muscular groups for the vowel carrier and consonantal modulation events.

While the Dudley-Ohman model does not clearly identify which vowel (V₁ or V₂ in V₁CV₂ sequences) is more closely affiliated with the consonant, data from a number of researchers suggest that from a physiological standpoint, CV and not VC affiliation is the more profound (Kelso, et al., 1986; Kozhevnikov and Chistovich, 1966; Stetson, 1951). Grillner (1981) characterizes the CV syllable as a hierarchy of reflexively organized and learned
movements called a "coordinative structure." A single command from the head neural elements of the hierarchy of associated neurons results in a cascade of commands which smoothly execute the multi-structure movements contained within the CV syllable. Grillner argues that many basic coordinations between vocal structures are reflexive. For example, wails and cries involve simultaneous, phasic jaw opening, velar closure, vocal fold approximation, and pulmonary drive if a loud (voiced) cry is to emerge and coax help from caregivers. Articulation learning in such a scenario would consist of learning to amplify or suppress one or more movements in such a "hardwired" coordinative structure and to append new single or multiple movements to the command complex. Kelso and his coworkers (Kelso, et al., 1986) have postulated strict temporo-spatial phasing between C and V within a syllable regardless of the tempo and stress level of the utterance. Furthermore, the new stratificational or syllabic phonology has rediscovered a wealth of syllable-internal processes which guide interaction (and selection) between syllable onsets, peaks, and codas. Finally, Ohala (1990) intimates that the "sonority principle"
account of distributional constraints (i.e., what classes of consonants and vowels may occupy what slots in syllable onsets/offsets) is partially predicted by the articulatory patterning of vocal modulation implicit in Dudley-Ohman type models.

**Idealized Temporal Pattern of the CV Syllable:**

Syllables and their role as motor organization units for speech output bear close examination in the developmental patterns of infant vocalizations, both prior to and following the first word stage (Shatuck-Huffnagel, 1983). The acoustic precursors to syllables in the speech of pre first word children have been called proto-syllables, (PS). The PS has a general "CV" shape consisting of a closant-constrictive (i.e., consonant-like) episode released into an open, resonant, vowel-like tract shape. The PS, beginning at the five to six month period (Oller, 1980), displays temporo-spatial cohesiveness in terms of (1) a smooth transition, (2) continuity of voicing and pitch contour, and (3) covariation of C and V duration within the PS. The PS for the stop-type tokens studied herein begins either with a burst-release into a transition toward the vocalic element or with a quasi-steady voice bar (pre release of
constrictive) with or without a following burst, which leads smoothly into a transition to the vowel. Finally, the vowel-like resonant sound ends with cessation of voice.

**Temporal Structure of the PS:**

The PS has an overall duration, as well as constituent C and V durations. Fine temporal structure includes: (1) voicing onset time (VOT), (2) length of pre-voicing, (3) transition onset time (interval from onset of phonation to onset of formant transition), (4) transition duration, and (5) time from syllable onset to peak amplitude, point of maximum velocity of amplitude change, and point of maximum velocity of formant movement. Another time-derived temporal measure is a "directional" amplitude, or shimmer, computed from successive envelope amplitudes across the duration of the syllable. Given Fant's (1960) derivation of the spectrum of the vowel wave in the

\[ P(s) = H(s) * S(s) \]

or

\[ \log P(s) = \log H(s) + \log S(s) \]

S-plane, one can see that the vowel pressure
spectrum, \( P(s) \) is the multiplicative product of the Vocal Tract Transfer Function, \( H(s) \), and Glottal Source Spectrum, \( S(s) \). The shimmer measure is a gross estimate of how stably the articulatory organs and larynx cooperated to produce the vowel sound. Instability of either \( H(s) \) of \( S(s) \) would lead to irregularities in the amplitude envelope of the PS which are deviations from the double ogive amplitude contour typical of adult syllables.

The infant proto-syllable is not a linguistically derived (and perhaps not even a "coordinative structure" driven) series of movements. Nevertheless, the syllable, construed as a carrier modulation acoustic/physiological event, offers a number of parameters which would reflect critically upon the infant's ability to coordinate and execute the closed-open articulator movement sequences underlying adult syllabification. If we consider the syllable as an AM/FM modulated carrier wave (AM being the amplitude, FM being the power spectrum formants, and the carrier wave being the vocalic segment), then the consonant closure imposes a severe attenuation of amplitude and a fall in formant frequencies, especially when F1 falls to 0 at the point of full
tract closure. The modulation is most effective when the carrier is modulated on/off rapidly and deeply, as when obstruent and vowel are co-articulated. Ideally, the carrier being modulated (the vowel) should not drift in its power spectrum (formant pattern). Therefore, an ideally modulated syllable would reflect a short transition displaying a swift, monotonically smooth rise to peak velocity of amplitude and formant movement, terminating swiftly into a steady vowel wave. The shorter the transition duration, the more potent the modulation. Similarly, time-to-peak amplitude velocity and time-to-peak formant velocity also reflect the maximum swiftness of the "ideal" modulatory episode and should occur near the midpoint of the transition (Furui, 1986). The ratios of both time-to-peak amplitude (relative to syllable onset) and time-to-peak amplitude velocity to syllable duration also reflect the ideal modulation, and should be as small as possible (Crandall, 1925). Time to transition onset reflects the presence of a delay between initiation of consonant onset and initiation of tract opening. In this period, adult-like aspiration and VOT can occur for information purposes (i.e.,
signaling voicing or voicelessness). Syllabic data gathered over the five to fourteen month epoch should show acoustic/articulatory changes reflective of the above parameters and sensitive to anatomical and neurological maturation.

**Statement of Problem:**

Given that the CV syllable is a primary framework for motor planning and phonological patterning of speech, the acoustic patterns of infant proto-syllables should be explored in depth. Such descriptive work would lay the groundwork for task-dynamic evaluation of the developmental stages of syllable forms (task-dynamic being a descriptive explanation for viewing syllabic coordination of articulatory motion as a temporally scored series of phase-coupled articulatory gestures). The present study will be a detailed temporo/acoustic examination of the physical form of infant CV utterances selected across the five to fourteen month age range. The detailed acoustic examination will provide for sorting of CV tokens into categories based upon the degree of resemblance or divergence from adult CV forms, particularly with regard to the key elements of transition dynamics and syllable time structure.
The purpose of this investigation, then, is to create a data base of the temporal and spectral properties of syllables in pre-linguistic infant speech as a preliminary to finer-grained acoustic studies which would permit inference as to the developmental course of vocal tract motor control. If carefully extracted and longitudinally charted for phonetically similar CV forms, spectral information can provide insight into the patterning and developmental course of speech motor control - an acoustic means to a physiological end. Information regarding the physiological base underlying infant speech development is of interest from a diagnostic point of view as well. According to van der Stelt and Koopmans-van Beinum (1986), an increasing "delay time" in several successive motor functions should alert one to look for factors that explain the delay. Finally, it has been suggested that neuromotor maturation is strongly conditioned by experience and use. The very early presence of syllable forms with characteristics homologous with those of adult speakers may suggest that biological predisposition for sequencing and controlling sounds may be the foundation upon which speech motor learning is based.
General Research Plan:

In the general research plan, a corpus of CVs with stop-like or glide-like closants (closants being defined as sounds of less than 20 ms which are characterized by friction or impulse noise and/or no more than one visible formant in the region of F1) which were reliably labeled by two of three experimenters were selected for analysis. Following digitization, spectral and temporal analysis was performed on each token. Initially, CV tokens were be classified ordinally based upon envelope and source properties of adult syllables. Adult CV syllables begin with a burst, aspiration (if the C is voiceless), and smooth rising amplitude envelope which achieves a single broad amplitude peak on the vowel. Formants move smoothly and unidirectionally through the transitions into a quasi-static vowel nucleus. No intrusive nasalization, extraneous noise, or reversals of amplitude or formant trajectories are seen. Based upon the adult pattern,
the presence, number, and appropriateness of bursts; (2) the presence and length of aspiration; (3) the presence or absence of reversals of formant motion; (4) the occurrences of intrusive noise/nasality during transition or vocalic nucleus; and (5) the number of large amplitude peaks across the syllable nucleus. Acoustic analysis consisted of: (1) FFT estimates of F1/F2 from the onset of syllable acoustic vibration to the quasi-static steady state vocalic element at 15 ms intervals; (2) measures of amplitude envelope across the entire syllable; (3) measures of syllabic duration from the temporal waveform; (4) measures of transition duration; (5) measures of the duration of rise of amplitude envelope to peak syllable intensity; (6) measures of the duration of time-to-peak velocity of amplitude envelope; (7) computation of velocity of amplitude envelope; (8) computation of a directional perturbation quotient for syllable amplitude envelope; (9) classification of f0 contour; and (10) descriptive statistics. While a 15 ms sampling point may be thought to result in too grainy a time sampling, accurate examination of infant formant transitions, particularly F2, has been found to be
possible, reliable, and valid using Fast Fourier Transform (FFT) techniques (Miller, et al., 1991). However, because of the need to sweep back and forth along the transition in order to best estimate F2 at the specified time interval (using antecedent or subsequent F2 positions to reliably locate F2), the process is a time consuming one.

Subject and Stimuli Selection:

Stimuli for analysis were selected from the archival recordings of three first-born infants (1 male, 2 females) which were made in conjunction with earlier research on the acoustic and interactive aspects of infant language acquisition (Zlatin, 1976). All subjects reportedly possessed normal hearing acuity and unremarkable pre, peri, and postnatal histories. The data source primarily consisted of bi-weekly studio recordings that were made in a noise-controlled environment when the infant was alone, as well as dyadic and triadic interactions with the mother, father, or both parents present. Weekly home recordings, which were often noisy and of lower fidelity, served as a supplementary data source for both analysis and verification of the
representativeness of studio data relative to the infant's typical utterances.

Following Irwin (1948), data were analyzed in two month intervals from five to fourteen months of age (Irwin's levels 3 to 7), roughly corresponding to the onset of non-nasal or fully resonant vowels secondary to the disengagement of the velum and epiglottis, through the single-word level of language acquisition (Kent, 1981). Potential stimuli were be restricted to non-distress, non-vegetative speech-like vocalizations (Maskarinec, et al., 1981; Oiler, et al., 1976; Stark, et al., 1975). They consisted of oral stops and glides since these "closant" forms display maximum consonant/vowel coarticulation and manifest acoustically and perceptually crucial F1/F2 transitions amenable to straight-forward acoustic estimation. Vegetative and distress sounds such as cries, whimpers, coughs, grunts, breathing artifacts, swallows, etc., have, according to Oiler, et al. (1976), no scientific advantage and only serve to confuse the data. In addition, utterances were excluded that exhibited (1) hyperphonation (squeals), (2) glottal fry (growls), (3) vocal tremor, (4) form shift, (5) harmonic doubling, (6) labial trills, (7)
glottal ejectives, (8) nasal resonance, (9) poor signal-to-noise ratio, (10) simultaneous adult-child vocalizations, (11) and/or excessive breathiness. Also excluded were utterances where prevocalic stricture was formed at the glottis (usually [h] and [ʔ]) since consonantal status is not readily distinguishable from transient alteration of laryngeal source function for the vowel (Kent and Bauer, 1985).

