Perceptual Evaluation of Infant Articulatory Transitions.

Nancye Chaney Roussel

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/5407
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.

UMI

University Microfilms International
A Bell & Howell Information Company
800 North Zeib Road, Ann Arbor, MI 48106-1346 USA
313-761-4700 800-521-0600
Perceptual evaluation of infant articulatory transitions

Roussel, Nancye Chaney, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1992
PERCEPTUAL EVALUATION OF
INFANT ARTICULATORY TRANSITIONS

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
Sciences and Disorders

by

Nancye Chaney Roussel
B.A., Southeastern Louisiana University, 1974
M.A., Louisiana State University, 1986
August, 1992
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td># I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td># II REVIEW OF LITERATURE</td>
<td>7</td>
</tr>
<tr>
<td>Stages of Infant Sound Production</td>
<td>7</td>
</tr>
<tr>
<td>Inferences Concerning Articulatory Control for Infant Vocalizations</td>
<td>16</td>
</tr>
<tr>
<td>Articulatory-Acoustic Features of Adult CV Transitions</td>
<td>20</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>22</td>
</tr>
<tr>
<td># III EXPERIMENTAL DESIGN</td>
<td>24</td>
</tr>
<tr>
<td>Stimuli Selection</td>
<td>24</td>
</tr>
<tr>
<td>Stimuli Analysis</td>
<td>29</td>
</tr>
<tr>
<td>Stimuli Editing</td>
<td>31</td>
</tr>
<tr>
<td>Listening Tapes Preparation</td>
<td>33</td>
</tr>
<tr>
<td>Subjects</td>
<td>34</td>
</tr>
<tr>
<td>Procedure</td>
<td>34</td>
</tr>
<tr>
<td># IV RESULTS</td>
<td>36</td>
</tr>
<tr>
<td>Identification Functions</td>
<td>37</td>
</tr>
<tr>
<td>Perceptual Critical Points</td>
<td>43</td>
</tr>
<tr>
<td>Critical Interval</td>
<td>44</td>
</tr>
<tr>
<td>Rated Listener Agreement</td>
<td>48</td>
</tr>
<tr>
<td>Error Analysis</td>
<td>56</td>
</tr>
<tr>
<td>Acoustic Analysis</td>
<td>58</td>
</tr>
<tr>
<td># V DISCUSSION</td>
<td>62</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>85</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>94</td>
</tr>
<tr>
<td>VITA</td>
<td>160</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Distribution of Stimuli by Closant Type and Vowel</td>
<td>30</td>
</tr>
<tr>
<td>2.</td>
<td>Mean Duration Between Syllable Onset and Perceptual Critical Point as a Function of Developmental Level</td>
<td>47</td>
</tr>
<tr>
<td>3.</td>
<td>Mean Duration Between Syllable Onset and Perceptual Critical Point as a Function of Place of Articulation</td>
<td>49</td>
</tr>
<tr>
<td>4.</td>
<td>Pairwise Comparisons of the Proportion of 4 and 5 Responses at Each Age Level</td>
<td>53</td>
</tr>
<tr>
<td>5.</td>
<td>Confusion Matrices for Initial and Final Truncation Conditions</td>
<td>57</td>
</tr>
<tr>
<td>6.</td>
<td>Comparison of Acoustic Measurements for Syllables Receiving High Perceptual Ratings vs. All Other Syllables</td>
<td>61</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Percent of Identification Functions Matching Predicted Patterns</td>
<td>39</td>
</tr>
<tr>
<td>2.</td>
<td>Percent Correct Psychometric Functions for the Syllable /du/</td>
<td>40</td>
</tr>
<tr>
<td>3.</td>
<td>Percent Correct Identification Values for Consonants, Vowels and Syllables Truncated at the Same Relative Distance from the Perceptual Critical Point</td>
<td>45</td>
</tr>
<tr>
<td>4.</td>
<td>Percentage of Stimuli at Each Developmental level receiving rankings of 4 or 5</td>
<td>51</td>
</tr>
<tr>
<td>5.</td>
<td>Percentage of Stimuli at Each Place of Articulation Receiving Rankings of 4 or 5</td>
<td>55</td>
</tr>
</tbody>
</table>
ABSTRACT

The transition portion of the speech signal has been identified as critical to the perception of both consonants and vowels. Furui (1986) demonstrated that for adult Japanese speakers a 10 ms segment of the transition centered on the area of maximum spectral movement contained the most critical information for joint consonant and vowel perception. The distribution of phonemic cues contained within consonant-vowel (CV) non-reduplicated (NRB) syllables of infants ages 6 months to 15 months was compared to the distribution observed in adult productions to examine the extent of coproductive overlap of infant consonant and vowel gestures. Truncated versions of infant CV syllables in which initial and final portions of the syllable were systematically deleted were presented to adult listeners for syllable identification. Perceptual critical points (truncation point at which syllable identification fell below 60% for the first time) were determined for both initial and
final truncation conditions, and critical intervals were calculated. Results revealed significant differences in the temporal distribution and strength of phonemic cues within the infant NRB CV syllable. First, perceptual critical points could only be determined in approximately 2/3 of the 60 NRB syllables included in the perceptual study, and fewer than 1/3 of the infant syllables demonstrated normal, adult-like identification functions under both initial and final truncation conditions. Obviously, the infant syllables did not manifest the level of cue redundancy or cue constancy seen in adult syllables. Second, infant critical intervals were on the average 2 1/2 times longer than those of adult Japanese speakers reflecting an overlap of the separate consonant and vowel gestures that is less compact than in adult syllables. Results also failed to identify a clear developmental progression towards either more adult-like critical interval durations or identification functions in the NRB syllables studied.
CHAPTER I
INTRODUCTION

The prevailing theory in language acquisition for many years was that of Jakobson (1941, 1968) who believed that a sharp discontinuity existed between the early, essentially random articulatory activity of the human infant and later speech production. This theory of discontinuity has been largely rejected on the basis of numerous studies beginning with the work of Irwin (Irwin, 1947, 1948, 1949; Irwin and Chen, 1946a, 1946b). Irwin's work demonstrated a developmental trend in the appearance of consonant and vowel types which began with early pre-speech vocalizations and extended through the production of initial word-like segments. Later studies confirmed this continuous pattern of phonetic development (Lieberman, 1980; Smith and Oller, 1981; Locke, 1983). Additionally, several investigators have shown that the earliest speech-like vocalizations of infants (often referred to generically as babbling) exhibit many of the same phonetic preferences and sequences found in the production
of early meaningful words (Oller, Wieman, Doyle and Ross, 1976; Kent and Bauer, 1985; Vihman, Ferguson and Elbert, 1986; Bauer and Kent, 1987). These findings highlighted the need for in-depth investigations of the vocalizations of infants prior to the appearance of the recognizable sequences of speech sounds that are interpreted by adults as "first words". As a result, a picture of infant oral-verbal development unfolding in a series of stages has emerged. While infant vocalizations at the earliest of these stages appear to be reflections of anatomical and neurophysiological constraints placed on the infant vocal tract (Locke, 1983; Kent, 1981; Holmgren, et al., 1986; Stark, 1986; Koopmans-van Beinum and Van der Stelt, 1986), later vocalizations provide evidence of the infant’s developing control over articulatory movements prior to the production of meaningful speech. While speech is normally produced under the control of cognitive and linguistic knowledge, it is, at its most basic level, a motor phenomenon governed by a speech motor control system (Baken, 1983). The development of the motor control component underlying the successful sequencing of
individual articulatory movements is a vital step in the infant's acquisition of phonetic competence. Unfortunately direct physiological evidence as to the level of motor control exhibited by the infant is difficult to attain. Investigators currently infer articulatory movement and motoric control from other types of evidence.

One of the skills that infants must possess in order to produce meaningful vocalizations is the ability to combine successive articulatory gestures into linear sequences of speech sounds. Research has shown that these gestures are not strung together end-to-end but overlapped temporally and spatially to produce a smearing of the acoustic signal to such an extent that it is difficult to tell where one gesture ends and the next begins (Fowler, 1985). This process has been called coproduction. The overlap of gestures is the greatest in the transitions between adjacent segments as articulators accelerate from one quasi-static articulatory configuration to the next. Because the transitional movements require the precise temporal co-ordination of multiple articulatory structures, the transition portions
of the speech signal strongly reflect the dynamic variables (phasing, timing, forcing) of speech motor control. Fine-grained analyses of the spectral/temporal properties of CV-like transitions in infant vocalization at various developmental stages can yield data that reflect the stage of articulatory control attained at each developmental level.

Studies of infant speech production have focused on, among other things, the inventory of sounds produced, the development of intonation and syllable structure, different types of phonation and the time-course for the maturation of phonetic gesturing in general (Oller, 1992). However, there has been little research investigating the relationship between the spectral, temporal, and perceptual properties of CV-like transitions in infant vocalizations, primarily because of the formidable acoustic problems associated with infant vocalizations. An alternative method for assessing developmental level of infant vocalizations is to assess the perceptual information conveyed by infant utterances, particularly the CV transition, at developmental epochs during the child’s development. Such
perceptual analyses can be readily correlated with spectral and linguistic analyses.

Recent research in speech perception (Mattingly and Liberman, 1989), speech physiology (Kelso, et al., 1986) and phonology (Halle, 1988) suggests that phonemes may be most realistically modelled as "cohorts" of primary articulatory gestures. The acoustic manifestations of these gestures, when presented to mature listeners, trigger perception of the intended phoneme (Browman and Goldstein, 1990). At an abstract level, linguistic features of a phoneme manifest motor (articulatory) as well as sensory (auditory, somesthetic) properties, separately accessible to the speaker-listener. Comparison of listener identification functions for syllables spoken by infants and adults offers an opportunity to assess how infant production diverges from the adult in terms of feature encoding. Feature encoding within a syllable may be most "dense" at the point of CV transitions for both vowels and consonants (Furui, 1986; Strange, Jenkins and Johnson, 1983). Studies of windowed syllables, beginning with Tekieli and Cullinan (1979), have shown that such
transitions convey substantial phonemic cues for both consonants and vowels.

