

Acoustical study of the development of stop consonants in children

by

Annika Karin Karlsson Imbrie

B.S., Cornell University (1999)

S.M., Massachusetts Institute of Technology (2002)

Submitted to the Harvard-MIT Division of Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2005

© Massachusetts Institute of Technology 2005. All rights reserved.

Signature of Author _____
Department of Health Sciences and Technology
May 12, 2005

Certified by _____
Kenneth Noble Stevens, Sc.D.
Clarence J. LeBel Professor of Electrical Engineering
Professor of Health Sciences and Technology
Thesis Supervisor

Accepted by _____
Martha L. Gray, Ph.D.
Edward Hood Taplin Professor of Medical and Electrical Engineering
Director, Harvard-MIT Division of Health Sciences and Technology

This page is intentionally left blank.

Acoustical study of the development of stop consonants in children

by

Annika Karin Karlsson Imbrie

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on April 12, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This study focuses on the acoustic patterns of stop consonants and adjacent vowels as they develop in young children (ages 2;6-3;3) over a six month period. Speech is generated using a series of articulatory, laryngeal, and respiratory gestures that children must learn to reproduce. As a child's speech develops, the gestures become more precise and coordinated, and the resulting acoustic patterns are refined. To explore their development, over forty different acoustic measurements were made on each of 1049 recorded utterances from ten children, including durational, amplitude, spectral, formant, and harmonic measurements. These acoustic data are interpreted in terms of the supraglottal, laryngeal, and respiratory actions that give rise to them. Data show that some details of the child's gestures are still far from achieving the adult pattern. Children have acquired appropriate positioning of their primary articulator for producing a stop consonant, but are still learning to adjust the tongue body during the consonant production. At constriction release, children have a high incidence of multiple bursts and a short burst duration, interpreted as a reflection of increased articulator compliance, smaller articulator size, and high subglottal pressure. Children are also still acquiring correct adjustment of vocal fold stiffness and glottal spreading as well as intraoral pressure, as evidenced by long voice onset times and highly variable fundamental frequencies. Additionally, amplitude changes over the course of the utterance and high amplitude variability reveal that children have not yet gained full control over subglottal pressure. Overall, results indicate that children are less consistent than adults in controlling and coordinating various gestures and with finding the ideal respiration and vocal tract postures, including the stiffness of their articulators. Certain aspects of child speech are found to become more similar to adult values over the six month period of the study.

Thesis Supervisor: Kenneth N. Stevens, Sc.D.

Title: Clarence J. LeBel Professor of Electrical Engineering and Professor of Health
Science and Technology

This page is intentionally left blank.

Acknowledgements

I would like to express my gratitude and appreciation to my thesis advisor, Professor Ken Stevens. It has been an honor and inspiration to work with him over the past six years. His insight and advice have been invaluable in the course of this research, and his genuine interest and excitement have been a great motivation. I would also like to thank my committee members, Ralph Ohde, Joe Perkell, and Jesse Snedeker for their guidance and helpful suggestions, and for the time they spent reading several drafts of this thesis. I am deeply appreciative of the Harvard-MIT HST/SHBT program and the collaboration with the MGH Institute of Health Professions, which has allowed me to pursue simultaneous interests in speech science and technology and communication disorders, and I would like to thank Bob Hillman for making this all possible.

A special thank you to Helen Hanson for sharing her technical knowledge and Ariel Salomon for his computer support, particularly with Linux. Thank you also to Arlene Wint for her continued help and support over the years with many administrative matters. I would also like to thank all the members of the Speech Communication Group, past and present, who have helped create such a positive and inspiring laboratory environment, including Lan Chen, Marilyn Chen, Steven Lulich, Sharon Manuel, Nicole Marrone, Xiaomin Mou, Tony Okobi, Stefanie Shattuck-Hufnagel, Janet Slifka, Ellen Stockmann, Jennell Vick, Virgilio Villacorta, and Majid Zandipour. Thank you also to all the subjects and their parents for their participation in the study, and their commitment to returning month after month for repeated recordings.

To my parents, whose commitment to education throughout the years inspired me to pursue this degree, and who have been very supportive throughout my graduate years. I would also like to thank Gary and Kris Imbrie for their support.

My final, and biggest, thank you is to my husband, Greg, for his unwavering support and encouragement, as well as his belief in me.

This work has been supported in part by NIH/NIDCD and HST.

This page is intentionally left blank.

Biographical Note

Annika Karin Karlsson Imbrie was born in Providence, Rhode Island on November 22, 1977. Her parents are Bengt Ingvar Karlsson and Kathryn Anne Karlsson. She spent the majority of her childhood in Providence, and graduated from Lincoln School in 1995. She received a Bachelor of Science in Computer Science from Cornell University in Ithaca, NY, in May 1999. The following September, she entered the Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program. She married Gregory Adair Imbrie in March, 2000. Within the SHBT Program, Annika chose to concentrate in communication disorders, and completed the academic qualifications for Speech-Language Pathology certification at the Massachusetts General Hospital Institute of Health Professions. In June 2002, she completed her Masters of Science in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology. She received a Doctor of Philosophy degree in Health Sciences and Technology in June 2005, from the Harvard University – Massachusetts Institute of Technology Division of Health Sciences and Technology Speech and Hearing Biosciences and Technology (HST / SHBT) Program.

This page is intentionally left blank.

Table of Contents

List of Figures.....	11
List of Tables.....	13
1. Introduction.....	15
2. Stop Consonant Production.....	17
2.1. Adult Stop Consonant Production	17
2.2. Background on Coarticulation	21
3. Literature Review.....	22
3.1. Respiration	23
3.2. Phonation	24
3.3. Articulation and Coarticulation.....	26
3.4. Variability	29
4. Specific Aims.....	31
5. Research Questions.....	35
6. Methods.....	42
6.1. Children’s Speech Experiment	42
6.2. Adult Speech Experiment	44
7. Measurements	45
8. Statistical Analyses and Results.....	48
8.1. Introduction.....	48
8.2. Respiration	48
8.2.1 Hypothesis 1: Intensity variation	48
8.2.2. Hypothesis 2: Burst duration and intensity.....	51
8.2.3. Hypothesis 3: Multiple bursts and intensity.....	56
8.2.4. Hypothesis 4: VOT and intensity.....	57

8.2.5.	Hypothesis 5: First formant and intensity	58
8.2.6.	Summary	59
8.3.	Phonation	60
8.3.1.	Hypothesis 6: Voiced stops – stop release, voicing onset, spectral tilt.....	60
8.3.2.	Hypothesis 7: Voiceless stops – fundamental frequency, spectral tilt.....	66
8.3.3.	Hypothesis 8: Fundamental frequency, intensity, VOT, spectral tilt.....	69
8.3.4.	Summary	71
8.4.	Articulation	72
8.4.1.	Hypothesis 9: Velar pinch.....	72
8.4.2.	Hypothesis 10: Tongue height and jaw movement.....	73
8.4.3.	Hypothesis 11: Back-front tongue body movement	75
8.4.4.	Hypothesis 12: Burst spectrum peaks: articulator placement	78
8.4.5.	Hypothesis 13: Primary articulator placement.....	82
8.4.6.	Hypothesis 14: Multiple bursts	83
8.4.7.	Summary	90
8.5.	Summary of Results.....	91
9.	Individual Differences	93
9.1.	Amplitude changes: Respiratory control	95
9.2.	Voiced stops – VOT: Laryngeal control.....	99
9.3.	Voiceless stops – F0 and VOT: Laryngeal and respiratory control.....	102
9.4.	Multiple Bursts: Articulatory properties and control; respiratory control.....	105
9.5.	Burst Duration: Articulatory properties and control; respiratory control.....	107
9.6.	F2 Transitions: Articulatory posture.....	110
9.7.	Discussion.....	113
10.	Summary and Conclusions	114
11.	Appendix.....	118
11.1.	Appendix 1: Target word list.....	118
11.2.	Appendix 2: Statistical Analyses Tables.....	119
11.3.	Appendix 3: Spectrograms.....	130
11.4.	Appendix 4: Data review for each child subject.....	132
12.	Bibliography	142

List of Figures

Figure 1. Sample schematized spectrogram	32
Figure 2. Average burst durations for single burst utterances	52
Figure 3. Scatterplot of burst duration vs. normalized burst amplitude	55
Figure 4. Scatterplot of F1 vs. amplitude at vowel center across all 10 children	58
Figure 5. Spectrogram and corresponding waveform showing a child’s production of the word “Gus” and VOT-lag	62
Figure 6. VOT for each voiced stop, broken down to show burst duration and lag before voice onset	63
Figure 7. Spectrogram and spectra of a child’s production of the word “puppy,” showing weak F3 values	68
Figure 8. F2 density plots for children and adults	79
Figure 9. Average burst spectrum prominence	81
Figure 10. Spectrogram and waveform of the word “Gus” showing multiple bursts.	85
Figure 11. Spectrogram and waveform of the word “tub” showing multiple bursts.	86
Figure 12. Average number of bursts for each place of articulation	88
Figure 13. Average amplitude change (burst amplitude – vowel center amplitude) for /k/- and /g/-initial words for child and adult groups	95
Figure 14. Amplitude change (voice onset amplitude – vowel center amplitude) for /k/- and /g/-initial words for child and adult groups	96
Figure 15. Variations in amplitude change for /k/- and /g/-initial utterances, and at vowel center	98
Figure 16. VOT for each voiced stop, broken down to show burst duration and lag before voice onset, for three individual subjects	101
Figure 17. Change in F0 over the first 20 ms of voicing following a voiceless stop for subjects C1, C5, and C7, and average child and adult results	103

Figure 18. VOT averages for the child group at each time point, compared with adult group averages	104
Figure 19. VOT variability averages for the child group at each time point, compared with adult group average	104
Figure 20. Average number of bursts across both voiced and voiceless stops for each place of articulation.	106
Figure 21. Overall average burst duration for single burst utterances	108
Figure 22. Burst duration for single burst utterances	109
Figure 23. F2 transition ratio (F2 (burst) / F2 (vowel center))	110
Figure 24. Average F2 at burst and vowel center for each place of articulation	112
Figure 25. Five example spectrograms of the word “bug” with a long VOT lag.	130
Figure 26. Five example spectrograms of the word “bug” with a short VOT lag.	131

List of Tables

Table 1. Overview of interpretations of acoustic data	33
Table 2. Total number of stop consonant initial target words collected, and the number of words analyzed for each stop consonant for subjects C1, C5, and C7	93
Table 3. Voiced stop VOT average and VOT-burst duration average for each place of articulation for subjects C1, C5, and C7 at Time 1	99
Table 4. Number of utterances with multiple bursts out of the total number of utterances analyzed for each stop place of articulation for subjects C1, C5, and C7 at the first recording session and six months later	105
Table 5. Amplitude differences (in dB) from voice onset to vowel center	119
Table 6. Average standard deviation in amplitude (dB) at vowel center	120
Table 7. Average values of VOT and difference between VOT and burst duration for voiced stops for each place of articulation	121
Table 8. Spectral tilt measurements (A1-A3) measured at VOT and 20 ms later, averaged across all tokens for each subject	122
Table 9. A1-A3 values at VOT and at VOT+10 ms averaged across all voiceless stops for each subject	123
Table 10. F2 and F3 in the burst and at vowel center averaged across all words with word-initial velar stop consonants, in child and adult subjects	124
Table 11. Average F1 and F3 values at vowel center for each subject, as well as the average F1/F3 ratio	125
Table 12. F2 at burst, in the following vowel, and the ratio of F2(burst) / F2(vowel) for all three stops	126
Table 13. Average F2 values for each stop place of articulation for each subject	127
Table 14. Average frequency of the spectral prominence in alveolar and velar stop bursts, and standard deviations of these averages for each child subject.	128
Table 15. Average frequency of the spectral prominence in alveolar and velar stop bursts, and standard deviations of these averages for each adult subject.	128
Table 16. Average number of bursts for by each child for each place of articulation.	129

This page is intentionally left blank.

1. Introduction

Children first begin producing stop consonant-like sounds during babbling in their first year of life. Over the next few years, these babbled sounds evolve into recognizable bilabial, alveolar, and velar stop consonants. Although these productions are perceived as acceptable stops by age three in most children, it is likely that they are still developing—that spatial and temporal modifications in the articulatory pattern are being made, as the production continues to approach the adult target.

This idea is supported by current research, which indicates that children’s speech production is much more variable than adults (Nittrouer, 1993; Smith and Kenney, 1994; Lee, Potamianos and Narayanan, 1999; Assman and Katz, 2000), possibly because they are still constantly adjusting their articulators, attempting to fine-tune articulatory placement. Nittrouer (1993) proposes that this variability is related to immaturity of articulatory neuromotor control. It has been suggested that the inconsistency in children’s productions arises from lack of temporal coordination (Ferguson and Farwell, 1975). Additionally, evidence from a pilot study indicated that children at this age were still developing increased coordination of articulation, phonation, and respiration.

The purpose of this study is to examine in more depth the nuances of the development of stop consonants over a 6-month period beginning from age 2;6 to 3;3. At this stage in their speech development, children are fairly consistent in using the same phonemes across multiple repetitions of the same word. However, although some basic articulatory movements are used to produce the stops, these movements are somewhat primitive and variable, and they lack some of the detailed attributes normally observed in adult speech. There are several aspects of stop consonant production that need refinement and that lack consistency, which can be revealed by acoustic measurements of

the speech. This study constitutes a detailed examination of stop consonant production in 2-3-year-old children and how this production develops over a six month period, and presents hypotheses about the developmental factors that could give rise to the observed modifications.

2. Stop Consonant Production

2.1. Adult Stop Consonant Production

Adult stop consonant production requires a complex sequence of articulatory and laryngeal movements, as reviewed this section. The production of stop consonants in a CV sequence is described in terms of the various motor actions required, along with their acoustic consequences. These are the gestures that a child is expected to acquire during the first few years of life.

- 1) As the CV syllable is initiated, a closure is created in the vocal tract by the lips, tongue blade, or tongue body. The articulator that creates the closure is called the primary articulator. The soft palate is raised, closing the velopharyngeal port. This results in an increase in the intraoral pressure, as air from the lungs passes into the closed vocal tract. The rate of intraoral pressure increase will depend on the laryngeal configuration (i.e., the positioning and stiffness of the vocal folds), as well as subglottal pressure.
- 2) The consonant closure is formed by the primary articulator with a force sufficient to maintain closure despite the increasing intraoral pressure. This closing gesture results in a distortion or flattening of the articulator surface against the opposing surface (i.e. the upper lip or the hard or soft palate).
- 3) During the stop closure interval for labial, alveolar, and velar consonants, articulators other than the primary articulator in the vocal tract can be repositioned

in preparation for the upcoming vowel. Depending on the consonant, there are limitations on the extent of anticipatory movements of the tongue body, tongue blade, or lips. For the velar consonant, the tongue body is the primary articulator, but the front-back position of the tongue, and the resulting location of the constriction, can be adjusted in anticipation of the upcoming vowel.

- 4) During the stop consonant closure interval, the glottal configuration, vocal fold stiffness, and pharyngeal wall stiffness, are adjusted depending on whether the stop consonant is voiced or voiceless. For a voiceless stop, the glottis is spread, and the vocal folds and vocal tract wall are stiffened. If the stop consonant is voiced, the glottis is not spread, and there is slackening of the vocal folds and vocal tract walls. There is also possibly an active expansion of the vocal tract volume, including a lowering of the larynx. Depending on whether the consonant is voiceless or voiced, these gestures result in either inhibition or facilitation of vocal fold vibration during the closure interval and the following consonant release.

- 5) Following the consonant closure interval is the stop consonant release. The consonant release occurs as the primary articulator is displaced in a ventral direction. For the first 10-20 ms following the release, the ventral movement of the articulator is influenced by forces from the heightened intraoral air pressure. The length of time that the intraoral air pressure influences the articulator movement depends on the articulator: the time is shortest for labial releases and longest for velar releases. During this 10-20 ms interval, the cross-sectional area of the constriction increases at a slower rate than it does otherwise throughout the release, most likely due to the Bernoulli Effect. A burst of noise is created in this interval, and this noise burst acts as a sound source that is then filtered by the vocal tract -- primarily by the section of the vocal tract downstream from the constriction.

- 6) Following the release, in the next few tens of milliseconds, the articulators move towards their target configurations for the upcoming vowel. This includes:
 - movement of the tongue body, pharyngeal walls, and mandible, causing movements of F1, F2, and possibly F3.
 - change in the laryngeal configuration, for voiceless stops, resulting in aspiration noise and breathy voicing, as well as delayed onset of glottal vibration, and spectral changes after vibration begins.
 - changes in vocal fold stiffness, causing modification of the fundamental frequency.

- 7) For utterance-initial CV syllables following an inhalation, lung pressure may continue rising following consonant release until it reaches a maximum during the vowel.

- 8) For utterance-internal CV syllables with no pause preceding the consonant, the subglottal pressure may remain fairly constant throughout the syllable.

In summary, production of a CV syllable involves various articulatory, laryngeal, and respiratory actions, the majority of which are represented in the acoustic signal. Some actions (e.g. choosing which articulator to use for a given speech sound) are fundamental to distinguishing place of articulation, or to distinguishing a voicing contrast. Among these various gestures, some might be considered as primary gestures that produce a distinctive sound attribute that distinguishes one place of articulation from another or that distinguishes a voiced from a voiceless consonant. Other gestures may be added to enhance the distinction between place of articulation or voicing for different consonants. Acoustic analysis of a child's speech can help to document how the child is producing these actions. Does the child initially implement one principal correlate of a feature and later refine secondary correlates of the feature? For example, in implementing an alveolar stop consonant, does a child initially position the tongue blade to produce an approximation to the proper spectrum and later refine secondary aspects of the consonant, such as timing of the tongue blade release, and configuring and timing the tongue body

movement to produce formant transitions that help to enhance the distinction between an alveolar stop and labial and velar stops? Questions such as these will guide the interpretation of the acoustic patterns produced by the children.

2.2. Background on Coarticulation

Coarticulation is the effect one speech sound has on another, and is caused by an overlap in articulator gestures. Coarticulation refers to the acoustic consequence of the gestural overlap, and thus some researchers (e.g. Goodell and Studdert-Kennedy, 1993; Nittrouer, Studdert-Kennedy, et al., 1996) prefer to discuss the gestural overlap itself, rather than coarticulation. Stop consonant production is often influenced by adjacent vowels, mainly due to the tongue body moving in preparation for the vowel. The amount of tongue body movement (and hence coarticulation) is different for each stop consonant, since different articulators form each constriction. The labial stop is formed with the lips, so the tongue body has substantial freedom to move around inside the mouth in preparation for the upcoming vowel. Therefore, in adults, labial stops tend to display the greatest amount of coarticulation. The alveolar stop is formed by raising the tongue blade to the alveolar ridge. While there is no set position for tongue body location in order to make an intelligible alveolar stop, it is likely that speakers learn to adjust the tongue body to help differentiate alveolar stops from other stop places of articulation, as noted above. Most often, alveolar stops in English are produced by adults with the tongue body in a forward position, likely an enhancing gesture, which creates more distinct formant transitions but prevents a high degree of coarticulation. Velar stops are made with the tongue body itself. The location of the constriction can be affected by the following vowel: before a front vowel, the constriction is often located more anteriorly, whereas before a back vowel, it has a more posterior location (Stevens, 1998). Additionally, as discussed above (see section 2.1, p. 17), the laryngeal configuration for the stop may affect the following vowel (e.g. resulting in breathy voicing following a voiceless stop).

3. Literature Review

Children typically develop the stop consonants within the first few years of life. According to Locke (1983), at 12 months the infant's babble repertoire contains all 6 stops (/b p d t g k/), while Leonard, Newhoff and Mesalam (1980) found all three voiced stops (/b d g/) dominant in children from 1;4-1;10, during the first 50-word stage. Prather and Hedrick (1975) found that 75% of children had developed /p d g k/ by age 2;4 and /b t/ by age 2;8. This agrees with data reported by Smit, Hand, Freilinger, Bernthal et al. (1990) showing that 76-99% (depending on the consonant and place of articulation) of children had developed all six stop consonants by age 3;0.

Studies determining when different phonemes are acquired are based on auditory-perceptual data, using transcription of children's speech to determine which sounds they are producing (e.g. Smit, Hand et al., 1990). Several researchers have suggested that children are still developing and perfecting the consonant gestures long after the consonants themselves are understood by listeners (Nittrouer, 1993). This theory is supported in studies by Kuijpers (1993) and Nittrouer, Studdert-Kennedy and Neely (1996).

There are two different theories of speech development: 1) children first develop speech sound segments and then learn to sequence increasing numbers of them (Prather and Hedrick, 1975; Kent, 1983), 2) children initially develop approximate gestural patterns for syllables or words, which become more precise and coordinated as the child's speech develops (Menn, 1983; Ferguson, 1986; Studdert-Kennedy, 1987; Nittrouer, Studdert-Kennedy and McGowan, 1989; Studdert-Kennedy, 1990; Nittrouer, 1995). Several studies have examined acoustic properties of consonant-vowel coarticulation (see p. 21 for a description of coarticulation) as a function of age, in order to help determine which of these theories holds true. If children coarticulate less than adults, it has been suggested that this indicates that coarticulation is a process that is acquired after the

development of the involved speech sounds, the segments of which are acquired separately, supporting the first theory. If children coarticulate more than adults, it would indicate that they are developing an approximate gestural pattern for an entire syllable or word, which is refined as they develop, supporting the second theory. Coarticulation studies have mostly used the speech of 3-7-year-old children, primarily focusing on the second formant frequency, comparing its changes from vowel to consonant to vowel in children vs. adults. Some researchers have concluded that there is increased coarticulation in children (Nittrouer, Studdert-Kennedy et al., 1989; Nittrouer, 1993; Siren and Wilcox, 1995), while others have concluded that there is decreased coarticulation (Kent, 1983; Sereno and Lieberman, 1987). Still others have concluded that there is no change in coarticulation with age (Turnbaugh, Hoffman, Daniloff and Ashber, 1985; Sereno, Baum, Marean and Lieberman, 1987; Katz, Kripke and Tallal, 1991), while Goodell and Studdert-Kennedy (1993), Sussman, Duder, Dalston and Cacciatore (1999), Gibson and Ohde (2003) found different coarticulation patterns depending on the stop consonant.

3.1. Respiration

A study of laryngeal and respiratory measurements during speech in 4-year-old and 8-year-old children (Stathopoulos and Sapienza, 1993) found that children generate higher tracheal pressures during speech than adults. Additionally, the open quotient in children, while similar to women at a comfortable intensity level, did not decrease as intensity increased, as occurs in adults, possibly due to immaturity of fine-tuned control of the larynx. Furthermore, the fundamental frequency increased in children as vocal intensity increased. Also, the children used a higher percentage of their vital capacity and had greater rib-cage displacement, despite their smaller thorax, as compared to adults. Overall, Stathopoulos and Sapienza (1993) concluded that the adult speech mechanism model should not be used for describing children's speech, and that age-specific models are needed.

A similar study on development of laryngeal and respiratory function for speech production found that preschool children (ages 3;3-4;3) have higher subglottal pressure, higher resistance, and less airflow through the glottis during production of /pi/ and /pa/ syllables than adults (Netsell, Lotz, Peters and Schulte, 1994). The researchers developed a hypothetical function that predicted that while adults balance inspiratory and expiratory muscle forces for speech breathing, preschool children use mostly expiratory muscle force. Thus, the children must do more expiratory work during speech breathing.

3.2. Phonation

Few studies have examined vocal fundamental frequency (F0) onset behavior in children. Ohde (1985) found that both 8- to 9-year-old children and adult males produced a high F0 at voice onset that significantly decreased over the following glottal periods. For adult male productions, Ohde (1984) found that F0 contrasted voiced from voiceless aspirated stop productions at voice onset. This differentiation did not occur at voice onset in the children (Ohde, 1985), but was present during subsequent vocal cycles. Additionally, F0 was highly variable in 8- to 9-year-old children compared with adults.

In a study on younger children, Robb and Saxman (1985) found that F0 was not significantly higher at voice onset than overall average F0 in 11- to 25-month-old children. Robb, Saxman and Grant (1989) suggest a relative uniformity of vocal F0 across preword and meaningful speech in early vocalizations.

More recently, Robb and Smith (2002) studies F0 changes following voiceless obstruents in 4-year-olds, 8-year-olds, and adults. While all subjects exhibited a high F0 at voice onset immediately following a voiceless obstruent, which subsequently decreased, adults showed a significantly higher F0 at voicing onset than 4-year-olds, with F0 values adjusted for age differences. Additionally, they found that adult females demonstrated a higher F0 at voice onset, and greater F0 change, than males. This study did not include voiced stop consonants.

Many studies have focused on voice onset time (VOT) as a measure of characterizing stop consonant voicing in children, with inconsistent results. While some have reported short mean VOT values in children ages 1;6-4;6 (Kewley-Port and Preston, 1974; Macken and Barton, 1980), others have found relatively long VOT values in children under 4-years-old (Menyuk and Klatt, 1975; Smith, 1978). Ohde (1985) found that the primary distinction of VOT for voiced versus voiceless stop consonants was strongly developed in 8- to 9-year-old children. One consistent finding across these studies is high variability in VOT for voiceless aspirated voiceless stops.