Of the remaining utterances, CV like proto-syllables with perceptual evidence of stop-like or glide-like qualities and an identifiable vowel were selected for analysis. Potential stimuli were filtered through an Allison Labs Model 2BR variable band-pass filter at 150 Hz and 10 K Hz and digitized through a MBC MetraByte Model STA-20 A/D converter at a 20K Hz sampling rate using a Zenith ZDH-1217-AO microcomputer. Digitized samples were stored on diskettes for waveform analysis prior to D/A conversion.

After a period of joint training in a modified transcription system (Zlatin, 1976) two phonetically trained listeners independently transcribed the corpus of tokens. These transcriptions were compared.
to those of the present investigator; only those stimuli for which there was agreement among two of the three listeners as to vowel identity and place of articulation of consonantal forms were considered for further analysis. Of this potential corpus, approximately 60 tokens per child will be analyzed for a total of 182 tokens.

**Stimuli Analysis:**

Tokens were digitized at a sampling rate of 20K Hz. All spectral analysis (i.e., FFT) were performed in a 2048 point data buffer, allowing optimum frequency resolution at 9.9 Hz across the range of 0 - to - 5K Hz. Symmetric windowing (Hanning) was applied to minimize "edge effects," (i.e., spurious frequency components arising from signal truncation at the edges of the data window") (Miller, 1989, p. 13). High-frequency preemphasis was applied, the effect of which introduced a 6 dB/octave spectral "tilt," thereby enhancing the visibility of mid-to-upper frequency components of the spectral analysis. Using a custom-made wave analysis system, WAVEDIT (Miller, 1989), the beginning, end, and vowel center of each token was determined following the method described by Furui (1986). At
the same time spectral analyses (FFTs) were calculated at 15 ms intervals across the consonantal interval (including the transitions) to the point at which there was less than ten percent change in F2. Additional spectra were calculated at various points across the vocalic portion of the syllable. When present, spectra were generated for all burst-release events.

Spectral Measures:

Expected infant formant onset and vowel steady-state values were derived from adult values for onset (Rabiner, 1968) and steady states (Lieberman, 1980). Given the judges labeling of C and V, expected direction and length of transitional movements of F2 were estimated (Cooper, et al., 1952; Holmes, et al., 1964; Lieberman, et al., 1959; Rabiner, 1968). Each infant transition was compared with the expected transition in terms of onset frequency and terminal frequency.

Following A/D sampling, the following measures of developmental change were made:

1. Vowel F1/F2 steady states were plotted in F1/F2 space.
2. All F1/F2 onset frequencies were plotted as above and compared to reported children's values.
3. All transition durations (F2 onset to first steady state F2 value) were plotted across time.
4. All CV durations were plotted across time.
5. All maximum F2 velocities were plotted across time.
6. Given the labeling of infant CV proto-syllables, the Rabiner data (Rabiner, 1968) was used to establish where F2 should have started vs. where it actually started.
8. Variances of syllable, vowel, consonant, and transition durations were plotted across time.
9. Variances of F1/F2 onsets and steady states were plotted across time.

Measures of fundamental frequency (f0) were extracted from the recorded corpus of proto-syllables via the Visipitch analyzer interfaced with an IBM microcomputer. The frequency range was set at 200-760 Hz with a 2 second display window in which only f0 data was present. The internal software frequency range was set at 0-800 Hz, a range specified as compatible with that of the manually set range for the analyzer. The display of the ongoing f0 waveform
was continually monitored to insure that the minimum and maximum f0 values produced by each subject did not exceed the preset range values.

**Measures of Phonatory Instability:**

Of the acoustic measures of phonatory instability generally available, shimmer was selected due to the fact that it reflects relatively "short-term" amplitude instability based on a short duration of fundamental periods (Ramig, et al., 1990). Shimmer is defined as amplitude perturbation of the vocal signal and is analogous to fundamental frequency perturbation or jitter (Baken, 1987). As long-term changes in intensity are not at issue in measures of shimmer (Baken, 1987), a number of investigators have attempted to eliminate the effects of amplitude "drift" by trendline smoothing (Kitajima and Gould, 1976; Koike, et al., 1977).

Based on the work of Ramig, et al. (1988;1990), shimmer was measured using the formula:

\[
\text{Shimmer} = \frac{100}{(N-2)\bar{X}} \sum_{i=2}^{N-1} |.5(X_{i-1}+X_{i+1})-X_i|\
\]

where \(X_i\) is the amplitude of a specific voicing cycle measurement (by peak picking routine); \(\bar{X}\) is the
average measured amplitude for all cycles; N is the number of cycles analyzed; and $X_{i-1}$ and $X_{i+1}$ are the values immediately preceding and following $X_i$, respectively. The resulting value, expressed as a percentage, effectively removes linear amplitude trends or drifts (Ramig, et al., 1990).

While measures of shimmer are generally based upon central segments of sustained vowels, for the purpose of this study analysis of tokens were conducted at 10 ms increments from the transition onset through the end of the phonatory event (physiological rest). This procedure was used because infants do not produce sustained vowels naturally and because of our desire to evaluate the smoothness of amplitude envelope across the transition, as well as the vowel steady state. As no comparable normal adult data were available for comparison with infants, an adult subject, JP, was evaluated using the same procedure across a representative sampling of 36 CV syllables. Target utterances were embedded in the carrier phrase "Say ____ again" with the target syllable (open) followed by an identical but closed syllable (e.g., "Say /pi pip/ again"). This was done to provide a more
natural means for an adult to produce CV and CVC utterances and to establish a slow and steady pace (Harris, Haskins Labs, unpublished data on syllable affiliation).

As part of the evaluation of amplitude contours for proto-syllables, a simple rating scale was devised to rate the visible unsteadiness of the syllable amplitude. Additionally, the number of well defined amplitude peaks was noted. Each token was rated on a scale of 1 to 4 (one being the smoothest amplitude contour and 4 being the roughest) by each of two raters, and the number of amplitude peaks was recorded. These subjective measures of amplitude instability were compared with objective measures of shimmer obtained with the Ramig, et al. (1990) formula.
CHAPTER III
RESULTS

Preliminary:

When the entire recorded corpus of utterances (4600+) was scanned, more than half were multisyllabic. Of the monosyllabic utterances, the acoustic conditions (noise, microphone placement, and simultaneous voicing by observers) required that all but 450 utterances be eliminated. This corpus was eventually reduced to 182 analyzable utterances free of squeals, growls, hyperphonation, and signal-to-noise ratios which rendered analysis impossible.

Temporal Measures:

A major finding is that the duration of children's proto-syllables (X = 364, s.d. = 140) is approximately 1.6 times that of adults (X = 233.7, s.d. = 59.83). (See Figure 1 for the frequency histogram of adult and infant syllable durations) The infant CVs take proportionately longer than adults to achieve peak amplitude velocity (1.9/1). Otherwise, the infants had shorter onsets than adults (.89/1), shorter time-to-peak amplitude (.73/1), and less of
FIGURE 1. SYLLABLE DURATION FREQUENCY HISTOGRAM
Infants and Adult

Percentage of Occurrence

Syllable Duration in 20 ms. Increments

Adult Subject JP  Infant Subjects
the syllable duration occupied by the time to peak amplitude (Adults = 49% of syllable duration; Infants = 23% of syllable duration) than adults. Thus, the swiftness of frequency and amplitude modulation is roughly comparable to adults, reflecting precocious control capabilities. Friedman's Two-Way ANOVA by Ranks ($X_{rs}^S$) was conducted on all intrinsic temporal measures (See Table 1). Contrary to expectation, for all infants there was a lack of a directional developmental progression for intrinsic temporal measures across sampling periods. These included time to onset of vocalic transition ($X_{r^2} = 1.67$, d.f. = 4, NS), transition duration ($X_{r^2} = 3.2$, d.f. = 4, NS), time-to-peak velocity of amplitude rise ($X_{r^2} = .8$, d.f. = 4, NS), time-to-peak amplitude ($X_{r^2} = 2.67$, d.f. = 4, NS), and time to maximum F2 velocity during vocalic transitions ($X_{r^2} = 2.13$, d.f. = 4, NS). The absence of a developmental progression was unexpected given the marked restructuring of the vocal tract, the neurological maturation observed over the age range investigated (Crelin, 1987), and the orderly development of vowels and the shifting redistribution of frequency of occurrence of place of articulation of stops. With the exception of
<table>
<thead>
<tr>
<th>Measure</th>
<th>Xr²</th>
<th>d.f.</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable Duration</td>
<td>1.33</td>
<td>4</td>
<td>.45*</td>
</tr>
<tr>
<td>Transition Onset</td>
<td>1.67</td>
<td>4</td>
<td>.40*</td>
</tr>
<tr>
<td>Transition Duration</td>
<td>3.2</td>
<td>4</td>
<td>.35*</td>
</tr>
<tr>
<td>Time to Peak Amplitude</td>
<td>2.67</td>
<td>4</td>
<td>.35*</td>
</tr>
<tr>
<td>Time to Peak Velocity</td>
<td>.8</td>
<td>4</td>
<td>.475*</td>
</tr>
<tr>
<td>Shimmer</td>
<td>4.53</td>
<td>4</td>
<td>.25*</td>
</tr>
<tr>
<td>Consonants With Burst Release</td>
<td>7.27</td>
<td>4</td>
<td>.10*</td>
</tr>
<tr>
<td>+ Voice Consonants With Burst Release</td>
<td>8.33</td>
<td>4</td>
<td>.05</td>
</tr>
<tr>
<td>Time to Formant Maximum Velocity</td>
<td>2.13</td>
<td>4</td>
<td>.40*</td>
</tr>
<tr>
<td>Transitions With Aberrancies</td>
<td>3.79</td>
<td>4</td>
<td>.25*</td>
</tr>
</tbody>
</table>

* Not Significant
transition duration, all intrinsic temporal events are positively correlated with syllable duration across infants. (Please see Table 2 for Student t approximations of the Spearman Rank Correlation Coefficients.) There was not, at any age level, a relationship between transition duration and syllable duration. Transition durations do not become faster as the children get older - the general figure is stereotypical and just slightly longer than for adults (43.3 ms and 35.38 ms for children and adults respectively). Further, close scrutiny of all 182 formant trajectories failed to reveal a relationship between presence or absence of aberrancies (abnormal onset frequencies or target overshoot) and absolute transition duration (t = .456, d.f. = 13, NS).