Furui's (1986) research on the information content of Japanese CV syllables demonstrated that a relatively short temporal window (critical interval) centered at the point where CV formant transitions are at maximum rate of frequency change contained the most critical information for both consonant and vowel identification. Stop consonants, because of shorter transitions and larger coproductive overlap have smaller critical intervals while less coarticulated consonants e.g. apical fricatives or slowly-moving glides have much larger intervals.

Purpose:

The purpose of this study, then, is to compare the distribution of phonemic information within infant protosyllabic CV vocalizations with the distribution observed in adult CV syllables (Furui, 1986). By including CV-like vocalizations from a range of developmental levels, changes in the underlying articulatory control mechanisms may also be detected.
CHAPTER II
REVIEW OF LITERATURE

This review of pertinent literature is organized so that the initial portion focuses on the development of prelinguistic sound production with emphasis on the development of the motor component underlying oral-verbal behavior. This is followed by a review of data concerning the distribution of consonant and vowel information within the transition region of the speech of adults. The chapter concludes with a statement of the problem and a rationale for perceptual inferences concerning the development of oral-verbal control in the prelinguistic infant.

Stages of Infant Sound Production

Although the continuity hypothesis as it is applied to the development of oral-verbal behavior in the infant implies a smoothly continuous unfolding of progressively complex behaviors without sharp discontinuities (Kent and Hodge, 1989), this developmental process has generally been described using various stage models that
highlight developmental milestones (Zlatin, 1975; Stark, 1980; Oller, 1980, 1992; Holmgren, et al., 1986; Koopmans-van Beinum and van der Stelt, 1986; Roug, Landberg and Lundberg, 1989). Although the models differ with regard to the number, duration and labels applied to the various stages, there are striking similarities in the features of infant vocal production described by the models. Oller (1992) has proposed a synthesis of both American and European models that contains five distinct stages. His model will guide the following discussion of prelinguistic infant sound production.

In the earliest stage (0-2 months), referred to as the phonation stage, infant sound production is primarily reflexive with vocalization under the control of brainstem level structures (Stark, 1986). The vocal tract of the infant differs both in size and configuration from that of adult humans, (Lieberman, 1975). The larynx is high enough in the neck that the epiglottis articulates with the velum. Effects of this configuration include the virtual elimination of the pharyngeal space as a resonator, obligate nose-breathing, and reduced oral cavity space (Lieberman, 1975;
Crelin, 1987). The sounds the infant makes at this stage fall into one of two categories: vegetative sounds (cries, burps, coughs etc) and what Oller refers to as "quasi-resonant nuclei". While cries and other vegetative sounds are distinctly non-speechlike in both phonation and timing patterns (Stark, 1986), the quasi-resonant sounds exhibit sound source and temporal features resembling those of adult speech (Oller, 1992). Since these sounds occur when the vocal tract is relatively closed i.e. with the larynx high and the tongue in apposition to the soft palate, airflow is primarily nasal. Acoustically, the sounds are characterized by broad, low amplitude resonances below 1200 Hz and may be identified perceptually as syllabic nasals or nasalized vowels (Oller, 1980). Zlatin and Stark both originally identified glottal stops and fricatives in the vocalizations of infants in the first 2 months of life. However, based on more recent data, these may actually be better classified as onsets, offsets, and interruptions in quasi-vocalic utterances reflecting the infants awkward control of phonation (Roug, Landberg and Lundberg, 1989; Oller, 1992).
During the second or primitive articulation stage (1-4 months), a new class of sounds emerges which Stark (1978) refers to as "comfort" sounds. The appearance of comfort sounds may coincide with the beginning of functioning of the limbic cingulate area (Ploog, 1979) which is associated with emotional expression. This is also the epoch when laughter first appears. Maturation of phonatory control by the infant during this stage is evidenced by the development of both glottal and velaric interruptions of the stream of vocalic phonation, providing a primitive CV framework for sustained utterances (Holmgren, et al., 1986; Roug, et al., 1989). Zlatin (1975) referred to the latter as "primitive syllables" or "protosyllables" and felt that they resulted from the infant's inability to suppress the vegetative activity of the vocal tract. Because of this lack of suppression, the acoustic features of constrictions in the oropharyngeal region and subsequent articulation of tongue dorsum (or epiglottis) with the velum are superimposed on the pleasure phonations (quasi-resonant nuclei) producing what have been referred to alternately as the "goos" of this stage of vocalization.
Although these utterances reflect the ability of the infant to produce significant variation in the power spectrum as the result of primitive articulation, Oller (1986) warns that infants at this stage are far from achieving full phonetic competence. He notes that these primitive articulations have neither fully resonant nuclei nor varied, well-timed and tightly constricted consonantal closures.

The third developmental stage (3-8 months) is the "expansion or exploratory" stage, during which a wide variety of infant vocal productions emerge. The new sound classes appear to be elaborations or expansions of the earlier vocal repertoire, and reflect both anatomical changes, vocal learning, and inventiveness on the part of the infant (Stark, 1986). Anatomically, downward and forward growth of the facial skeleton increases the size of the oral cavity and gives the tongue more room to move. In addition, the infant ceases obligate nose-breathing sometime between the age of 4 - 6 months and adopts oral tidal respiration (Crelin, 1987). Maturation of the intrinsic muscles of the tongue may allow for more precise adjustments (Kent, 1981) and when combined with expansion of
the oral and pharyngeal resonating spaces may account for the diversification of vowel types and appearance of fully-resonant nuclei seen at this stage (Lieberman, 1980; Buhr, 1980; Kent and Murray, 1982; Oller, 1986). Infants in this period appear to be exploring the limits of their vocal mechanisms as squealing, growling, yelling and whispering all occur frequently (Laufer and Horii, 1977; Oller, 1980). The velar closant sounds seen previously are supplemented by the inclusion of bilabial and labiolingual trills and friction noises. Infants may also produce combinations of fully-resonant nuclei and consonant-like elements resulting in what Oller refers to as "marginal syllables". These utterances are immature in that their transitions are slow and irregular. Predominant closures at this time are stop-like with a brief trill or friction noise as the stop is released (Stark, 1986).

The fourth stage (5-10 months) marks the appearance of well-formed or "canonical" syllables in both single and reduplicated forms. The onset of this stage is often sudden and is a robust milestone in the development of speech sounds;
the appearance of reduplicated forms are particularly salient to parents. This is the first instance in infant speech production in which closant-vocant transition is well-timed. Vowel nuclei at this stage are primarily fully-resonant, and transitions are smooth with durations (25-120 ms) that fall within the range of transition durations seen in mature natural languages (Stark, 1986; Oller, 1986). The most frequently appearing consonant-like elements are labial and alveolar stops, nasals and glides. The canonical syllables are most often produced in reduplicated form i.e "bababa" but are also extremely common in non-reduplicated forms (Roug, et al., 1989). The degree of articulatory precision may vary from syllable to syllable within the string as motor control necessary for the fine timing of laryngeal adjustment and vocal tract posturing is not available to the infant (Baken, 1983). Place of articulation is often difficult for listeners to discern as the tongue blade is still relatively large with respect to the mouth resulting in imprecision of constriction placements. Stark (1986) attributes the appearance of reduplicated babble to the
development of general rhythmic motor development seen also in limb and torso behaviors. The appearance of this stage is well-timed with peak occurrence of repetitive motor behaviors in general (Thelen, 1981) as well as with the first expression of unimanual hand preference (Ramsey, 1984). Van der Stelt and Koopmans-van Beinum (1986) provide evidence that the distribution of onset of babbling in a group of 51 infants is similar to the distribution of onsets of other motor behaviors such as smiling, rolling over, crawling, sitting, and pulling up. There is also evidence that auditory feedback is crucial to the timely emergence of reduplicated babble and for progression through the final two stages of prelinguistic sound development. Oller and colleagues have shown that while deaf babies may babble, the appearance of canonical syllables is significantly delayed when compared to normal hearing babies (Oller, et al., 1985; Oller and Eilers, 1988). In addition, Stoel-Gammon and Otomo (1986) found smaller consonant repertoires and fewer multisyllabic utterances containing true consonants in hearing impaired infants when compared to age-matched normally hearing infants.
In the final stage of infant prelinguistic vocal development (9-18 months), infants begin to display the beginnings of meaningful speech with some mixing of babbled utterances and true words. Vihman, et al. (1986) estimate that the period of coexistent babbling and words may extend until the child’s recognizable vocabulary reaches 30 words. Although there is substantial evidence to suggest that significant similarities exist across late babbling and early speech (Locke, 1986; Stoel-Gammon and Cooper, 1984), the work of Vihman and colleagues (Vihman, et al., 1986; Vihman and Miller, 1988) highlight the diverse paths taken by infants as they move from babble to speech. Infants’ babbled strings at this stage have been referred to as "gibberish" or "jargon". Menyuk, Menn and Silber (1986) note that these jargon stretches are often uttered with communicative intent and eye contact, gesture and intonation contour appropriate to that intent. These utterances seem to be transitional between the final babbling stages and the emergence of first words.
Inferences Concerning Articulatory Control from Infant Vocalizations

Because direct physiological means of assessing motor activity and control are generally inappropriate for infants, investigators have historically turned to other means of estimating articulatory skill. The earliest studies relied on phonetic transcription, diary studies and anecdotal evidence to construct inventories of the sounds produced at various stages of development (Lewis, 1936; Irwin and Chen, 1946a, 1946b). While phonetic transcriptions have the advantage of highlighting relationships between infant utterances and adult productions, they tend to obscure fundamental differences by forcing phonetic categories on nonmeaningful vocalizations (Lynip, 1951; Kent and Bauer, 1985; Holmgren et al., 1986). Acoustic analyses on the other hand allow the description of infant sounds in terms of parameters common to all sounds. Acoustic studies, though more objective and quantifiable than earlier phonetic studies, also have limitations. In general, acoustic approaches provide descriptions of the behaviors under investigation, but fail to show specific
relationships between the infant vocalizations and meaningful speech. Contemporary studies, therefore, rely on combined acoustic-phonetic approaches in the investigation of infant vocalizations utilizing the full IPA, an augmented IPA with special child specific symbols or feature based descriptive systems developed specifically for infant sounds.