A recent study by Koenig (2000) compared the voiceless aspirated stops /p,t/ to /h/ as produced by 5-year-olds and adults. This study included data on oral airflow, intraoral pressure, and acoustic measurements. Through comparisons of the voicing characteristics of the voiceless stop consonants with the continuant consonant /h/, which is produced with a fairly open vocal tract and little upper articulator movement, inferences were made about laryngeal movements. Koenig found a high variability in VOT for /h/ production in 5-year-olds, despite the absence of upper articulator movements, and determined that the children still had not acquired adult-like control over the larynx. Koenig concluded that VOT variability for stop consonants is attributable not just to interarticulator timing control, but also to purely laryngeal factors, such as abduction degree and vocal fold tension.

In a study of the voiced/voiceless distinction in stop consonant production, Kuijpers (1993) found that there was a significant difference in “degree of voicing” of voiced plosives between four- and six-year old Dutch-speaking children in word-medial plosives spoken during elicited spontaneous speech. The majority of four-year-olds’ productions in his study contained continuous voicing throughout the plosive and no noise burst, while six-year-old productions were most often realized as a voiced period followed by a short period of silence and then a noise burst. In studying the differentiation of the voiced/voiceless pairs, Kuijpers also found that the closure duration of the velar stop /k/ was shorter in the 4-year-old children than in the 6-year-olds, while /p/ and /t/ durations were similar in both ages. The author suggested that this could be due to the fact that /k/ has no voiced cognate in Dutch, and thus the children were not attempting to contrast /k/ from /g/ as they are /p,t/ from /b,d/.

3.3. Articulation and Coarticulation

Several researchers have studied stop consonant-vowel coarticulation in intrasyllabic contexts in children and adults (Turnbaugh, Hoffman et al., 1985; Repp, 1986; Sereno and Lieberman, 1987; Hodge, 1989; Goodell and Studdert-Kennedy, 1993; Nittrouer, 1993; Nittrouer, 1995; Sussman, Minifie, Buder, Stoel-Gammon et al., 1996; Sussman, Duder et al., 1999). However, many of these studies have contradictory results. Hodge (1989) (as cited in Goodell and Studdert-Kennedy (1993)) looked at repetition of CV syllables with an initial voiced alveolar stop consonant in 3-, 5-, 9-year-olds and adults. She examined F2 onset values, and found that variation with following vowel context was greatest in the youngest age group, indicating greater coarticulation in children. The study by Turnbaugh, Hoffman et al. (1985) used speech by 3½-year-olds, 5-year-olds and adults and inferred stop-vowel coarticulation from second formant frequency measurements in CVC syllables. In contrast to Hodge (1989), this study found no significant effect of age on coarticulation.

Nittrouer (1993) performed a study of stop consonant production using older children (3½-, 5-, and 7-year-olds) and compared the results to productions by adults. Results indicated longer stop closures in children than adults, and in 5- vs. 7-year-olds, and also revealed a significantly longer duration of voice onset time (VOT) for /k/ than for /t/ in 3- and 5-year-olds. Additionally, she found that there is a vowel effect on VOT for children but not adults, and the vowel gesture gains prominence earlier in the children's syllables than in the adults. It was found that children's gestures have a similar shape to those of adults, but they are produced more slowly, and with greater temporal variability. Due to the finding that the first formant frequency, which reflects tongue height and can be controlled by jaw movements, does not display increased variability in different phonetic contexts with age, Nittrouer concluded that gestural patterns are acquired at different rates depending on the articulator involved: jaw movements for articulation are acquired before tongue movements. The results of the study also indicate that there is increased coarticulation in children relative to adults.

In another study, Nittrouer (1995) looked at spectral moments of the burst and following vowel in alveolar and velar stop-vowel syllables in 3½-, 5-, and 7-year-olds and adults. She analyzed the first 20 ms after stop consonant constriction release, and found no differences between children and adults in the first spectral moments of /t/ and /k/ productions. However, she did find that for the velar stop consonant /k/ (but not /t/), children's constriction location was more greatly affected by the upcoming vowel production than adults. Thus, results indicated increased coarticulation in children for the velar stop consonant, but not for the alveolar stop.

Nittrouer, Studdert-Kennedy et al. (1996) looked at fricative-vowel syllables spoken in the carrier phrase "It's a ___ Bob" in 3½-, 5-, and 7-year-olds and adults. This study replicated the results of (Nittrouer, Studdert-Kennedy et al., 1989), finding that differentiation of /ʃ/ and /s/ increased with age, and coarticulation of the fricative with the following vowel decreased with age. Furthermore, study results indicated that children develop intersyllabic gestural organization earlier than intrasyllabic gestures—that is, the children refine articulatory gestures between syllables before refining articulatory patterns within syllables. The researchers concluded that "learning to coordinate articulatory gestures as required to impose adult-like phonemic structure on utterances is a long process, extending well into childhood" (p. 388).

Sereno, Baum et al., (1987) studied anticipatory coarticulation of /t/, /d/, and /s/ in CV syllables with two vowel environments (/i/, /u/) in 8 children ages 3-7 years, and 4 adults. Although they found an acoustic effect of coarticulation (lip-rounding) in both adults and children, perceptual experiments indicated that for the children's productions, listeners were often unable to make use of the coarticulatory cues of lower frequency spectral energy peaks during consonants preceding /u/ caused by anticipatory lip rounding. The researchers speculated that the acoustic parameters investigated in their study may not be those that listeners use in the perception of coarticulation. They found that children display less coarticulation than adults, and the perceptual data indicated that children show much greater coarticulation for /t/ than for /d/ or /s/, leading to their suggestion that anticipatory coarticulation may generalize to other consonants with age.

Sussman, Duder et al. (1999) used the locus equation, a straight-line regression fit to data points formed by plotting frequency onsets (Hz) of second formant frequency (F2)

transitions vs. F2 at midvowel, to define the amount of coarticulation. They followed a single female child from age 7 months to age 40 months, and made monthly recordings of her productions for the duration of the study. They used locus equations to measure the degree of CV coarticulation as the child progressed from babbling to words. This study found that the three voiced stop consonants /b,d,g/ each followed a distinct pattern of articulation as they progressed towards adult norms. The locus equations of bilabial CVs initially had very low slopes, indicating minimal coarticulation, but during months 10-13, these slopes increased sharply toward adult values. The alveolar CVs initially displayed high slopes, indicating a high degree of coarticulation, which quickly decreased between months 7-12, moving toward adult values. The velar CV productions initially displayed extreme fluctuations in degree of coarticulation, which converged toward a relatively constant value slightly above the adult target around 18 months of age. These differences are likely due to the different articulatory sequences required for each stop (See Background on Coarticulation, p. 21). The researchers concluded that coarticulation patterns are dependent on the consonant involved.

In another study of stop-vowel coarticulation, Goodell and Studdert-Kennedy (1993) found similar results. They recorded six 22-month-old females imitating nonsense syllables and then rerecorded them again at 32 months of age. They found that for lingual stops, both at stop closure and stop release, there was greater coarticulation between /g/ and a stressed vowel than between /d/ and a stressed vowel at both ages, but the amount of coarticulation decreased with age. For the labial stop, coarticulation decreased with age at closure and increased with age at release, indicating a growing capacity to differentiate successive gestures. Thus, whether coarticulation increased or decreased with age depended on the stop place of articulation, and the landmark studied (e.g. closure vs. release).

Gibson and Ohde (2003) performed a similar study, looking at coarticulation of stop consonant-vowel productions in ten children, ages 17-21 months. Their findings indicated that the degree of coarticulation depended on place of articulation. Using the locus equation metric, as described above, they found that the voiced velar stop productions had consistently more coarticulation than either the labial or alveolar stop syllables.

3.4. Variability

Several researchers have reported on variability of speech productions by children as compared to adult productions. Sussman, Minifie et al. (1996) found that an infant had greater production variability with respect to constriction location and vowel formant frequencies for each stop consonant at 21 months than at 12 months. They speculated that this could be due to the precise articulatory targets necessary for intelligible production of words, as opposed to meaningless babble. Lee, Potamianos et al. (1999) performed a large-scale cross-sectional acoustic study of the development in speech, recording 436 children, ages 5 to 17 year, and 56 adults. With increased age, the study found a trend of decreased magnitude and within-subject variability of segmental durations, fundamental and formant frequencies, and spectral envelope variability. The researchers speculate that the reduction in segmental durations and within-subject duration variability could be due to improvements in neuromuscular control of the articulators, resulting in increased speed and timing control of speech gestures. Furthermore, they suggest that there is a reduction in variability of gestural configurations for any given sound, causing the reduced within-subject and spectral variability.

In a smaller, longitudinal study of variability, Smith, Kenney and Hussain (1996) recorded a group of 12 children between the ages of 7 and 11 producing multiple repetitions of two different words (“saucer” and “sissy”), and then rerecorded them on the same stimuli 1½ years later. The researchers measured temporal variability as well as segmental durations of different portions of the stimuli. They found that although there was a general trend of decreasing duration and variability, several children displayed little or no decrease in at least one of these measures over the 1½-year period of the study. Furthermore, the younger children did not necessarily show a greater decrease in the measured parameters than the older subjects. This supports the idea that each child develops at a different rate, reinforcing the need for longitudinal studies of development.

Although many of the acoustic studies of children's speech have used cross-sectional designs, there have been several longitudinal studies of child speech, including Robb and Saxman (1990), Goodell and Studdert-Kennedy (1993), Snow (1994), Smith, Kenney, and Hussain (1996), Sussman, Minifie et al. (1996), Sussman, Duder et al. (1999). All of these studies have added insight into results of cross-sectional designs, as well as into the development of different speech parameters in individual children. The longitudinal design of this study allows us to look at patterns of development of speech over time, rather than assuming that all of the children are at the same stage of development at the same time.

Another observation is that most studies have used imitated and/or nonsense syllables. Siren and Wilcox (1995) found that anticipatory coarticulation is greater in production of nonmeaningful stimuli than meaningful words, in individuals of all ages. Thus, this study focuses on elicited productions of real words that are already in the vocabulary of typically developing 2-3-year-olds.

Overall, recent research indicates that children continue to develop segmental organization, decreasing the domain from syllable size segments to smaller speech sound units, throughout the first 3-5 years of life (Nittrouer, Studdert-Kennedy et al., 1989; Goodell and Studdert-Kennedy, 1993). This study takes a more in-depth look at the specific developments and changes that are taking place in children during their third year of life.

4. Specific Aims

Despite the large number of acoustic studies of children's speech, a limited range of acoustic measurements have been made, including mainly first and second formant frequency measurements (Turnbaugh, Hoffman et al., 1985; Hodge, 1989; Nittrouer, 1993; Nittrouer, Studdert-Kennedy et al., 1996; Sussman, Duder et al., 1999) and durations of various speech segments (Kewley-Port and Preston, 1974; Macken and Barton, 1979; Nittrouer, 1993), and the data have been used primarily to analyze the presence and extent of coarticulation. Most of these studies are aimed at supporting one of the two theories of speech development (see p. 22), yet they focus on only a small subset of the speech gestures by basing conclusions on a few acoustic measurements. None of these studies have provided a thorough examination of the development of the speech system as a whole, including refinement and coordination of the motor actions for articulation, phonation, and respiration, as is proposed in this study. This investigation of speech development, with the main goal of using acoustic measurements to provide an in depth examination of motor speech production in 2-3-year-old children and its development over a six month period, may provide additional insight into the validity these theories.

The study involves in depth acoustic analysis of the utterances and an interpretation of the data in terms of articulation, phonation, and respiration. Measurements have been made in the vicinity of acoustic landmarks, such as consonant closure and release, as illustrated in Figure 1 (p. 32), and the data are analyzed in terms of their relation to acoustic cues to phonological features. The acoustic measurements include:

1. burst: duration, spectral prominences, amplitude
2. aspiration: duration, spectral prominences, amplitude

3. acoustic landmark labeling and durations of events determined by these landmarks
4. formant frequency measurements and formant movements from and into adjacent vowels
5. fundamental frequency changes
6. vowel amplitude changes.

From these measurements, inferences are made about the primary articulator and its placement and movement, tongue body position and movement, laryngeal configuration, vocal fold stiffness, and respiration, as explained below, and summarized in Table 1 (p. 33).

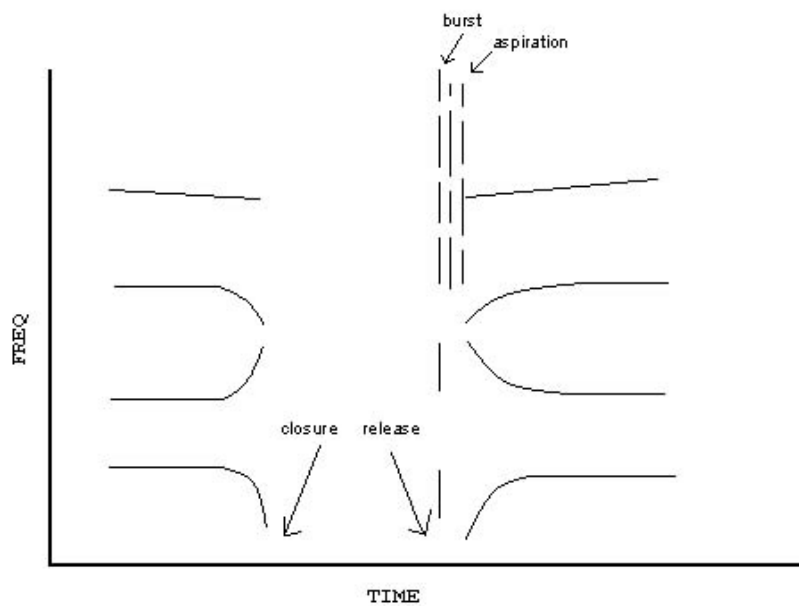


Figure 1. Sample schematized spectrogram showing the closure and release of a consonant, the burst, the aspiration noise, and the formant transitions in the following vowel.

	Articulation		Phonation (Larynx)	Respiration
	Articulator	Tongue body		
Burst prominences	X	X		
Aspiration prominences		X		
Amplitudes				X
Durations	X		X	X
Formant measurements	X	X		
F0			X	

Table 1. Overview of interpretations of acoustic data. The top row shows the categories of speech actions being studied, while the left-most column shows the resulting acoustic measurements. From the acoustic measurements, actions relating to articulation, phonation, and respiration are inferred.

Primary articulator placement can be inferred from frequencies of the spectral prominences of the burst, the time course of the burst, and formant frequency movements before closure and following release of the constriction. For alveolar and velar stop consonants, the frequency of the spectral prominence of the burst reflects the resonance of the cavity anterior to the constriction, from which the location of the constriction can be calculated. Additionally, lower spectral prominences in the burst reflect the configuration of the cavity behind the constriction.

Tongue body position can be estimated from the formant frequencies and their movements. Tongue height can be inferred from the first formant frequency, while the front-back position of the tongue body can be inferred from the second formant frequency.

Vocal fold movement can be inferred in part from the time course of glottal vibration in the consonantal closure. Before the consonant closure, if the glottal opening increases in preparation for an unvoiced consonant, the amplitude of the second harmonic and the amplitude of the first formant peak will decrease, starting 20-30 ms prior to voicing offset. As voicing resumes, the amplitude of second harmonic and first formant

peak amplitude will rise until they reach modal vibration after ~20 ms. Immediately following a voiceless stop, the fundamental frequency tends to be high, likely due to vocal fold stiffness, but the frequency quickly drops back to a typical value (Ohde, 1984).

Control over respiration is evident from amplitude measurements and changes. During adult stop consonant production, respiratory control produces the correct subglottal pressure. For utterance-initial stop consonants, if the subglottal pressure is high at burst release and then falls, the amplitude of the burst relative to the amplitude during the onset of the following vowel will be high.

This study tests the hypothesis that children's stop consonant productions are still developing at the age of 2;6-3;3, and that this development is apparent within a six month period as progressing toward adult speech patterns. Furthermore, the study involves an in depth examination of child motor speech production as inferred from acoustic measures, including respiration, phonation, and articulation. Past studies have indicated that speech in this age range is still very different from adult speech. Given the rate of development in young children, six months should be sufficient time to observe significant changes.

5. Research Questions

This study focuses on the relationship between linguistic representation, acoustics, articulation, and perception of stop consonants by children. Acoustic measurements are made to uncover details of stop consonant development in children ages 2;6-3;3. The details of the acoustic properties are used to infer articulatory movements, and determine if the child is performing the basic gestures correctly and consistently. This includes coordination of respiration, phonation, and articulation in the combination used for the adult target.

In the age range around 2;6, most children appear to have acquired the concept of a stop consonant in contrast to other consonants such as nasals and fricatives. Within the class of stop consonants, these children also distinguish stop consonants produced with three different articulators—lips, tongue blade, and tongue body. For word-initial position, the children usually distinguish between voiced and voiceless stop consonants in their production; the voiceless stops are produced with aspiration in the initial part of the following vowel, and voiced stops have no aspiration. At this stage in their development of speech, although some basic articulatory movements are used to produce voiced and voiceless stops, these movements are somewhat primitive and variable, and they lack some of the detailed attributes normally observed in adult speech. There are several aspects of stop consonant production that need refinement and that lack consistency. Described here are several of these attributes that are expected to show development over the six months of observation of ten children in the proposed research. Some of the expected inadequacies in the children's speech are described, as well as how these inadequacies are estimated based on acoustic measurements of the speech, and how correction of these inadequacies is expected to develop. Additionally, hypotheses are presented on the developmental factors that could give rise to the observed modifications.

Respiration

Information about the development of respiration for speech can be obtained from estimates of changes in subglottal pressure, which can be inferred from the acoustics. In normal adult speech production, the subglottal pressure remains approximately constant throughout the consonant and vowel, and thus the variation in lung pressure throughout an utterance is relatively small, except near phrase boundaries and as influenced by certain other prosodic effects. Children in the 2-3 year age range have not necessarily acquired this relatively constant subglottal pressure within an utterance. Previous studies indicate that children generate higher tracheal pressures than adults for any given intensity output (Stathopoulos and Sapienza, 1993), and that children have higher intraoral air pressures during voiceless stop production (Stathopoulos and Weismer, 1985; Netsell et al., 1994). Acoustic data for short utterances like “a duck” suggest that the subglottal pressure may be high during the initial /d/ and then decreases during the vowel. Thus, it is hypothesized that children will use a higher pressure at the initiation of an utterance (Hypothesis 1). An acoustic consequence is that the intraoral pressure during /d/ closure is relatively high and hence the airflow during the frication release burst may be high. This could lead to a noise burst with an amplitude that is large relative to the amplitude of the following vowel. Another possible consequence is that the time course of the burst may differ from that for a normal adult production. It is hypothesized that a high subglottal pressure at utterance initiation will push the constriction open quickly, reducing burst duration (Hypothesis 2). An additional consequence of a high intraoral pressure is that following the initial constriction release, the pressure will be followed by a negative pressure (Bernoulli’s effect), which will pull the constriction back closed again, resulting in a second burst. Thus, it is hypothesized that a high burst intensity will be evidence of conditions that result in a higher total number of bursts (Hypothesis 3). Another possible consequence of high initial subglottal pressure, is that it will decrease more rapidly following burst release, resulting in a shorter VOT (Hypothesis 4). Additionally, since speaking at a louder volume is correlated with an increase in mouth opening in adults (which results in an increase in F1), it is hypothesized that this also occurs in children (Hypothesis 5).

Acoustic parameters such as duration of burst, its amplitude relative to the vowel amplitude in particular frequency ranges, and the time course of amplitude change in the vowel following the stop consonant can all provide evidence about control of lung pressure during a consonant vowel sequence. It is hypothesized that some children in the proposed study will acquire a more normal pattern of control over lung pressure, and this will be reflected in these and possibly other acoustic parameters. More consistent control of lung pressure could also be reflected in a more stable acoustic pattern with less variability.

Phonation

Development of phonation, including adjustment of laryngeal configuration and stiffness, and vocal tract movement and expansion, will also be examined through measurements of the amount of voicing present during the various stop consonants. The acoustic measurements will include presence and duration of prevoicing, the time course of voicing termination at stop closure and voicing onset at stop release, and fundamental frequency changes at closure and release. These measurements will reflect what the children are doing with their glottis during voiced vs. voiceless stops. For example, during an intervocalic voiced stop consonant, the expansion of the vocal tract must be adjusted and the walls must be slackened to allow voicing to continue throughout the stop. It is hypothesized that for voiced stops children will have difficulty coordinating the slackening of the vocal folds and vocal tract wall, and hence, the onset of voicing with stop release (Hypothesis 6). For voiceless stop consonants, the glottis is spread until the release of the consonant, and then the vocal folds begin to adduct so voicing is initiated. Children are expected to have difficulty consistently coordinating the onset of voicing with the offset of aspiration in voiceless stops, resulting in a different pattern of F0 movement from that seen in adults (Hypothesis 7). Additionally, coordination of the spreading and stiffening the vocal folds may take longer in children's speech.

Due to the effect of respiratory and laryngeal actions on one another, several acoustic measurements are expected to interact. For example, in order to increase

fundamental frequency, the vocal folds must be stiffened. This stiffening results in an increase in the threshold of phonation, and possibly a compensating increase in subglottal pressure. A high initial subglottal pressure might be expected to decrease more rapidly following burst release, resulting in a shorter VOT. Another possibility is that a high initial subglottal pressure pushes the vocal folds apart for the first portion of the release, increasing the delay until voicing onset (Hypothesis 8). Just as a breathy voice results in increased spectral tilt, a louder voice results in less spectral tilt, and thus spectral tilt is expected to decrease with high intensity.

Articulation

Development of articulation of the labial stop consonant will be inferred from acoustic measurements, including the acoustic correlates of tongue body movements. The labial stop consonant is produced by a constriction being formed at the lips. Thus, the vocal-tract resonances are all posterior to the constriction site. In adult productions, as the constriction is formed, all formant frequencies tend to decrease, or at least to remain constant. At the release of the constriction, formant frequency changes depend on the following vowel, since the tongue is already positioned close to the shape for the vowel. If a front vowel follows the labial stop, a sharp increase in the second formant frequency will be seen, since this is the front cavity resonance for the vowel. If the bilabial stop precedes a back vowel, then it is the first formant frequency that rises rapidly at the release of the constriction, since this is usually the front cavity resonance, whereas the back cavity, and hence the second formant frequency, remains essentially unchanged (Stevens, 1998). Are these patterns the same in children's speech, and do the transitions occur at the same rate, or is there a difference in tongue body placement during the stop consonant production, or tongue movement following release of the stop? Or are the formant frequency-cavity affiliations different for children?

Acoustic measures reflecting articulation of the alveolar stop consonant and its development over six months will be examined. Adults produce alveolar stop consonants by raising the tongue blade up to the alveolar ridge to form the constriction, and also adjusting the tongue body to a more fronted position. In this study, the place of

constriction, as inferred from the prominent spectral peak in the burst, will be compared to the position of the tongue body, which is estimated from the formant frequency values at the release of the stop. In adult productions, after the release of the alveolar constriction, the second formant frequency (F2) begins to increase before a front vowel and decrease before a back vowel (Stevens, 1998). Does this also occur in productions by children, and does it occur to the same degree? For the voiceless stops, the aspiration noise will also be examined for changes in prominent spectral peak frequency. Furthermore, in adult productions, it is theorized that there is a division in F2 frequencies between front and back vowels, caused by a subglottal resonance that the formant transitions are less likely to cross over. Is this division evident in children's productions?

Development of articulation of the velar stop consonant will also be examined. The production of velar stops in adults involves creating a constriction that is typically 5-6 cm from the lips. For adults, the constriction is released at a slower rate than either bilabial or alveolar constrictions, and the length of the constriction is greater. As the constriction is formed, the back and front cavity resonances will move towards each other, as reflected in the proximity of the second and third formant frequencies. At the release of the constriction, F2 and F3 are initially very close, and then shift away from each other toward the vowel targets. This velar pinch is caused by the tongue body moving up towards (and then away from) the palate and changing the configuration of resonant cavities in the vocal tract, and therefore the acoustic output. Since children have acquired this speech sound, it is assumed that they are moving the tongue body to the palate, which acoustically results in the velar pinch, and thus the same pattern of formant movement is expected (Hypothesis 9). The location of the velar constriction in English is more anterior when the following vowel is a front vowel than for a following back vowel, causing F2 and F3 to be higher in frequency, with F3 nearing F4 when a high front vowel follows the velar stop (Stevens, 1998).

It is known that both the burst and the formant transitions (reflecting the movement of the tongue body and mandible) can be cues for place of articulation for all of the syllable-initial stop consonants. It is presumed that since jaw movements are

visible to children, who have a natural tendency to imitate facial movements, sounds relying on this motion are acquired earlier than other speech sounds. The parts of speech relying on these gestures (i.e. tongue height and jaw movements) are expected to be comparable to adult values (Hypothesis 10). However, front-back tongue body movements are not generally visible to the child, who must therefore rely solely on auditory feedback to acquire these gestures. It is therefore expected that children in this age group may not have fully developed the back-front tongue body movements (Hypothesis 11). Additionally, front-back tongue body movements (as estimated by second formant frequencies) are expected to be different for the three different places of articulation, given the different tongue body placements during stop production.