**Burst:**

Of all syllables beginning with a stop-like closant, 54% exhibited an adult-like burst release. The highest frequency for individual subjects occurred for subject AD at 80% of all closants, followed by MO and JF with 49% and 8.6%, respectively. While the number of bursts observed as surprising, (particularly for subject AD), the
<table>
<thead>
<tr>
<th>Measure</th>
<th>Subject</th>
<th>rho</th>
<th>Student t</th>
<th>d.f.</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable Duration and</td>
<td>AD</td>
<td>.31*</td>
<td>3.14*</td>
<td>73</td>
<td>.005</td>
</tr>
<tr>
<td>Transition Onset</td>
<td>MO</td>
<td>.46*</td>
<td>5.8*</td>
<td>65</td>
<td>.0005</td>
</tr>
<tr>
<td></td>
<td>JF</td>
<td>.61*</td>
<td>6.01*</td>
<td>36</td>
<td>.0005</td>
</tr>
<tr>
<td>ADULT</td>
<td>-.29</td>
<td>-1.51</td>
<td>34</td>
<td></td>
<td>.10**</td>
</tr>
<tr>
<td>Syllable Duration and</td>
<td>AD</td>
<td>.01*</td>
<td>.32</td>
<td>73</td>
<td>**</td>
</tr>
<tr>
<td>Transition Duration</td>
<td>MO</td>
<td>.02</td>
<td>.14</td>
<td>65</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>JF</td>
<td>.06</td>
<td>.4</td>
<td>38</td>
<td>**</td>
</tr>
<tr>
<td>ADULT</td>
<td>Not Evaluated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable Duration and</td>
<td>AD</td>
<td>.63</td>
<td>8.88</td>
<td>73</td>
<td>.0005</td>
</tr>
<tr>
<td>Time-to-Peak Amplitude</td>
<td>MO</td>
<td>.56</td>
<td>6.79</td>
<td>65</td>
<td>.0005</td>
</tr>
<tr>
<td></td>
<td>JF</td>
<td>.66</td>
<td>7.01</td>
<td>36</td>
<td>.0005</td>
</tr>
<tr>
<td>ADULT</td>
<td>.43</td>
<td>3.34</td>
<td>34</td>
<td></td>
<td>.0005</td>
</tr>
<tr>
<td>Syllable Duration and</td>
<td>AD</td>
<td>.43</td>
<td>4.8</td>
<td>73</td>
<td>.0005</td>
</tr>
<tr>
<td>Time-to-Peak Velocity</td>
<td>MO</td>
<td>.34</td>
<td>3.37</td>
<td>65</td>
<td>.0005</td>
</tr>
<tr>
<td></td>
<td>JF</td>
<td>.51</td>
<td>4.49</td>
<td>38</td>
<td>.0005</td>
</tr>
<tr>
<td>ADULT</td>
<td>-1.3</td>
<td>-2.73</td>
<td>34</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Time-to-Peak Amplitude and</td>
<td>AD</td>
<td>.72</td>
<td>11.6</td>
<td>73</td>
<td>.0005</td>
</tr>
<tr>
<td>Time-to-Peak Velocity</td>
<td>MO</td>
<td>.40</td>
<td>4.18</td>
<td>65</td>
<td>.0005</td>
</tr>
<tr>
<td></td>
<td>JF</td>
<td>.62</td>
<td>6.21</td>
<td>38</td>
<td>.0005</td>
</tr>
<tr>
<td>ADULT</td>
<td>.34</td>
<td>2.43</td>
<td>34</td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Syllable Duration and</td>
<td>AD</td>
<td>-.58</td>
<td>-3.93</td>
<td>73</td>
<td>.0005</td>
</tr>
<tr>
<td>and Shimmer</td>
<td>MO</td>
<td>-.33</td>
<td>-2.31</td>
<td>64</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>JF</td>
<td>-.17</td>
<td>- .97</td>
<td>38</td>
<td>**</td>
</tr>
<tr>
<td>ADULT</td>
<td>-.24</td>
<td>-1.25</td>
<td>34</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Syllable Duration and</td>
<td>AD</td>
<td>.36*</td>
<td>3.98*</td>
<td>73</td>
<td>.0005</td>
</tr>
<tr>
<td>Time-to-Maximum</td>
<td>MO</td>
<td>.35*</td>
<td>3.51*</td>
<td>64</td>
<td>.0005</td>
</tr>
<tr>
<td>Transition Velocity</td>
<td>JF</td>
<td>.57</td>
<td>5.12</td>
<td>34</td>
<td>.0005</td>
</tr>
<tr>
<td>ADULT</td>
<td>Not Evaluated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transition Duration and</td>
<td>AD</td>
<td>-.50*</td>
<td>-3.50*</td>
<td>73</td>
<td>.0005</td>
</tr>
<tr>
<td>Percent of Transition to</td>
<td>MO</td>
<td>-.02</td>
<td>- .18</td>
<td>64</td>
<td>**</td>
</tr>
<tr>
<td>Maximum Velocity</td>
<td>JF</td>
<td>-.32*</td>
<td>-1.63</td>
<td>34</td>
<td>.10**</td>
</tr>
<tr>
<td>ADULT</td>
<td>Not Evaluated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Corrected for Ties
** Not Significant
results are consistent with the findings of Kewley-Port and Preston (1974), and suggests that laryngeal control necessary to delay vocal fold closure (re., articulatory release) is not well developed in early infancy compared to other intrinsic timing characteristics of proto-syllables. This observation leads one to hypothesize that there should be an increase in the percentage of bursts as a function of increasing age and the language being learned (i.e., L1). The results of a Friedman's One-Way ANOVA suggest that while the largest percentage (64%) occurs at thirteen to fourteen months for all tokens, the trend is not significant ($X^2 = 7.267$, d.f. = 4, NS). Particularly interesting is the fact that when ranks were assigned based upon frequency of occurrence at each age range, the second largest proportion of bursts occurred at the seven to eight month range for all tokens. When proto-syllables initiated with tokens judged to be voiced stops are examined, (these being the most frequent for the current infants and for English speakers in general), the highest level of occurrence actually shifts to the seven to eight month range with the thirteen to fourteen month range dropping to second position.
For the voiced proto-syllables only, there is a statistically significant relationship between age and presence of burst releases ($X_r^2 = 8.333$, d.f. = 4, p. = .05).

In 4 of the 182 tokens analyzed, a double stop-burst release was evidenced. The 2 burst-like events were of essentially equal amplitude and duration separated by fairly uniform intervening periods of 22-25 ms. The double bursts occurred on both voiced and voiceless velar closants and is consistent with those observed in adult velar productions (Ladefoged, 1982). Only one such token displayed obvious failure of the velo-pharyngeal seal, exhibiting approximately 14 ms of nasal murmur between bursts. A similar intrusion was, on rare occasions, found during quasi steady-state vowel production, the occurrence of which, however, did not appear to be related to the age of the subject.

**Measures of Phonatory Instability:**

Friedman's Two-Way ANOVA by Ranks was conducted on shimmer measures across age levels. There was no significant difference observed ($X_r^2 = 4.533$, d.f. = 4, NS) suggesting that on-line control of vocal
intensity (which is mediated by midline compression and resulting vocal fold resistance, as well as pulmonary pressure) does not change across the sampling period, in contrast with the rapid neuro-muscular maturation seen in gross motor activities (e.g., postural stability against gravity, protective extension, weight bearing and shift, and ambulation). Evaluation of standard deviation as a measure of dispersion across ages was also not significant ($Xr^2 = 1.0667, \text{d.f.} = 4, \text{NS}$). Despite the apparent lack of effect of age on measures of shimmer and the associated standard deviation, the results of a Spearman Rank Correlation Coefficient showed a statistically significant positive correlation between the mean for shimmer at each age level and standard deviation ($r = .44, p = .05$). That is, as amplitude instability increased, the dispersion of measures about the mean also increased. Analysis of adult tokens with the Ramig, et al. (1990) analysis yielded a mean shimmer value of 15.2 with a standard deviation of 8.87. When compared to the average across all infant tokens, adult values were somewhat higher than those of the infant's (Table 3: mean child value = 14.2, std. = 6.5).
# TABLE 3. MEAN VALUES FOR ADULT AND INFANTS

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>O</th>
<th>TD</th>
<th>TP</th>
<th>TV</th>
<th>V</th>
<th>P/D</th>
<th>V/D</th>
<th>V/P</th>
<th>SHIM</th>
<th>%T to Vm</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD X</td>
<td>345.100</td>
<td>35.413</td>
<td>46.810</td>
<td>127.800</td>
<td>78.500</td>
<td>0.470</td>
<td>0.370</td>
<td>0.236</td>
<td>0.714</td>
<td>13.983</td>
<td>0.712</td>
</tr>
<tr>
<td></td>
<td>171.070</td>
<td>33.060</td>
<td>28.700</td>
<td>102.240</td>
<td>74.850</td>
<td>0.470</td>
<td>0.200</td>
<td>0.148</td>
<td>0.549</td>
<td>6.847</td>
<td>0.355</td>
</tr>
<tr>
<td>MO X</td>
<td>333.600</td>
<td>29.100</td>
<td>40.580</td>
<td>97.940</td>
<td>55.470</td>
<td>0.327</td>
<td>0.307</td>
<td>0.177</td>
<td>0.680</td>
<td>12.522</td>
<td>0.490</td>
</tr>
<tr>
<td></td>
<td>150.220</td>
<td>24.700</td>
<td>24.690</td>
<td>60.850</td>
<td>46.760</td>
<td>0.163</td>
<td>0.149</td>
<td>0.137</td>
<td>0.655</td>
<td>5.340</td>
<td>0.320</td>
</tr>
<tr>
<td>JF X</td>
<td>365.260</td>
<td>81.000</td>
<td>42.350</td>
<td>152.160</td>
<td>118.570</td>
<td>0.390</td>
<td>0.406</td>
<td>0.325</td>
<td>0.795</td>
<td>16.005</td>
<td>0.648</td>
</tr>
<tr>
<td></td>
<td>99.260</td>
<td>68.110</td>
<td>20.390</td>
<td>74.740</td>
<td>68.040</td>
<td>0.716</td>
<td>0.154</td>
<td>0.213</td>
<td>0.518</td>
<td>7.217</td>
<td>0.282</td>
</tr>
<tr>
<td>X Child</td>
<td>364.000</td>
<td>48.000</td>
<td>43.300</td>
<td>84.300</td>
<td>126.000</td>
<td>0.396</td>
<td>0.360</td>
<td>0.250</td>
<td>0.730</td>
<td>14.200</td>
<td>0.730</td>
</tr>
<tr>
<td>s.d. Child</td>
<td>140.000</td>
<td>42.000</td>
<td>24.700</td>
<td>70.000</td>
<td>79.300</td>
<td>0.450</td>
<td>0.170</td>
<td>0.170</td>
<td>0.580</td>
<td>6.490</td>
<td>0.580</td>
</tr>
<tr>
<td>X Adult</td>
<td>233.700</td>
<td>54.220</td>
<td>35.380</td>
<td>115.280</td>
<td>65.139</td>
<td>0.264</td>
<td>0.502</td>
<td>0.299</td>
<td>0.593</td>
<td>15.200</td>
<td>0.609</td>
</tr>
<tr>
<td>s.d. Adult</td>
<td>59.830</td>
<td>35.520</td>
<td>22.237</td>
<td>44.260</td>
<td>37.192</td>
<td>0.177</td>
<td>0.154</td>
<td>0.171</td>
<td>0.279</td>
<td>8.669</td>
<td>0.327</td>
</tr>
</tbody>
</table>

| Ratio  | 1.558 | 0.885 | 1.224 | 0.731 | 1.934 | 1.500 | 0.717 | 0.836 | 1.231 | 0.934 | *        |
| Ratio  | 2.340  | 1.182 | 1.111 | 1.582 | 2.132 | 2.542 | 1.104 | 0.994 | 2.079 | 0.732 | *        |
| s.d.C/s.d.A |       |       |       |       |       |       |       |       |       |       |         |