Based on a number of recent acoustic-phonetic investigations, the following facts are known about the acoustic features of infant vocalizations. Formant frequencies for vocalic sounds are approximately two times the values reported for adult males (Peterson and Barney, 1952). Kent and Murray (1982) describe an infant formant space having as its center $F_1=1$ kHz, $F_2=3$ kHz and $F_3=5$ kHz. While vowel space tends to increase in $F_1 - F_2$ dimensions with age and approach the acoustic vowel space of adult speakers, the center point remains fairly stable (Buhr, 1980; Lieberman, 1980; Kent and Murray, 1982). Vowel space in the $F_1$ dimension (jaw opening) appears to approximate the typical adult pattern at an earlier age than the $F_2$ dimension perhaps due to better control of jaw versus front-
back tongue movements. Central, low-mid front and low-front vowels predominate with high vowels produced infrequently. The mean fundamental frequency (f0) falls within a range of 300 Hz - 500 Hz with excursions as high as 1000 Hz or as low as 100 Hz (Laufer and Horii, 1977; Kent and Murray, 1982; Murry, Hoit-Dalgaard and Gracco, 1983; Kent and Bauer, 1985; Bauer and Kent, 1987). Phonatory characteristics such as vocal tremor (vibrato), harmonic doubling, abrupt f0 shifts, noise components and breathiness indicating immature control of laryngeal valving are often present in voiced segments (Kent and Murray, 1982).

Stops, nasals, and fricatives are the predominant consonant-like segments in infant vocalizations with stops occurring more frequently in CV syllables, and fricatives and nasals dominating VC syllables (Kent and Bauer, 1985). Spectral energy and prominent spectral peaks of fricative-like segments are higher in frequency than comparable adult fricatives with spectral energy extending to 14 kHz or more (Bauer and Kent, 1987). The frequency range of friction noise widens and lowers as place of articulation
moves from front-to-back. Frontal articulations for all consonant types are more frequent than back articulations except in the primitive and marginal syllables seen in the primitive articulation and expansion stages (Kent and Bauer, 1985; Stark, 1986; Koopmans-van Beinum and van der Stelt, 1986). Bilabial closures are the largest category of consonant-like articulations seen in infant productions (Kent and Bauer, 1985). In the early stages, intensity differentials between consonant-like elements and adjacent vocalic segments are not great due perhaps to incomplete consonantal closures and/or lack of full resonance in the vocalic nuclei. The opening/closing gesture from vocal tract constriction into or out of vocalic segments is initially not well-timed, with slow, unsteady transitions (Oller, 1980). Around the age of 7 to 10 months, articulatory control improves such that vocal tract closures produce intensity differentials of equal to or greater than 10 dB in the power spectrum. Transitions into and out of closures are smooth with durations in the range of 25-120 ms. In addition, changes in the size and configuration of the vocal tract allow for fully-resonant vocalic
nuclei. Oller (1986) refers to vocalizations that meet these requirements as canonical syllables.

Articulatory-Acoustic Features of Adult CV Transitions

Articulatory transitions are the result of coproduction processes employed by speakers to produce swift, transient-less, temporally sequenced strings of phones. During coproduction, articulatory gestures for adjacent phonemes are temporally overlapped causing a reduction in the spatial distance between target positions for each articulatory gesture. More extensive multiphone overlappings are called "coarticulations". As these gestures are overlapped the acoustic features that are associated with each gesture are also overlapped in time (Browman and Goldstein, 1990). As a result, acoustic features for the perception of adjacent segments reside along the margins of and within the trajectory of interphonemic transitions. Numerous perceptual studies have investigated the recovery of features from various portions of the speech signal including the transition. Importance of the transition area in the perception of stops, liquids, glides and
nasals has been documented (Stevens and House, 1972; Kewley-Port, 1982; Repp, 1986a; Kurowski and Blumstein, 1987). There is also strong evidence that listeners extract information for vowel identification from the syllabic transitions (Lehiste and Meltzer, 1973; Strange, Verbrugge, Shankweiler and Edman, 1976; Tekieli and Cullinan, 1979; Cullinan and Tekieli, 1979; Strange, Jenkins and Johnson, 1983; Furui, 1986; Benguerel and McFadden, 1989). Cullinan and Tekieli (1979) presented temporally segmented stop-consonant CV syllables to listeners for identification. They found that clues for better than chance identification of place of articulation, tongue advancement and tongue height occurred within the first 60 ms of the syllable. Strange et al. (1983) presented listeners with edited versions of CVC syllables in which all steady-state information had been removed. They found that vowel identification based on transitions alone did not differ significantly from the identification of the unedited syllables. Furui (1986) attempted to determine the duration of the transition necessary for accurate identification of complete Japanese CV syllables. Syllables
which had been edited to systematically remove either initial or final portions were presented to listeners for identification in an attempt to determine the critical part of the speech wave for consonant, vowel and syllable identification. Results indicated that for stop consonants a segment of the transition approximately 10 ms in length and centered on the time point of maximum spectral movement during the transition contained sufficient information to permit percent correct identification greater than or equal to 80% for both consonant and vowel. Moreover, CV syllables truncated from the rear exhibited consonant and vowel identification scores which jointly declined as the transition was approached, suggesting that crucial information for both consonant and vowel identification is contained in the same initial part of the syllable.

Statement of the Problem

Given the importance of well-timed, smoothly articulated consonant-vowel transitions in productions of adult speakers, detailed descriptions of the properties of CV transitions in the speech-like vocalizations of infants are
needed. Since CV transitions are the most heavily coproduced and most carefully phased portions of the speech stream, the appearance of adult-like control of articulatory transitions is a significant step in the development of speech motor control. The purpose of this study was to characterize this development through investigation of the perceptual properties of CV transitions in infant vocalizations. Based on the work of Furui, acoustically well-specified infant CV protosyllabic transitions were compared to those of adults with regard to the temporal distribution of features across the transitions. The perceptual properties of the transitions were then compared to acoustic and temporal measures in an attempt to describe developmental changes in speech motor control.
CHAPTER III
EXPERIMENTAL DESIGN

Stimuli Selection:

Stimuli for inclusion in the perceptual tasks were selected from the archival recordings of two first-born infants (1 male, 1 female) which were made in conjunction with a larger study on the interactive aspects of language acquisition (Zlatin, 1976). Both subjects were reported to possess normal hearing acuity and unremarkable pre, peri and postnatal histories. Bi-weekly audio recordings made in noise-controlled laboratory conditions were the primary source of stimuli. These recordings, made in a naturalistic playroom setting, typically included interactions between the infant and one or both parents, although there were also periods when the infant was alone. Print through and strong background noise prevented more than 2/3 of all utterances from being analyzed. Weekly home recordings were also available for use but were generally of even poorer auditory quality and were used primarily to
assure that the studio tapes were representative of the children's verbal behavior at each age level.

Potential stimuli were selected from recordings spanning a ten month period from age six months to age fifteen months with the stimuli grouped at two month intervals into 5 developmental levels. This age range was chosen to include both the landmark emergence of "canonical" syllables and the attainment of the single word level of language acquisition (Oller, 1992). In the initial scanning of the corpus of acoustically acceptable utterances, all non-distress, non-vegetative CV-type vocalizations were identified and, when possible, labelled using broad phonetic transcription. Utterances which exhibited any of the following characteristics were excluded: (1) hyperphonation (squeals) (2) glottal fry (growls) (3) perceptually apparent hypernasal resonance (4) poor signal-to-noise ratio or (5) simultaneous adult-child vocalizations.

From the remaining utterances, those exhibiting perceptual evidence of stop-like qualities and an identifiable vowel-like segment
were selected for further screening. Stop-like closants (closant being defined as a consonant-like segment) are the predominant closant type in infant vocalizations for the time period studied (Kent and Bauer, 1985). They also display maximum consonant/vowel coarticulation and perceptually crucial F1/F2 transitions. Vocalizations selected were isolated CV-like syllables as opposed to those occurring in multisyllabic strings. Non-reduplicated consonant babbling (NRB) was chosen because the utterances are easy to detect, emerge prior to reduplicated forms and persist to months 14-15 and offer the opportunity to follow the development of simple CV types without the complicating issues of rhythm, rate, stress and utterance position.

The CV-like syllables thus identified were then extracted from the original recordings and a master tape for transcription purposes was made. Two phonetically expert listeners with extensive familiarity listening to infant speech jointly transcribed the corpus of tokens using the IPA alphabet and enhanced diacritics (Shriberg and Kent, 1982). This will be referred to as the master transcription. Additionally, a confidence
rating of 1 - 5 (1 denoting little or no confidence; 5 denoting extreme confidence) was assigned to each token expressing the transcribers' confidence in their transcription. These transcriptions were then compared to the original broad transcriptions made by the principle investigator. Only those syllables for which there was agreement between the two transcriptions as to vowel identity and place of articulation for the closant form and a confidence rating of 3 or higher were included in the stimulus set. An attempt was made to include 20 tokens per child for each of the 5 developmental levels. However this was not possible for the earliest levels. In level 1 (6-7 months) only 14 tokens from one child and 9 from the other met the selection criteria, while at level 2 (8-9 months) there were 17 and 10 tokens respectively. The total number of CV-like tokens available for the perceptual studies was 170.