Articulation of the stop consonants will also be studied via acoustic measures of articulator placement. The major spectral peak during the burst represents the front cavity resonance, from which the location of the constriction can be inferred. Since it has been found that the burst spectrum carries the most weight of all cues to stop place of articulation (Suchato, 2004), it is hypothesized that children have learned how to produce the correct burst spectrum, in order for their stop productions to be understood. The frequency of the principal acoustic prominence in the burst is a reflection of the lowest resonance in the oral cavity anterior to the location of the consonant closure. Measurement of this frequency for alveolars and velars indicate of the location of the consonant closure, which is expected to be accurate in children. Additionally, the frequencies of the burst spectrum peaks corresponding to the starting frequency of F2, which are also cues to place of articulation, are expected to be distributed in the correct order in children (Hypothesis 12). For a burst preceding a back vowel, labial F2 is lower than velar F2 and alveolar F2 is highest. Furthermore, although a child might use the correct articulator to make the closure for a stop consonant, it is expected that there will be some variability in the placement of this articulator, particularly for alveolar and velar stops, which would be represented in terms of the noise burst spectrum at the consonant release (Hypothesis 13).

The release for mechanism of the primary articulator is another area of study. Stop burst release requires coordination of articulator movement, opening of the glottis, and changes in stiffness of the vocal tract, glottis, and articulator. All of this happens over just a few milliseconds. Children are still learning to speak, and are still learning to control respiration, phonation, and articulation for speech. Differences in the compliance of the articulator, specifically decreased stiffness, could result in a decrease in fine-tuned control over the articulator and increase the probability of multiple bursts. Additionally, children's smaller articulators will have less mass and therefore a higher "natural frequency" of vibration. Thus, it is hypothesized that children will produce a greater number of bursts during stop release than adults (Hypothesis 14).

6. Methods

6.1. Children's Speech Experiment

For the proposed study, fifteen monolingual American English-speaking children, who were between the ages of 2;2-3;6 at the beginning of the study, were recruited. Data from ten of these children (ages 2;6-3;3) were analyzed for this study. Fifteen children were initially recruited to allow for the possibility that some children may not complete the full six months of recordings. While all fifteen children did in fact complete the study, some children were much more active and less cooperative during the recording sessions than others, resulting in fewer usable target word productions. Thus, these children's productions were not used for this study.

Each child attended a recording session (also referred to as a play session, and described below) once every 4-6 weeks over a 6 month period. Only data from the first and last play sessions were analyzed for this study. Three different screenings were used to determine that the children are typically developing: a hearing screening, The Goldman-Fristoe Test of Articulation-2 (Goldman and Fristoe, 2000), and calculation of mean length of utterance (MLU) from 45-50 utterances of spontaneous speech recorded during a short play session. Some of the children were unable to complete the hearing screening at the initial visit; this was attempted a second time during the fifth or sixth recording session.

A target word list was developed (see Appendix 1: Target word list, p. 118), consisting of 20 words, aimed at eliciting each stop consonant in initial, medial, and final position. When possible, for each stop consonant, a word was chosen in which the stop is produced as a singleton and is followed by a neutral or slightly backed vowel (for word-initial stops), or preceded by a neutral or slightly backed vowel (word-final stops). Thus,

all stop consonants were produced in similar vowel environments, minimizing the effect of surrounding vowels on stop consonant production differences. For stops in the medial position, it was not possible for all stops to have the same preceding vowel, but they are all followed by the vowel /i/. To collect data on each child's vowel space, several words with non-neutral vowel targets were included, but were not used in stop consonant analysis. Due to the large amount of data collected, in order to reduce the data analysis phase of this study, only word-initial stop consonants were analyzed for this study.

For the experimental study, the children came either to the sound attenuated booth in the Speech Communication Group lab at MIT or to a quiet room at the First Congregational Church in Milton, and were recorded using a digital audio tape (DAT) recorder with a 48 kHz sampling rate. The first task of the session was the reading of a story by the researcher using the target words. During this reading, the researcher gave prompts ("She drinks from ...") and asked questions ("What's this girl's name?"), eliciting 2-3 productions of each target word from the child. The story pictures were actual photographs of toys. After the reading of the story, there was an 8-10 minute play session, during which the child played with the researcher using the toys from the story pictures. The play session began with the researcher taking each toy from a pillowcase one at a time, and asking the child to name the toy before giving to him/her. This was followed by 8-10 minutes of free play, and then clean-up, during which the child was asked to name each toy before the researcher put it back into the pillowcase. The free play generally resulted in the elicitation of several spontaneous productions of each target word. Thus, 5-15 tokens of each target word were collected during the session. This sequence was repeated with two more stories.

The digitized signals were downsampled for investigation of different aspects of the speech. For example, when the high frequencies of the stop burst, frication, and aspiration noise are examined, a 24 kHz sampling rate is used, and the recording is low pass filtered at 12 kHz. When studying the vowels and formant frequency movements, a 16 kHz sampling rate is used with low-pass filtering at 8 kHz. All spectrograms are produced using a 6.4 ms Hamming window. Spectrogram and spectra analysis are described in the results.

6.2. Adult Speech Experiment

The primary adult female caretaker for each child in the study was requested to participate in one experimental session to record adult speech. Ten adults participated in the recording, nine of whom were parents of the children whose data are analyzed for this study. These subjects were recorded speaking ten tokens of each word from the word list. They were recorded either in the sound attenuated booth at the Speech Communication Group at MIT, or in a quiet room at the First Congregational Church in Milton. Their speech was recorded to a digital audio tape (DAT) with a sampling rate of 48 kHz.

The adult subjects were shown pictures of the toys using a PowerPoint presentation, and completed both naming and story-telling tasks. The first task of the recording session was the training, in which the subjects were re-familiarized with the pictures and associated target words. Note that since all subjects were present during at least two of their child's recording sessions, they were already familiar with the pictures. During the training, they were asked to name each picture starting with the article "a", pronounced /ə/ (e.g. "a dog"). Thus each token was preceded by a neutral vowel. Following the training, the subjects were shown a series of pictures taken from each of three storybooks that had been read to the children. The words were removed from these pictures, and the adults were asked to tell what was happening in each picture (e.g. "the bug is on the dog's head"). The subjects were told that this was practice, to familiarize them with the stories. After these readings, the subjects then completed one naming task (in which each picture is shown only once), then retold each of the stories, and after each one, the picture-naming task was repeated. Thus, each subject completed a total of five naming tasks, including the training, resulting in five tokens of each word, and two readings of each story, resulting in six or more tokens of each word. For analysis purposes, only the first five tokens recorded of each word were used. In order to minimize analysis of tokens for which the subject may have been overly focused on their speech, the words were collected primarily from the initial story telling, followed by the training and then the second story telling if not enough tokens were collected.

7. Measurements

The word-initial stop consonant measurements were divided into four categories: duration and burst, amplitude, formant, and harmonic. Recordings were all downsampled to 24 kHz for analysis.

Duration and burst measurements were made via visual inspection of the waveform and the spectrogram. Following are all the duration and burst measurements (1/0 indicates a binary yes/no type of measurement):

- Presence of prevoicing before burst release (1/0)
- Prevoicing duration
- Voicing during stop (1/0)
- Burst duration
- Aspiration duration
- Voice onset time
- Vowel duration
- Word duration
- Number of bursts
- Utterance initial word? (1/0)

Amplitude measurements were made using a 3 ms Hamming window, centered over the relative maximum in the area of the waveform being measured. A short Hamming window was chosen because the burst in the children's speech can be very brief. Amplitude during the aspiration was measured in at three consecutive 1 ms intervals and averaged for the final measurement. The amplitude of the burst spectrum prominence was often measured a second time with increased gain if the overall utterance

amplitude was low, obscuring higher frequency peaks. Following are all the amplitude measurements:

- Burst spectrum prominence frequency
- Burst spectrum prominence amplitude
- Burst spectrum prominence amplitude (spectrum gain = 6000 dB)
- Burst amplitude: overall RMS (root-mean-squared) amplitude of the burst
- Aspiration amplitude (onset)
- Aspiration amplitude (prior to voice onset)
- Vowel amplitude (at voice onset)
- Vowel amplitude (at voice onset + 10 ms)
- Vowel amplitude (center)
- Vowel amplitude (end of voicing – 10 ms)

Formant frequency measurements were taken using a 6.4 ms Hamming window with pre-emphasis. Visual inspection of the spectrogram was used to validate the spectral measurements. Following are all the formant measurements:

(Hamming window: 6.4 ms; with pre-emphasis)

- Burst
 - All stops: F1, F2
 - Velar stops only: F3
- Aspiration onset
 - Voiceless stops: F1, F2
 - Velar voiceless stop only: F3
- Aspiration offset
 - Voiceless stops: F1, F2
 - Velar voiceless stop only: F3
- Voice onset
 - All stops: F1, F2
 - Velar stops only: F3
- Vowel middle
 - All stops: F1, F2, F3

Harmonic measurements were taken using a 10 ms Hamming window. This window size was chosen because it was large enough to ensure that all harmonic information was displayed, but small enough to focus on a specific part of the utterance. Following are all the harmonic measurements:

- Voice Onset
 - First harmonic amplitude
 - Second harmonic amplitude
 - First formant amplitude
 - Third formant amplitude
 - Third formant amplitude (spectrum gain = 6000 dB)
 - Fundamental frequency
- Voice Onset + 20 ms
 - First harmonic amplitude
 - Second harmonic amplitude
 - First formant amplitude
 - Third formant amplitude
 - Third formant amplitude (spectrum gain = 6000 dB)
 - Fundamental frequency

8. Statistical Analyses and Results

8.1. Introduction

In this section, each hypothesis is presented individually and described in detail, followed by the statistical methods used to analyze it, and the subsequent results. The principle findings from these results fall into three categories: respiratory control (inferred from acoustic measurements of amplitude change, voice onset time and burst duration), laryngeal control (inferred from acoustic measurements of voice onset time and fundamental frequency), and articulatory control and articulator placement (inferred from acoustic measurements of burst duration, spectral prominences, and formant frequencies).

8.2. Respiration

8.2.1 Hypothesis 1: Intensity variation

Children will have inconsistent control over lung pressure and difficulty coordinating respiration, and thus subglottal pressure, with phonation and articulation. Specifically, children will use a higher pressure at the initiation of an utterance.

- Children are still learning to speak, and are likely still learning to control respiration for speech in a way that balances elastic recoil with inspiratory and expiratory muscular forces. It is therefore hypothesized that they will begin their utterances with a high subglottal pressure that they are unable to sustain throughout the utterance.

Acoustic consequences of inconsistent control over respiration and an initial high subglottal pressure that is not maintained include a decrease in overall amplitude over the course of the utterance or word, and a high overall fluctuation in amplitude at vowel center from one utterance to the next. Thus, there are several steps necessary for analysis of this hypothesis.

First, it must be determined whether there is a *decrease in the overall amplitude* from voice onset to vowel center over the course of the word that is significantly different from that seen in adults. Burst amplitude varies depending on place of articulation, and different voicing characteristics; therefore, amplitude change was measured only over the course of the vowel.

Using an independent measures t-test, amplitude at voice onset minus amplitude at vowel center in children was compared to this difference in adults. This test found that the amplitude increased during the first half of the vowel for both children and adults, opposite from the hypothesized decrease. This is most likely caused by the presence of some breathiness at voice onset, as vocal folds and vocal tract adjust to their target positions. While results for voiceless stops were not statistically significant, for voiced stops child amplitudes increased significantly more than adult amplitudes ($p = 0.049$), and the pattern of greater increase occurred in 8 of the 10 children (for individual results, see Table 5 in Appendix 2, p. 119). Thus, it is likely that children are still learning to coordinate voicing onset for vowels following voiced stops, and that this difference is caused by difficulties with laryngeal adjustment, not respiration. Amplitude differences were not found to change significantly over the 6-month course of the study ($p > 0.1$).

The next analysis for evaluating control over subglottal pressure was to measure the *range of intensities at vowel center*. For this hypothesis, the mean standard deviation in amplitude over all stop-consonant-initial words for each subject was computed. In order to account for the differences in overall amplification settings for each recording, the amplitude measurements are normalized by the mean amplitude at vowel center for each

subject. Amplitudes were measured in the logarithmic dB scale, so normalization is completed by calculating the difference between vowel center amplitude in each word and mean amplitude at vowel center across all utterances. The standard deviation of the normalized vowel center amplitude values was computed for each subject. An independent measures t-test was used to compare these normalized results, and a significant difference was found between the child and adult populations ($p < 0.0001$), with the children having greater variability than the adults ($n_{\text{child}} = 5.96$ dB, $n_{\text{adult}} = 3.46$ dB). All ten children showed greater variability than adults. Individual results are shown in Appendix 2 (Table 6, p. 120). The child population was also compared to the six-month follow-up child data, and no significant change was found in intensity fluctuations between these two recordings ($p > 0.1$).

Overall, the results showing amplitude increase over the course of the vowel are interpreted to be a reflection of laryngeal adjustment, not respiration, and indicate that children are still acquiring optimal glottal positioning and vocal tract posture for voicing onset. Additionally, results showing high utterance-to-utterance amplitude variability indicate that children have inconsistent control over respiration and hence subglottal pressure.

8.2.2. Hypothesis 2: Burst duration and intensity

A high burst intensity will co-occur with in a short burst duration.

- Children are learning to speak, which includes learning to control respiration for speech. Previous studies have found that children generate higher tracheal pressures than adults for any given intensity output (Stathopoulos and Sapienza, 1993), and that children have higher intraoral air pressures during voiceless stop production (Stathopoulos and Weismer, 1985). It is therefore hypothesized that they will begin their utterances with a high subglottal pressure that will push the constriction open quickly, reducing burst duration.

Analysis of this hypothesis involves several steps. First, it must be determined whether burst duration averages are shorter in children than in adults, and then whether burst duration is correlated with burst intensity. Since many of the children's stops were produced with multiple bursts, which prolonged overall burst duration, there are three burst duration measurements of interest here: average total overall burst duration (from the beginning of the first burst to the end of the last burst), average overall burst duration of single burst utterances only, and average duration per burst (the total overall burst duration divided by the number of bursts).

To test whether burst durations are shorter in children, a two-way analysis of variance (ANOVA) is used to compare the children versus the adults for each stop place of articulation. This test revealed no significant difference between the child and adult populations for average total overall burst duration ($p > 0.1$). However, because there was such a high prevalence of multiple bursts (for which burst duration is measured from the beginning of the first burst to the end of the last burst), this ANOVA was repeated, using only single burst stop productions. A significant difference was found between the two groups for average burst duration for single burst utterances, with adults having longer burst durations than children (see Figure 2, p. 52).

ANOVA

Independent variable: Age (adult, child)

Independent variable: Stop consonant (b, d, g, p, t, k)

Dependent variable: average overall burst duration for single burst utterances

Source	Degrees of freedom	F-ratio	p-value
Age	1	12.611	= 0.001
Stop	5	24.272	< 0.001
Age * stop	5	1.316	= 0.263

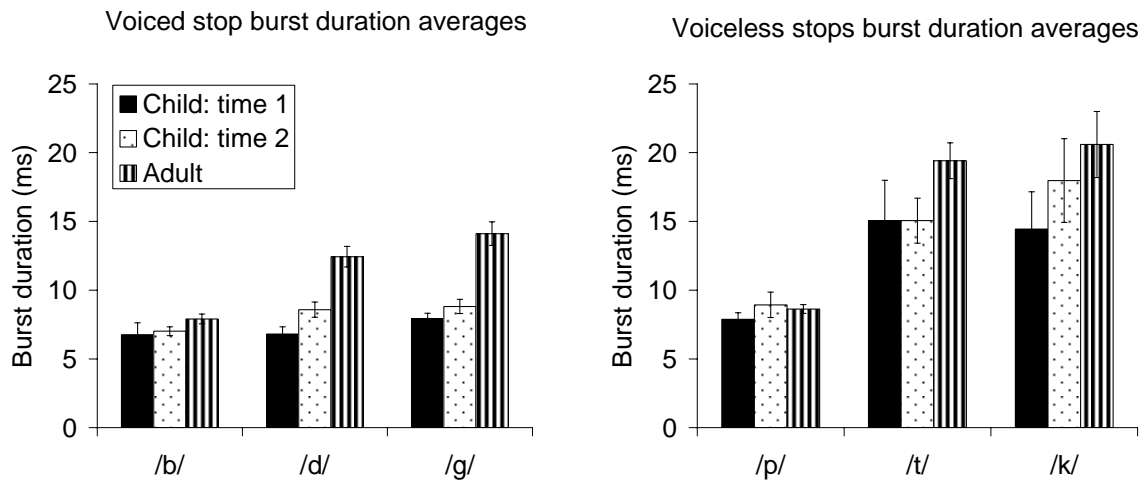


Figure 2. Average burst durations for single burst utterances only for the child group at Time1 (black bars), six months later at Time 2 (dotted bars), and the adult group (striped bars). Error bars show standard error within each subject group.

As seen in Figure 2, there is no significant difference in bilabial stop burst duration between the two groups; only alveolar and velar stop burst durations differ significantly. To test whether burst durations increase significantly towards adult values over the six-month period of the study, another ANOVA was performed to compare the children at Time 1 versus the children at Time 2 for each stop place of articulation. This test revealed no significant difference in average burst duration for single burst utterances over the period of the study ($p > 0.1$).

Because there was such a prevalence of multiple bursts, average duration per burst was also analyzed, in order to include these productions. Average duration per burst is the overall burst duration for each utterance (including single burst utterances) divided by the total number of bursts. A significant difference was found between the two populations for average duration per burst, with adults having longer durations per burst than children (child average duration per burst = 8.62 ms, adult average duration per burst = 12.1 ms). Additionally, duration per burst was found to vary significantly depending on stop place of articulation, with velar stops having longer durations than alveolar stops, which had longer durations than bilabial stops.

ANOVA

Independent variable: Age (adult, child)

Independent variable: Stop consonant (b, d, g, p, t, k)

Dependent variable: average duration per burst

Source	Degrees of freedom	F-ratio	p-value
Age	1	55.821	< 0.001
Stop	5	27.080	< 0.001
Age * stop	5	3.243	< 0.01

Regressions were performed to test for the effect of speaking rate, as measured by vowel duration, on burst duration, and no correlation was found between burst duration and vowel duration.

After finding a significant difference between the child and adult populations for the duration per burst measurement, the child measurements at the first recording session were compared with the child measurements from recordings performed six months later. A two-way analysis of variance (ANOVA) to compare the children at the two recording sessions for each stop place of articulation revealed a significant difference between the two populations for average duration per burst.

ANOVA

Independent variable: Age (child time 1, child time 2)

Independent variable: Stop consonant (b, d, g, p, t, k)

Dependent variable: average duration per burst

Source	Degrees of freedom	F-ratio	p-value
Age	1	9.155	< 0.003
Stop	5	20.900	< 0.001
Age * stop	5	1.191	> 0.1

Thus, duration per burst is found to increase toward adult values over the six-month period of the study (child (time 1) average duration per burst = 8.62 ms, child (time 2) average duration per burst = 10.1 ms).

A significant difference was found between the child and adult populations for the average burst duration of single burst utterances and duration per burst measurements. In order to determine if high burst intensity co-occurs with a short burst duration, it must next be determined if there is a correlation between either of these measurements and overall amplitude at stop release. To account for the differences in overall amplification settings for each recording, the burst amplitude measurements are normalized by the mean amplitude at vowel center for each subject. Amplitudes were measured in the logarithmic dB scale, so normalization is completed by calculating the difference between burst amplitude and mean amplitude at vowel center. A scatterplot of the data revealed a weak positive correlation between duration per burst and normalized overall amplitude of the burst (see Figure 3), with a Pearson correlation of 0.15 (with outliers have a standardized residual value greater than 2.5 removed from data set), p-value = 0.002. This is opposite from the hypothesized effect of greater amplitudes resulting in shorter burst durations. Because average burst duration differs significantly depending on place of articulation, this test was repeated, separating the data into six groups: one group for each place of articulation and voicing distinction (/p,b,t,d,k,g/). The only group that displayed a significant correlation was the group containing voiceless velar stops (/k/), which had a correlation coefficient of 0.286 (p = 0.013). Again, this is a positive

correlation, opposite from the hypothesized effect. No correlation was found between average burst duration for single burst utterances only and burst amplitude.

Child time 1: duration per burst vs burst amplitude

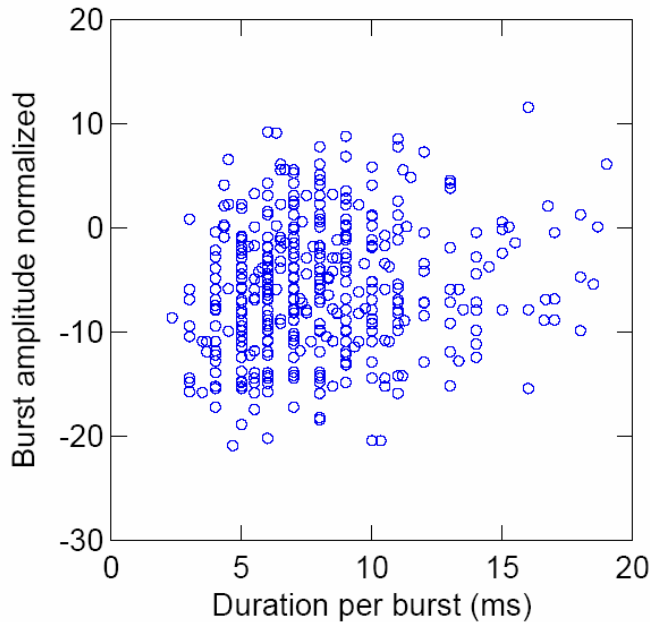


Figure 3. Scatterplot of burst duration (in ms) versus normalized burst amplitude (in dB). Normalized burst amplitude is overall burst amplitude minus average amplitude at vowel center.

Overall, these results indicate that children do produce a significantly shorter burst duration. However, within the children's utterances, shorter bursts are not correlated with higher burst amplitude measures, and thus variations in burst duration are not correlated with variations in subglottal pressure. Another possible cause of a short burst duration is that the children's small articulator size forms a short constriction which opens more quickly at release.

8.2.3. Hypothesis 3: Multiple bursts and intensity

A high burst intensity will be correlated with a higher total number of bursts.

- A high burst intensity is caused by a high subglottal pressure. This pressure will push the constriction open quickly, but will be followed by a negative pressure (Bernoulli's effect), which will pull the constriction back closed again, resulting in a second burst. This may occur several more times before the constriction is finally released.

This hypothesis is analyzed by grouping the utterances depending on how many bursts were present at stop release. One group contains all utterances produced with only one burst, while utterances produced with two or more bursts were all placed in the other group. Then, an independent measures t-test analyzing burst intensity was performed on these two groups, with the prediction that the group with only one-burst utterances would have lower overall burst intensities. However, the independent measures t-test found no significant difference between the two groups ($p = 0.284$), indicating that burst intensity is not related to total number of bursts.

8.2.4. Hypothesis 4: VOT and intensity

Variations in intensity will be correlated with variations in VOT, with high intensities associated with short VOTs, since a high initial subglottal pressure will decrease more rapidly following burst release.

- Changes in intensity are caused by changes in subglottal pressure. A high initial subglottal pressure will decrease more rapidly following burst release, resulting in a shorter VOT.

This hypothesis is analyzed by measuring the correlation between VOT and amplitude at vowel center. The amplitude must be normalized by average overall amplitude at vowel center for each child to account for differences in pre-amplification settings during the recordings. Amplitudes were measured in the logarithmic dB scale, so normalization is completed by calculating the difference between burst amplitude in each word and mean amplitude at vowel center across all utterances. A Pearson correlation was computed for this data, separating out voiced versus voiceless stop consonants. For both voiced and voiceless stops, this resulted in a slightly negative, but non-significant correlation. Thus, variations in subglottal pressure are not correlated with variations in VOT.

8.2.5. Hypothesis 5: First formant and intensity

A high amplitude will co-occur with an increase in F1 (adult studies show F1 increases when shouting).

- Speaking at a louder volume is correlated with an increase in mouth opening in adults, which results in an increase in F1. It is hypothesized that this also occurs in children.

To test this hypothesis, the Pearson correlation coefficient was computed for normalized overall amplitude at vowel center versus first formant frequency. Amplitudes were measured in the logarithmic dB scale, so normalization is completed by calculating the difference between vowel center amplitude in each word and mean amplitude at vowel center across all utterances. No correlation was found between amplitude and first formant frequency measurements in children, as seen in the scatterplot below (Figure 4). This could be due to the high variability in amplitude in children's speech. Another possible reason why these two measures are not correlated in children is that the that children might already be using a wider mouth opening, and therefore it does not increase significantly with higher intensities.