**D = Total syllable duration**

**O = Release or onset of phonation to first visible F2**

**TD = Transition duration**

**TP = Time-to-peak amplitude**

**TV = Time-to-peak velocity of amplitude envelope**

**V = Peak velocity of amplitude envelope**

**P/D = Percent of syllable duration occupied by time-to-peak amplitude**

**V/D = Percent of syllable duration occupied by time-to-peak amplitude velocity**

**V/P = Time-to-peak amplitude velocity:time-to-peak amplitude**

**SHIM = Amplitude perturbation (shimmer)**

**% T to Vm = Percent of transition occupied by time-to-peak formant velocity**

*Not Evaluated*
When shimmer was correlated with syllable duration (Table 2) using the Student t approximation of $r_S$, a strong negative correlation was obtained for two of the subjects, MO ($t = -3.93$, d.f. = 65, p. = .0005) and AD ($t = -2.31$, d.f. = 73, p. = .025). While there was no statistically significant correlation for JF ($t = -.97$, d.f. = 38, NS) or the adult subject ($t = -1.25$, d.f. = 34, NS), the negative trend was evident. The data suggest, then, that as syllable duration increases, perturbation decreases, and vice versa.

When subjective measures of amplitude stability (perceived roughness and number of amplitude peaks) were correlated with shimmer, neither measure produced a significant correlation ($t = -.102$, d.f. = 13, NS and $t = -.299$, d.f. = 13, NS, for roughness and amplitude peaks, respectively). These results suggest that the metric used for shimmer was insensitive to gross visible shifts in the pressure waveform, (i.e., those over a longer time scale than adjacent time points [10 ms apart] on the amplitude waveform.)

**Spectral Structure:**

As previously noted, measures of $f_0$ were
extracted from the recorded corpus of proto-syllables. The resulting mean f0 values of 314 Hz, 303 Hz, and 322 Hz for subjects AD, MO, and JF, respectively, were consistent with published values for speaking in eleven to twenty-five month old subjects (X = 375) (Robb and Saxman, 1985).

Analysis of infant transitions identified a substantial percentage of transitions, roughly 32%, which were aberrant in that: (1) F2 began the transition at an unexpectedly high/low onset value given the phonetic label applied to the consonant (82% of aberrant transitions displayed this); and (2) F2 overshot the steady state value and then returned to the F2 target (41% of aberrant transitions displayed this). (Please see Table 4 for breakdown of aberrancies by infant.) In approximately 28% of transitions showing aberrancies, both inappropriate onset values and overshoot occurred simultaneously.

For all 182 tokens, whether there was aberrancy or not, the F2 transitions appeared to increase or decrease monotonically either to steady state or to the point of overshoot prior to steady state. Some transitions were as short as 25-30 ms and typically reflected F2 onset frequencies much closer to F2
TABLE 4. TRANSITION ABERRANCIES

<table>
<thead>
<tr>
<th>Subject</th>
<th>AD</th>
<th>JF</th>
<th>MO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tokens</td>
<td>75</td>
<td>40</td>
<td>65</td>
</tr>
<tr>
<td>Number With Aberrancies</td>
<td>29</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Percentage of Transitions Which Were Aberrant</td>
<td>39%</td>
<td>30%</td>
<td>26%</td>
</tr>
<tr>
<td>Percentage of Transitions with Onset Aberrancies</td>
<td>79%</td>
<td>92%</td>
<td>76%</td>
</tr>
<tr>
<td>Percentage of Transitions With Overshoot Aberrancies</td>
<td>62%</td>
<td>33%</td>
<td>29%</td>
</tr>
<tr>
<td>Percentage of Transitions With Aberrant Onset and Overshoot</td>
<td>41%</td>
<td>25%</td>
<td>6%</td>
</tr>
</tbody>
</table>
terminal frequencies than was predicted. These short transitions were not typical of the adult. Given a 15 ms sampling interval with a 10 ms Hanning window, the transitions were typically sampled by 4 measures for F2, but in no case fewer than 3 (onset, end and middle of the F2 transition). Frequency increase (for rising F2) or decrease (for falling F2) was always noted between successive F2 estimates across the length of the transition. Despite the aberrancies noted, F1 and F2 always achieved the quasi steady state values appropriate to the perceived vowel, suggesting that while the vowel and consonant spatial target values are apparently not well established or not readily accessible for on-line control of transitioning, acoustic targets are well established and ultimately achieved after transitions. The above findings were typical of the entire five to fourteen month age range studied. Finally, in cases of dipthongal movement (12.2% of the 182 cases studied), infants were similar to adults in that a rapid secondary transition occurred between the quasi steady-state F2 values appropriate to the first and second vocalic elements (Holbrook and Fairbanks, 1962).
The results of an analysis of quasi steady-state F1/F2 target frequencies across subjects and ages found that with the exception of /i/ (which appeared slightly centralized, with lower F2 values than expected), infant's vowel spaces closely approximated those reported by Lieberman (1980). (Figures 2a, 2b, and 2c show scattergrams of F1/F2 steady state values for the infants while Figure 3 shows the mean F1/F2 values for the current infants and those values reported by Lieberman, [1980].)

Developmental Course of Infant Vocalizations:

The present study sought to scrutinize the developmental trends in the infants' CV prosyllables. Frequency of occurrence data for consonantal and vocalic utterances by type were compiled across the five to fourteen month interval. Table 5 lists frequency of occurrence of vowels (in percentages) across developmental stages for the three infants studied as well as for those 60-80 infants studied by Irwin (1948). Of the vowels recorded at level 3 for the present infants, 69% were more centralized /æ, ʌ, a/ versus 27.5% for Irwin's subjects at the same level. With the exception of /o/, both groups expanded their vowel
O Reported Lieberman (1980) Values
* Infant Values From This Study
FIGURE 2B. F1/F2 SCATTERGRAM FOR i AND I
All Subjects

O Reported Lieberman (1980) Values
* Average Infant Values From This Study
FIGURE 2C. F1/F2 SCATTERGRAM FOR ^ AND a

All Subjects

O Reported Lieberman (1980) Values
* Average Infant Values From This Study
FIGURE 3. COMPARISON OF REPORTED LIEBERMAN* VALUES WITH THOSE OF CURRENT SUBJECTS

* Lieberman Values in Parentheses
From Lieberman (1980)
<table>
<thead>
<tr>
<th>Age Level</th>
<th>l</th>
<th>I</th>
<th>e</th>
<th>E</th>
<th>ae</th>
<th>^</th>
<th>a</th>
<th>o</th>
<th>U</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>7.69</td>
<td>7.69</td>
<td>7.69</td>
<td>15.38</td>
<td>38.46</td>
<td>15.38</td>
<td>0</td>
<td>0</td>
<td>7.69</td>
</tr>
<tr>
<td>Irwin</td>
<td>1.77</td>
<td>24.82</td>
<td>1.58</td>
<td>31.27</td>
<td>2.62</td>
<td>23.44</td>
<td>1.52</td>
<td>0.17</td>
<td>8.2</td>
<td>4.17</td>
</tr>
<tr>
<td>4</td>
<td>3.7</td>
<td>11.11</td>
<td>7.4</td>
<td>14.81</td>
<td>3.7</td>
<td>40.74</td>
<td>3.7</td>
<td>0</td>
<td>11.11</td>
<td>3.7</td>
</tr>
<tr>
<td>Irwin</td>
<td>1.8</td>
<td>24.96</td>
<td>2.04</td>
<td>31.4</td>
<td>3.13</td>
<td>19.65</td>
<td>2.17</td>
<td>0.3</td>
<td>10.12</td>
<td>3.88</td>
</tr>
<tr>
<td>5</td>
<td>8.11</td>
<td>5.41</td>
<td>0</td>
<td>0</td>
<td>10.81</td>
<td>40.54</td>
<td>16.22</td>
<td>8.11</td>
<td>8.11</td>
<td>2.7</td>
</tr>
<tr>
<td>Irwin</td>
<td>2.09</td>
<td>22.22</td>
<td>2.85</td>
<td>32.35</td>
<td>2.72</td>
<td>16.8</td>
<td>5.34</td>
<td>0.68</td>
<td>10.24</td>
<td>4.06</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>14.63</td>
<td>2.42</td>
<td>12.2</td>
<td>9.76</td>
<td>24.39</td>
<td>21.95</td>
<td>7.32</td>
<td>4.88</td>
<td>2.42</td>
</tr>
<tr>
<td>Irwin</td>
<td>4.15</td>
<td>23.02</td>
<td>3.25</td>
<td>27.56</td>
<td>4.13</td>
<td>16.17</td>
<td>6.48</td>
<td>1.33</td>
<td>7.71</td>
<td>5.36</td>
</tr>
<tr>
<td>7</td>
<td>6.52</td>
<td>10.87</td>
<td>6.52</td>
<td>10.87</td>
<td>10.87</td>
<td>21.74</td>
<td>13.04</td>
<td>13.04</td>
<td>4.35</td>
<td>2.17</td>
</tr>
<tr>
<td>Irwin</td>
<td>4.32</td>
<td>23.13</td>
<td>3.25</td>
<td>27.03</td>
<td>2.25</td>
<td>16.58</td>
<td>6.64</td>
<td>1.17</td>
<td>8.79</td>
<td>5.69</td>
</tr>
</tbody>
</table>
space to include a full complement by age level 4.
Thereafter, at levels 5, 6, and 7, Irwin's infants preferred the mid-high front vowels /I/ and /£/ twice as often as our infants, whereas our infants preferred the low-mid vowels /ae/, /\ tape/ and /a/ substantially more often. If any discernible trends existed, it was the later appearance and lower preference for mid- and high-front vowels /£, e, I, i/ which distinguished our infants from Irwin's. It should be noted that the differences between percentage of occurrence in the two studies may be artifactual, related to sampling. In particular, by limiting ourselves to 3 subjects and using only monosyllables, we may have produced a percentage of occurrence of vowels both idiosyncratic to the three infants, and/or peculiar to mono-syllabic utterances, which at levels 5-7 comprised the minority of total infant utterances. Had the home recordings been of better acoustic quality, a larger and more representative sampling of vowel preference could have been produced. As it stands, however, our infants compared well with Irwin's infants in terms of the distribution of vowels produced at various developmental levels. The mean F1/F2 plots of the
infant's vowels (Figure 3) show that, in general, the infants' vowels fell in the expected locations, with the sole exception of /i/, which was unexpectedly low for F2. In fact, the mean F1/F2 values for [I, , a, ∧] were almost perfectly congruent with those observed by Lieberman (1980) at the same developmental levels.