Before the set of tokens was edited further for this experiment, it was necessary to determine if the complete syllables could be identified consistently in a standard identification paradigm. A computer program was used to
randomize and output the 170 syllables through headphones to listeners seated in a sound booth. Three graduate students with phonetic training and no known hearing loss served as listeners. They were instructed to record the syllables they heard on prepared answer forms using standard IPA form. Therefore, the response set was the open set of all English consonants and vowels. Listener responses were compared to the master transcriptions described above, and percent correct agreement for each syllable was determined. Consonants were considered correct if place and manner of articulation matched the master transcription. Voicing was not considered. A binary classification system described by Cullinan and Tekieli (1979) in which tongue height and advancement are specified was employed for vowel judgements. In this system, high front vowels include /i, I, e/; low front vowels include /E, ae /; high back vowels include / u, U, o /; low back vowels include /a, ɔ, ʌ /. A vowel was judged as correct if it matched the height and advancement characteristics of the vowel segment of the master transcription. Only those tokens in which percent agreed upon scores for the full
syllable (both the closant and vocalic segment correctly identified as specified above) fell at or above 80% were included in the experiment. There were 60 tokens that met this criterion. The distribution of tokens by developmental level as well as by closant type and vowel is given in Table 1.

Stimuli analysis:

Each token was filtered through an Ithaco bandpass filter (36 dB/octave) with cutoff points of 150 Hz and 10 kHz and digitized through a Metrabyte DAS 20 digital conversion system (12 bit amplitude resolution, 100 kHz maximum sampling rate) at a 20 kHz sampling rate using a Zenith Data Systems advanced desktop microcomputer (16 MHz, 80386 cpu with 80387 math coprocessor). A custom-made wave analysis system, WAVEDIT (Miller, 1989) was used for all waveform measurements and spectral analysis. Digitized samples were normalized (largest waveform in the signal was brought up to maximum buffer size) to ensure maximum amplitude resolution for each signal and stored. Duration measurements were made both for the release, transition and vocalic nuclei of the
Table 1. Distribution of stimuli by closant type and vowel

<table>
<thead>
<tr>
<th>Level</th>
<th>Bilabial</th>
<th>Alveolar</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HF</td>
<td>LF</td>
<td>HB</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

* HF=high-front LF=low-front HB=high-back LB=low-back
complete syllable. Beginning and end of the transition were determined visually from the waveform or "oscillographic" display. The beginning of the transition was set at the first glottal pulse, while the vowel steady-state (transition end) was determined to be the first waveform whose subsequent waveforms did not differ in shape except in minor amplitude fluctuations.

Spectral analysis (FFT) was performed using a 10 ms sample centered within a 512 point FFT analysis buffer with symmetrical Hanning windowing. This protocol has been shown to provide accurate formant frequency estimates for infant vocalizations within transition regions (Miller, et al., 1991). Spectral analyses (FFTs) were calculated at seven temporally equidistant points across the transition portion of the syllable. Additional spectra up to a maximum of four were calculated at 20 ms intervals across the vocalic portion of the syllable. Estimates of F1 and F2 center frequency were made for each FFT.

**Stimuli Editing:**

Truncated syllables for use in the listening task were prepared by digitally editing stored
waveforms using the MacSpeech Lab II system (G.W. Instruments, 1988) residing on a MacIntosh IIGS computer. The 60 syllables included in the study were digitized at a 20 kHz sampling rate, low-passed filtered at 9 kHz through the internal digital antialiasing filter of the MSLII system, normalized, and stored on disk. Truncation points for both initial and final truncations varied in 15 ms steps between the beginning of the syllable and the midpoint of the vowel. This point was determined by calculating the time period between the first glottal pulse and the offset of the vowel and dividing by two. Nine of the syllables contained vowels of long duration (>400 ms). To minimize the number of truncated stimuli from these syllables, they were edited to fall within a syllable duration range of 250 - 400 ms. Care was taken that this editing did not change the perceptual identity of the vowel. For the initially truncated syllables henceforth referred to as initial truncations, the portion of the waveform preceding the truncation point was deleted, while for final truncations that portion of the waveform following the truncation point was deleted. Most actual truncations were made at
zero crossings to prevent the introduction of transients in the edited stimuli. As a result, step sizes varied +/- 1 ms from the 15 ms figure given above. Because of the variations in syllable length, the number of transition steps per stimulus ranged from 6 to 15. The total number of truncated stimuli produced from the 60 stimulus items was 1,186.

Listening Tapes Preparation

Twenty-two sets of truncated stimuli were constructed: eleven for the initial truncation condition and eleven for the final truncations. Each test set consisted of 60 tokens with all tokens within a given set having the same truncation point. In the cases where there were fewer than 60 tokens at a given truncation point, stimuli were repeated to complete the set of 60. Because of the small number of syllables having 10 or more truncation points, stimuli with truncation points 10 and 11 were combined as test set number 10, and those with truncations at the 12th, 13th, 14th and 15th points were combined to make up set number 11 for both stimulus conditions. The order of presentation of stimuli within a test set was
randomized. Three randomizations of each test set were prepared for a total of 66 test sets.

Stimuli were presented in groups of three identical repetitions with an interstimulus interval of 800 ms. The clustered stimuli were then output from the MacSpeech Lab II into a Nakamichi cassette tape deck and recorded on Memorex dBS normal bias cassette tapes with a 5 second delay between each stimulus cluster and a 15 second delay between test sets.

**Subjects:**

Two phonetically expert listeners (NR, RD) with extensive experience transcribing infant vocalizations participated as listeners for this study.

**Procedure:**

Audio tapes of the truncated stimuli were presented over loudspeakers at a comfortably loud presentation level, approximately 80 dB A at a distance of 2 meters from the loudspeakers to the listener's head. Listeners recorded the stop-like closant and vocalic element heard on prepared answer sheets using IPA notation. The response set
for the closant element was the closed set of six English stops while all American English vowels remained a response alternative for the vocalic segment. Each subject heard each tape 3 times for a total of nine trials per stimulus. The total number of stimulus trials, including filler stimuli, was 11,880 (66 test sets X 3 repetitions X 60 stimuli per set). Listeners first heard all initial truncations beginning with the longest stimuli and progressing to the shortest condition. The order of tape presentation for final truncations was also ordered such that listeners heard the longer stimuli and progressively moved to the shorter.
CHAPTER IV
RESULTS

Listener responses were scored using the criteria discussed in the methods section. Percent correct (PC) identification scores for each syllable and its constituent consonant and vowel were calculated for each stimulus and plotted as a function of truncation point for both the initial and final truncation stimulus conditions. Individual identification functions for each listener and for results averaged across the two listeners were prepared. All data were subjected to two-point smoothing before plotting. Perceptual critical points for each stimulus were identified using the averaged data. The perceptual critical point was defined as the truncation point prior to the interval in which syllable identification fell below 60% for the first time (initial truncations) and that point at which syllable identification initially exceeded 60% (final truncations). The critical interval, which Furui defines as the interval between the
two critical points, was also calculated by subtracting the mean duration between syllable onset and perceptual critical points for both initial and final truncation conditions.

Identification Functions

Based on the results of Furui (1986), the general form of the percent correct psychometric functions could be predicted, in so far as the infant CV's manifested adult-like acoustic properties. In the initial truncation condition, as the truncation point moved from syllable onset toward syllable offset, initially high consonant identification scores should fall abruptly as the perceptual critical point is reached and then remain low. Vowel identification scores should remain high for all left to right truncations. For final truncations, as the truncation point moves from syllable offset (right) toward syllable onset (left), both consonant and vowel identification should remain high until the transition is reached and fall rapidly and simultaneously as the perceptual critical point is passed.

Of the 240 identification functions based on
individual listener data, only 32% (77/240) matched the predicted patterns (see Figure 1). Forty-six percent (55/120) of initial truncation functions matched the predicted pattern, while 18% (22/120) of final truncation functions matched the prediction. When identification data was averaged for the two listeners, 38% (46/120) of the functions showed the predicted pattern (51.6% and 23% for the initial and final truncations respectively). Figure 2 shows the psychometric functions for a stimulus having normal functions for both truncation condition. Graphs of percent correct psychometric functions for each stimulus (averaged data) are arranged according to their general shape and included in the Appendix.

The percent correct psychometric functions for the averaged data were inspected post hoc to identify deviations from the expected patterns described above. Of the 30 initial truncation functions failing to match the predicted pattern, twelve (40%) had predicted consonant identification functions but showed a decline in vowel identification as the truncation point moved into the vowel. In some cases, this drop in vowel identification could be attributed to
Figure 1. Percent of identification functions matching predicted patterns.
Figure 2. Percent correct psychometric functions for the syllable /du/.
insufficient vowel duration. Furui (1986) found that listeners needed approximately 40 ms of the steady-state vowel to attain 80% or better identification of adult vowels. Ten of the 60 infant syllables contained vowels of such a short duration that when the truncation point approached the midpoint of the vowel, the remaining portion of the stimulus was only 40-50 ms. These results may also reflect unsteadiness of vocal tract musculature, yielding unstable formant patterns during the "steady-state" portion of the vowel. Vowel perception would be expected to shift as early portions of the vowel are removed, leaving the more variable vowel remnants.

Another group of 7 stimuli (23% of abnormal functions) had identification functions that fell below 60% at the first or second truncation point, rose above 60% at one or more subsequent points and then fell once more (5 had aberrant consonant functions only, 1 had an aberrant vowel function and 1 contained aberrant vowel and consonant functions). In addition, eight stimuli (27%) contained a poorly identified consonant and/or vowel i.e. percent identification fell below 60% for all truncation points. Thus, despite reliable
transcription of the intact syllables, the very same listeners could not reliably label a subset of syllables shortened as little as 15 ms in several cases.