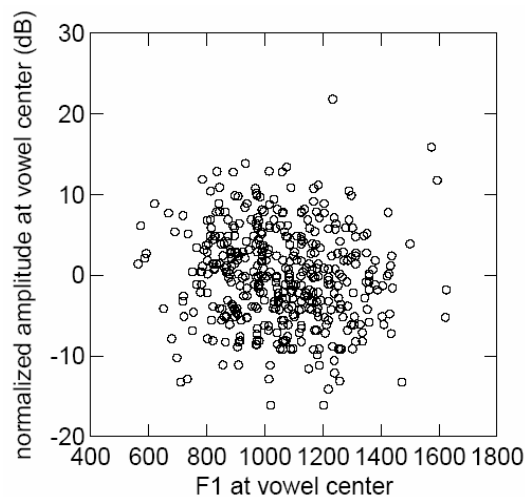


Figure 4. Scatterplot of first formant frequency (in Hz) versus amplitude at vowel center (in dB) across all 10 child subjects.

8.2.6. Summary

Analyses of hypotheses and acoustic measurements relating to respiration have led to several significant results. Children have inconsistent control over respiration, as seen in the high variability in intensity at vowel center. Additionally, children's stop productions are found to have short burst durations, which could be caused by their smaller articulators, or by use of a high subglottal pressure at burst release, though no correlation was seen between burst duration and amplitude. No correlation was found between production of multiple bursts and burst amplitude either. One possibility is that all of the children's utterances are produced with a high subglottal pressure (as has been found in other studies, e.g. Netsell, et al., 1994), and small variations in this pressure are not significantly correlated with small changes in burst duration or with the production of multiple bursts.

8.3. Phonation

8.3.1. Hypothesis 6: Voiced stops – stop release, voicing onset, spectral tilt

Children will have difficulty coordinating onset of voicing (slackening of vocal folds and vocal tract wall) with stop release for voiced stops.

- For voiced stops, release of the vocal tract constriction and onset of voicing occur within a few milliseconds of each other. Since the children have only recently begun using these sounds for speech, it is hypothesized that they are still learning to fine-tune some aspects of the production. This is one instance in which several different actions (release of the stop constriction, and slackening of vocal folds and vocal tract wall) are completed within a few milliseconds. The children are expected to have some difficulty either coordinating these actions or forming the correct vocal fold and vocal tract posture before stop release.

Coordination of voicing onset and stop release for voiced stops is analyzed using several different acoustic results. First, coordination of voicing onset with stop release is analyzed by measuring voice onset time (VOT), and comparing VOT between the two groups. It is expected that children may show delayed onset of voicing for voiced stops, due to difficulty with coordination of constriction release and onset of glottal vibration (i.e. the glottal state may not be correct at the time of release). This hypothesis is tested using an independent measures t-test. Both overall VOT and the difference between VOT and burst duration were analyzed. For overall VOT, children were found to have a significantly longer duration than adults only for bilabial stops. When burst duration was subtracted from VOT (since VOT is measured from burst release, these values are typically roughly equal in adults), children were found to have a significantly delayed VOT for all three places of articulation ($p < 0.05$) (see Figure 5, p. 62 for a sample spectrogram, Figure 6, p. 63 for average group values, and Appendix 2, Table 7, p. 121, for average values for each subject). This VOT lag indicates that children are having difficulty coordinating voicing onset with stop release. Evaluation of children's 6-month

follow-up data using an independent measures t-test found a significant decrease in the VOT-lag duration over the period of the study ($p < 0.1$).

Since children appear to be having difficulty providing conditions necessary for beginning voicing with stop release, it is hypothesized that they are not appropriately slackening their vocal folds and vocal tract wall for voicing, or that they have increased glottal spreading. This would be reflected in spectral tilt measurements, with greater spectral tilt reflecting greater glottal spreading and increased breathiness. It is expected that after voicing begins, spectral tilt will decrease. Spectral tilt can be measured in several different ways. One method is to analyze the difference in amplitude between the first harmonic (H1) and the third formant (A3), and another is to analyze the difference in amplitude between the first formant (A1) and the third formant (A3). Because children have higher fundamental frequencies than adults (and thus the first harmonic can be both the fundamental frequency and the first formant), it is likely that the (A1-A3) spectral tilt measurement will be more accurate for comparison with adult values, but both values were calculated for analysis of this hypothesis.

The hypothesis is tested using an independent measures t-test to compare child versus adult values for change in tilt over the first 20 ms of voicing. The difference between tilt (A1-A3) at voice onset and tilt 20 ms later is computed for each utterance, and results for each subject were averaged, resulting in one value for each subject. This is repeated using the H1-A3 tilt measurement. For (H1-A3) values, no significant difference was found ($p > 0.5$), but for (A1-A3) values, children's tilt differences were found to be significantly less than adults' ($p < 0.02$), opposite from predicted results. Closer inspection of the data reveals that for half of the children, tilt actually increased over the first 10 ms of the utterance (see Appendix 2, Table 8, p. 122 for average values for each subject at the first recording session). Thus, there is high variability in this measurement within the child group. Children's (A1-A3) mean values are found to increase significantly ($p < 0.1$) over the 6-month period of the study toward adult values.

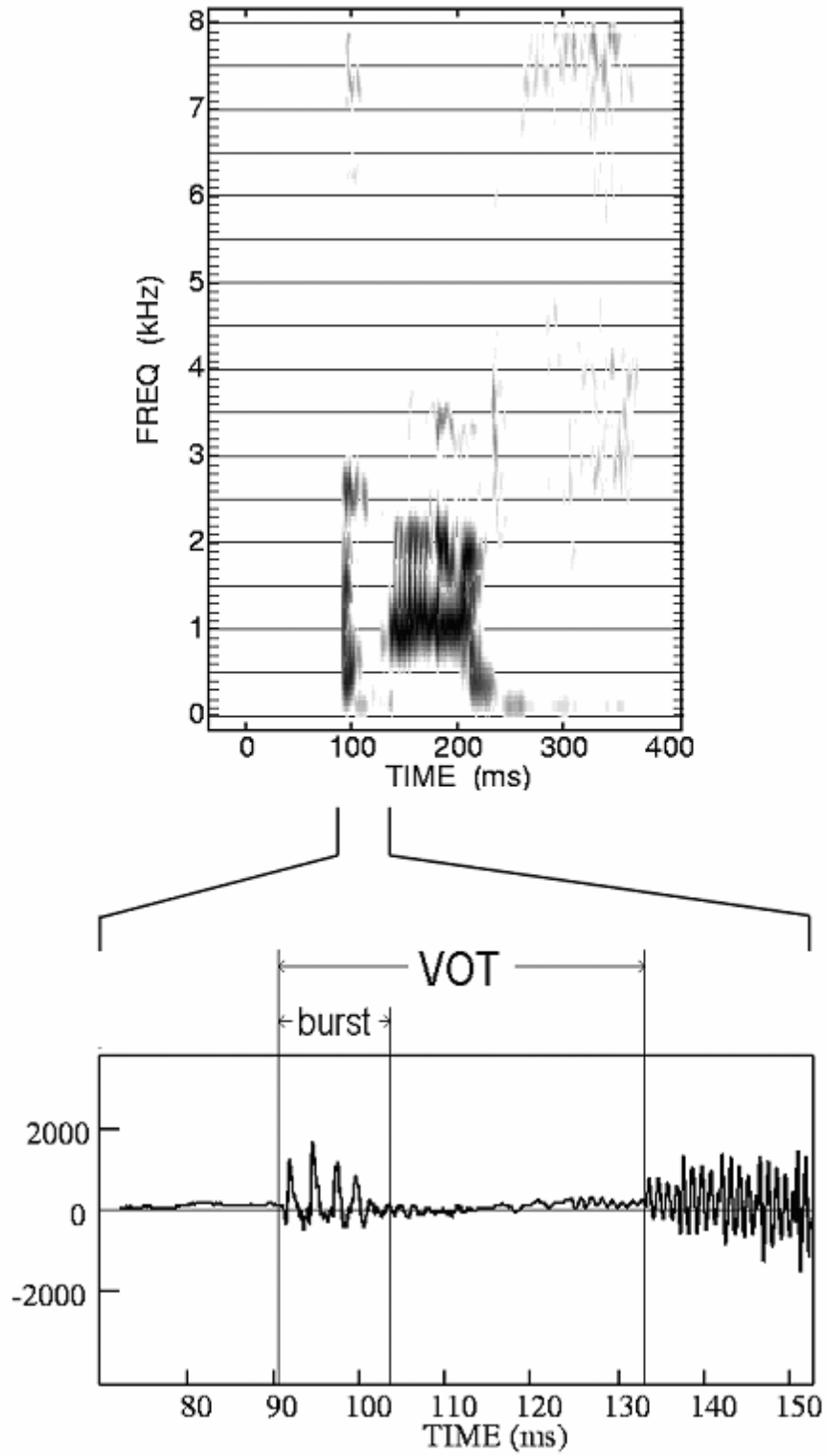


Figure 5. Spectrogram and corresponding waveform of a 2;6-year-old child's production of the word "Gus". Note the VOT-lag.

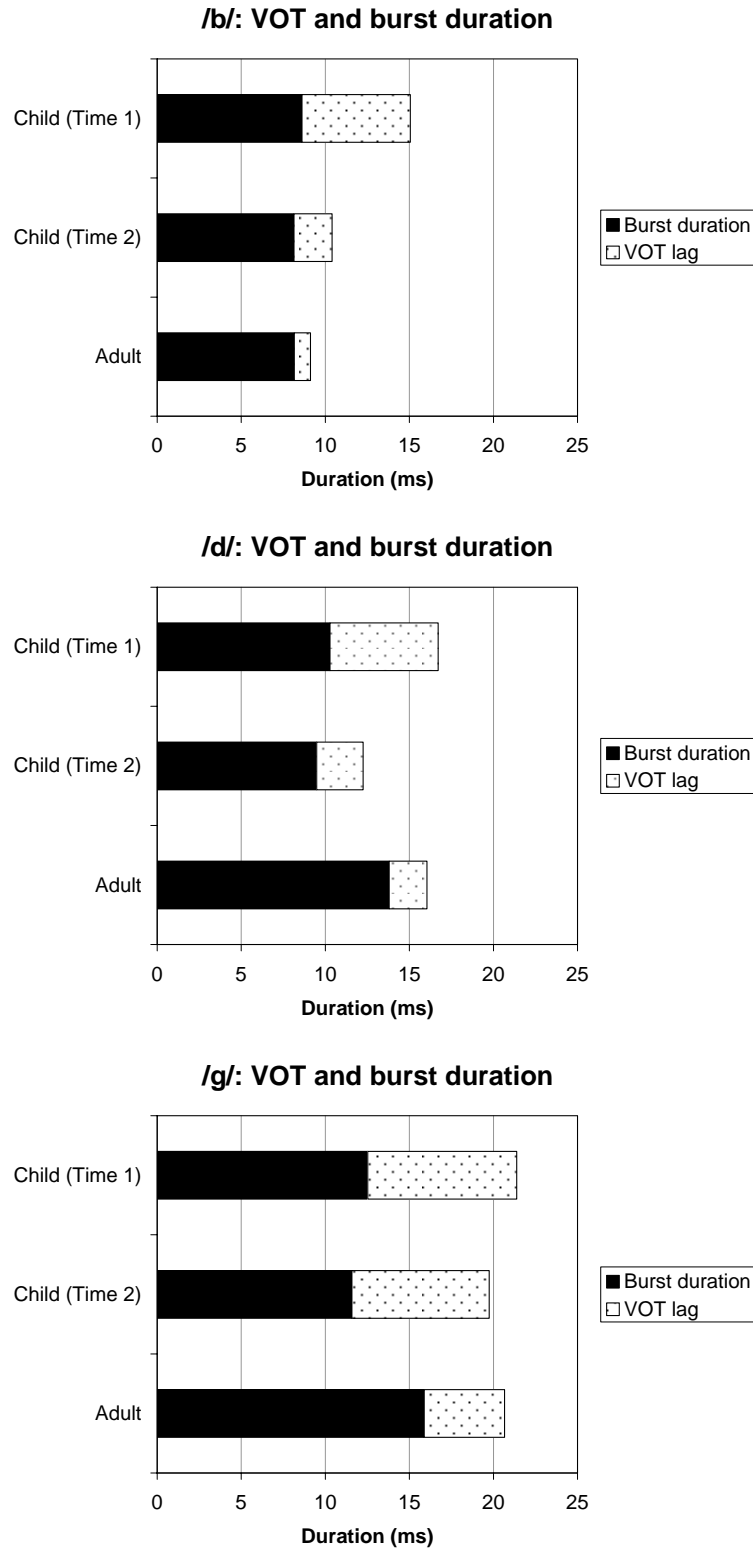


Figure 6. VOT for each voiced stop, broken down to show burst duration and lag before voice onset. The graphs show the overall child average at each recording time, as well as the adult average.

The duration of the VOT-lag or the ‘gap’ between the end of the burst and the onset of voicing discussed above is also hypothesized to be correlated with fundamental frequency (F0) differences, such that the utterances with the ‘gap’ or short-lag VOT tokens will have a higher F0 than those with no VOT-lag, comparable to the F0 differences seen in adults after production of voiced versus voiceless stops. Similar to spectral tilt expectations, this difference would be due to increased stiffness in the vocal folds for the short-lag VOT stops, in which stiffness would cause the delay in voicing onset. For this hypothesis, one ‘gap’ or ‘lag’ value was computed for each token spoken by each child, and these values were then labeled as “no-lag” (i.e. the ‘gap’ value = 0 ms) or “short-lag” (i.e. the ‘gap’ value \geq 1 ms). An independent measure t-test was performed to test this hypothesis, and it was found that the short-lag VOT F0 values were significantly lower than the no-lag VOT F0 values ($n_{\text{short-lag}} = 311$ Hz, $n_{\text{no-lag}} = 348$ Hz), opposite from the predicted result ($p < 0.005$). One possibility is that for the no-lag utterances, the children are adjusting the glottis for earlier onset of voicing following stop release despite residual vocal tract and vocal fold stiffness, resulting in a higher fundamental frequency.

These results lead to the question of whether children’s F0 values following voiceless stops are higher than those following voiced stops, as seen in adults. An independent measures t-test was used to test this hypothesis. Children’s voiced stop F0 values were found to be significantly lower than their voiceless stop F0 values, as expected ($n_{\text{voiced}} = 322$, $n_{\text{voiceless}} = 355$, $p = 0.004$).

In summary, a VOT-lag is present in the children’s voiced stop initial utterances that is significantly greater than that seen in adults. This indicates that the children are having difficulty adjusting their larynx for voicing onset following voiced stop release. They may be having difficulty coordinating voicing onset with stop release, or they may not yet have developed the correct vocal tract posture for immediate voicing onset following release of the constriction. The VOT-lag is correlated with decreased F0 values. This could indicate that children are learning to adjust the larynx for immediate onset of

voicing in such a way that results in an increased F0 in the no-lag utterances. Spectral tilt results are found to be inconsistent within the child group.

8.3.2. Hypothesis 7: Voiceless stops – fundamental frequency, spectral tilt

For voiceless stops, fundamental frequency may not be relatively high at VOT and it may not decrease over the following 20 ms, as is seen in adults.

- During voiceless stop production in adults, the vocal tract and vocal folds are spread and held stiff during stop release, frication, and aspiration to prevent the onset of voicing. The children may not be able to consistently coordinate the onset of voicing with the offset of aspiration, resulting in a different pattern of F0 movement.

Fundamental frequency (F0) movements are analyzed by measuring F0 at VOT and at VOT + 20ms, and finding the difference between these two values. Since children typically have a higher average F0 than adults, due to their smaller larynx, the differences for each token were normalized by average F0 during the vowel for each subject. The difference between F0 at VOT and F0 at VOT + 20ms for each utterance is divided by average F0 in the vowel (across all tokens), resulting in a normalized F0 difference value for each token. Then, all normalized F0 difference values for each speaker are averaged, resulting in one value per subject. It is hypothesized that the F0 difference will be less in children than in adults, indicating that children have reduced control over vocal fold stiffness. This hypothesis is analyzed using an independent measures t-test, comparing the child versus adult F0 differences. The t-test found no significant difference between the child and adult populations ($p > 0.1$). However, the child values included two outliers (standardized residual values greater than 2.5), and when these two values were excluded, the child differences were significantly less than those for adults ($p < 0.1$). This indicates that child F0 decreased less over the first 20 ms of the utterance than adult F0. The original child data were then compared with the six-month follow-up data, and a paired t-test found no difference over the six-month period, regardless of whether or not the two outlier values were included ($p = 0.215$).

Another measurement used to analyze the laryngeal movements during voiceless stop production is the spectral tilt. Specifically, the spectral tilt is expected to be high at voice onset and to decrease during voicing in vowel production. The difference between tilt (A1-A3) at voice onset and tilt 20 ms later is computed for each utterance, and results for each subject were averaged, resulting in one value for each subject. This is repeated using the H1-A3 tilt measurement. This hypothesis (spectral tilt will be high at voice onset and decrease during voicing in vowel production) is analyzed using an independent measures t-test, comparing the child versus the adult spectral tilt differences. One spectral tilt difference value was computed for each subject (shown in Appendix 2, Table 9, p. 123). This test showed a significant difference between the child and adult populations for the A1-A3 tilt difference measurement only, with children having lower tilt difference values than adults ($p < 0.001$) as predicted. However, the (A1-A3) difference values are higher 20 ms after VOT than at VOT for both adults and children, indicating greater spectral tilt in the vowel than at stop release. This is an unusual and unexpected finding, possibly caused by the wide spacing of the harmonics caused by the high F0 in children's utterances. It is also possible that F1 is still rising during the 20 ms of the vowel, which would influence the tilt value. Additionally, the A3 measurement may be contaminated by noise at higher frequencies (see example: Figure 7, p. 68).

The last measurement on voice onset for voiceless stops is overall voice onset time. Using t-tests to compare average VOT values for each place of articulation, children were found to have a significantly longer VOT than adults for all three places of articulation ($p \leq 0.002$). While average VOT does decrease over the period of the study, this decrease was not found to be significant ($p \geq 0.125$). Additionally, children are found to have much greater variability in VOT than adults, which does not change over the period of the study. These findings indicate that children are still learning to coordinate voice onset with stop release and the offset of aspiration for voiceless stop production.

Overall, the main findings of this hypothesis are a smaller F0 decrease after voice onset following voiceless stop release in children, and a long and highly variable VOT. These results indicate that children are still developing coordination of voice onset with the end

of aspiration, or they have not yet acquired the necessary laryngeal adjustments and vocal tract posture for voicing onset.

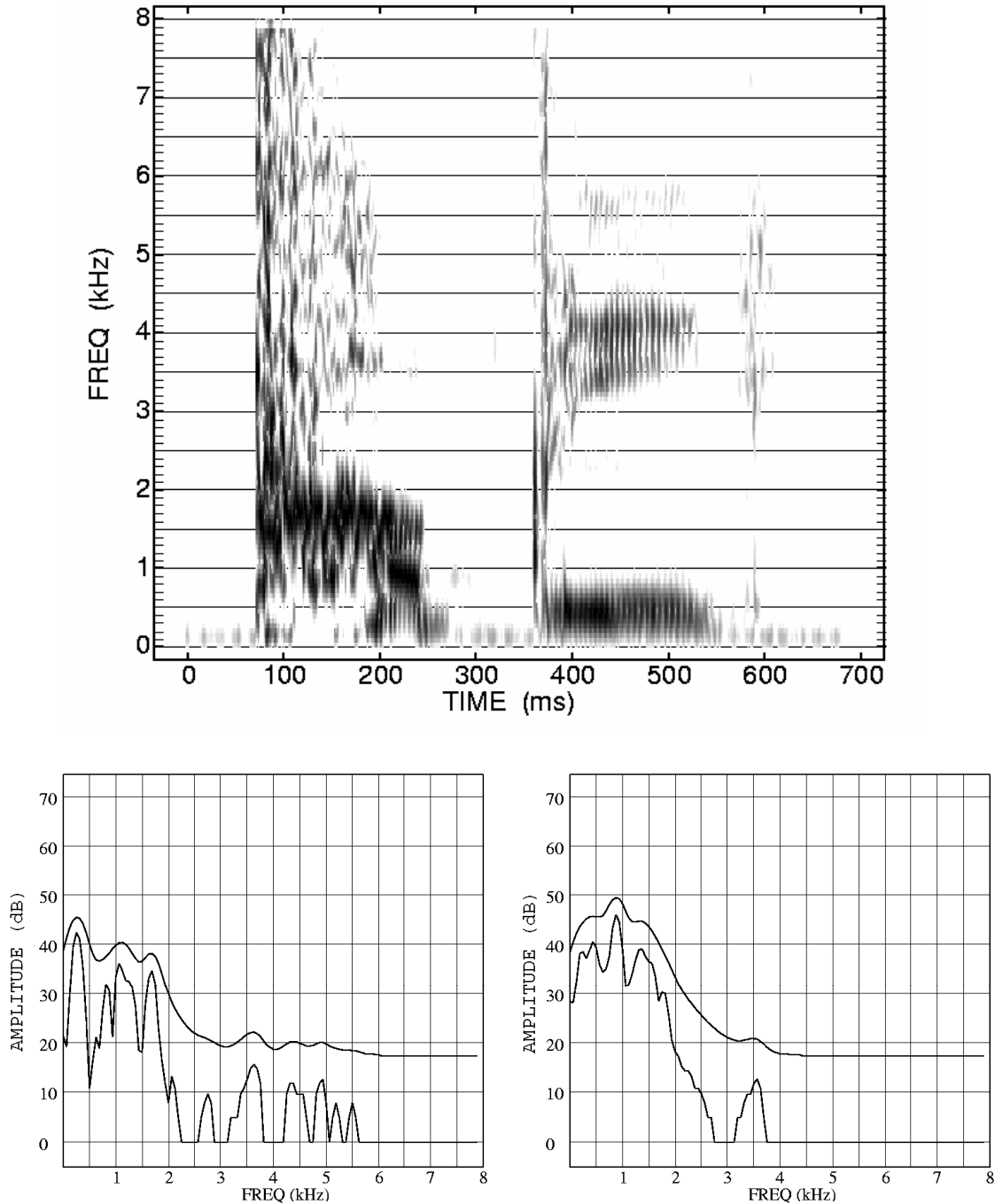


Figure 7. Spectrogram and spectra of the word “puppy” spoken by a 3;0-year-old boy. The spectrum on the left was produced using a window centered on the first glottal period at voice onset (~195 ms in the spectrogram shown). The spectrum on the right was produced using a window centered on the glottal period that occurred 25 ms later. Both spectra were produced using a 10-ms window. Note the weak F3 in the spectrogram, and the corresponding low A3 values in the spectra.

8.3.3. Hypothesis 8: Fundamental frequency, intensity, VOT, spectral tilt

F0 will be correlated with intensity, such that a high intensity will co-occur with a high F0 and a short VOT. Also, spectral tilt will decrease with high intensity.

- In order to increase fundamental frequency, the vocal folds must be stiffened. This stiffening results in an increase in the threshold of phonation, and possibly a compensating increase in intensity. A high initial intensity will decrease more rapidly following burst release, resulting in a shorter VOT. Just as a breathy voice results in increased spectral tilt, a louder voice results in less spectral tilt, and thus spectral tilt is expected to decrease with high intensity.

The first part of this hypothesis, that F0 will be correlated with subglottal pressure, is analyzed by measuring the correlation between fundamental frequency (F0) and amplitude at vowel center. The amplitude must be normalized by average overall amplitude at vowel center for each child to account for differences in pre-amplification settings during the recordings. Amplitudes were measured in the logarithmic dB scale, so normalization is completed by calculating the difference between vowel center amplitude in each word and mean amplitude at vowel center across all utterances. A Pearson correlation was computed comparing F0 and amplitude at vowel center, resulting in a correlation of 0.274 ($p < 0.001$). This computation was repeated excluding outlying data (outlying data points are those with a standardized residual value greater than 2.5), with a similar result of a 0.271 correlation coefficient ($p < 0.001$), supporting the hypothesis that a high F0 co-occurs with a high subglottal pressure.

This hypothesis also speculates that there will be a negative correlation between fundamental frequency (F0) and voice onset time (VOT). A Pearson correlation was used to compare F0 and VOT for both voiced and voiceless stops. While no correlation was found for voiceless stops, voiced stops were found to have a negative correlation with a Pearson coefficient of -0.212 ($p < 0.001$), as predicted. When outliers were removed (outlying data points are those with a standardized residual value greater than

2.5), the voiced stop correlation coefficient value remained negative, at -0.219 ($p < 0.001$). One possible reason no correlation is found between F0 and VOT for voiceless stops is the high variability in VOT for voiceless stops.