Consonants sampled in this study were limited to stops and glides in order to emphasize the phonetically vital transitions. Irwin (1946) and Oller (1980) have shown that an early preference for velar stops and voiceless stops yields to more nearly equal proportions of labial, alveolar, and velar stops by twelve to fourteen months. For one of our subjects (Table 6), there were more alveolars (38% vs. 5%) and labials (14% vs. 8%) than Irwin's subjects at level 3. At level four, alveolar preference peaked and labial preference grew. By levels 5-7, labial preference had stabilized at 50% (vs. approximately 38% for Irwin) and alveolar preference at 30%, so that there was fairly good match of the present data to Irwin's data for frequency of occurrence of place of articulation for stops. It should be noted that the over-preference
TABLE 6. Stop-Consonant Percentages By Age Level

<table>
<thead>
<tr>
<th>Age Level</th>
<th>Labial</th>
<th>Alveolar</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>C</td>
<td>14</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>25</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>55</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>21</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>50</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>32</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>Unavailable</td>
<td></td>
</tr>
</tbody>
</table>
for labial stops was caused by subject JF who in levels 5-7 produced labial stops almost to the exclusion of the others (>80%). On the whole, however, given the few subjects used and the sampling limited to isolated CVs, the infants studied were showing phonetic preferences similar to those for Irwin's infants.

Temporal Patterning:

Figure 1 presents an adult and infant syllable duration histogram, while Tables 2 and 7a,b,c present means and standard deviations from this measure. Observe that modal adult and infant syllable duration overlap at approximately 240ms, which is roughly what Ladefoged (1982) states to be the "typical" open syllable duration of adults. These data compare well with those of Lehiste and Peterson (1960) who reported approximately 250 ms as the average duration for stop-vowel combinations in closed CVC syllables. The adult distribution, however, displays a slight skew to shorter durations, whereas the infant distribution is skewed toward longer durations (with several in the 800-1200 ms region). The infant skew accounts for much, if not all, of the difference in overall modal syllable durations between the adult
<table>
<thead>
<tr>
<th>LEVEL</th>
<th>D</th>
<th>O</th>
<th>TD</th>
<th>TP</th>
<th>TV</th>
<th>V</th>
<th>P/D</th>
<th>V/D</th>
<th>V/P</th>
<th>SHIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>464.333</td>
<td>27.500</td>
<td>43.333</td>
<td>88.483</td>
<td>46.063</td>
<td>0.420</td>
<td>0.220</td>
<td>0.126</td>
<td>0.559</td>
<td>13.005</td>
</tr>
<tr>
<td>s.d.</td>
<td>284.342</td>
<td>16.008</td>
<td>16.997</td>
<td>36.241</td>
<td>18.218</td>
<td>0.131</td>
<td>0.106</td>
<td>0.068</td>
<td>0.212</td>
<td>6.639</td>
</tr>
<tr>
<td>4</td>
<td>251.500</td>
<td>22.500</td>
<td>67.500</td>
<td>69.700</td>
<td>48.267</td>
<td>0.325</td>
<td>0.276</td>
<td>0.200</td>
<td>0.774</td>
<td>16.034</td>
</tr>
<tr>
<td>s.d.</td>
<td>46.668</td>
<td>7.500</td>
<td>20.966</td>
<td>53.538</td>
<td>44.017</td>
<td>0.104</td>
<td>0.188</td>
<td>0.241</td>
<td>0.505</td>
<td>5.346</td>
</tr>
<tr>
<td>5</td>
<td>348.588</td>
<td>40.706</td>
<td>30.000</td>
<td>116.112</td>
<td>79.285</td>
<td>0.278</td>
<td>0.314</td>
<td>0.233</td>
<td>0.920</td>
<td>12.614</td>
</tr>
<tr>
<td>s.d.</td>
<td>142.065</td>
<td>39.833</td>
<td>13.878</td>
<td>85.486</td>
<td>53.633</td>
<td>0.196</td>
<td>0.156</td>
<td>0.129</td>
<td>0.902</td>
<td>3.895</td>
</tr>
<tr>
<td>6</td>
<td>336.955</td>
<td>26.045</td>
<td>40.409</td>
<td>106.260</td>
<td>54.898</td>
<td>0.338</td>
<td>0.327</td>
<td>0.163</td>
<td>0.642</td>
<td>12.564</td>
</tr>
<tr>
<td>s.d.</td>
<td>123.527</td>
<td>29.675</td>
<td>24.112</td>
<td>56.303</td>
<td>48.431</td>
<td>0.164</td>
<td>0.157</td>
<td>0.139</td>
<td>0.871</td>
<td>5.554</td>
</tr>
<tr>
<td>7</td>
<td>287.313</td>
<td>24.063</td>
<td>40.938</td>
<td>81.313</td>
<td>37.181</td>
<td>0.328</td>
<td>0.318</td>
<td>0.149</td>
<td>0.487</td>
<td>10.081</td>
</tr>
<tr>
<td>s.d.</td>
<td>113.918</td>
<td>16.509</td>
<td>27.399</td>
<td>39.269</td>
<td>29.282</td>
<td>0.102</td>
<td>0.109</td>
<td>0.087</td>
<td>0.240</td>
<td>3.719</td>
</tr>
<tr>
<td></td>
<td>333.600</td>
<td>29.100</td>
<td>40.580</td>
<td>97.940</td>
<td>55.470</td>
<td>0.327</td>
<td>0.307</td>
<td>0.177</td>
<td>0.680</td>
<td>12.522</td>
</tr>
<tr>
<td>s.d.</td>
<td>150.220</td>
<td>24.700</td>
<td>24.690</td>
<td>60.850</td>
<td>46.760</td>
<td>0.163</td>
<td>0.149</td>
<td>0.137</td>
<td>0.655</td>
<td>5.340</td>
</tr>
</tbody>
</table>

D = Total syllable duration
O = Release or onset of phonation to first visible F2
TD = Transition duration
TP = Time-to-peak amplitude
TV = Time-to-peak velocity of amplitude envelope
V = Peak velocity of amplitude envelope
P/D = Percent of syllable duration occupied by time-to-peak amplitude
V/D = Percent of syllable duration occupied by time-to-peak amplitude velocity
V/P = Time-to-peak amplitude velocity:time-to-peak amplitude
SHIM = Amplitude perturbation (shimmer)
TABLE 7B. Subject AD

<table>
<thead>
<tr>
<th>RANGE</th>
<th>D</th>
<th>O</th>
<th>TD</th>
<th>TP</th>
<th>TV</th>
<th>V</th>
<th>P/D</th>
<th>V/D</th>
<th>V/P</th>
<th>SHIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>X</td>
<td>390.250</td>
<td>41.250</td>
<td>42.500</td>
<td>145.275</td>
<td>106.975</td>
<td>0.153</td>
<td>0.388</td>
<td>0.257</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>233.142</td>
<td>37.312</td>
<td>25.617</td>
<td>73.416</td>
<td>78.724</td>
<td>0.056</td>
<td>0.092</td>
<td>0.103</td>
<td>0.191</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>264.176</td>
<td>33.235</td>
<td>51.412</td>
<td>89.612</td>
<td>55.732</td>
<td>0.737</td>
<td>0.366</td>
<td>0.226</td>
<td>0.620</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>86.566</td>
<td>34.383</td>
<td>21.265</td>
<td>43.992</td>
<td>41.100</td>
<td>0.613</td>
<td>0.190</td>
<td>0.143</td>
<td>0.258</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>312.417</td>
<td>22.667</td>
<td>40.667</td>
<td>121.583</td>
<td>64.092</td>
<td>0.313</td>
<td>0.368</td>
<td>0.195</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>110.372</td>
<td>7.685</td>
<td>25.705</td>
<td>85.320</td>
<td>35.351</td>
<td>0.127</td>
<td>0.176</td>
<td>0.070</td>
<td>0.280</td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>287.357</td>
<td>30.571</td>
<td>36.643</td>
<td>77.325</td>
<td>66.621</td>
<td>0.306</td>
<td>0.284</td>
<td>0.239</td>
<td>1.115</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>136.897</td>
<td>25.701</td>
<td>21.400</td>
<td>65.988</td>
<td>53.523</td>
<td>0.138</td>
<td>0.127</td>
<td>0.127</td>
<td>1.043</td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>430.679</td>
<td>43.786</td>
<td>52.357</td>
<td>176.275</td>
<td>100.357</td>
<td>0.503</td>
<td>0.406</td>
<td>0.255</td>
<td>0.607</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>194.329</td>
<td>38.689</td>
<td>34.795</td>
<td>127.278</td>
<td>100.291</td>
<td>0.508</td>
<td>0.235</td>
<td>0.181</td>
<td>0.289</td>
</tr>
</tbody>
</table>

| X     | 345.100 | 35.413  | 46.810  | 127.800 | 76.500  | 0.470   | 0.370   | 0.236   | 0.714   | 13.983  |
| s.d.  | 171.070 | 33.060  | 28.700  | 102.240 | 74.850  | 0.470   | 0.200   | 0.148   | 0.549   | 6.847   |