Of the 46 final truncation identification functions that failed to match the expected pattern twenty-two (48%) showed identification scores above 60% for all truncation points for the consonant and/or vowel (13 consonant only, 2 vowel only, 7 both consonant and vowel). A possible explanation for this is that listeners were required only to identify place of articulation for consonants and tongue height/advancement for the vowels. Tekieli and Cullinan (1979) demonstrated that tongue advancement and height features are identified earlier in temporally segmented vowels and CV syllables than the tense-lax feature. They also found that while cues for better than chance identification of place of articulation could be found in the initial 10 ms of CV syllables, 22 ms were needed for the voicing cue. Since the shortest stimulus condition in the present study was 15 ms, listeners may well have had sufficient acoustic information to make accurate perceptual judgements for all stimulus
conditions in a substantial proportion of the stimuli.

There were also ten stimuli (22% of abnormal functions) that showed highest identification scores for the shortest segments (15-30 ms) with the accuracy of identification paradoxically falling as the stimulus was increased in length. This finding applied mainly to consonant identification; perhaps it is the result of a release of the backward masking effect of the vowel on the less intense preceding consonant. As with the initial truncation results, there were eight stimuli (17%) which exhibited identification scores below 60% across all truncation points for the consonant and/or vowel.

**Perceptual Critical Points**

Perceptual critical points were established based on averaged data. For those syllables in which a critical point could be established, the percent correct identification for consonants, vowels and syllables truncated the same relative distance from the perceptual critical point were averaged and plotted to produce the results shown in Figure 3. Results for the initial truncation
condition show that consonant identification dropped from 85% to 41% as the truncation point moved through the perceptual critical point. Identification scores dropped further to 29% in the interval 30 ms after the critical point. Vowel identification scores showed only slight change. As a result, syllable identification scores coincide with consonant identification. This profile is in general agreement with Furui's results with adult speakers. In the final truncation condition, syllable and vowel percent correct identification scores rose sharply from 40% - 50% to approximately 80% - 90% when the truncation point passed through the critical point. Identification scores for the consonant show a more modest increase, from 74% to 91%. These results also resemble those of Furui for adult speakers with the exception of the higher than expected consonant identification in the intervals both prior to and following the perceptual critical point.

**Critical Interval**

Mean duration between the syllable onset and the perceptual critical point for initial and
Figure 3. Percent correct identification values for consonants, vowels and syllables truncated at the same relative distance from the perceptual critical point.
final truncations was calculated for those syllables in which a critical point could be established. Means and standard deviations as a function of age level are contained in Table 2. The temporal interval between these perceptual critical points (critical interval) was also calculated and included in Table 2. Correlation between age level and the duration of the critical interval was positive but not significant, revealing no clear trend or developmental progression (Spearman’s rho = 0.30, Kendall’s tau = 0.20). The critical interval for the earliest age level (6-7 months) was substantially smaller than other critical intervals and comparable to that observed for adult Japanese speakers, but this may be due to the small sample size at this age level (only 4 stimuli compared to 10 at subsequent levels) and the fact that 3 of the 4 stimuli were bilabial (see results of relationship between critical interval and place of articulation). Critical intervals at subsequent age levels varied from a high of 48 ms to a low of 30 ms. When data was averaged over developmental levels, the critical interval for all syllables was 31.98 ms compared to 13 ms for adults (Furui,
Table 2. Mean Duration between syllable onset and perceptual critical point for initial and final truncations as a function of developmental level.

<table>
<thead>
<tr>
<th>Develop Level</th>
<th>Initial</th>
<th>Final</th>
<th>Critical Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>41.00</td>
<td>25.62</td>
<td>52.50</td>
</tr>
<tr>
<td>2</td>
<td>60.00</td>
<td>47.96</td>
<td>100.71</td>
</tr>
<tr>
<td>3</td>
<td>64.09</td>
<td>43.00</td>
<td>94.29</td>
</tr>
<tr>
<td>4</td>
<td>49.50</td>
<td>46.93</td>
<td>97.50</td>
</tr>
<tr>
<td>5</td>
<td>43.33</td>
<td>27.50</td>
<td>73.33</td>
</tr>
<tr>
<td>All Syllables</td>
<td>54.20</td>
<td>40.69</td>
<td>86.10</td>
</tr>
</tbody>
</table>
1986). As shown in Table 3, critical interval duration did vary as a function of place of articulation with shortest intervals associated with bilabial place (14.49), followed by velar place of articulation (28.00) and alveolar (59.25). Results taken from Furui's study (1986) with adult speakers indicate a bilabial < alveolar < velar ordering for duration of the critical interval.

**Rated Listener Agreement**

To further ascertain the degree to which individual infant tokens exhibited adult-like perceptual properties, the concordance of perceptual performance between the two listeners was rated. Individual listener identification functions for each stimulus were used to rate the stimuli as to how well they matched the predicted patterns described previously. A rating of 5 indicated that both listeners showed the predicted identification functions for both the consonant and the vowel. A rating of 4 indicated that one listener showed the predicted pattern for both the consonant and vowel identification functions, and the other listener showed the predicted pattern on
Table 3. Mean Duration between syllable onset and perceptual critical point for initial and final truncations as a function of place of articulation.

<table>
<thead>
<tr>
<th>Place of Artic.</th>
<th>Initial</th>
<th>Final</th>
<th>Critical Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Bilabial</td>
<td>61.76</td>
<td>36.32</td>
<td>76.25</td>
</tr>
<tr>
<td>Alveolar</td>
<td>33.75</td>
<td>32.06</td>
<td>93.00</td>
</tr>
<tr>
<td>Velar</td>
<td>62.00</td>
<td>47.73</td>
<td>90.00</td>
</tr>
<tr>
<td>[p]</td>
<td>50.00</td>
<td>22.91</td>
<td>70.00</td>
</tr>
<tr>
<td>[t]</td>
<td>90.00</td>
<td>42.42</td>
<td>135.00</td>
</tr>
<tr>
<td>[k]</td>
<td>61.67</td>
<td>43.37</td>
<td>98.33</td>
</tr>
<tr>
<td>[b]</td>
<td>64.28</td>
<td>38.77</td>
<td>78.33</td>
</tr>
<tr>
<td>[d]</td>
<td>22.50</td>
<td>14.58</td>
<td>82.50</td>
</tr>
<tr>
<td>[g]</td>
<td>62.50</td>
<td>58.03</td>
<td>71.25</td>
</tr>
</tbody>
</table>
either the consonant or the vowel. A rating of 3 was used to denote that both listeners showed the predicted pattern on one of the syllabic elements, either the consonant or the vowel, while a rating of 2 indicated that one listener exhibited the predicted pattern on the consonant and the other listener showed the predicted pattern for the vowel. A rating of 1 was given to stimuli in which neither listener presented the predicted identification functions for either syllabic element.

Ratings of 4 or 5 were taken as an indication of adult-like encoding of the infant syllables, because at least one listener exhibited the predicted identification function for both the consonant and the vowel. The proportion of syllables at each age level receiving rankings of 4 or 5 was calculated for both the initial and final truncation conditions. As shown in Figure 4, the proportions for the initial truncation condition exceeded those for final truncations at each age level. Again, no clear developmental progression was apparent in the results of either initial or final truncation conditions. However, visual inspection of the proportions revealed a U-
Figure 4. Percentage of stimuli at each developmental level receiving rankings of 4 or 5.
shaped function with level 3 (10-11 months) showing the lowest proportion of well-perceived syllables for both initial and final truncation conditions. For the initial truncation condition, 100% of syllables at the earliest developmental level received ratings of 4 or 5 with the next highest proportion (83%) occurring at the latest developmental level. Syllables from the Level 3 had the lowest level of high ranking responses (41%). Proportions were relatively consistent across developmental levels for the final truncation condition with the exception of the 10-11 month level with only 12% of the syllables at this level receiving ranks of 4 or 5. Pairwise comparisons of the proportions using the Two-Sample Proportions Test were made to determine the significance of this trend with the proportions at each developmental level compared to the proportion for Level 3 (see Table 4). For initial truncations, the proportions at Levels 1 and 5 were statistically different from Level 3 indicating a significant decline in the proportion of well-formed syllables at age 10-11 months. Results for final truncations did not clearly support this trend however, with only Level 2
Table 4. Pairwise comparisons of the proportion of 4 and 5 responses at each age level.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Condition</th>
<th>z value</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 vs. Level 3</td>
<td>Initial only</td>
<td>2.67</td>
<td>0.008*</td>
</tr>
<tr>
<td></td>
<td>Final only</td>
<td>1.70</td>
<td>0.04*</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>1.70</td>
<td>0.04*</td>
</tr>
<tr>
<td>Level 2 vs. Level 3</td>
<td>Initial only</td>
<td>0.69</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Final only</td>
<td>2.45</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>2.01</td>
<td>0.02*</td>
</tr>
<tr>
<td>Level 4 vs. Level 3</td>
<td>Initial only</td>
<td>0.69</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Final only</td>
<td>1.29</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>0.29</td>
<td>0.39</td>
</tr>
<tr>
<td>Level 5 vs. Level 3</td>
<td>Initial only</td>
<td>2.27</td>
<td>0.01*</td>
</tr>
<tr>
<td></td>
<td>Final only</td>
<td>1.41</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>0.93</td>
<td>0.18</td>
</tr>
</tbody>
</table>

* significant
showing significantly greater proportions of high ranking syllables.

To further investigate the perceptual stability of the infant syllables a more stringent criteria was applied. Stimuli receiving rankings of 4 or 5 in both initial and final truncation conditions were identified. Only 23% (14/60) of the stimuli fell in this category with more than half (8/14) of these syllables coming from the earliest two age levels. As can been seen in Table 4, while proportions at Levels 1 and 2 were statistically greater than Level 3, proportions at later developmental levels did not differ significantly.

The proportion of syllables receiving rankings of 4 or 5 across place of articulation was also investigated (see Figure 5). As before, proportions for initial truncations were greater than for final truncations in each condition. Pairwise tests of significance using the Two Sample Proportions Test revealed no significant differences in proportions as a function of place of articulation.
Figure 5. Percentage of stimuli at each place of articulation receiving rankings of 4 or 5.
Error Analysis

Consonant confusion matrices for the initial and final truncation conditions are presented in Table 5. Errors made during the interval 15 ms after the critical point for initial truncations and 15 ms before the critical point for final truncations (intervals in which identification scores showed the most rapid change) were analyzed. The error rates are expressed as a percentage of the total number of responses during each interval.