The last part of this hypothesis is that spectral tilt will decrease with increased intensity. Analysis requires computation of the correlation between spectral tilt and intensity. Vowel amplitude was normalized as described above. As mentioned previously, spectral tilt can be measured in several different ways. One method is to analyze the difference in amplitude between the first harmonic (H1) and the third formant (A3), and another is to analyze the difference in amplitude between the first formant (A1) and the third formant (A3). Because children have higher fundamental frequencies than adults (and thus the first harmonic can be the fundamental frequency or even the first formant), it is likely that the (A1-A3) spectral tilt measurement will be more accurate for comparison with adult values, but both values were calculated for analysis of this hypothesis. Scatterplots and regressions of spectral tilt and normalized vowel center amplitude yielded no correlation between these sets of data in children, with R^2 values of 0.016 for (H1-A3) spectral tilt measurements, and 0.001 for (A1-A3) spectral tilt measurements.

In summary, F0 and vowel center amplitude are found to be positively correlated, while F0 and VOT for voiced stops are found to be negatively correlated, and there is no correlation between vowel center amplitude and spectral tilt. The negative correlation between F0 and VOT could indicate that the glottis is being adjusted to allow for earlier onset of voicing.

8.3.4. Summary

Analyses of hypotheses and acoustic measurements relating to phonation and laryngeal adjustments have led to several significant results. First, results indicate that children are still developing coordination of voicing onset for vowels following voiced stops, or are having difficulty achieving the correct vocal tract posture before burst release. This is inferred from the finding of a lag following the burst release before voicing onset, and is supported by results showing an amplitude increase over the course of the vowel in children that is significantly greater than that measured in adult speech. The children's increase in amplitude over the course of the vowel may be related to breathiness at voice onset caused by a more spread glottis at consonant release. Next, children are inferred to have reduced control over vocal fold stiffness or glottal opening, as indicated by a smaller decrease in F0 following voicing onset, and high variability in VOT for voiceless stop-initial productions. These results all indicate that children are still developing appropriate glottal adjustments and intraoral pressure for stop consonant production, including vocal fold stiffness and glottal spreading.

8.4. Articulation

8.4.1. Hypothesis 9: Velar pinch

A typical velar pinch in the children's /k,g/-initial words is expected.

- The velar pinch is caused by the tongue body moving up towards (and then away from) the palate and changing the configuration of resonant cavities in the vocal tract, and therefore the acoustic output. Since children have acquired this speech sound, it is assumed that they are moving the tongue body to the posterior portion of the palate, which acoustically results in the velar pinch. Children do have a relatively shorter pharyngeal cavity compared to their vocal tract length, which may change the acoustic consequences of this action.

The velar pinch is analyzed by measuring the divergence of the second and third formant frequencies after stop release. For this hypothesis, we look at the ratio of F3 to F2 at stop release divided by the ratio of F3 to F2 at vowel center in order to account for the effects of vocal tract length differences on formant measurements. It is hypothesized that the average value of this ratio in child group will be equal to the average value of this ratio in the adult group. The hypothesis is tested using an independent measures t-test to compare the children versus the adult values. This test revealed no significant difference in formant ratios for the velar pinch measurement between the child and the adult populations ($p = 0.206$). The mean formant measurements are shown in Appendix 2, Table 10 (p. 124). These results support the hypothesis that a typical velar pinch is seen in children's speech acoustics. This indicates that children have acquired the correct articulatory movement (tongue body to palate) for velar stop production.

8.4.2. Hypothesis 10: Tongue height and jaw movement

Tongue height and jaw movements are expected to be comparable to adults.

- Jaw movements are visible to children, who have a natural tendency to imitate facial movements. Sounds relying on this motion are acquired earlier than other speech sounds, and the parts of speech relying on this movement (i.e. first formant frequencies) are expected to be more precise.

Tongue height and jaw movements are analyzed by measuring the movements of the first formant over the course of the stop release. The first formant frequency reflects the resonance of the oral cavity, and thus it is possible to infer tongue and jaw movements from this data. To control for vocal tract length differences when testing this hypothesis, we look at the ratio of F1 at stop release to F1 at vowel center.. Thus it is hypothesized that F1 at stop release divided by F1 at vowel center in children will equal F1 at stop release divided by F1 at vowel center in adults. The hypothesis is tested using an independent measures t-test to compare the children versus the adults. This test revealed no significant difference between the child and adult populations for the average first formant ratio of F1 at stop release and vowel center ($n_{\text{child}} = 0.643$, $n_{\text{adult}} = 0.674$, $p = 0.19$).

To complete the analysis of this hypothesis, it is also of interest to measure whether measurements of the first formant itself at vowel center are comparable to adult values. For testing this hypothesis, we use the ratio of F1 at vowel center to F3 at vowel center in order to account for the effects of vocal tract length differences on formant measurements. It is hypothesized that the average value of this ratio in child group will be equal to the average value of this ratio in the adult group. An independent measures t-test is used to compare the child versus the adult values. This test revealed a significant difference between the child and adult populations, such that children's values were lower than adults ($p = 0.013$). This could indicate less mouth opening or tongue lowering

in children. Another possibility is that children's oral cavities are longer compared to the length of their vocal tract than adults', resulting in comparatively lower F1 values. See Appendix 2 (Table 11, p. 125) for individual subject averages.

Overall, the results of this hypothesis show that children's speech does display the expected F1 movement from burst into vowel, indicating that they have acquired the correct tongue height and jaw movements for velar stop consonant production and movement into the following vowel.

8.4.3. Hypothesis 11: Back-front tongue body movement

Some difficulty coordinating back-front tongue body movements with consonant release is expected.

- Back-front tongue body movements are not generally visible to the child, who must therefore rely solely on auditory feedback to acquire these gestures. Sounds relying on these motions are acquired later than visible tongue height and jaw movements, and are expected to be more variable, as they are still being fine-tuned. Front-back tongue body movements (as estimated from second formant frequencies) are expected to be different for the three different places of articulation, given the different tongue body placements during stop production.

Back-front tongue movements are analyzed by measuring the movements of the second formant from stop release to vowel center. Because children and women have shorter pharyngeal cavities than oral cavities (Goldstein, 1980), the second formant frequency will reflect the resonance of the pharyngeal cavity. Thus it is possible to infer front-back tongue movements from the second formant data.

The second formant data are analyzed as ratios in order to control for the effect of vocal tract length on formant values. Specifically, F2 at stop release is divided by F2 at vowel center for each utterance, and an average value is computed for each subject. The hypothesis is tested using an independent measures t-test to compare the child group versus the adult group. This test revealed a significant difference between the child and adult groups in second formant frequency movements after stop release for bilabial ($p = 0.004$) and alveolar ($p < 0.0001$) places of articulation, but not after velar stops ($p = 0.108$). Following bilabial stops, the children's productions had greater increases in F2 than the adult's ($n_{\text{child}} = 0.745$, $n_{\text{adult}} = 0.878$), indicating less gestural overlap. However, following alveolar stops (in which F2 is expected to decrease following stop release into a back vowel) the children's productions displayed less F2 decrease than the adult's (n_{child}

= 1.207, $n_{\text{adult}} = 1.329$), indicating more gestural overlap. See Appendix 2, Table 12 (p. 126) for average F2 values for each subject.

These results may lead one to question if children's F2 values at vowel center might be greater than adult values in relation to their vocal tract length, causing greater bilabial release increases and lesser alveolar release decreases. For this hypothesis, we look at the ratio of F2 at vowel center to F3 at vowel center in order to account for the effects of vocal tract length differences on formant measurements. It is hypothesized that the average value of this ratio in the child group will be greater than the average value of this ratio in the adult group. A t-test performed this data showed that these corrected F2 values are actually lower in children than in adults ($n_{\text{child}} = 0.513$, $n_{\text{adult}} = 0.571$, $p = 0.012$), opposite from the proposed higher F2 values. Thus F2 value at vowel center does not affect the different F2 movement findings found for labials versus alveolars. The different results for different places of articulation are likely related to the fact that the tongue body itself is used to produce alveolar stops, and must then be moved to the subsequent vowel position, whereas for bilabial stops, the tongue is not used in stop production and thus the child can place it appropriately in anticipation of the upcoming vowel. More specifically, during alveolar stop production, adults often use a fronted tongue body position, which brings the tongue farther away from the vowel target of a slightly backed tongue body, resulting in less gestural overlap between the alveolar stop and the backed vowel. The results imply that the children may not yet have acquired this tongue fronting gesture, and their tongue bodies are in a neutral or slightly backed position during alveolar stop production.

In addition to measuring second formant frequency movements, the variability of F2 at vowel center was analyzed. This measurement was normalized by dividing the standard deviation of F2 at vowel center by F2 mean at vowel center. A t-test was used to compare this measurement between children and adults. This test showed a significant difference in F2 variability between the two populations, with children having more variable F2 values than adults ($p < 0.0001$).

In summary, the results of this hypothesis show that children have less gestural overlap than adults in the production of bilabial stops, but more gestural overlap in the production of alveolar stops, while gestural overlap in velar stop production is similar to adult values. Additionally, children's F2 values at vowel center are more variable than adults'. Overall, these results indicate that children are still fine-tuning front-back tongue body placement and movements. Future work is needed to analyze front-back tongue body movements between stop consonants and other vowel environments.

8.4.4. Hypothesis 12: Burst spectrum peaks: articulator placement

The frequencies of the peaks corresponding to the starting frequency of F2 are expected to show the following order for a burst preceding a back vowel:

/ labial < velar < alveolar /

This is the F2 frequency order seen before back vowels in adult speech. Additionally, primary articulator placement is expected to be correct, as reflected in the spectral shape. For bilabial bursts, the burst spectrum is expected to slope downwards at high frequencies with no major spectral prominence, while a spectral peak in the F5-F6 region is expected for alveolars, and a prominent peak in the F2 region is expected in the velar stop spectrum.

- Children's stop consonant productions are intelligible and therefore differentiable. Since it has been found that the burst spectrum carries the most weight of all cues to stop place of articulation (Suchato, 2004), it is hypothesized that children have learned how to select the correct articulator and place of articulation, which would produce the correct burst spectrum in their stop productions.

The distribution of the starting frequency of the second formants is analyzed by measuring the differences between the F2 means for each place of articulation. For this hypothesis, it is expected that there will be a difference in the sample means of the F2 burst peaks such that labial F2 will be less than velar F2 and alveolar F2 will be greatest, with adult values being lower than the children's. The hypothesis is tested using a 2-way analysis of variance (ANOVA) to compare the children versus the adults. This test revealed a significant difference for place of articulation, as predicted, and for age, with adults having lower F2 values than children.

ANOVA

Independent variable: Age (adult, child)

Independent variable: Place of articulation (bilabial, alveolar, velar)

Dependent variable: F2 at burst

Source	Degrees of freedom	F-ratio	p-value
Age	1	120.343	< 0.001
Place of artic	2	226.123	< 0.001
Age * place of artic	2	9.120	< 0.001

Thus, children have the same F2 peak distribution pattern as adults, as predicted, with a significant difference in F2 frequency based on stop place of articulation. This can be seen in the F2 peak density plots below.

Child time 1: F2 density plot by place of articulation

Adult: F2 density plot by place of articulation

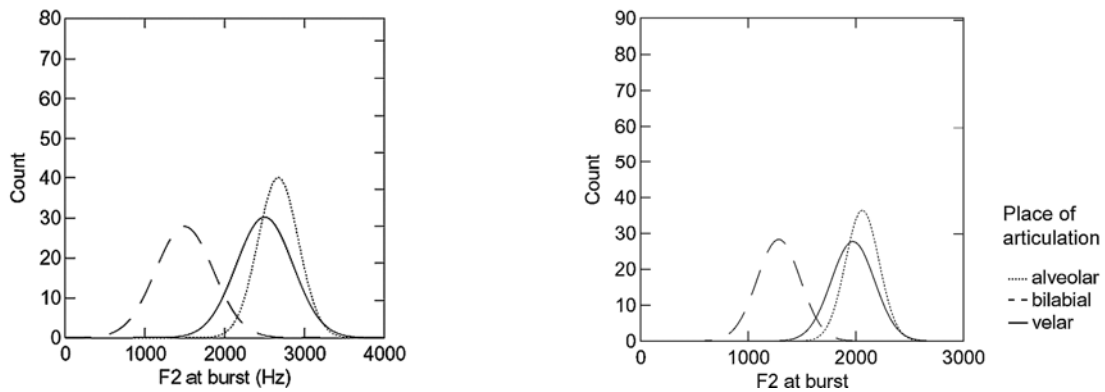


Figure 8. Second formant frequency density plots for children (left) and adults (right). The dashed line represents bilabial F2, the dotted line represents alveolar F2, and the solid line represents velar F2.

Although the overall average results indicate a slightly higher burst F2 value for alveolar than velar stops, for two of the child subjects and three of the adult subjects, average F2 results did not fall in this order, but rather with the velar F2 value slightly higher than the alveolar value. Average F2 values for each subject for each stop place of articulation are listed in Appendix 2 (Table 13, p. 127).

In addition to measuring F2 in the burst, the spectral peak corresponding to primary articulator placement is analyzed. This frequency reflects the length of the cavity in front of the constriction. For bilabials, no spectral peak is expected, and none was noted in the children's stop bursts. For alveolar stops, a spectral peak is expected in the F5-F6 region, while for velars, the spectral prominence is expected to be in the F2 region. There are several steps to determining what frequency range these formant regions correspond to.

First, average F3 in the vowel is computed for each subject. Because F3 remains relatively constant across different vowel productions, it can be used to estimate vocal tract length. Based on vocal tract length information, the formant regions corresponding to a neutral vocal tract are computed, including F2, F5, and F6. The average F3 across all ten children was found to be 4030 Hz. The length of the vocal tract is equal to five times the velocity of sound divided by four times F3. Thus, average vocal tract length is approximately 11 cm for the children. Using this length, neutral vowel formant values for F2, F5, and F6 were computed. ($F2 = (3*c)/(4* l)$, $F5 = (9*c)/(4* l)$, $F6 = (11*c)/(4* l)$, where c = the velocity of sound). Thus the burst spectral prominence for alveolars is expected to be between 7300 and 8900 Hz, and velar burst spectral prominence is expected to be around 2400 Hz. These numbers correspond with the measured values: average alveolar spectral prominence across all 10 child subjects is 8500 Hz, while average velar spectral prominence across all 10 child subjects is 2400 Hz, as predicted (see Figure 9). See Appendix 2, Table 14, p. 128, for individual subject averages. This indicates that children have acquired correct primary articulator placement for stop consonant production.

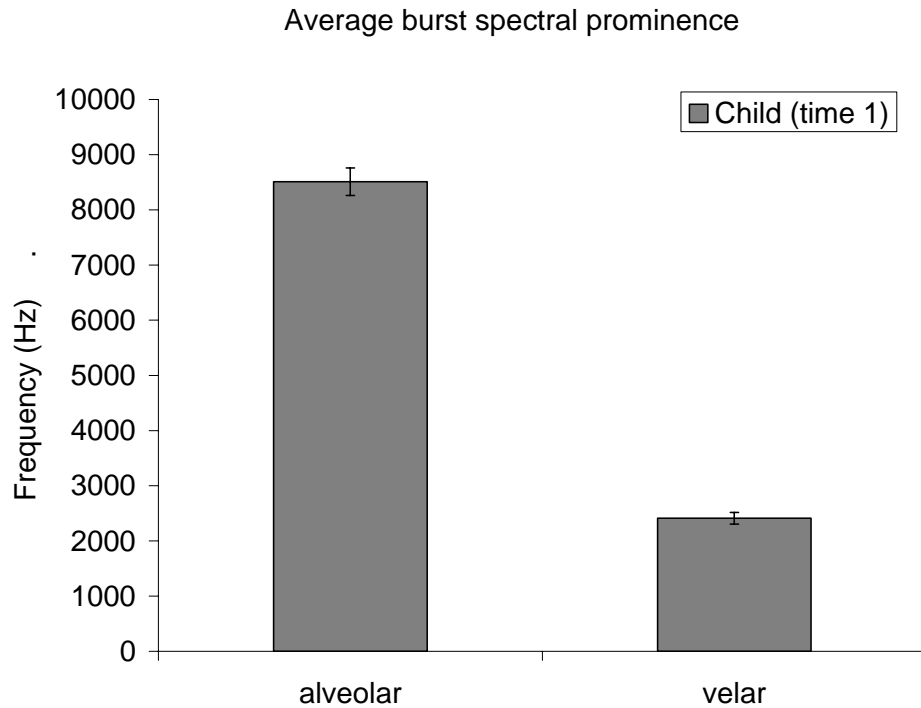


Figure 9. Average prominence in the burst spectrum across all ten children for alveolar and velar stop productions. Data are from the first recording session. Error bars show standard error.

Overall, the results of this hypothesis indicate that the children are producing the correct burst spectrum shape for all three places of articulation. The F2 distribution pattern for children corresponds to the adult pattern, while burst spectral prominences indicate correct primary articulator placement.

8.4.5. Hypothesis 13: Primary articulator placement

Primary articulator placement will be more variable in children (measured for alveolar and velar stops only).

- Since the speech sounds have only recently been acquired in children, it is hypothesized that tongue placement will be more variable as the sounds continue to be fine-tuned. Previous research has found increased variability in children's speech relative to adults' speech.

Primary articulator variability is analyzed by measuring the mean standard deviation in the frequency of the spectral prominence during the burst across all tokens for each place of articulation. Both voiced and voiceless stops consonants are grouped together for each place of articulation. In order to account for the differences in frequency of the burst spectral prominence in adults vs. children, the standard deviation measurements will be normalized by the mean for each subject, such that standard deviation divided by the mean equals the normalized standard deviation.

The hypothesis is that mean normalized standard deviation in spectral prominence across all ten children will be greater than the mean normalized standard deviation in spectral prominence across all ten adults. It is tested using an independent measures t-test to compare the children versus the adults. For both the alveolar and velar places of articulation, $p > 0.1$. Thus, this test revealed no significant difference between the child and adult populations for the normalized standard deviation measurement for either place of articulation. The individual mean, standard deviation, and normalized values are listed in Appendix 2 (see Table 14 and Table 15, p. 128). The subjects are ordered by the normalized value of standard deviation. Some subjects have very different standard deviation values for the two places of articulation. Overall, these results indicate that children's articulator placement is not more variable than adults'.

8.4.6. Hypothesis 14: Multiple bursts

Children will produce a greater number of bursts during stop release than adults.

- Stop burst release requires coordination of articulator movement, opening of the glottis, and changes in stiffness of the vocal tract, glottis, and articulator, all over a period of just a few milliseconds. Children are still learning to speak, and are still learning to control respiration, phonation, and articulation for speech. Differences in the compliance of the articulator, specifically decreased stiffness, could result in a decrease in fine-tuned control over the articulator and increase the probability of multiple bursts. Additionally, children's smaller articulators will have less mass. There will be a threshold of vibration of the articulator depending on stiffness, articulator separation (from the opposing surface), mass, pressure behind the constriction, and rate of movement.

For this hypothesis, the average overall number of bursts produced by children is expected to be greater than the average overall number of bursts produced by adults. Two examples of multiple burst waveforms are shown with their corresponding spectrograms in Figure 10 (p. 85) and Figure 11 (p. 86). This hypothesis is tested using a 2-way analysis of variance (ANOVA) to compare the children versus the adults. This test revealed a significant difference in average overall number of bursts, with children having a greater average number of bursts than adults ($n_{\text{child}} = 1.76$, $n_{\text{adult}} = 1.25$). Additionally, a significant effect was seen for place of articulation, with the greatest number of bursts produced during velars, and the least number of bursts for bilabials (see Appendix 2, Table 16, p. 129).

ANOVA

Independent variable: Age (adult, child)

Independent variable: Place of articulation (bilabial, alveolar, velar)

Dependent variable: Number of bursts

Source	Degrees of freedom	F-ratio	p-value
Age	1	67.005	< 0.001
Place of artic	2	31.873	< 0.001
Age * place of artic	2	4.546	= 0.011

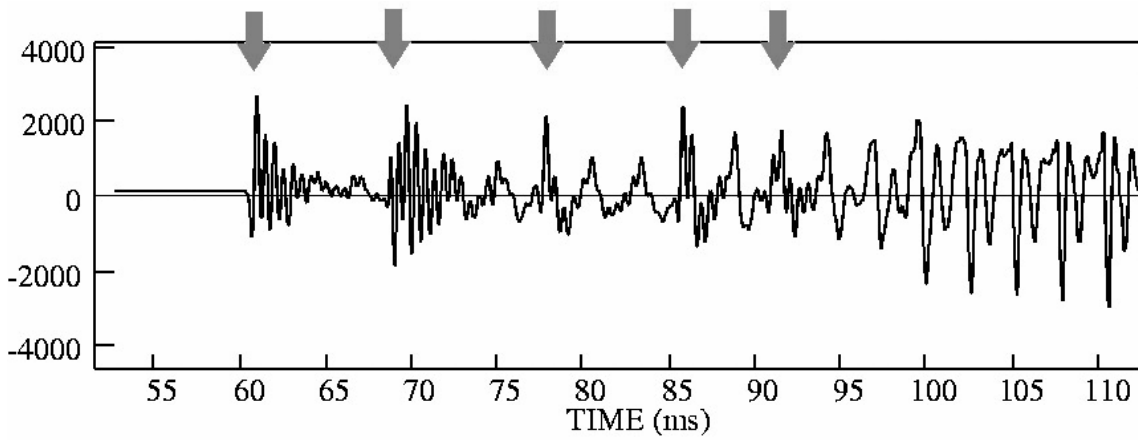
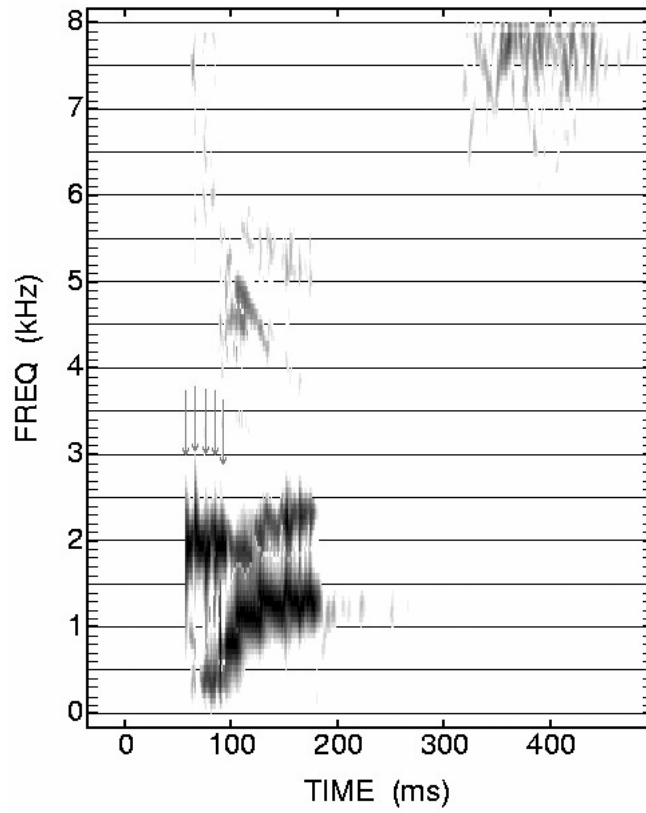


Figure 10. Spectrogram and waveform of a child speaking the word "Gus". Arrows point to each burst recorded for this utterance.

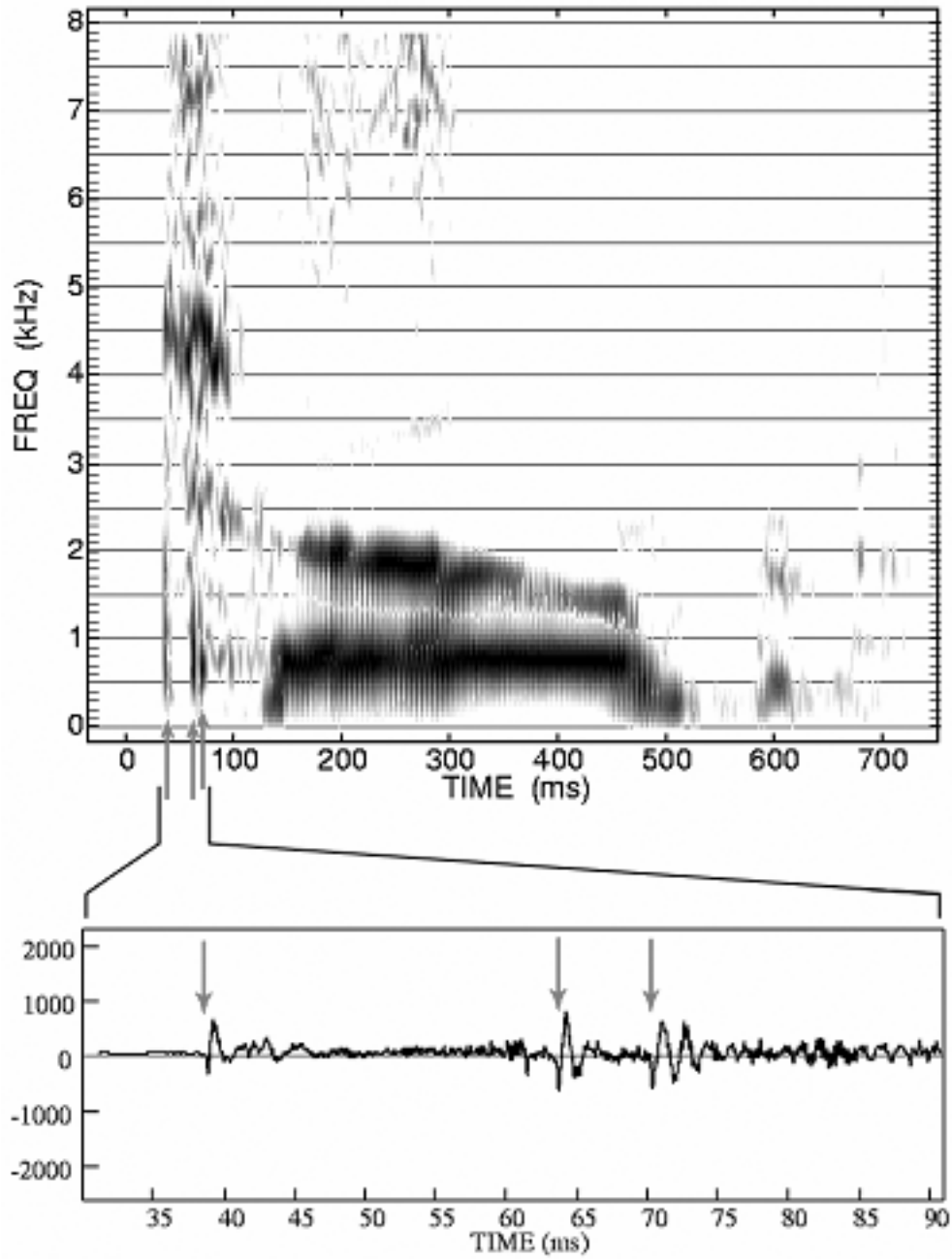


Figure 11. Spectrogram and waveform of the word “tub” spoken by a child. Arrows point to each burst recorded for this utterance.