D = Total syllable duration
O = Release or onset of phonation to first visible F2
TD = Transition duration
TP = Time-to-peak amplitude
TV = Time-to-peak velocity of amplitude envelope
V = Peak velocity of amplitude envelope
P/D = Percent of syllable duration occupied by time-to-peak amplitude
V/D = Percent of syllable duration occupied by time-to-peak amplitude velocity
V/P = Time-to-peak amplitude velocity time-to-peak amplitude
SHIM = Amplitude perturbation (shimmer)
<table>
<thead>
<tr>
<th>LEVEL</th>
<th>D</th>
<th>O</th>
<th>TD</th>
<th>TP</th>
<th>TV</th>
<th>V</th>
<th>P/D</th>
<th>V/D</th>
<th>V/P</th>
<th>SHIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>343.000</td>
<td>60.000</td>
<td>25.000</td>
<td>122.000</td>
<td>100.200</td>
<td>0.230</td>
<td>0.355</td>
<td>0.302</td>
<td>0.904</td>
<td>13.766</td>
</tr>
<tr>
<td></td>
<td>15.122</td>
<td>53.385</td>
<td>7.071</td>
<td>15.591</td>
<td>70.082</td>
<td>0.062</td>
<td>0.038</td>
<td>0.219</td>
<td>0.733</td>
<td>1.942</td>
</tr>
<tr>
<td>4</td>
<td>388.667</td>
<td>86.333</td>
<td>35.000</td>
<td>193.000</td>
<td>242.667</td>
<td>0.186</td>
<td>0.496</td>
<td>0.705</td>
<td>1.208</td>
<td>10.668</td>
</tr>
<tr>
<td></td>
<td>47.225</td>
<td>69.043</td>
<td>21.213</td>
<td>24.386</td>
<td>113.374</td>
<td>0.026</td>
<td>0.011</td>
<td>0.154</td>
<td>0.494</td>
<td>3.215</td>
</tr>
<tr>
<td>5</td>
<td>378.214</td>
<td>96.357</td>
<td>39.714</td>
<td>168.971</td>
<td>113.729</td>
<td>0.557</td>
<td>0.425</td>
<td>0.285</td>
<td>0.802</td>
<td>18.940</td>
</tr>
<tr>
<td></td>
<td>72.886</td>
<td>66.427</td>
<td>22.095</td>
<td>84.848</td>
<td>79.717</td>
<td>1.179</td>
<td>0.176</td>
<td>0.162</td>
<td>0.627</td>
<td>7.231</td>
</tr>
<tr>
<td>6</td>
<td>367.500</td>
<td>54.000</td>
<td>54.750</td>
<td>110.600</td>
<td>98.875</td>
<td>0.405</td>
<td>0.295</td>
<td>0.280</td>
<td>0.759</td>
<td>17.852</td>
</tr>
<tr>
<td></td>
<td>146.420</td>
<td>82.867</td>
<td>10.146</td>
<td>77.711</td>
<td>74.590</td>
<td>0.102</td>
<td>0.129</td>
<td>0.199</td>
<td>0.399</td>
<td>9.052</td>
</tr>
<tr>
<td>7</td>
<td>349.250</td>
<td>85.000</td>
<td>43.333</td>
<td>157.575</td>
<td>110.917</td>
<td>0.275</td>
<td>0.449</td>
<td>0.312</td>
<td>0.681</td>
<td>13.243</td>
</tr>
<tr>
<td></td>
<td>104.965</td>
<td>54.199</td>
<td>20.548</td>
<td>61.908</td>
<td>77.428</td>
<td>0.122</td>
<td>0.134</td>
<td>0.194</td>
<td>0.270</td>
<td>5.113</td>
</tr>
<tr>
<td></td>
<td>365.260</td>
<td>81.000</td>
<td>42.350</td>
<td>152.160</td>
<td>118.570</td>
<td>0.390</td>
<td>0.406</td>
<td>0.325</td>
<td>0.795</td>
<td>16.005</td>
</tr>
<tr>
<td></td>
<td>99.260</td>
<td>68.110</td>
<td>20.390</td>
<td>74.740</td>
<td>88.040</td>
<td>0.716</td>
<td>0.154</td>
<td>0.213</td>
<td>0.518</td>
<td>7.217</td>
</tr>
</tbody>
</table>

D = Total syllable duration  
O = Release or onset of phonation to first visible F2  
TD = Transition duration  
TP = Time to peak amplitude  
TV = Time to peak velocity of amplitude envelope  
V = Peak velocity of amplitude envelope  
P/D = Percent of syllable duration occupied by time to peak amplitude  
V/D = Percent of syllable duration occupied by time to peak amplitude velocity  
V/P = Time to peak amplitude velocity time to peak amplitude  
SHIM = Amplitude perturbation (shimmer)
and infants: 233.7 ms vs. 345.1 ms, respectively. The infants mean standard deviation of 171 ms is about triple that of the adult. Given that adult and infant mean transition durations differ modestly (Figure 4), it is the vocalic quasi steady-state that differentiates the syllable durations of the two groups. These data conform with those of Smith (1978) who observed both absolutely longer syllables and relatively longer vowels in the speech of young children. Both Smith (1976) and Kent (1981) have observed that timing, rate, and temporal control are perhaps the last-maturing aspect of articulatory control during speech development and both stress the greater variability of all temporal measures for young children. Consistent with these findings, review of Tables 3, 7a, 7b, and 7c reveals that adult variance was less than that for any infant on measures of syllable duration (D), time-to-peak amplitude (TP), and time-to-peak velocity (TV). For onset time (O) and transition duration (TD), one infant posted a standard deviation which was less than the adult, suggesting that the occasional infant could be less variable than the adult on syllable transition duration. Nevertheless, in 15 of 18
FIGURE 4. INFANT AND ADULT TRANSITION DURATIONS

Frequency Histogram

FREQUENCY OF OCCURRENCE BY PERCENTAGE

DURATION IN MILLISECONDS

INFANTS

ADULT

0
5
10
15
20
25
30

0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100
101+
comparisons over the six measures (D, O, TD, TP, TV, and V [peak amplitude velocity]), the adult subject manifested the smaller standard deviation.

The most surprising result of the study was the fact that the mean transition duration of the infants, (43.2 ms, s.d. = 40.2), was only 22% longer than the adult mean (35.38, s.d. = 22.23). The adult F2 transition data are taken from Kewley-Port (1982) and represent the mean over all stops in all vowel contexts. Knowing that absolute transition duration depends upon tempo in adults, it is problematic to establish what the "tempo" of the infant utterance was. Since all measures were made on vocalizations elicited while playing in a toy-filled room, we may assume that the infants were relaxed and unstressed as they vocalized, so that utterance tempo would perhaps be on the slow side, with correspondingly longer than usual stop transition durations. Regardless, the small standard deviations for infants' transitions and the near-adult durations reflect quasi-adult motor preference at even the very early level 3 (five to six months) of pre-linguistic development.

As previously noted, initiation of voicing
typically preceded or was almost simultaneous with oral release. For two of the three infants this resulted in average onset times which were less than for the adult speaker, reflecting very short VOT or brief prevoicing prior to F2 onset. Clearly this is indicative of the infants' predispositions toward beginning voicing very nearly simultaneously with consonantal onset. Infant JF, who presented the largest onset time mean (81 ms), tended to produce long voicing leads prior to release into the F2 transition.

In two infants, the time-to-peak velocity of amplitude rise was longer than for the adult. Simple calculation shows that, on the average, the maximum amplitude envelope velocity occurred 92%, 65%, and 88.7% of the way through the transition for the three infants, vs. 31% of the way through the transition for the adult. In addition, the time-to-peak amplitude for the three infants and the adult (36, 28.2, 28.8, and 25.7 ms, respectively), occurred beyond the end point of the formant transition. Further, the maximum amplitude velocity itself was higher for the infants (.47, .327, .390) than for the adult (.264). As a result, one can deduce that
the velocity of amplitude envelope growth was greater and occurred later in the transition for the infants than for the adult. For both the adult and infants, however, peak amplitude, on the average, occurred approximately 30 ms beyond the end of the transition. Tables 7a, b, c show that, on the average, F2 velocity was highest at about 61% of the duration of the transition for infants, which compares favorably with Furui's (1986) estimate of maximum F1/F2/F3 conjoint peak velocity at a point between 50-60% of the duration of the stop-consonant transition. As the maximum rate of change of vocal tract cross sectional area coincides with the point of most rapid formant movement (Fant, 1960), the slightly lagging point of maximum envelope amplitude velocity for infants cannot be attributed to the a delayed point of maximum rate of opening of the vocal tract. Rather, the infant larynx may interact more strongly with vocal constriction such that the maximum rate of glottal wave amplitude increase occurs later and at larger vocal tract openings for the infant than for the adult. Additionally, the infants' glottal wave may have an initially slow rise toward maximum source spectrum amplitude. The significance of this is that
infant transition dynamics are such that despite the slightly longer infant transitions, infants and adults exhibit formant velocity maxima at about 50-60% of transition duration and an amplitude envelope peak at about 30 ms beyond the end of the transition. On the other hand, infant amplitude velocity maxima occur relatively late (re., the adult), at an average of 81% of the way through the transition for infants and earlier (31% of the way through), for the adult. The infants' higher peak velocity of amplitude growth may underlie the catch-up of the infants with adult waveform envelope at a common point about 30 ms beyond the end of the transition. Thus it can be argued that infant and adult transitions resemble each other far more than they differ. Infant transitions may be a bit longer, but are not much more variable in length. The infant achieves maximum speed of opening at about the same point in the transition as the adult and attains maximum envelope amplitude at the same relative time point as the adult.
Prior to an in-depth discussion of the results of this study, it should be noted that a number of opposing theories presently forwarded by researchers in the fields of speech science, linguistics, and motor control could, in whole or part, be postulated as theoretical underpinnings for the current findings. Given the limited scope of the sample, it would appear rather parsimonious to suggest any single explanation which may account for the results. This is not to say that the personal biases of the investigator will not be presented. Rather, whenever possible, opposing views will be discussed in order to allow the reader to form his or her own opinion as to the most plausible explanation for the data.

Since the present methods of scientific investigation in the area of motor control are not viable for infants and children (these being invasive or requiring a level of cooperation beyond the scope of the infant), acoustic analyses offer the most quantitative means of examining potential motor
infants. Given the present knowledge of the relationship between vocal-tract anatomy and physiology and acoustic output in adults, it appears reasonable to infer motor control capabilities from the acoustic output in infants if differences in tract parameters are accounted for. I believe that this has been done within the current scope of knowledge of acoustic theories of speech production (Fant, 1960; Stevens and House, 1955). With this in mind, discussion of the results follows.

**General Aspects of Syllable Morphology:**

Power spectra of infant tokens displayed a set of formants congruent with those of Oller (1980) and Kent's (1976) Quasi-Resonant Nuclei, in the five to six month stage, and fully resonant, quite vowel-like spectra at the twelve to fourteen month stage. The proto-syllable nucleus began with either a burst and/or aspirative/fricative noise preceding the voiced transitions or a low frequency quasi-static resonance, which rapidly moved into a transitional movement to the vowel.

Contrary to expectations, the well formedness of the infant proto-syllables was quite evident.
Specific acoustic events indicative of loss or reduction of articulatory control were rare. These included double burst stop releases, prolonged or intrusive frication, intrusive nasality, voice breaks, pervasive vocalic dipthongization, and double or triple syllable amplitude peaks (Kent, 1981; Oller, 1980). Similarly, when syllable amplitude contours were rated on a 1-4 scale for smoothness/irregularity of the amplitude envelope, the rating did not correlate significantly with the Kamig, et al. (1990) shimmer quotient used to estimate intensity fluctuation. Typically, the first acoustically identifiable sign of syllable onset, quasi-periodic oscillation characteristic of laryngeal pulsing, began with either 5-10 ms of identifiable burst or, more often, voicing, suggesting that initiation of voicing either preceded or was almost simultaneous with oral release in the great majority of infant utterances (Kewley-Port and Preston, 1974). These general aspects of syllable morphology suggest that vocal tract modulation is relatively smooth, rapid, and unidirectional with little or no post-oscillation of tract shape once maximum opening is achieved. Therefore, proto-
syllables exhibit precociously skilled patterns of articulatory and laryngeal control for vocal production. The "play-situation" monosyllabic utterances display a smoothness of muscle forcing and sequencing characteristic of adult syllables. Being neither vegetative or distress vocalizations, the pattern of temporal smoothness and coherence displayed suggests the presence of "coordinative-structure" schemata for vocal motor control (Grillner, 1981).