For initial truncations, favored error responses differed for voiced/voiceless cognate pairs e.g. p -> /k,g/, b -> /t,d/. For final truncations, results similar to those for initial truncations held for p/b only. Initial truncations showed labials being heard as velars most frequently, alveolars as either labial or velar and velars as either labial or alveolar depending on voicing. For final truncation, labials were heard as velars or alveolars (depending on voicing), alveolars favored velars and velars favored alveolars. Clearly, the aspirative noise excitation of voiceless stop transitions and the vocalic excitation of
Table 5. Confusion matrices for initial and final truncation conditions based on error rates in the interval 15 ms after the critical point for initial truncations and 15 ms before the critical point for final truncations. Actual number of responses are given in parentheses.

<table>
<thead>
<tr>
<th>Initial Truncations</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>In\Out</td>
<td>p,b</td>
<td>t,d</td>
<td>k,g</td>
</tr>
<tr>
<td>p (108)</td>
<td>12% (13)</td>
<td>41% (44)</td>
<td></td>
</tr>
<tr>
<td>b (324)</td>
<td>10% (31)</td>
<td>4% (13)</td>
<td></td>
</tr>
<tr>
<td>t (72)</td>
<td>11% (8)</td>
<td></td>
<td>42% (30)</td>
</tr>
<tr>
<td>d (270)</td>
<td>42% (114)</td>
<td></td>
<td>3% (7)</td>
</tr>
<tr>
<td>k (162)</td>
<td>12% (20)</td>
<td></td>
<td>60% (98)</td>
</tr>
<tr>
<td>g (108)</td>
<td>44% (47)</td>
<td></td>
<td>27% (29)</td>
</tr>
<tr>
<td>Total</td>
<td>189</td>
<td>171</td>
<td>94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Final Truncations</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>In\Out</td>
<td>p,b</td>
<td>t,d</td>
<td>k,g</td>
</tr>
<tr>
<td>p (108)</td>
<td>10% (11)</td>
<td></td>
<td>37% (40)</td>
</tr>
<tr>
<td>b (324)</td>
<td></td>
<td>23% (69)</td>
<td>21% (63)</td>
</tr>
<tr>
<td>t (72)</td>
<td>11% (8)</td>
<td></td>
<td>38% (27)</td>
</tr>
<tr>
<td>d (270)</td>
<td>21% (56)</td>
<td></td>
<td>21% (58)</td>
</tr>
<tr>
<td>k (162)</td>
<td>18% (29)</td>
<td></td>
<td>27% (43)</td>
</tr>
<tr>
<td>g (108)</td>
<td>15% (22)</td>
<td></td>
<td>20% (26)</td>
</tr>
<tr>
<td>Total</td>
<td>115</td>
<td>149</td>
<td>188</td>
</tr>
</tbody>
</table>
transitions in voiced stops provided differing cues for stop labelling in initial truncations and to a lesser degree in final truncations. Overall, labials most frequently evoked velar errors, alveolars evoked a more mixed response and velars evoked more alveolar than labial responses. Given that labials (n=432) outnumbered alveolar (n=342) and velar (n=270) tokens in the test syllables, a response bias towards labials is possible. An analysis of error responses show comparable numbers of labial (n=189) and alveolar (n=171) error responses for initial truncation suggesting no particular bias for the most frequently occurring labial consonants. For final truncations, if anything, velar error responses (n=188) outnumbered the labial (n=115) and alveolar (n=149) error responses suggesting perhaps a mild bias toward velars, the least frequently occurring stop-consonant type in the English speaking community.

Acoustic Analysis

The results of acoustic analysis were used to probe the relationship between acoustic measurements and perceptual judgements for the
infant syllables included in the study. Measurements of transition duration and F1 and F2 onsets and offsets for syllables receiving high perceptual rankings for both the initial and final truncation conditions i.e. those that were judged to be well-encoded were compared with the remaining syllables. The overall transition duration for all syllables was 71 ms (s.d.=44, min.=7 ms, max.=237 ms) as compared to 35.38 ms (s.d.=22.23) mean transition duration for adult speakers (Kewley-Port, 1982). The mean transition duration was 63.9 ms for the group of syllables receiving high perceptual rankings compared to 68.2 ms for all other syllables. Formant measurements (F1F2 onsets and offsets) were compared to expected values predicted by multiplying Kewley-Port’s (1982) adult values by a factor of two. Percent deviation for each stimulus was then calculated. The mean absolute deviations for F1 and F2 onsets and offsets for both sets of stimuli are included in Table 6. Results of a Mann-Whitney Two Sample Test revealed no differences between the two groups for transition duration (z=0.782, NS), F1 onset (z=0.021, NS), F1 offset (z=0.0105, NS), F2 onset
(z=0.294, NS) or F2 offset (z=0.600, NS). These acoustic measurements, therefore, failed to differentiate the perceptually different syllable groupings.
Table 6. Comparison of acoustic measurements for syllables receiving high perceptual ratings vs. all other syllables in this study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>High Perceptual Rank</th>
<th>All Other Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Trans. Dur.</td>
<td>63.92</td>
<td>40.19</td>
</tr>
<tr>
<td>F1 Onset</td>
<td>21.77</td>
<td>24.49</td>
</tr>
<tr>
<td>F1 Offset</td>
<td>22.77</td>
<td>24.95</td>
</tr>
<tr>
<td>F2 Onset</td>
<td>15.62</td>
<td>14.81</td>
</tr>
<tr>
<td>F2 Offset</td>
<td>32.08</td>
<td>49.20</td>
</tr>
</tbody>
</table>
CHAPTER V
DISCUSSION

Given that the purpose of this investigation was to document developmental changes in the encoding of consonants and vowels within a syllabic frame, it is obvious that the scope of the study is limited by the small number of infant vocalizations at each age level meeting criteria for inclusion in the perceptual tasks. Although the original corpus of infant vocalizations was quite extensive, reliable identification of infant CV syllables extracted from their natural context proved to be more difficult to achieve than expected. Because of the design of the experiment, it was extremely important that the infant syllables be reliably identified. Several investigators of early infant vocalizations have cited high intersubject agreement among transcribers using an augmented IPA with child specific symbols (Roug et al., 1989; Oller, 1992). In the present study, however, even though listeners were only required to make place of
articulation judgments for the consonants and height/advancement judgements for the vowels, only 60 of a set of 170 syllables (35%) met the 80% identification criteria. Additionally, our use of single syllable nonreduplicated consonantal babble (NRB) as the vocalization type of choice for inclusion in the perceptual tasks limits our ability to generalize the results. NRB utterances have received less research attention than canonical reduplicated and variegated forms and represent a small percentage of the verbal output of children in the latter stages of babbling. It may be the case that NRB is a "stranded" form, an early occurring register of child vocalization that shows little developmental change and is eventually abandoned by the infant. We chose to use this form of babbled vocalization both because it was easier to isolate from the string of infant vocalizations and because NRB syllables most nearly matched the single CV syllables spoken by the adults in Furui's (1986) study. In retrospect, we may have unintentionally missed strong developmental trends by excluding multisyllabic forms from the vocalizations sampled. Representativeness of the sample may
also have been limited by the small number of infants studied. A larger experimental group might have yielded greater diversity in consonant-vowel combinations and a more equal balance of labial, alveolar and velar CV syllables at each developmental level. Because of these factors the conclusions drawn in the following discussion must be guarded.

**General Considerations**

In general, results indicate that infant syllables are not nearly as adult-like acoustically as previous research might indicate. Piroli, et al. (1992), investigating the same developmental time period, found well-formed formant trajectories across the 5-14 month span. Transitions moved smoothly and unidirectionally from onset to steady-state with few acoustic events indicative of loss of articulatory control. In addition, the points of maximum F2 rate-of-frequency change and maximum syllable intensity were comparable to those seen in adult syllables. Infant syllable onsets were 11% shorter than adult forms, while infant transitions were 22% longer than adult transitions, but both were no more
variable than the adult form. While there appears to be some congruence between surface features of adult and infant CV syllables, the perceptual evaluation of the current study revealed significant differences in the temporal distribution and strength of phonemic features within the infant NRB CV syllable reflecting underlying adult-infant differences in articulatory mechanics not so readily apparent through acoustic analyses.

First, and most important, of the approximately 200 NRB syllables which were acoustically satisfactory, only 60 could be reliably labelled. Second, the perceptually critical interval for the infant tokens studied was on the average 2 1/2 times longer than that reported by Furui (1986) for adult Japanese speakers, reflecting an overlap of the separate consonant and vowel gestures that is less compact than in adult syllables. Additionally, fewer than 1/3 of the infant syllables demonstrated normal, adult-like identification functions under both initial and final truncation conditions. It is obvious that the infant syllables did not manifest the level of cue redundancy or cue constancy seen
in adult syllables. The encoding of infant CV segments was tenuous at best. Despite the fact that the adult listeners (author and dissertation director) were highly motivated veteran judges, perception of the truncated versions of the infant syllables was a difficult task.

Critical Interval Durations

The critical interval, which by definition contains the most important acoustic information for both consonant and vowel identification, was 31.98 ms long compared to the mean of 13 ms reported for adult Japanese stop consonant CV syllables (Furui, 1986). Longer critical intervals are an indication that the acoustic cues for the infant segments are spread over a longer period of time. One explanation for the wider critical intervals of infant syllables is increased transition durations. Although the transition durations reported in the present study are roughly double that of the average adult (71 ms vs 35 ms), it is likely that the method used to determine transition duration i.e. visual determination of the point at which the waveform stabilizes greatly over-estimated underlying
transition length. Pirola, et al. (1992) by defining the end of the transition as that point where successive (5 ms apart) F2 values changed less than 10% determined that infant transitions in 160 of the same type of NRB CV stimuli sampled in the present study averaged only 22% longer than the adult. We feel this is a more accurate measure of the infant transitions. While we would expect infant critical intervals to show some lengthening due to the slightly longer transitions, the magnitude of the difference in reported intervals for adult syllables and those of the infants suggests that other factors besides swiftness of transitional movement contribute to critical interval width.