Because there is a significant difference between the child and adult populations for the average number of bursts measurement, next it must be determined if there is a significant change in this measurement over the six-month period of this study. It is expected that this measurement will decrease toward adult values. That is, the average overall number of bursts produced across all 10 children at Time 1 is expected to be greater than the average overall number of bursts produced across all 10 children at Time 2. This hypothesis was tested using a 2-way analysis of variance (ANOVA) to compare the child populations at these two different points in time. Place of articulation was also included as an independent variable in the analysis to analyze its affect on number of bursts. This test revealed a significant difference in average number of bursts between the two groups ($n_{\text{child1}} = 1.76$, $n_{\text{child2}} = 1.37$), as well as a significant effect for place of articulation.

ANOVA

Independent variable: Age (child time 1, child time 2)

Independent variable: Place of articulation (bilabial, alveolar, velar)

Dependent variable: Number of bursts

Source	Degrees of freedom	F-ratio	p-value
Age	1	53.703	< 0.001
Place of artic	2	46.000	< 0.001
Age * place of artic	2	7.006	= 0.001

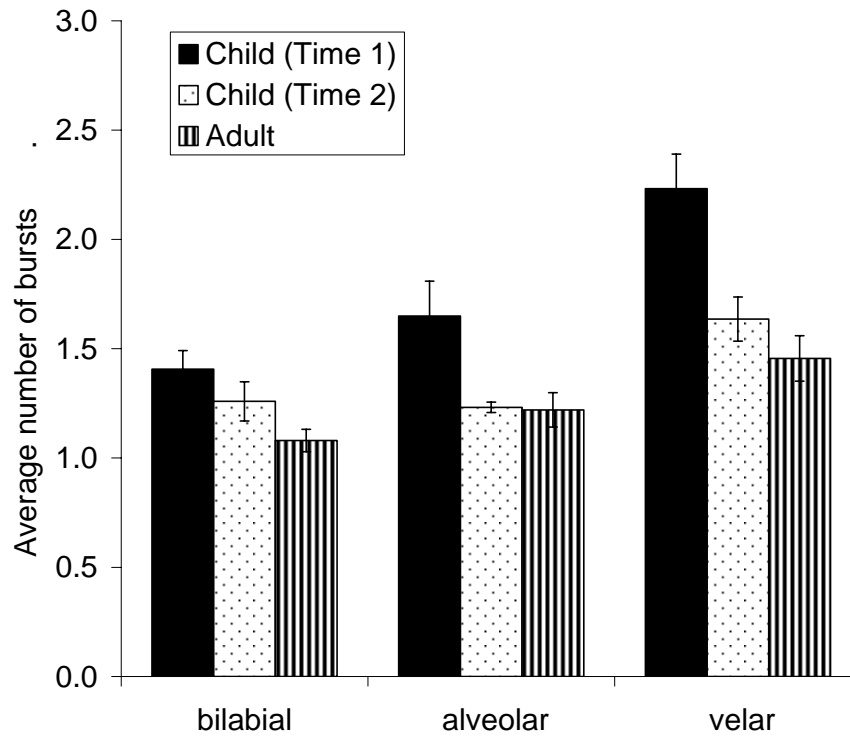


Figure 12. Average number of bursts for each place of articulation for the children at Time 1, six months later at Time 2, and the adult group.

These results indicate that children’s productions have a significantly higher incidence of multiple bursts than adult productions. This could be caused by the physical characteristics of the surface that is contacting the palate (or upper lip for labials), such as increased compliance of the articulator. Another possibility is that the children use a high subglottal pressure at burst release (supported by Netsell, Lotz et al., 1994, who found that preschool children use a higher subglottal pressure than adults during /p/-production.). This high pressure would be followed by a negative pressure that could pull the constriction closed again (Bernoulli’s effect). Additionally, the children’s small vocal tract size and correspondingly short constriction length, small articulator size, and low articulator mass, might result in a higher “natural frequency” of vibration, which would increase the likelihood of multiple bursts, since there would be less time for the child to pull the articulator away from the opposing surface before the next closure. The incidence of multiple bursts is found to decrease significantly over the six-month period of the study. The mass and size of the articulators are not expected to increase greatly

over the 6 month period, indicating that the change in the incidence of multiple burst production must be due to factors other than simple physical growth. Burst duration also decreases, but this change is not significant.

8.4.7. Summary

Analyses of hypotheses and acoustic measurements relating to articulation have led to several significant findings. First, results indicate that children have acquired correct placement of the primary articulator, as inferred from the burst spectrum. However, children are still developing the release mechanism for the primary articulator in the initial part of the release, as evident in the high incidence of multiple bursts and short burst duration in children's productions. Possible causes of these acoustic results include increased compliance of the articulator, a high subglottal pressure, a small articulator size with a corresponding short constriction length and low mass that results in a higher "natural frequency" of vibration, and a vibration of the articulator that can be initiated at a relatively low intraoral pressure. The incidence of multiple bursts is found to decrease significantly over the six-month period of the study. Additionally, children are found to still be acquiring correct tongue body placement and movements. Acoustic results on coarticulation indicate that children use less gestural overlap during bilabial stop production, but more gestural overlap during alveolar stop production, than adults. Adults often use a fronted tongue body position for alveolar stop production, which would bring the tongue further away from the vowel target of a backed tongue body, resulting in less gestural overlap. The results imply that the children have not yet acquired this tongue fronting gesture, and their tongue bodies are in a neutral or backed position during alveolar stop production. These results all indicate that children while children have acquired correct primary articulator placement, they are still developing other features of articulator adjustment for stop consonant production.

8.5. Summary of Results

The analysis of the hypotheses in this section has led to several significant findings, summarized here. The respiration section examines intensity as it may be related to respiration, i.e. subglottal pressure or variation in subglottal pressure. This includes analysis of possible relations of subglottal pressure and intensity to other parameters such as F1 (jaw opening), VOT (influenced by intensity and therefore subglottal pressure), burst duration (influenced by subglottal pressure and intensity), and multiple bursts. No correlation was found between F1 and intensity, VOT and intensity, burst duration and intensity, or number of bursts and intensity. However, children were found to have greater fluctuations in subglottal pressure than adults, as inferred from higher variability in intensity measures. Children were also found to have short burst durations in their production of alveolar and velar stops. Additionally, for voiced stops, child amplitudes increased more than adult amplitudes over the first half of the following vowel, indicating that the children are having some difficulty adjusting the larynx for voicing onset.

In the phonation section, acoustic measurements relating to phonation and laryngeal adjustments revealed several significant results. First, results showing a short VOT-lag for voiced stops in children's productions indicate that children are having some difficulty finding the optimal glottal position and vocal fold tension or stiffness for prompt voice onset following voiced stop release. This is supported by results showing a high amplitude increase over the course of the vowel in children, which may be caused by breathiness at voice onset. Next, for voiceless stop production, children are found to have reduced control over vocal fold stiffness or glottal opening, indicated by a smaller decrease in F0 following voicing onset, and high variability in VOT. These results all indicate that children are still developing appropriate glottal adjustments and intraoral pressure for stop consonant production, including vocal fold stiffness and glottal spreading.

With regard to articulation, children have acquired correct primary articulator placement, as inferred from the burst spectrum, but are still developing the constriction release

mechanism. This is evident in the high incidence of multiple bursts and short burst duration in children's productions. There are several possible causes for the difference in the child's release mechanism, including increased compliance of the articulator, a high subglottal pressure, and a small articulator size with a low mass, which results in a higher "natural frequency" of vibration for the articulator. The high "natural frequency" would allow the child less time to move the articulator away from the opposing surface before the next closure. Additionally, children are found to still be acquiring correct tongue body placement and movements for labials and alveolars. Acoustic results indicate that, compared to adults, children use less gestural overlap in the production of bilabial stops, but more gestural overlap in the production of alveolar stops, into a slightly backed vowel. Since production of alveolar stops often involves a fronted tongue body, these results imply that the children have not fully acquired this tongue fronting gesture. Overall, acoustic analysis of children's articulation indicates that while children have acquired correct primary articulator placement, they are still developing other features of articulator adjustment for stop consonant production, including the articulator release mechanism and tongue body position.

9. Individual Differences

In this section, we perform more in-depth investigation of the most significant findings, analyzing three individual children at each time point, in addition to child and adult averages.

The three children we have chosen to analyze are C1, C5, and C7. C1 is a young girl who was 2;6 at the first recording session, C5 is a young boy who was 2;11 at the first recording session, and C7 is a young boy, who was 3;0 at the first recording session. All three were very cooperative during the sessions and not overly active, resulting in a high number of tokens collected during each session (see Table 2).

Subject	Session	# words extracted	p	b	t	d	k	g
C1	Time 1	66	7	10	10	10	9	10
	Time 2	77	8	6	10	10	9	10
C5	Time 1	66	10	10	10	10	10	9
	Time 2	72	9	10	10	9	9	8
C7	Time 1	68	9	10	8	10	10	10
	Time 2	75	10	10	8	10	10	10

Table 2. Total number of stop consonant initial target words collected, and the number of words analyzed for each stop consonant for subjects C1, C5, and C7.

Some features of the stop consonant-initial words were found to have already been acquired by the children. The most significant of these is that children's positioning of the primary articulator was found to be accurate. This position is estimated from the burst spectrum: the spectral peak reflects the length of the cavity in front of the

constriction. The data from children's speech is similar to adult values in that F2 is excited for velars, F5 or F6 is excited for alveolars, and no major spectral peaks are noted for the labials.

9.1. Amplitude changes: Respiratory control

Respiration is analyzed via acoustic measurements of amplitude and spectral tilt. No significant differences were found between the child and adult populations for spectral tilt. However, changes in amplitude were significantly different in the children as compared with the adult group. Amplitude changes from burst to vowel center are analyzed here for velar stops only, since the spectral shape of the velar burst is more similar to that of the vowel than bilabial or alveolar spectra. Adults have a higher vowel amplitude than burst amplitude for both voiced and voiceless velar stops (see Figure 13). Overall, the child average shows that the children have much less amplitude change for the voiceless stops, although this varies widely among the three individual subjects shown, with C1 and C7 having an overall greater burst amplitude, and C5 having a greater vowel center amplitude similar to the adult average. For voiced stops, the children have less amplitude increase from burst to vowel center than the adult average, but all do show a similar pattern of overall increase.

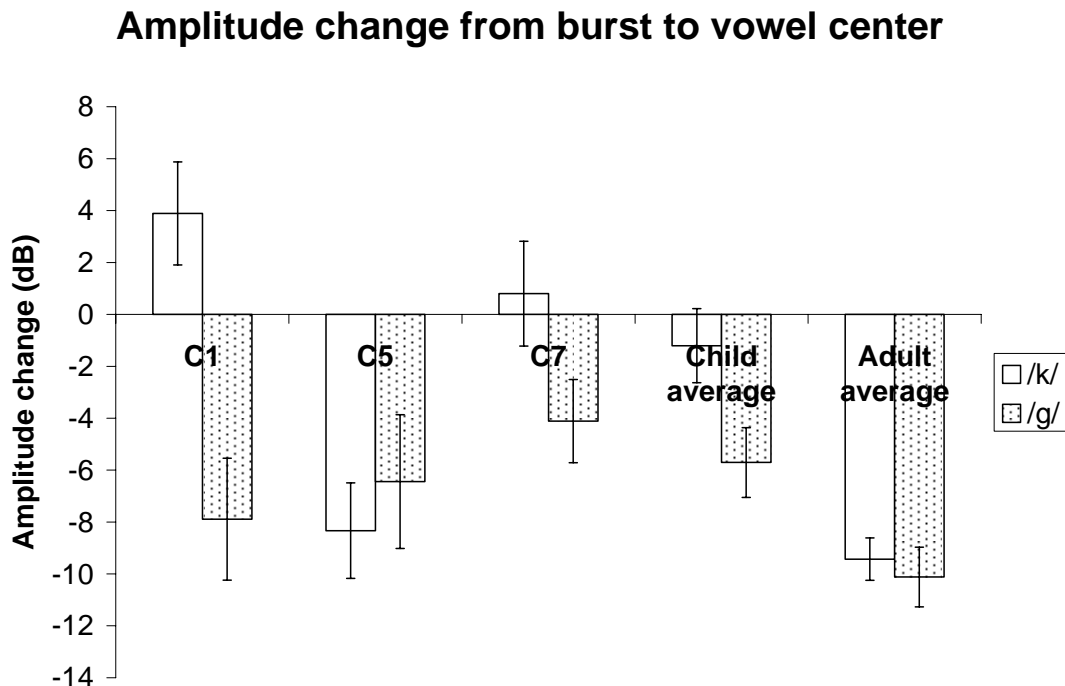


Figure 13. Average amplitude change (burst amplitude – vowel center amplitude) for /k/- and /g/-initial words. Child and adult averages are across all ten subjects in each category. Bars show standard error. Child data from Time 1.

In addition to burst-to-vowel center amplitude measurements, voice onset-to-vowel center amplitude measurements were also made. As reported under Hypothesis 1, an independent measures t-test found that the amplitude increased over the course of the vowel for both children and adults. While voice-onset to vowel center amplitude changes for voiceless stops were not statistically significant, for voiced stops child amplitudes increased significantly more than adult amplitudes ($p = 0.049$), (see Figure 14). Thus, it is likely that children are still developing coordination of voicing onset with stop release for vowels following voiced stops, or that they are not yet able to achieve the correct vocal tract posture to allow for voicing onset at burst release.

C1, C5, and C7 were all found to have a greater amplitude increase from voice onset to vowel center for voiced stops than the adult group. These findings may indicate that for voiced stops, the children have a high subglottal pressure at stop release, resulting in increased airflow and possibly increased breathiness at voice onset.

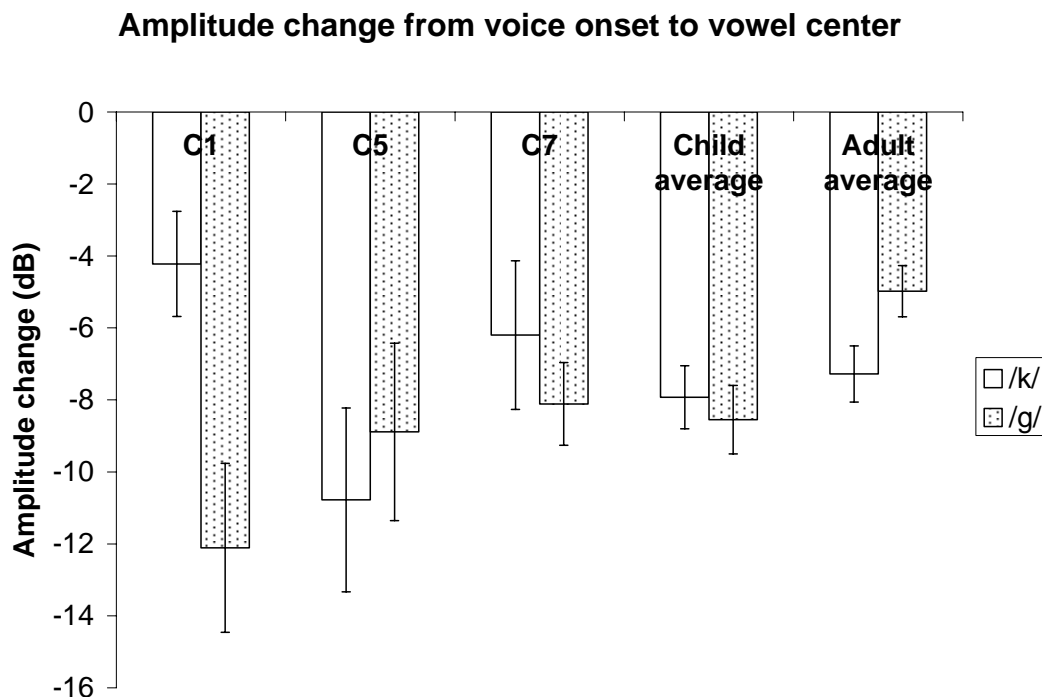


Figure 14. Amplitude change (voice onset amplitude – vowel center amplitude) for /k/- and /g/-initial words. Child and adult averages are across all ten subjects in each category. Bars show standard error. Child data from Time 1.

The last analysis that was performed to evaluate control over subglottal pressure was to measure the range of intensities at vowel center across all tokens, as well as the variability in intensity change from burst to vowel center and from voice onset to vowel center for velar stop-initial words only. For this hypothesis (Hypothesis 1), the mean standard deviation in amplitude and amplitude change for each subject was computed. As reported under Hypothesis 1, an independent measures t-test used to compare the results of standard deviations in vowel center amplitude found a significant difference between the child and adult populations ($p < 0.0001$), with the children having greater variability than the adults. The child population was also compared to the six-month follow-up child data, and no significant change was found in intensity fluctuations between these two recordings ($p > 0.1$). It was also found that children have greater variation in amplitude change from burst to vowel center and from voice onset to vowel center for velar stop-initial words, as shown in Figure 15. These results indicate that the children have less control over subglottal pressure than the adult group.

Amplitude variations

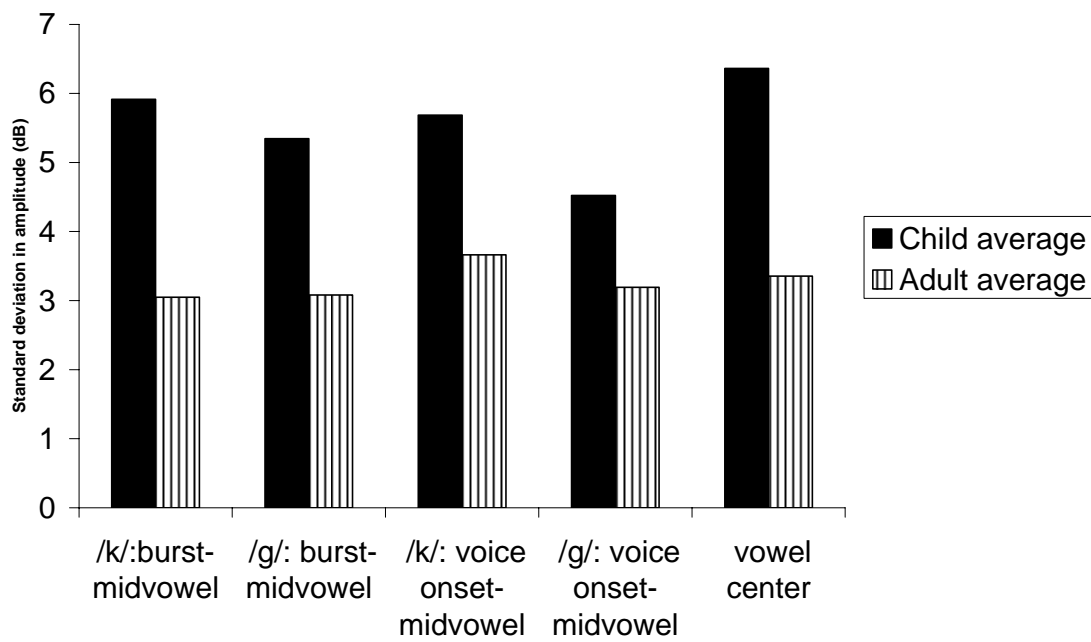


Figure 15. Variations in amplitude change for /k/- and /g/ initial utterances, as well as at vowel center.

9.2. Voiced stops – VOT: Laryngeal control

The first significant finding on phonation is the large difference between burst duration and VOT for voiced stop production in children (see Figure 16, p. 101). Under Hypothesis 6, children were found to have a significantly delayed VOT for all three places of articulation ($p < 0.05$). This “gap” indicates that children are having difficulty coordinating voicing onset with stop release. This is supported by the finding of a high amplitude increase from voice onset to vowel center in children, which may indicate breathiness at voice onset. Evaluation of children’s 6-month follow-up data using an independent measures t-test found a significant decrease in the “gap” duration over the period of the study ($p < 0.1$).

Voiced VOT averages for C1 and C5 increase with place of articulation as expected (see Table 3). Bilabial VOT in C1 is shorter than most from the child group, similar to adult values, while her velar VOT is shorter than most children and adults. The VOT-burst duration differences for C1 for all three places of articulation are shorter than those seen in most of the children, but longer than those recorded for most adults. C5’s VOT-burst duration differences are typical of the child group.

Subject	Bilabial		Alveolar		Velar	
	VOT	VOT-burst dur.	VOT	VOT-burst dur.	VOT	VOT-burst dur.
C1	7.5	0.9	12.6	3.8	15.4	4.2
C5	12.8	3.8	10.9	4.1	20.4	9.8
C7	24.7	13.3	18.7	11.7	20.7	8.7

Table 3. Voiced stop VOT average and VOT-burst duration average for each place of articulation for subjects C1, C5, and C7 at Time 1.

For subject C7, voiced VOT averages do not increase with place of articulation as expected (see Table 3). Bilabial VOT in C7 is much longer than for any other child, and greater than adult values. VOT-burst duration difference for bilabial and alveolar voiced stops in C7 is also longer than for most children. Closer inspection of this child’s individual VOT values for voiced bilabial stops reveals that while half of his 10 /b/-initial

utterances have VOT values between 10 and 12 ms, the other half range from 23 to 52 ms (see Appendix 3, p. 130, for spectrograms of these utterances). This indicates that C7 is able to produce the voiced bilabial stop with the typical child VOT, but often produces it differently, possibly implying that C7 is in the process of learning a new more adult-like vocal tract posture for production of /b/-initial utterances, but is not yet able to achieve this posture consistently.