Temporal Patterning:

The ratios of infant and adult temporal measures are not especially revealing. Given the longer infant syllable duration, peak amplitude and peak velocity of amplitude occur relatively earlier (percentage wise) in the syllables of infants than the adult. Further, given the later occurrence of peak velocity of envelope amplitude in the transition in the infants, the TV/TP ratio results in the adult achieving maximum velocity earlier in the time interval to peak amplitude than the infants.

The shimmer measure is remarkable for not differentiating gross amplitude envelope stability between the infants and the adult, despite the
inclusion of trendline correction which factors out much of transition rise/fall in the envelope. (See Figure 5 for frequency histogram of adult and infant shimmer values.) That is, as amplitude instability increased, the dispersion of measures about the mean also increases. This finding is opposite to what one might expect relative to other measures of phonatory instability, notably the Amplitude Perturbation Quotient or APQ (Takahashi and Koike, 1975). For APQ, a strong negative correlation exists: as phonatory instability increases, absolute deviation of each measure from the mean decreases. Analysis of adult tokens with the Ramig, et al. (1990) analysis yielded a mean shimmer value of 15.2 with a standard deviation of 8.87. When compared to the average across all infant tokens, adult values were somewhat higher than those of the infant's (Table 3: mean child value = 14.2, std. = 6.5). This is consistent with the trend noted by Takahashi and Koike (1975) in which the APQs for adult males were highest, women were in the intermediate range, and infants (from this study) in the lowest range.

The strong negative correlation between syllable duration and shimmer suggests that shimmer reflects
FIGURE 5. ADULT AND INFANT SHIMMER VALUES

Frequency Histogram

Shimmer Values Represent Ranges
the transition period, which occupies a greater proportion of syllable duration as syllable duration shrinks. Given that the adult trend was similar but non significant suggests that 2 of 3 infants had instability of amplitude output not solely attributable to rapid amplitude rise, but to unsteadiness of glottal/pulmonary origin, at least beyond the levels typical for the adult. This is as one might expect in such young subjects.

Spectral Patterning:

Perhaps the most compelling finding was the swiftness and monotonicity of formant two trajectory during transitions. It was expected that the smoothness and monotonicity of the F2 transition would vary across developmental levels, revealing a non-monotonic and highly variable transition at early stages which would slowly fade into a smooth trajectory at later stages. Instead, Table 4 shows that only 17-29\% of transitions across infants showed any "aberrancies." All 182 F2 transitions monotonically and smoothly increased or decreased across the span of the transition. Surprisingly, when post hoc digital spectrograms were computed with WAVEDIT (Miller, 1989) were made on 40\% of the
syllables selected equally from all 5 developmental levels, all presented smooth appearing F2 transitions on visual inspection. This finding confirms the smooth, unidirectionality characteristic of normal adult transitions which was manifest in the series of FFT F2 spectra. The similar transition durations, similar point of maximum F2 velocity, similar point of maximum amplitude, and smooth, unidirectionality of F2 motion all support the conclusion that the "ballistic" transitional motions of the vocal tract are well established in the newly babbling five to six month old infant. Both the dynamic biomechanic impedance and temporal patterning of muscle forcing for infants are analogous to those for adults, even though the infant transitions did not carry phonemic or lexical information at the five to six month developmental interval. Table 1 shows that there was no statistically significant difference across developmental levels for any temporal or spectral measure. Only the number of stops labeled as voiced showed a slightly significant developmental difference: bursts occurred more frequently during the first and last developmental levels than at intermediate points.
These observations lead one to the conclusion that the rather adult-like transitions, short, swift, and smooth, were within the infant's motor capabilities at a very early age (zero to five months) and then changed little across the five to fourteen month interval! The absence of of change for temporo-spectral measures did not extend to distribution of stop and vowel phonemes. As expected, children showed an increased preference for labial and alveolar stops across levels and a pattern of increasing numbers of point and intermediate vowels beyond an initial cohort of low-mid central vowels.

As Table 4 shows, however, there were differences between our infants and the adult in the target values of formant onset (consonantal target) and transition end point (vowel target). These aberrancies were of two types: (1) unexpectedly high/low F2 onset values relative to the expected onset frequency (given the consonant label applied during transcription) and (2) overshooting the final vocalic F2 target value for the vowel. For high vowels like [i], the F2 transition moved past expected F2 to even higher values before swiftly
reversing to return to a less extreme steady state value. Similarly, for low vowels F2 transitions moved very low, only to rise slightly to appropriate F2 levels. In other words, the infants appeared to be expanding their vowel targets outward, away from more central or neutral tongue positions. However, consonant targets were aberrant more often than vowel targets for all 3 subjects, especially JF (3/1) and MO (2.5/1). Many aberrant transitions, especially for infants AD (41%) and JF (25%) had both onset and overshoot errors.

The 25-40% aberrancy rate of infant transitions suggests substantial variation around apical, labial, and velar points of stop articulation, and to a lesser degree, a preference for extreme tongue positions for vocalic utterances. While a Friedman's One-Way ANOVA showed a slight trend toward fewer trajectory errors in the last developmental periods, it was not statistically significant. Most comprehensive studies of prelinguistic vocalization (Kent, 1981; Kent and Bauer, 1985; Oller, 1980, 1986; Stark, 1980) stress that early attempts at closant articulations vary widely from archetypal places, with articulations being fronted, back, raised,
palatalized, velarized, nasalized, etc. It is no surprise that the infants showed consonant target variability. There was a small reduction of transition aberrancy across developmental periods, but as noted, it did not achieve significance. The vowel overshoot suggests several explanations. As noted, the vowel segments of the infant syllables are longer than for adults. If we presume that the vowels (being long, powerful, and distinct) are highly salient to the infant, and recognizing that newborns can readily discriminate vowels (Kuhl and Meltzoff, 1984), both acoustically and visually, then we may deduce that infants have well developed auditory images for certain vowels by five to six months of age. If we assume that the tactile-kinesthetic vowel target parameters are not as well organized as the sensory image of vowel targets, then it is possible that the infant may move the vocal organs past ideal tactile-kinesthetic targets in order to increase the amount of feedback which arises from a maximally raised/lowered jaw and raised/lowered - fronted/backed tongue. However, the infant is probably incapable of reacting swiftly enough (15-30 ms) to acoustic/somesthetic feedback in
order to correct tongue/lip placement for F2 location. A more parsimonious explanation is that the infants' tongue and lip were biomechanically underdamped, which could result in the overshoot and multiple oscillations beyond targets observed in many of the aberrant transitions. While the infant vowels were highly variable, Figures 2a, 2b, and 2c show that the F1/F2 mean values fall within the central areas of predicted F1/F2 space (Lieberman, 1980). If 3.5 Bark smoothing defines vowel quality (Syrdal-Lasky and Gopal, 1986), then the initial overshoot could be readily tolerated by the infant's listeners insofar as the vowel spectra need only assume a gross f0-F1-F2-F3 spacing to achieve natural class distinctions.

Examination of transition durations revealed that they do not become faster as the children get older - the general figure is stereotypical and just slightly longer than for adults (43.3 ms and 35.38 ms for children and adults respectively). Further, close scrutiny of all 182 formant trajectories across transitions failed to reveal a relationship between presence or absence of aberrances (abnormal onset frequencies or target overshoot) and absolute
transition duration (t = .456, d.f. = 13, NS). These data are compatible with the putative motor organization suggested by Harris, et al. (1986) and Kelso, et al. (1983, 1986) where internal temporal score and phasing of gestures is preserved in the face of scalar changes in absolute duration, amplitude, velocity, etc.

Theoretical Implications:

In a review of theories of speech acquisition, Ingram (1989) cites universal, articulatory learning, maturational, and refinement theories. The maturational (Browman and Goldstein, 1989; Locke, 1983) and refinement theories (Oller, 1980; 1981) are capable of accommodating the results of this study. Locke (1983) asserts that infants possess a repertoire of "biologically predetermined" sounds, determined largely by anatomical and acoustic contingencies which predispose (Wood, 1982) infants to produce certain vowels and consonants. If we extend this "biological predisposition" to include transitions, then transitioning as well as targeting are parts of a central core of speech motor behaviors which need not be learned in the traditional sense,
but only "potentiated" during first social interactions. In the refinement model, infants interactively (with the environment) explore and develop loudness, pitch, vocal quality, resonance and timing. It is plausible that "stereotypic" transitions are one of the "core ensembles" of vocal motor behavior to which Locke refers. Logically, since quasi phonemic targets first occur in the CV proto-syllabic framework, transitioning capabilities must be concomitants of targeting behavior, that is, one does not exist without the other. Quite obviously, even the production of a single, continued vocal sound demands that the lip-jaw-larynx system engage in transitional movement away from physiological rest to target, and back to the silence and articulatory "neutral" state of vegetative rest (i.e., jaw slightly open, tongue relaxed within the mandibular arch, and larynx fully abducted). If voicing onset is at the point of the movement away from articulatory rest, transitional formant motion will occur haphazardly. The infant has only to listen to and then refine this "inadvertent" transition to create a consonantal perception preceding the vocalic continuant being attempted.
This scenario is not far fetched given that the infants vocalize intensely at birth, emitting tract-open wails and cries and tract-closed, nose-open hums and comforting sounds.

The "refinement theory" of Oller (1980) suggests that infants at the zero to one month stage produce "QRNs" (quasi resonant nuclei), i.e., nasalized CV "types" with poorly defined transitions and no clear consonantal percept. By age two to three months (in the "Goo" stage) QRNs may couple with transitions which are sufficiently well defined to support velo-uvular closants coupled to the vocalic nuclei, producing nasalized CV proto-syllables. Although not explicitly argued, Oller's scenario posits poorly defined transitions from physiological rest in month one, shifting to velar/uvular transitions by three months, and to multiple consonantal/closant transitions by seven months. Contrary to expectation, Oller (1980) argues that the infant's use of fully resonant nuclei shows an abrupt onset, perhaps reflecting a cognitively motivated predisposition. Oller's observed progression of proto-syllabic forms (QRN to QRN plus velar/uvular constriction to Fully Resonant Nuclei
[FRN]) implies rapid incremental learning of transitioning C and V targets.