It is quite possible that infant control of the coproduction of "tract open" and "tract closed" gestures is unpolished, less skilled and therefore, likely to be both less accurate and less successfully overlapped. This lack of tightly planned and/or executed coproduction of syllabic elements in the infant productions could result in increased critical interval durations. In Browman and Goldstein's (1990) theory of gestural specification, "gestural coordination is
specified in terms of the relative phasing of gestures". In other words, a specified point in one gesture coincides temporally with a point in another gesture resulting in a spatial and temporal overlap of the separate gestures. Each gesture, however, is generally composed of a set of specifications for an ensemble of individual articulator movements which in principle require different points of overlap and therefore different relative phases. Perhaps the prelinguistic infants do not produce the phase patterning for key movements that allow for the tight overlap of critical cues in the adult speaker. In Evart's (1982) terminology, perhaps much of the motor control pattern for early CV productions is appropriated from portions of vegetative co-ordinative structures, which have phase patterns not specifically adapted for speech. Only after socio-linguistic feedback can the infant learn to shape the syllable in a more adult-like fashion. Infant coproduction strategies also may be hampered by lack of smooth forcing function in the control of the individual articulators because of the infant's neuromuscular immaturity, particularly their inability to effect
swift on-line control of movement through fast sensori-motor feedback.

In addition to wider critical intervals, we found that the relationship between critical interval length and place of articulation differed in the infant syllables. Furui (1986) found that the closer to the front of the vocal tract the point of articulation was, the shorter the duration between syllable onset and perceptual critical point. This did not hold true for the infant syllables (refer to Table 3). In all cases, alveolar place of articulation produced the longest durations and largest critical intervals. These results may be a reflection of the later development of alveolar stops in infant productions. Typically velars appear first, followed by bilabials with alveolars appearing last. The production of alveolar stops requires greater fine motor control especially of the tongue tip region which is quite immature in morphology until well past month 10 (Zemlin, 1991; Crelin, 1987). The infant's productions of alveolar stops are likely to be more variable with respect to actual placement of the tongue upon release and on subsequent movements into the
adjacent vowel. The perceptual result may be ambiguous place cues within both the burst and transitions. Listeners, then, would require larger segments of the less accurately cued stimulus to identify the syllable, effectively increasing the width of the critical window. Our results are in agreement with those of Chapin, Tseng and Lieberman (1982) who reported that adult listeners had greater difficulty identifying infant alveolar consonants than labials or velars. One-word utterances with initial voiced stop consonants were isolated from the spontaneous discourse of two 16 month old infants. The utterances were edited to produce 15 ms, 25 ms, 40 ms, and 135 ms samples of the initial portion of the syllables which were presented to adult listeners for consonant identification. While identification was better than chance at all stimulus conditions, they found that velars and bilabials were better perceived at all durations than alveolars. This is not the case with adult speakers. Kewley-Port, Pisoni and Studdert-Kennedy (1983) found close to 100% identification of /b/, /d/ with only 20 ms of stimulus with /g/ identification improving to 90% at the 40 ms
point. Clearly, the infants in the present study, as well as the slightly older infants of the Chapin et al. study, are not producing alveolar stops with adult competence.

**Perceptual Critical Point**

For syllables in which a perceptual critical point could be established, the syllable, consonant, and vowel percent correct identification in the intervals 15 ms prior to and following the critical point follow the pattern reported by Furui (1986) for adult speakers (see Figure 3). In initial truncations, consonant and syllable identification drop simultaneously in the interval following the critical point while vowel identification remains high. The decline in identification scores during this interval for both infant and adult syllables is approximately 40%. In the final truncation condition, in the interval prior to the critical point, we find that consonant and vowel scores for the infant syllables show a smaller 15% and 30% increase respectively compared to increases of 30% for consonants and 40% for vowels in the adult syllables. These results indicate that while some
of the infant syllables resembled adult syllables in the perceptually critical region of the acoustic signal i.e. that area immediately surrounding the critical point, the infant cues were less temporally concentrated than the adults.

**Identification Functions**

One of the most striking results of the investigation was the small number of infant syllables producing percent correct identification functions that matched the predicted patterns and the fact that this failure rate did not lessen developmentally. Of the 60 infant CV syllables, only 14 (23%) produced the predicted percent correct psychometric functions for both initial and final truncation conditions. Examining only the initial truncations, we find that one half (30) of the psychometric functions deviated from the expected pattern. Vowel identification fell below the criterion in 40% of these functions, while in 27% of the functions consonant and vowel identification was well below 60% for all truncation points. Another 23% exhibited psychometric functions in which the consonant or vowel demonstrated an erratic identification
pattern with several peaks and valleys. For the final truncation condition, 46 of the 60 syllables failed to match the predicted psychometric function. Of these, 48% exhibited identification scores greater than 60% for all truncation points and an additional 22% showed highest identification scores for the shortest segments with identification becoming worse as the stimulus increased in duration. Another group of syllables (17%) exhibited identification scores below 60% for all truncated stimuli.

It is clear that the spatio-temporal encoding of consonant and vowel information in the infant syllables differed markedly from adult syllables. Examination of the mean psychometric functions (see Figure 3) for both types of truncation does not reveal substantial differences in the percent correct functions for consonant and vowel except where expected in the initial truncations. Likewise, examination of aberrant psychometric functions does not clearly favor less accurate encoding of either consonant or vowel, although the fact that half of the syllables that failed to produce predicted functions in the initial truncation condition contained only aberrant vowel
functions is an indication that vowel encoding, even in the steady-state portions of the syllable, is not adult-like.

The high incidence of aberrant PC functions points toward both immature consonant and vowel targets and immature coproduction "phasing" skills. Such immaturity may not only weaken or neutralize certain features, but may introduce conflicting perceptual cues. The weakened and/or conflicting cues may not be potent enough to alter the perception of intact syllables, but truncation removes sufficient redundancy to uncover latent ambiguity in the infant syllables. Neutralized, conflicting and temporally diffuse feature overlap must contribute heavily to the anomalous psychometric functions which oscillate with increasing truncation size or commence their decline when as little as 15 ms of the syllable is deleted. This tenuous cuing of syllables is not a linguistic anomaly since the infants have no phonemic categories, although they do have "preferences" for certain sound classes (Kuhl and Meltzoff, 1984). Perhaps what the Piroli, et al. (1992) data show is that speech appropriate tract opening/closing is a ballistic movement pattern
that is biologically specified and organized but must be "tuned" for verbal communication through socio-linguistically appropriate verbal interaction. The tuning may in large part be slow mastery of the Browman-Goldstein phasing strategies during the course of auditory-motor matching of the infant's productions to adult phonetic forms. Certainly, the present data suggest that the majority of NRB utterances remain untuned out to one year or more. What is puzzling are the data from Oller (1992) which suggest that canonical forms of babble show rapid development of prosodic cues as well as broadening and attunement of phonetic space and phonetic cues. Why then are the NRB syllables not part of the mainstream of developmental phonetic change? At any rate, the NRB syllables under investigation are for the most part very tenuously cued despite the surface resemblance of their transitions to adult forms.

The F2 onset is a major cue for place of articulation for adults, though not the only one for stop perception. We know from the work of Piroli and colleagues (1992) that F2 onsets fail to match predicted values in 30% of their CV
syllables while overshoot of F2 steady-state occurred in up to 24% of the tokens. In the present study, we attempted to investigate directly, the effect of F1F2 irregularities on perception. We found no significant differences, however, in the F1F2 onsets and offsets for a group of well-perceived syllables when compared to other syllables that did not receive high perceptual rankings. Although F1F2 values in the present investigation and in the work of Piroli, et al. (1992) ended near to values predicted given the adult label of the vowel, the high fundamental frequency of the infant productions and the subsequent wide spacing of harmonics must impact negatively on accurate perception of the formant location and spacing. In addition, the higher infant formants (F3, F4) typically are not visible in the 36 dB dynamic range of our FFT acoustic analysis, therefore we have no way to compare our infant phonetic spaces to the F0, F1, F2, F3, "linear contrast" vowel space postulated by Syrdal and Gopal (1986) to govern vowel perception. We do know that because of the anatomical constraints placed on the infant vocal tract, production of vowels, particularly the extreme vowels such as
/i, I, u, U/ requires the infant to hyperarticulate in order to produce reasonable approximations of the adult forms (Crelin, 1987). It is probable that the higher formants in these hyperarticulated segments, if present at all, do not lie in the same relationship with F1F2 as in vowels produced by adult speakers. Additionally, infants at this stage of their development have little or no linguistic knowledge to guide production of vowels that match the acoustic constraints of adult phoneme categories. Given all of the factors above, high perceptual variability for NRB CV syllables is not unexpected.