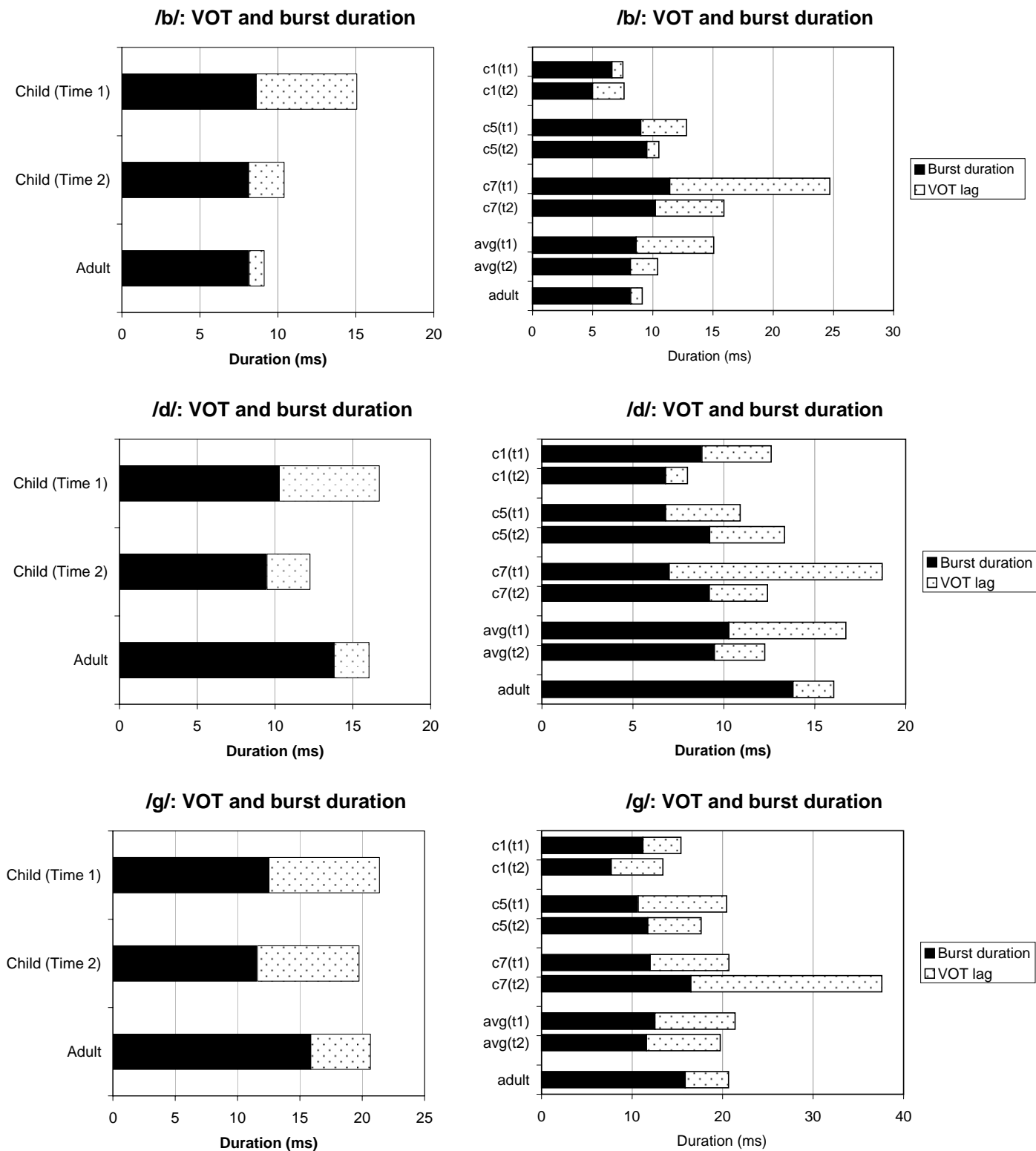


Figure 16. VOT for each voiced stop, broken down to show burst duration and lag before voice onset. The graphs on the left show the overall child average at each recording time, while the graphs on the right include data on individual subjects C1, C5, and C7 (t1 = Time 1, t2 = Time 2). Note the different scales of the x-axes.

9.3. Voiceless stops – F0 and VOT: Laryngeal and respiratory control

Another significant finding relating to phonation in the child group is a smaller average decrease in fundamental frequency following voice onset for voiceless stop-initial words (see Figure 17, p. 103). Following voiceless stops, adults were found to have a high F0 that falls over the first few periods of voicing, reflecting previous findings (Ohde, 1984). The t-test comparing F0 change over the first 20 ms of the utterance found a significant difference between the child and adult populations, when outliers were removed from the data set ($p < 0.1$). While there were no outliers in the adult group, the child values included two outliers (see Hypothesis 7). C1 was one of the outliers, with a large decrease in F0 following VOT. Her F0 decrease is greater than those seen in most adults. The comparatively small VOT-burst duration difference, the large F0 decrease, and the high amplitude increase from voice onset to vowel center all indicate that C1 is adjusting her larynx for earlier onset of voicing despite a high subglottal pressure and high airflow through the glottis. These adjustments result in a higher F0 and increased breathiness at voice onset.

Subjects C5 and C7 also had greater F0 decrease following voice onset than most other children, as can be seen by the low child average value in Figure 17. While C7's values are close to the adult average at both time points, C5's value decreases over the period of the study, away from the adult average (see Figure 17). This is an indication of the high variability in this measure within the child group.

F0 (voice onset) - F0 (voice onset +20 ms), normalized

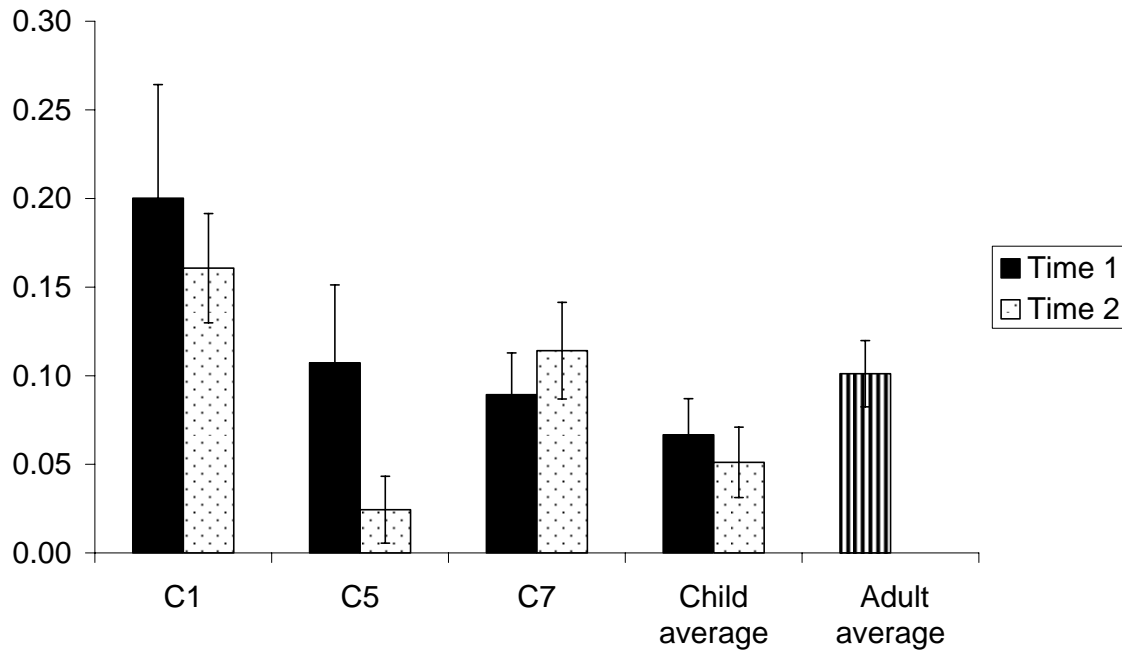


Figure 17. Change in fundamental frequency over the first 20 ms of voicing following a voiceless stop for subjects C1, C5, and C7, as well as the average results. The child and adult averages are across all ten subjects. Bars show standard error.

Another significant finding for voiceless stops is a longer VOT in children for all three places of articulation. For alveolar and velar stops, this value decreases toward adult values over the course of the study (see Figure 18). Children were also found to have high VOT variability at both time points (see Figure 19). These findings both indicate that the children are still learning to coordinate voicing onset with stop release and the offset of aspiration for voiceless stops.

VOT: Voiceless stops

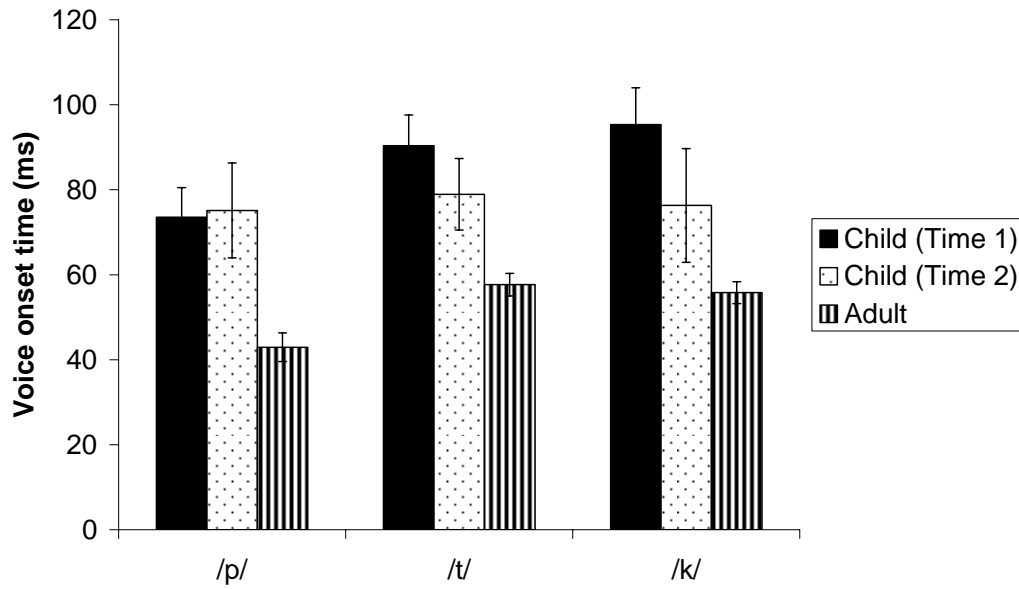


Figure 18. Voice onset time averages for the child group at each time point, compared with adult group averages. Bars show standard error.

VOT variability: Voiceless stops

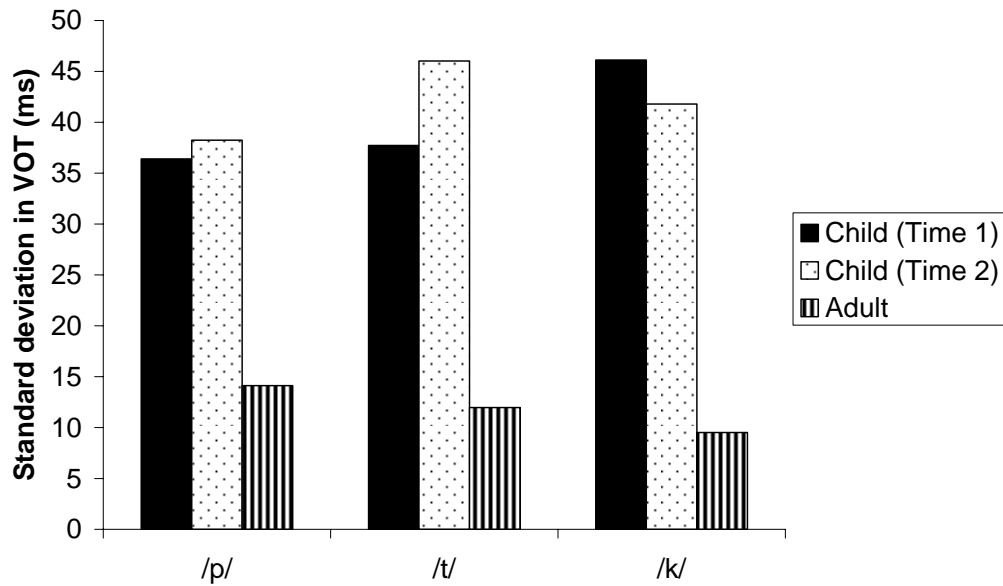


Figure 19. Voice onset time variability averages for the child group at each time point, compared with adult group averages.

9.4. Multiple Bursts: Articulatory properties and control; respiratory control

The first significant finding about articulation was the high incidence of multiple bursts in children as compared with the adult group (see Figure 20, p. 106). The average number of bursts for the child group increased with place of articulation from bilabial to alveolar to velar, as is seen in C1 and C5. C7 had a high average number of bilabial bursts at both recording sessions. The average number of bursts for all three places of articulation decreased over the period of the study toward adult values. The number of utterances for each stop place of articulation produced with a multiple burst are shown in Table 4.

Subject	Session	Bilabial	Alveolar	Velar
C1	Time 1	2/17	6/20	13/20
	Time 2	1/13	5/20	8/19
C5	Time 1	6/20	7/20	15/19
	Time 2	4/19	5/18	9/17
C7	Time 1	9/19	9/18	14/20
	Time 2	9/20	2/18	14/20

Table 4. Number of utterances with multiple bursts out of the total number of utterances analyzed for each stop place of articulation for subjects C1, C5, and C7 at the first recording session (Time 1), and six months later (Time 2)

One possible reason for the high incidence of multiple bursts in the children's speech is increased compliance of the articulator or opposing surface. Just as increased stiffness would help to hold the articulator and opposing surface apart as air flowed through the opening following burst release, increased compliance would increase the likelihood of the articulator's returning to contact the opposing surface after release. This closure then causes a momentary pressure build-up in the vocal tract, pushing the constriction open again, and resulting in a repeated burst release. Further research is needed to address this finding.

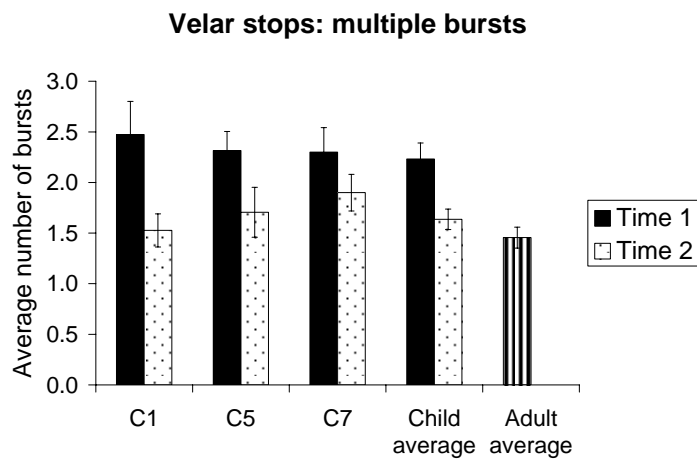
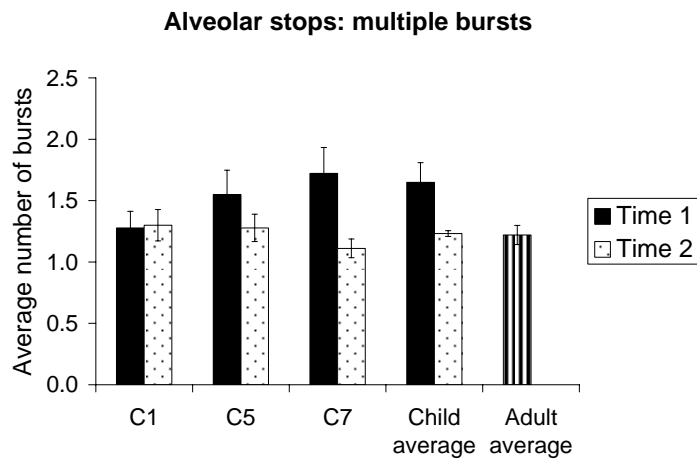
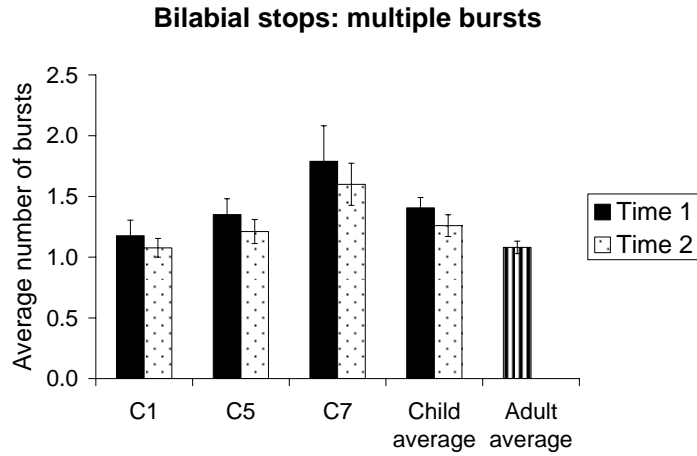


Figure 20. Average number of bursts across both voiced and voiceless stops for each place of articulation for subjects C1, C5, and C7, as well as the overall child and adult group averages. Bars show standard error.

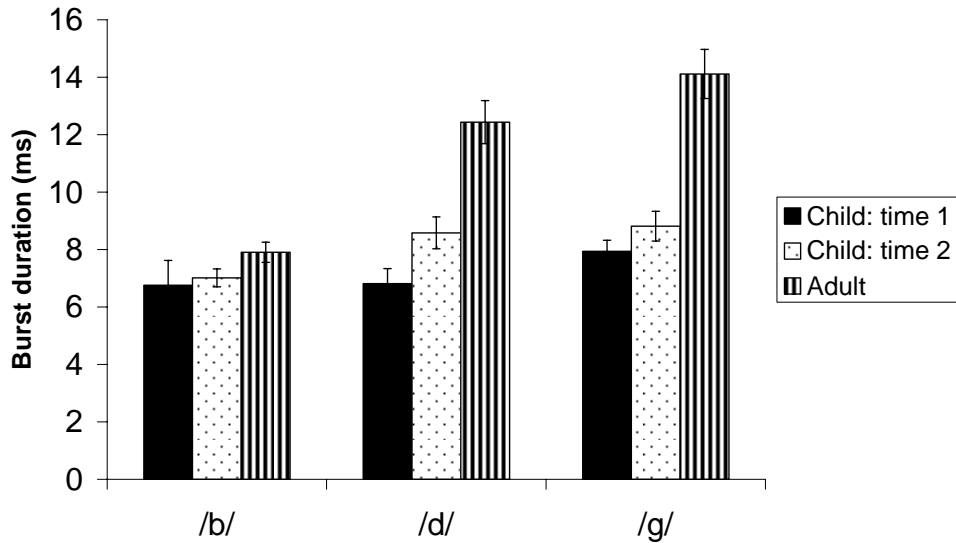
9.5. Burst Duration: Articulatory properties and control; respiratory control

Another significant finding about articulation is the short burst duration in children as compared to adults, for both voiced and voiceless stops (see Figure 21, p. 108). This is most evident for alveolar and velar stops. Note that child /d, g, k/ burst durations increase towards adult values over the 6-month course of the study.

One possible reason for the short child burst durations is their small vocal tract size, and correspondingly short constriction length. The shorter constriction would open more quickly at burst release, resulting in shorter burst durations. Another possibility is that a high subglottal pressure forces the burst open quickly. This is supported by Netsell, Lotz et al. (1994) who found that children use a higher subglottal pressure and have greater airflow than adults during /p/-production.

Burst duration was found to increase for each individual subject C1, C5, and C7 with place of articulation, from labial to alveolar to velar stops, as was the pattern in both the child and adult groups (see Figure 22, p. 109). While neither C1 nor C5 showed a significant change in burst duration toward adult values over the course of the study, C7's values increased toward adult values for all six stops. For /b/, /g/, and /t/-bursts, C1's burst duration actually decreases, away from adult targets, possibly a reflection of a decrease in burst duration measurements before a subsequent rise towards adult values. Subject C5 most notably has a very long /k/-burst duration, but it is important to note that due to a high incidence of voiceless velar bursts in this child, the Time 1 measurement is based on only one utterance, while the Time 2 measurement is based on only two utterances.

Voiced stop burst duration averages



Voiceless stops burst duration averages

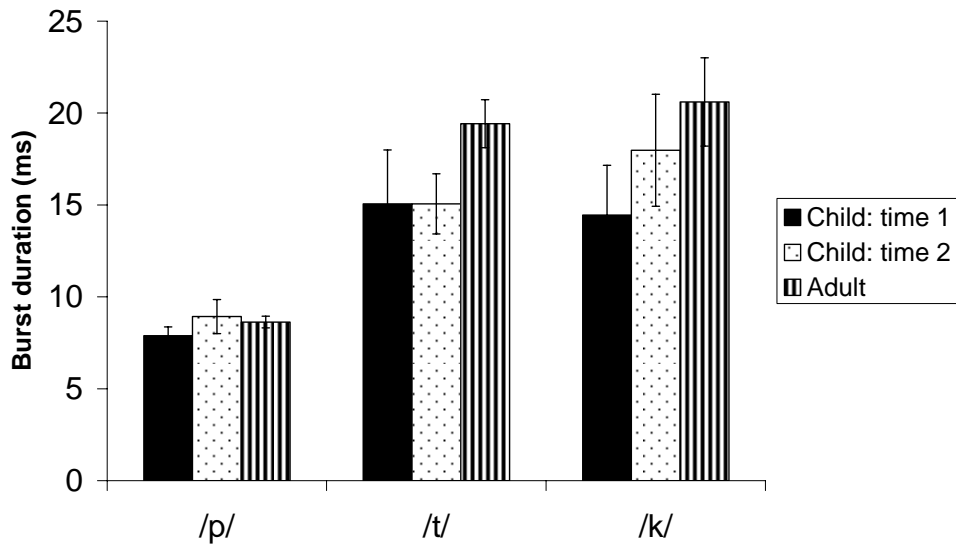


Figure 21. Overall average burst duration for single burst utterances only. Note the difference in y-axis scales. Bars show standard error.

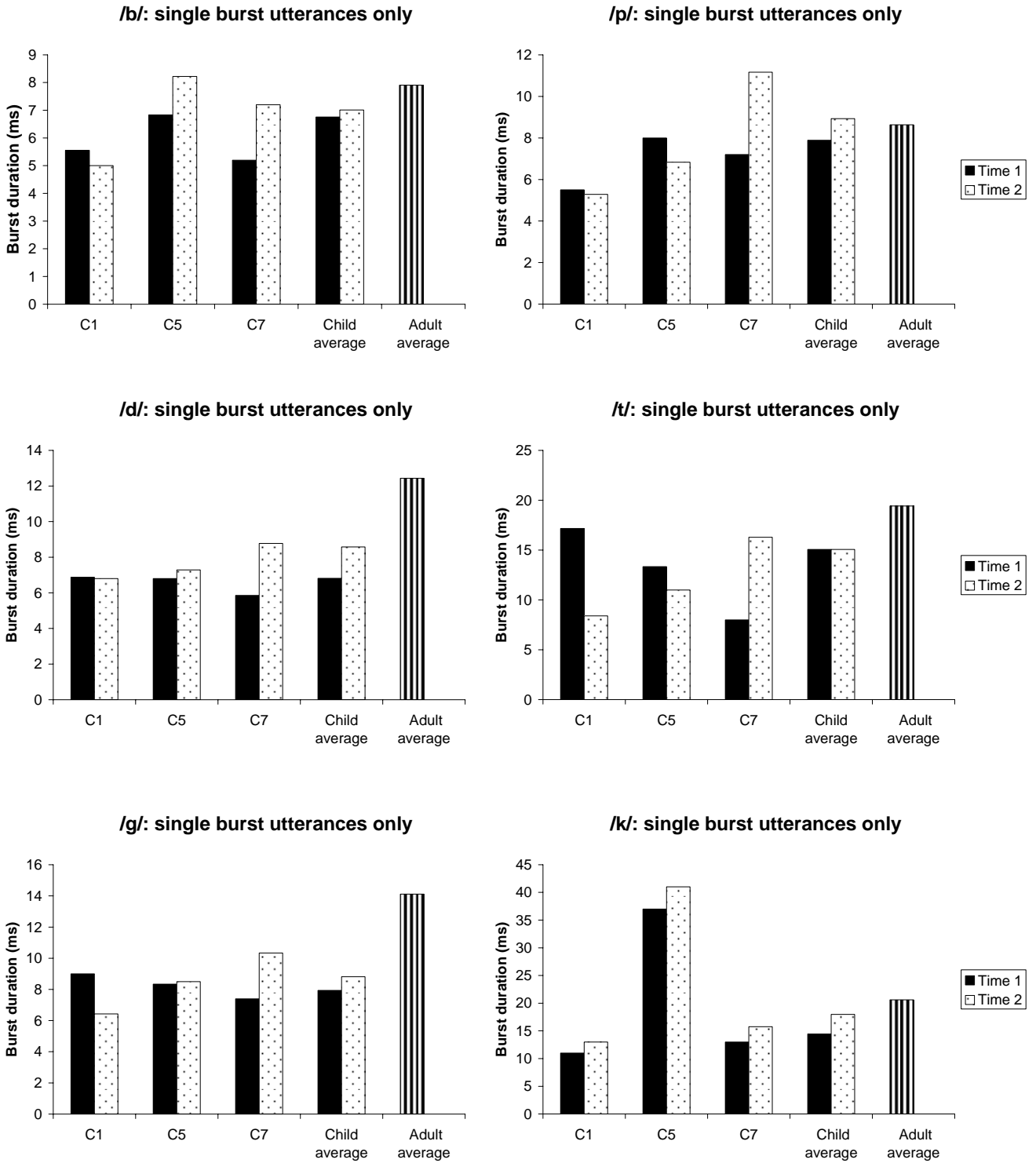


Figure 22. Burst duration for single burst utterances only, for each place of articulation.. Note that for some of the individual children, data is based on only a few utterances. Note the difference in y-axis scales.

9.6. F2 Transitions: Articulatory posture

The last significant finding about articulation is a difference between the child and adult groups in second formant frequency movements after stop release for bilabial and alveolar places of articulation. Analysis of Hypothesis 11 revealed that following bilabial stops, children's productions had a greater increase in second formant frequency transitions from burst to vowel center than adults did, while following alveolar stops (in which F2 is expected to decrease following stop release into a back vowel), the children's productions displayed less F2 decrease than the adult's (see Figure 23). Figure 24 (p. 112) shows the average F2 transitions for both voiced and voiceless stops for the child and adult groups. Note that at vowel center, F2 values for the adult group have converged, while these values are still slightly separated in the child group.