If Oller and Eiler's (1988) claim that canonical (FRNs which may be reduplicated) syllabic babbling emerges at six months is correct, then sophisticated control of transitioning is accomplished at some point antecedent to the six-month level. If the reader accepts the above assertion, then the question that follows is whether this control reflects: (1) skilled incremental motor learning which depends on the learners (infant's) close monitoring of somesthetic and acoustic feedback to gauge the nature and success of transitioning or (2) a biologically mediated predisposition to automatically effect rapid transitions. A mechanism which conflates both views is offered by Grillner (1981, p. 218). He argues that "learned and innate movements may not be distinctly separate classes of movements,...[but share] common neural mechanisms." That is, learned movements may consist of recombination of parts of different innate motor mechanisms, (i.e., "coordinative structures.") As learning through motor practice proceeds, higher neural centers would selectively assume control of individual parts of
hard-wired motor programs, thereby allowing "parcellation of the larger innate motor programs into smaller parts." In the vocal tract, reflexive acts (i.e., coordinative structures) such as criescreams, sucking, and coughing could be recruited. During cry, swift vocal tract opening at the lips and tongue, velar raising, vocal fold adduction, and expiratory air expulsion are recruited in a nicely phased time course along with body extension and grimacing to emit wails. For sucking, successive lip and tongue constrictions, jaw raising, tongue raising, and antero-posterior movements of the tongue, as well as velar closure, are entrained. If, for example, the rapid, reflexive gestures of sucking (bilabial closure, tongue tip elevation and retroflexion, and tongue base elevation) can be decoupled once the reflex is integrated, then the stage might well be set for recruiting these gestures for phonatory purposes. Is it the case that the articulatory dynamics of the CV transition are components of motor control hardwired into the crying and sucking reflexes? Or is this "transitioning" a "speech-special" module awaiting only decoupling of vocal targeting from reflexive behavior before being
coordinated with vowel targets to create the proto-syllable? In the "nativist" scenario, a specific module for CV transitions is not proposed, but rather, through readiness for learning from somesthetic, auditory, and social feedback, the child practices shaping "inadvertent" transitions into the smooth, well practiced, dynamic events observed at the six month level in this study. A choice between these two alternatives and other explanations can not be made, but the time frame has been narrowed insofar as considerable control is shown to exist for transitions at or slightly before the fifth month of life.

Developmental Issues:

It was unexpected that temporal and spectral properties of the proto-syllabic CV forms showed little, if any, developmental trends across the five to fourteen month interval studied. This finding was verified post-hoc through correlational analysis of the most representative infant, (AD). Development in C and V distributions occurred as expected, syllable reduplication emerged (as did the first word for all three children), as well as an accelerated tempo of social development. One must question why the
temporal parameters did not migrate to shorter, less variable, more highly correlated values. First, selected samples were limited to isolated CV tokens, which by month twelve were a small minority group in the population of utterances for all 3 infants. Perhaps syllables embedded in trains would have shown the developmental trends which were expected. Second, with only 3 children and 182 tokens, I may have failed, through smallness of sample and chance statistical variation, to achieve a representative sample of vocalization. This is rather unlikely since the infants truly enjoyed and thrived in the play-room situation in which the data were recorded and all continued to develop normally. A third explanation is that monosyllable temporal frames, once developed by month five, may be adequate vehicles for speech output until well into the second year of life. I prefer this interpretation but must hasten to add that had many more samples and even more powerful acoustic analyses been used, small developmental trends may have emerged. If Smith's (1978, 1983, 1986) contention that development of temporal control and its concomitant variability are extended over long periods developmentally is
correct, then it may be the case that the plateau of temporal measures seen herein reflect only a slight, statistically insignificant portion of a very slowly increasing pattern of enhanced temporal motor control.

**Clinical Implications and Need for Further Research:**

Given the general failure to find a smooth developmental progression across sampling periods for intrinsic temporal measures, the emergence of rather adult-like transitioning by the fifth month of life, and the positive correlation between intrinsic temporal events with overall syllable duration, one may posit that over the period from five to fourteen months, analysis of a small sampling of isolated CV productions may be used to identify subtle neuromuscular delays in at risk infants. This is particularly true of the adult-like rate of transition duration and time-to-peak velocity of formant movement, as well as the achievement of maximum envelope amplitude at the same point past the transition as adults.

If an increasing delay time in several successive motor functions can, in fact, be of
diagnostic use as van der Stelt and Koopmans-van Beinum (1986) suggest, then measures such as the ratio of maximum F2 velocity to transition duration could provide a more sensitive indicator of neuromotor delay than achievement of gross motor and speech milestones. This is particularly true for minimally involved infants who generally do not present significant delays until well into the second year of life. While more severely involved infants are easily identified by abnormal reflexive activity (either absence of a reflex or failure to integrate the reflex at an appropriate age), feeding difficulties, and abnormal body tone, the minimally involved infant often is not. Consequently, the minimally involved infant does not receive the timely intervention necessary to insure the best possible prognosis.

It is notable that current research on infant cry can be used as an indicator of neurological involvement. The coarseness of this reflexive crying behavior, however, may not be any more sensitive to involvement than more easily observed reflexive behavior and general body tone. The final analysis of the usefulness of this procedure must wait for the
results of fairly protracted longitudinal study of those infants identified as being at-risk by the procedure, particularly relative to the later emergence of abnormal reflexive, gross, and fine motor behaviors. Because of the unusual stability of the measures of time-to-peak formant velocity within the transition and achievement of maximum velocity of amplitude envelope over the five to fourteen month period, these measures may very well be a more sensitive indicator of later emerging motor delays.

If these findings are to be of clinical value, however, normative studies with a larger sampling will be necessary, particularly in light of the rather large standard deviations exhibited for these measures relative to adult values. Further, more sophisticated means of obtaining these values are needed. The fairly time-intensive methods utilized in this study simply are not practical for anything other than research endeavors. As a result of the universally acknowledged difficulties in the acoustic analysis of infant data (Miller, et al., 1991), general clinical use of infant spectral analysis must necessarily wait for refinement in instrumentation.
Summary:

1. The infants of the current study compared well with those of Irwin (1946, 1948) in terms of distribution of vowels produced at various developmental levels and frequency of occurrence of place of articulation for stops.

2. In general, the F1/F2 values fell in the expected location in the F1/F2 vowel space and were almost perfectly congruent with those of Lieberman (1980) at the same developmental level.

3. The orderliness and well formedness of infant proto-syllables was remarkable with few acoustic events indicative of loss of articulatory control, suggesting the presence of a "coordinative-structure" schemata for vocal motor control of transitions.

4. The modal adult and infant syllable durations overlapped at ~240 ms with the infants exhibiting a strong skew toward longer durations, accounting for the difference in overall syllable duration (233.7 vs. 345.1).

5. The mean transition duration of infants was only 22% longer than for adults, with adult-like values
for achievement of peak formant velocity in the transition (60%) and maximum envelope amplitude ~30 ms after the transition.

6. With the exception of transition duration, all intrinsic temporal measures were positively correlated with overall syllable duration.

7. Shimmer measures failed to differentiate adults from children, despite trendline correction, suggesting that the glottal pulse for CV frames may be quasi adult-like as early as the fifth month of life.

8. The uniformity of formant trajectory, despite onset and overshoot errors, suggest that the rather adult-like transitions emerged very early and changed hardly at all across the 5-14 month epoch.

9. The precociously developed transitioning skills developing at or before five months of age suggest a biologically mediated predisposition to automatically effect rapid transitions and/or a recruitment of the components of reflexive gestures for phonatory purposes.
10. Monosyllabic frames, once developed, may be stable temporal frameworks for speech output well into the second year of life (Shattuck-Hufnagel, 1983).
Bibliography


Irwin, O.C. (1946) chart from Ingram (ibid.).


VITA
JAMES ROBERT PIROLI, PH.D.
1020 Edgewood Avenue, NE
Atlanta, Georgia 30307
(404) 523-2005

PERSONAL DATA
Birth Date: January 12, 1950
Marital Status: Single
SSN: 362-52-5655

EDUCATION
1975 B.S. Eastern Michigan University
Ypsilanti, Michigan
Degree: Speech and Language
Impaired/English Language and Literature

1978 M.A. Eastern Michigan University
Ypsilanti, Michigan
Degree: Speech and Language Pathology

1991 Ph.D. Louisiana State University
Baton Rouge, Louisiana
Degree: Communication Disorders

PROFESSIONAL EXPERIENCE:
1986-present Senior Speech/Language Pathologist
Egleston Children's Hospital
at Emory
Atlanta, Georgia

1984-1986 Speech/Language Pathologist
Children's Hospital
New Orleans, Louisiana
PROFESSIONAL EXPERIENCE: (con't)

1984-1986  Speech/Language/Learning Specialist
            in private practice
            Remedial Specialists
            New Orleans, Louisiana

1984-1986  Teaching/Research Assistant
            (including clinical supervision),
            Communication Disorders
            Louisiana State University
            Baton Rouge, Louisiana

1982-1984  Instructor, Severe Language Impaired/
            Speech and Language Pathology
            St. Charles Parish Public
            Schools
            Luling, Louisiana

1977-1982  Instructor, Learning Disabilities and
            Mild-Moderate, Self-Contained/Resource
            Consultant
            St. Charles Parish Public
            Schools
            Luling, Louisiana

1976-1977  Instructor, Specific Learning
            Disabilities, Resource/Consultant
            Palm Beach County Public
            Schools
            West Palm Beach, Florida

1976       Guest Lecturer, Language/Psycho-
            linguistics
            Palm Beach Junior College
            West Palm Beach, Florida

CERTIFICATION AND LICENSURE:

Certificate of Clinical Competence, American
Speech/Language/Hearing Association

State of Georgia license in Speech Pathology
CERTIFICATION AND LICENSURE: (cont')

State of Louisiana Valid Type A Teacher Certification in Elementary Grades, Speech and Hearing Therapy, and Learning Disabled

COURSES TAUGHT:
COMD 1080  Survey of Communication  B.A.
COMD 4384  Basic Voice Disorders  B.A./M.A.
COMD 7383  Cleft Palate/ Orofacial Disorders  M.A./Ph.D.

PROFESSIONAL ORGANIZATIONS:
Acoustical Society of America
American Speech/Language/Hearing Association

ACTIVITIES:
1990-present Chairperson, Metropolitan Atlanta Administrators in Speech Pathology and Audiology
1990 Co-presenter, Assessment and Management of Pediatric and Adult Dysphagia, GSHA Continuing Education Program
1989-present Member, Dysphagia Interest Group of Georgia Atlanta, Georgia
1981-1982 Chairperson: Southern Association of Colleges and Schools Self-Evaluation Steering Committee, R.J. Vial Elementary School, St. Charles Parish Public Schools, Luling, Louisiana
1980-1981  Piloted Palm Beach County's "Systematic Approach to Managing Mathematics Instruction" for Phi Delta Kappa

1979-1984  Member, Participatory Management Team, Department of Special Education
           St. Charles Parish Public Schools, Luling, Louisiana

1979  Consultant, Act 10 Inservice on "Systematic Approach to Reading Improvement," Department of Special Education
      St. Charles Parish Public Schools
      Luling, Louisiana

1976-1977  Chairman, Nongraded Instructors, Lantana Elementary School, Lantana, Florida

HONORS AND AWARDS:

1978  Invited to Membership in Phi Kappa Phi for Outstanding Academic Achievement

1979-1980  Recipient of Board of Regent's Scholarship, Eastern Michigan University

1969-1970  Member, President's Roundtable (Invitational), Eastern Michigan University
Candidate: JAMES ROBERT PIROLI

Major Field: COMMUNICATION DISORDERS

Title of Dissertation: ACOUSTIC TYPOLOGY OF INFANT PROTO-SYLLABLES

Approved:

[Signatures of Major Professor and Chairman, Dean of the Graduate School]

EXAMINING COMMITTEE:

[Signatures of committee members]

Date of Examination: 3/18/91