**Absence of Developmental Trends**

The lack of a strong developmental trend of any type in the perceptual data was unexpected given that other aspects of infant vocalizations (such as phonetic detail, phonetic inventory, communicative intent and intonation) do undergo developmental change during this time period. Nevertheless, we could find no developmental progression as late as 14 months towards either more adult-like critical interval durations or identification functions. Our results are in
agreement with those of Piroli, et al. (1992) who also found no statistically significant developmental trends across months 5-14 for a number of spectral and temporal measurements. Certainly the small sample size as well as the heavy reliance on NRB as stimulus items could have contributed to the lack of a definite age-related progression. In fact, the only age-related trend evident in the data may lend credence to the latter possibility. When we compared the proportion of perceptually stable responses (as determined by how well individual listeners matched the idealized identification functions) across age levels, there was a statistically significant trend towards higher proportions at the earliest two developmental levels encompassing ages 6 months - 9 months. If the NRB are truly "stranded" forms, we would expect to see high proportions of these syllables at earlier age levels. This does not explain, however, why the infant would abandon this early mastered form for later productions that were less accurate and more tenuously encoded. Perhaps the NRB syllables are neglected by the infants who favor exploration and refinement of other aspects of verbal productions
such as intonation, rhythm, intensity, etc. (Papousek and Papousek, 1991). The end of the exploration-expansion stage at about 8 months coincides with the decline observed in the proportion of well perceived NRB syllables. If skilled motor movements are not massively practiced with keen attention paid to sensory feedback, the skilled movement will show deterioration (Schmidt, 1982). We may not be seeing refinement in the coproductive aspects of NRB because the infant has abandoned interest in this form in favor of the canonical forms of reduplicated and variegated babble. Although there is no evidence to cite it is possible that the infant receives little interactive feedback from caregivers to reinforce the occurrence of NRB. In other words, this isolated form of babbling may not be communicatively salient. As a result, the infants have little incentive to improve production, and performance actually shows some deterioration.

It is also possible that the absence of a developmental progression is the result of the lack of balance among the developmental levels with regard to consonant and vowel types. Given
the results of the relationship between critical intervals and place of articulation, it is not surprising that a comparison of Tables 1 and 2 reveals that age levels with a larger proportion of syllables beginning with alveolar consonants also have larger critical intervals.

**General Discussion**

Despite the absence of significant developmental trends, the data is consistent with the view that infant coproduction of syllabic elements is far from being the closely phased, temporally compact process exemplified by adult productions. As a result, coarticulatory cues are not as reliable as those of adults. This is reflected in the longer critical intervals, the lack of well-defined perceptual critical points in many of the syllables, and the large number of psychometric functions that failed to match the predicted patterns. Evidently, much of the fine phonetic-linguistic detail seen in adult CV syllables is missing from the infant NRB syllables. If we include the large number of NRB syllables we could not reliably label, we can assert, with some justification, that the typical
(median) NRB syllable is tenuously coproduced and the NRB CV syllables discussed herein represent the upper end of a distribution of "well-formedness" of this utterance type. Hence what we observe in NRB forms are the biomechanical "prototypes" of speech-like transitions and little more. Perhaps these NRB motor forms must be commandeered for socio-linguistically intended communicative acts before strong refinement can occur.

Two opposing theories concerning the development of speech motor sequencing skills in children are currently being debated in the literature. According to one theory, young children have a tendency to produce speech segment by segment. The ability to temporally order speech units is acquired first followed by the ability to smoothly overlap separate phonetic segments (Kent, 1983). Consequently, children will exhibit fewer or less consistent intrasyllabic coarticulatory effects (Sereno and Lieberman, 1987). The opposing view suggests that young children may rely on the syllable as a unit of speech motor planning and as a result exhibit coarticulation effects that either match or exceed
those of adult speakers (Nittrouer et al., 1989; Nittrouer and Whalen, 1989; Repp, 1986). We speculate that newborns are biologically disposed to eventually produce speech appropriate CV transitions given sufficient vocal practice. Thus infants have a basic crude coproduction/coarticulation control schema which must be tuned to accept a variety of acquired appropriate consonant and vowel targets, the endpoints of the transition. The infants must acquire the fine phonetic detail of phasing and forcing each articulator movement for each CV combination. Thus the dichotomy between these two views may be bridged.

The data of Katz, Kripke and Tallal (1991) though based on productions of much older children than the infants studied here appears to support our findings. They studied lingual anticipatory coarticulation in [sV] syllables of adults and 3, 5, and 8 year old children and found that while young children and adults displayed similar patterns of anticipatory coarticulation, coarticulatory cues were less perceptible in the 3-year-old children’s productions. This was attributed to either less regularity in the
production of coarticulatory cues or to difficulty perceiving regularly produced cues due to poorly produced fricatives. Thus, for the three-year-old group there was spectral evidence of coarticulation for some stimuli that lacked perceptually salient coarticulatory cues (Katz et al., 1991). There was no evidence, however, to show that any of the groups of children produced greater degrees of coarticulation than adult speakers.

In summary, we have found that despite acoustic evidence of smooth, well-timed CV transitions in NRB syllables, consonant and vowel target knowledge and coproduction are sufficiently tenuous to render the majority of such syllables unreliablely identifiable. Additionally, of the syllables that can be identified, the majority are very sensitive to temporal disruption. The lack of developmental trends for improvement of this form suggest that this utterance type does not come under environmental-linguistic control as do other forms of infant utterances. Clearly, more extensive investigations are needed in order to substantiate and extend our results. Replication of the present study with a larger sample of
infant syllables that includes the reduplicated and variegated forms and more equal balance across developmental level with regard to consonant and vowel type is indicated. It may also be interesting to extend the scope of the investigation to include slightly older infants (out to 18 months) and a broader range of speech sounds (e.g. fricative plus vowel syllables). In addition, methods such as Furui's spectral transition measure using cepstral techniques should be employed so that direct relationships between acoustic and perceptual results may be ascertained.
REFERENCES


APPENDIX
Percent Correct Identification Functions for all Stimuli
Group I
Stimuli with PC Functions that Match Predicted Patterns
du 22081

Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
60% Identification

Percent Correct

Truncation Point

Percent Correct

Truncation Point
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% Identification
Initial Truncation

Final Truncation

Legend:
- Consonant
- Vowel
- 60% Identification

Percent Correct

Truncation Point

Percent Correct

Truncation Point
ba 13146

Initial Truncation

Legend
- Consonant
- Vowel
60% Identification

Final Truncation

Legend
- Consonant
- Vowel
60% Identification
Initial Truncation

Final Truncation

Legend:
- Consonant
- Vowel
- 80% Identification
ga 24304

**Initial Truncation**

- **Legend**
  - Consonant
  - Vowel
  - 60% Identification

**Final Truncation**

- **Legend**
  - Consonant
  - Vowel
  - 60% Identification
ka 24268

Initial Truncation

Final Truncation
ka 24251

Initial Truncation

Final Truncation
ba 25360

Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% Identification
Group II
Stimuli with PC Functions that Show Better Than Expected Performance for Consonant and/or Vowel
Ba 22074

Initial Truncation

Final Truncation
pa 23172

Initial Truncation

Final Truncation
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% Identification
ba 24227

**Initial Truncation**

**Final Truncation**

Legend:
- Consonant
- Vowel
- 80% Identification
go 25365
paU 25362

**Initial Truncation**

**Percent Correct**

- Legend:
  - Consonant
  - Vowel
  - 60% Identification

**Final Truncation**

**Percent Correct**

- Legend:
  - Consonant
  - Vowel
  - 60% Identification

Truncation Point
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% Identification
Group III
Stimuli with Aberrant Consonant PC Functions
BAE 11013

Initial Truncation

Final Truncation

Legend
- Consequent
- Vessel
50% Identification
Initial Truncation

Percent Correct

Final Truncation

Legend
- Casement
- Vessel
60% Identification

Percent Correct

Legend
- Casement
- Vessel
60% Identification
ba 23204

Initial Truncation

Percent Correct

Legend
- Consonant
- Vowel
- 60% Identification

Final Truncation

Percent Correct

Legend
- Consonant
- Vowel
- 60% Identification

Truncation Point

Truncation Point
ti 24297

Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 90% Identification
Initial Truncation

Final Truncation

Legend

- Consonant
- Vowel
- 80% Identification
ba 14192
Group IV
Stimuli with Aberrant Vowel PC Functions
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
60% Identification
DU 21049

Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% Identification
dU 22078

Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
80% Identification
dU 22092

Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% identification

Truncation Point
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% Identification
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 60% Identification

Percent Correct

Truncation Point
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% Identification
ba 13147

Initial Truncation

Final Truncation
da 24263

Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 80% Identification

Percent Correct

Truncation Point

Percent Correct

Truncation Point
Initial Truncation

Final Truncation

Legend:
- Consonant
- Vowel
- 60% Identification

Percent Correct

Percent Correct

Truncation Point

Truncation Point
Group V
Stimuli with Aberrant Consonant and Vowel PC Functions
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
- 60% Identification
dE 21050

Initial Truncation

Final Truncation
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
60% Identification

Percent Correct

Truncation Point

Percent Correct

Truncation Point
Initial Truncation

Final Truncation

Legend:
- Consonant
- Vowel
- 80% Identification

Percent Correct

Truncation Point

Truncation Point
Initial Truncation

Final Truncation
pa 24228

Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
80% Identification
Initial Truncation

Legend
- Consonant
- Vowel
60% Identification

Final Truncation

Legend
- Consonant
- Vowel
60% Identification
du 15335

Initial Truncation

Final Truncation
Initial Truncation

Final Truncation

Legend
- Consonant
- Vowel
60% Identification
Ms. Roussel received her Bachelor of Arts degree in speech pathology from Southeastern Louisiana University in 1974. Following graduation, she was employed as a public school speech therapist in both East Baton Rouge and Tangipahoa Parish school systems. She received her Master of Arts degree from Louisiana State University in 1986 in communication disorders and holds the ASHA Certificate of Clinical Competence in speech/language pathology. Ms. Roussel currently holds the position of assistant professor in the Department of Communicative Disorders, University of Southwestern Louisiana. She resides in Lafayette, La. with her three sons.
DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Nancye C. Roussel

Major Field: Communication Disorders

Title of Dissertation: Perceptual Evaluation of Infant Articulatory Transitions

Approved:

[Signature]
Major Professor and Chairman

[Signature]
Dean of the Graduate School

EXAMINING COMMITTEE:

[Signature]
[Name]

[Signature]
[Name]

[Signature]
[Name]

[Signature]
[Name]

Date of Examination: 7/9/92