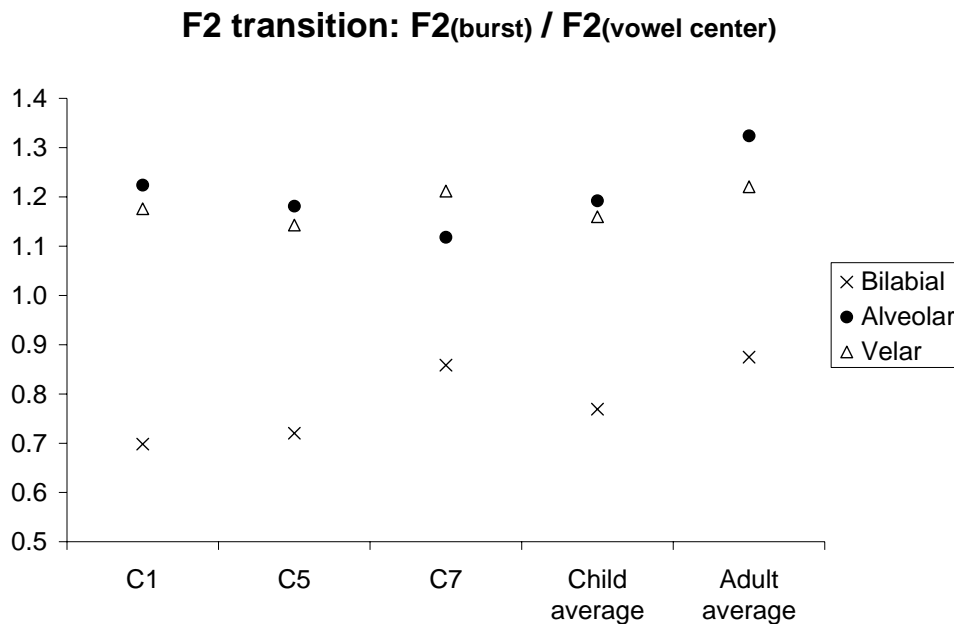
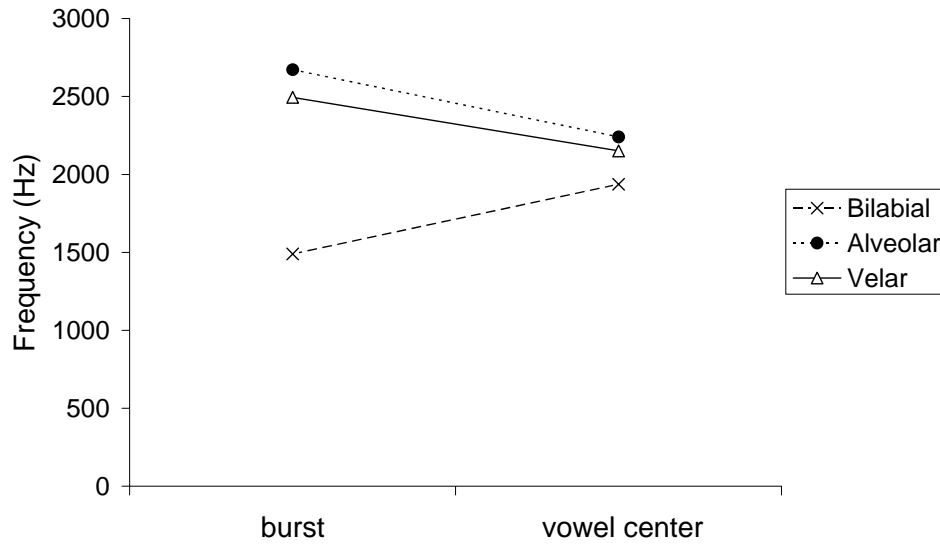


Figure 23. F2 transition ratio ($F2(\text{burst}) / F2(\text{vowel center})$) for each place of articulation for subjects C1, C5, and C7, as well as child and adult averages across all subjects. Words with both voiced and voiceless-initial stop consonants are combined for each place of articulation. Note that the closer the values are to 1.0, the less change there is in F2 from burst to vowel center

These results indicate that for bilabial stop production preceding a back vowel, there is overall less gestural overlap in child speech than in adult speech. However, for alveolar stop production preceding a back vowel, there is more gestural overlap in children. Adults often use a fronted tongue body position as an enhancement gesture for alveolar stop production, which would bring the tongue further away from the vowel target of a backed tongue body, resulting in less gestural overlap. The results imply that the children have not yet acquired this enhancement gesture, and their tongue bodies are in a neutral or backed position during alveolar stop production. These results are solely a reflection of F2 and F2 transitions however, and as reported earlier, the children have acquired correct primary articulator placement. The burst spectral prominence (for alveolars and velars) or spectral shape (for bilabials) does distinguish among the three different places of articulation.

F2 transitions: child average



F2 transitions: adult average

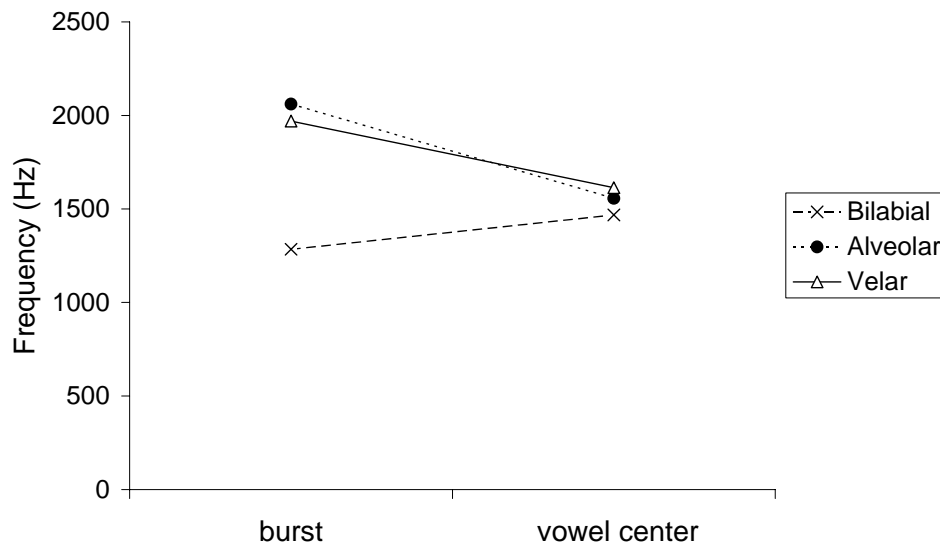


Figure 24. Average second formant frequency values at burst and at vowel center for each place of articulation. The top graph shows child averages; the bottom graph shows adult averages.

9.7. Discussion

The overall results of the study indicate that 2-3-year-old children have not yet completely acquired adult stop consonant production patterns, and the results presented in this section demonstrate some variability in developmental patterns between children. There are several possible causes for the differences seen in children's speech other than the possibility that children are learning how to adjust the larynx and various articulators to more closely approximate an acoustic goal. Young children are still undergoing neuromuscular development, which can affect control of inspiratory and expiratory lung muscles for speech breathing, as well as fine-tuned control of laryngeal, vocal tract, and articulator muscles. Variability in children's speech can also be caused by prosodic differences resulting from their increased excitement and activity level. Additionally, coarticulation may not be an acquired or learned process, but rather a by-product of fluent speaking produced in service of an acoustic goal. Thus, while children may still be acquiring correct stop consonant production, this may be due to a number of developmental factors and does not necessarily indicate that they are "learning" correct production. The pattern of development of these diverse factors may differ among children, resulting in the variability seen within the child population. Future research on children's speech in more controlled conditions may help distinguish between some of these possible factors in child speech production.

10. Summary and Conclusions

Analysis of acoustic data on the speech of young children indicates that the stop consonants are still developing at 2-3 years of age. The acoustic consequences of the speech gestures for stop consonant production have not yet reached adult target values, and the child continues to make spatial and temporal modifications in the articulatory pattern. The results of this study can be summarized in five principle findings, as described below. These findings fall into two main categories: the first two findings relate to managing the flows and pressures for obstruent production, while the last three findings relate to positioning the structures for filtering of the source.

First, acoustic measurements showed that the *positioning of the primary articulator* appears to be correct. From measurements of the burst spectrum, the frequency of the spectral peak in the stop burst is in the expected range: F2 is excited for velars, F5 or F6 is excited for alveolars, and no major spectral peaks were noted for labials.

The second finding relates to *tongue body adjustment* during stop production, as estimated from formant transitions, particularly F2. These transitions show some differences from normal adult values. For bilabial stops, children's F2 transitions indicate that there is less gestural overlap between the stop and the following vowel than is seen in adults—that is, the vowel is not anticipated as much in children. The F2 transition for alveolars in children is very similar to velar F2 transitions. Children's alveolar F2 transitions indicated that they use more gestural overlap between the stop production and the following backed vowel than adults. Adults often use a fronted tongue body position for alveolar stop production, which would bring the tongue further away from the vowel target of a backed tongue body, resulting in less gestural overlap.

The results imply that the children have not yet acquired this tongue fronting gesture, and their tongue bodies are in a neutral or slightly backed position during alveolar stop production.

The next significant finding is that children are still acquiring the *release mechanism for the primary articulator*, as evidenced by the temporal shape of the burst in the initial part of the release. This observation is supported by results showing a high incidence of multiple bursts and a short burst duration. This could be caused by the physical characteristics of the surface that is contacting the palate (or upper lip for labials), such as increased compliance of the articulator, or it could be due to higher subglottal pressure, or to the force of the articulator against the opposing surface. Other possible explanations for these findings include the smaller size of their articulators, and their lower mass, and thus their higher “natural frequency” of vibration, leading to multiple bursts. The incidence of multiple bursts is found to decrease significantly over the six-month period of the study. Assuming articulator mass and size are not changing appreciably over this time interval, this decrease may be attributed to changes in articulator compliance or subglottal pressure. Burst duration also decreases, but this change is not significant.

Another significant result of this study is the finding that children are still developing the appropriate *glottal adjustments and intraoral pressure* for stop consonant production, including vocal fold stiffness and glottal spreading. This was determined from voice onset time and fundamental frequency measurements. For children’s voiced stop productions, there was a lag following the burst release before voicing onset. This indicates that children are still acquiring the coordination of voicing onset with stop release for vowels following voiced stops (via vocal fold stiffness and separation), or are having difficulty achieving the correct vocal tract posture before burst release. This is supported by the finding that amplitude increases over the course of the vowel following voiced stops in children significantly more than in adult speech, which may be caused by breathiness at voice onset. For voiceless stops, it was found that children produce less F0 decrease over the first 20 ms of voicing than adults do. Additionally, children’s

productions have a long but highly variable VOT. During voiceless stop production in adults, the vocal tract and vocal folds are held stiff during stop release, frication, and aspiration to prevent the onset of voicing. This stiffness in the vocal folds results in a high F0 at voice onset, which rapidly decreases. The results indicate that children may not be able to consistently coordinate the onset of voicing with the offset of aspiration, or may have reduced control over vocal fold stiffness or glottal opening, resulting in a smaller decrease in F0 following voicing onset, and high variability in VOT.

The last principal finding of this study is that children have less *control over subglottal pressure*, as evidenced by a high variability in amplitude measurements. This is presumably a problem independent of which segments are being produced, especially at the beginning of an utterance. Subglottal pressure is controlled by balancing the expiratory muscles and recoil forces of the lungs with the inspiratory muscles and the glottal opening. Results indicate that children are still acquiring the correct balance of these different respiratory aspects for speech. Children use a high subglottal pressure at burst release for both voiced and voiceless stops, and there may be a fall in subglottal pressure following the initial burst for stops.

In addition to the findings listed above, acoustic measurements from children's utterances were found to have *increased variability* as compared to adult values. This has been found in several other studies of children's speech. Measurements found to have increased variability in children include VOT, fundamental frequency, F2 at vowel center, and intensity at vowel center. Possible reasons for the increase in variability include increased activity during the recording session, increased excitement level, and decreased control over respiration, phonation, and articulation.

In conclusion, interpretations of acoustic data have helped to determine which aspects of speech production the children are still acquiring. For example, the children had problems with control of subglottal pressure. This could be due to difficulties manipulating the balance between inspiratory and expiratory muscles necessary to maintain a constant pressure. Children also had trouble finding the optimal glottal

position and vocal fold tension or stiffness for prompt voice onset following voiced stop release. Additionally, children had difficulty with articulator control, resulting in a short burst duration and a high incidence of multiple bursts. These results indicate that further analysis of longitudinal data on young children will result in more refined models of the development of the coordination of articulation, phonation, and respiration for motor speech production.

11. Appendix

11.1. Appendix 1: Target word list

Initial Position:

/p/	/b/	/t/	/d/	/k/	/g/
puppy	bug	tub	duck	cup	Gus
	bee		dog	cat	goose

Final Position:

/p/	/b/	/t/	/d/	/k/	/g/
cup	tub	cat	mud	duck	bug

Medal Position:

/p/	/b/	/t/	/d/	/k/	/g/
puppy	baby	kitty	daddy	cookie	Maggie

Diphthong:

Vowel	
/ɔi/	boy
/ai/	pie
/au/	cow

11.2. Appendix 2: Statistical Analyses Tables

Subject	Voiced stop average (dB)	Voiceless stop average (dB)
C1	-8.41	-8.75
C2	-4.58	-5.35
C3	-9.81	-12.05
C4	-6.61	-7.05
C5	-6.45	-10.77
C6	-5.36	-8.41
C7	-8.50	-6.31
C8	-2.06	-5.21
C9	-10.71	-5.42
C10	-7.32	-10.96
Child average	-6.98	-8.03
A1	-5.73	-7.07
A2	-3.33	-7.93
A3	-5.07	-3.64
A4	-4.00	-7.13
A5	-5.67	-8.47
A6	-4.40	-8.07
A7	-2.80	-8.00
A8	-5.93	-10.13
A9	-7.73	-3.93
A10	-5.13	-10.80
Adult average	-4.98	-7.52

Table 5. Amplitude differences (in dB) from voice onset to vowel center, discussed in Hypothesis 1.

Subject	Standard deviation in amplitude (dB)
C1	5.49
C2	4.31
C3	4.82
C4	5.79
C5	6.67
C6	6.58
C7	6.85
C8	4.96
C9	7.55
C10	6.62
Child average	5.96
A1	3.51
A2	2.93
A3	3.65
A4	3.94
A5	2.27
A6	3.88
A7	3.36
A8	2.55
A9	5.45
A10	3.04
Adult average	3.46

Table 6. Average standard deviation in amplitude (dB) at vowel center, discussed in Hypothesis 1.

subject	bilabial		alveolar		velar	
	VOT	VOT-burst dur.	VOT	VOT-burst dur.	VOT	VOT-burst dur.
C1	7.5	0.9	12.6	3.8	15.4	4.2
C2	22.4	16.2	29.8	7.8	27.1	16.0
C3	19.7	3.8	19.6	8.3	13.0	2.0
C4	24.0	14.6	24.2	13.1	29.4	18.5
C5	12.8	3.8	10.9	4.1	20.4	9.8
C6	7.7	1.1	12.1	2.0	17.8	2.9
C7	24.7	13.3	18.7	11.7	20.7	8.7
C8	11.4	3.1	12.2	6.0	30.3	10.0
C9	13.9	6.3	10.3	4.8	14.2	4.9
C10	6.6	1.3	16.7	2.7	25.5	11.7
A1	11.2	0.8	9.6	1.2	16.4	2.0
A2	9.2	2.0	21.4	2.2	20.0	2.2
A3	8.2	0.8	19.4	2.6	36.6	12.6
A4	7.4	0.6	17.2	1.2	17.2	3.8
A5	8.2	0.6	17.0	3.0	24.4	4.8
A6	13.4	4.2	15.8	2.4	18.0	5.8
A7	8.8	0.2	19.4	6.0	22.8	6.4
A8	6.6	0.0	11.8	0.6	13.6	0.0
A9	7.6	0.0	11.8	0.0	18.8	1.2
A10	10.6	0.4	17.0	3.2	18.8	9.0

Table 7. Average values of voice-onset time (VOT) in ms, and difference between VOT and burst duration in ms for voiced stops for each place of articulation for each subject.

Subject	(A1-A3) at VOT	(A1-A3) at VOT+20 ms	Average tilt difference
C1	28.26	28.12	0.14
C2	30.62	34.50	-3.88
C3	26.44	26.30	0.14
C4	30.01	31.07	-1.06
C5	31.50	30.07	1.43
C6	29.63	26.52	3.11
C7	30.83	33.62	-2.79
C8	28.08	26.65	1.43
C9	25.27	25.79	-0.52
C10	26.28	28.43	-2.15
A1	19.14	19.44	-0.3
A2	18.71	17.68	1.03
A3	22.44	19.37	3.07
A4	21.77	18.66	3.11
A5	24.51	22.03	2.48
A6	19.89	20.28	-0.39
A7	22.46	19.71	2.75
A8	20.71	17.31	3.4
A9	19.88	20.32	-0.44
A10	24.63	21.56	3.07

Table 8. Spectral tilt measurements (A1-A3) in dB, measured at voice onset time and 20 ms later, averaged across all voiced stop consonants for each subject. Child data is from the first recording session.

Subject	(A1-A3) at VOT	(A1-A3) at VOT+20 ms	Average tilt difference
C1	23.91	27.07	-3.33
C2	18.40	23.26	-4.36
C3	24.47	29.79	-5.40
C4	22.69	27.69	-4.98
C5	18.72	25.74	-7.61
C6	18.35	21.56	-4.73
C7	20.74	31.22	-9.29
C8	20.76	29.14	-8.48
C9	28.81	30.23	-1.42
C10	24.94	31.31	-6.33
A1	18.69	21.26	-2.57
A2	8.01	14.91	-6.90
A3	16.73	13.65	3.08
A4	13.38	19.44	-6.06
A5	16.09	20.21	-4.13
A6	10.41	14.21	-3.45
A7	21.94	21.66	0.28
A8	14.50	18.85	-4.35
A9	20.79	19.18	1.61
A10	13.43	23.01	-9.58

Table 9. A1-A3 values at VOT and at VOT + 20 ms averaged across all voiceless stops for each subject. Tilt differences are calculated for each voiceless stop initial utterance and then averaged to result in the average tilt difference value shown.

Subject	Burst F2	Burst F3	Vowel center F2	Vowel Center F3
C1	2200	3951	1894	4713
C2	2771	3591	2266	4124
C3	2626	3619	2188	4118
C4	2269	3602	1948	4137
C5	2651	3528	2330	3918
C6	2399	4096	2233	4421
C7	2673	3270	2207	3671
C8	2316	3929	2011	4480
C9	2692	3642	2117	3869
C10	2351	3466	2316	4263
A1	1984	2343	1524	2634
A2	1926	2342	1599	2633
A3	1747	2217	1549	2582
A4	2120	2488	1651	2469
A5	2036	2410	1636	2714
A6	2125	2595	1834	2796
A7	1889	2191	1522	2743
A8	1862	2224	1681	2483
A9	2100	2389	1558	2652
A10	1911	2336	1584	2602

Table 10. Second and third formant values in the burst and at vowel center averaged across all words with word-initial velar stop consonants, in child and adult subjects.

Subject	F1 average	F3 average	F1/F3
C1	1051	4569	0.23
C2	1027	4094	0.25
C3	1237	4218	0.29
C4	935	4168	0.23
C5	1136	4022	0.29
C6	1149	4460	0.26
C7	943	3853	0.28
C8	932	4498	0.21
C9	1131	3962	0.29
C10	1040	4227	0.25
A1	717	2702	0.27
A2	729	2687	0.27
A3	791	2667	0.30
A4	772	2630	0.29
A5	803	2775	0.29
A6	816	2832	0.29
A7	806	2861	0.28
A8	802	2602	0.31
A9	794	2741	0.29
A10	769	2717	0.28

Table 11. Average F1 and F3 values at vowel center for each subject, as well as the average F1/F3 ratio.

	Bilabial stops			Alveolar stops			Velar stops		
	F2 burst	F2 vowel	ratio	F2 burst	F2 vowel	ratio	F2 burst	F2 vowel	ratio
C1	1235	1784	0.70	2514	2100	1.22	2200	1894	1.18
C2	1687	1964	0.87	2829	2268	1.26	2771	2266	1.23
C3	1393	1954	0.73	2800	2337	1.21	2626	2188	1.20
C4	1645	1876	0.87	2406	1997	1.22	2269	1948	1.17
C5	1617	2263	0.72	2851	2428	1.18	2651	2330	1.14
C6	1863	2150	0.61	2548	2332	1.10	2399	2233	1.08
C7	1563	1876	0.86	2623	2368	1.12	2673	2207	1.21
C8	1310	1687	0.78	2770	2201	1.27	2316	2011	1.15
C9	1261	1951	0.65	2604	2149	1.22	2692	2117	1.28
C10	1328	1870	0.71	2777	2231	1.27	2351	2316	1.02
A1	1296	1318	0.99	1941	1571	1.24	1984	1524	1.31
A2	1438	1448	1.00	2087	1566	1.34	1926	1599	1.21
A3	1261	1532	0.82	2106	1553	1.36	1747	1549	1.13
A4	1372	1459	0.94	2089	1582	1.32	2120	1651	1.29
A5	1270	1487	0.86	2034	1581	1.29	2036	1636	1.25
A6	1468	1568	0.94	2162	1732	1.25	2125	1834	1.16
A7	1165	1387	0.84	1909	1484	1.29	1889	1522	1.24
A8	1283	1519	0.85	2177	1595	1.37	1862	1681	1.11
A9	1218	1471	0.83	2066	1472	1.41	2100	1558	1.35
A10	1070	1488	0.72	2038	1428	1.43	1911	1584	1.21

Table 12. Second formant frequency measurements at burst, in the following vowel, and the ratio of F2(burst) / F2(vowel) for all three stop places of articulation.

subject	bilabial F2 at burst	alveolar F2 at burst	velar F2 at burst
C1	1235	2514	2200
C2	1687	2829	2771
C3	1393	2800	2626
C4	1645	2406	2269
C5	1617	2851	2651
C6	1863	2548	2399
C7	1563	2623	2673
C8	1310	2770	2316
C9	1261	2604	2692
C10	1328	2777	2351
A1	1296	1941	1984
A1	1438	2087	1926
A3	1261	2106	1747
A4	1372	2089	2120
A5	1270	2034	2036
A6	1468	2162	2125
A7	1165	1909	1889
A8	1283	2177	1862
A9	1218	2066	2100
A10	1070	2038	1911

Table 13. Average F2 values (in Hz) for each stop place of articulation for each subject.

alveolar stops				velar stops			
Child	burst prominence frequency (Hz)	std dev	normed value	Child	burst prominence frequency (Hz)	std dev	normed value
C8	8432	212	0.025	C1	2028	144	0.071
C6	8570	491	0.057	C6	2283	206	0.090
C9	8444	493	0.058	C3	2380	233	0.098
C3	8621	730	0.085	C8	2284	282	0.124
C7	8548	867	0.101	C4	2232	305	0.137
C4	8222	895	0.109	C7	2442	341	0.140
C2	8885	1000	0.113	C9	2734	417	0.153
C10	9056	1071	0.118	C5	2552	401	0.157
C5	8124	1002	0.123	C10	2316	401	0.173
C1	8216	1086	0.132	C2	2853	598	0.210

Table 14. Average frequency of the spectral prominence in alveolar and velar stop bursts, and standard deviations of these averages for each child subject. The normed value is the normalized standard deviation, which equals the standard deviation divided by the average. The subjects are listed in order from least to most variable for each place of articulation.

alveolar stops				velar stops			
Adult	burst prominence frequency (Hz)	std dev	normed value	Adult	burst prominence frequency (Hz)	std dev	normed value
A8	5700	328	0.058	A2	1938	109	0.056
A7	5644	423	0.075	A5	2006	181	0.09
A9	5085	425	0.084	A9	2250	200	0.089
A10	5684	492	0.087	A6	2041	201	0.099
A4	5328	531	0.100	A8	1853	193	0.104
A1	6250	657	0.105	A4	2282	244	0.107
A3	4600	567	0.123	A7	1888	220	0.117
A6	4991	672	0.135	A3	1701	243	0.143
A5	5572	790	0.142	A1	2010	303	0.151
A2	5753	1019	0.177	A10	2035	366	0.180

Table 15. Average frequency of the spectral prominence in alveolar and velar stop bursts, and standard deviations of these averages for each adult subject. The normed value is the normalized standard deviation, which equals the standard deviation divided by the average. The subjects are listed in order from least to most variable for each place of articulation.

subject	bilabial	alveolar	velar
C1	1.18	1.28	2.47
C2	1.57	2.67	2.4
C3	1.79	2.07	2.06
C4	1.41	1.33	1.68
C5	1.35	1.55	2.32
C6	1.1	1.83	2
C7	1.79	1.72	2.3
C8	1.6	1.1	3.18
C9	1.08	1	1.36
C10	1.2	1.94	2.55
A1	1.5	1.3	1.9
A2	1	1.7	1.6
A3	1	1	2
A4	1	1.6	1.2
A5	1.2	1.2	1.7
A6	1	1.1	1.2
A7	1	1	1.44
A8	1	1.1	1.11
A9	1	1	1.3
A10	1.1	1.2	1.1

Table 16. Average number of bursts produced by each child for each place of articulation.

11.3. Appendix 3: Spectrograms

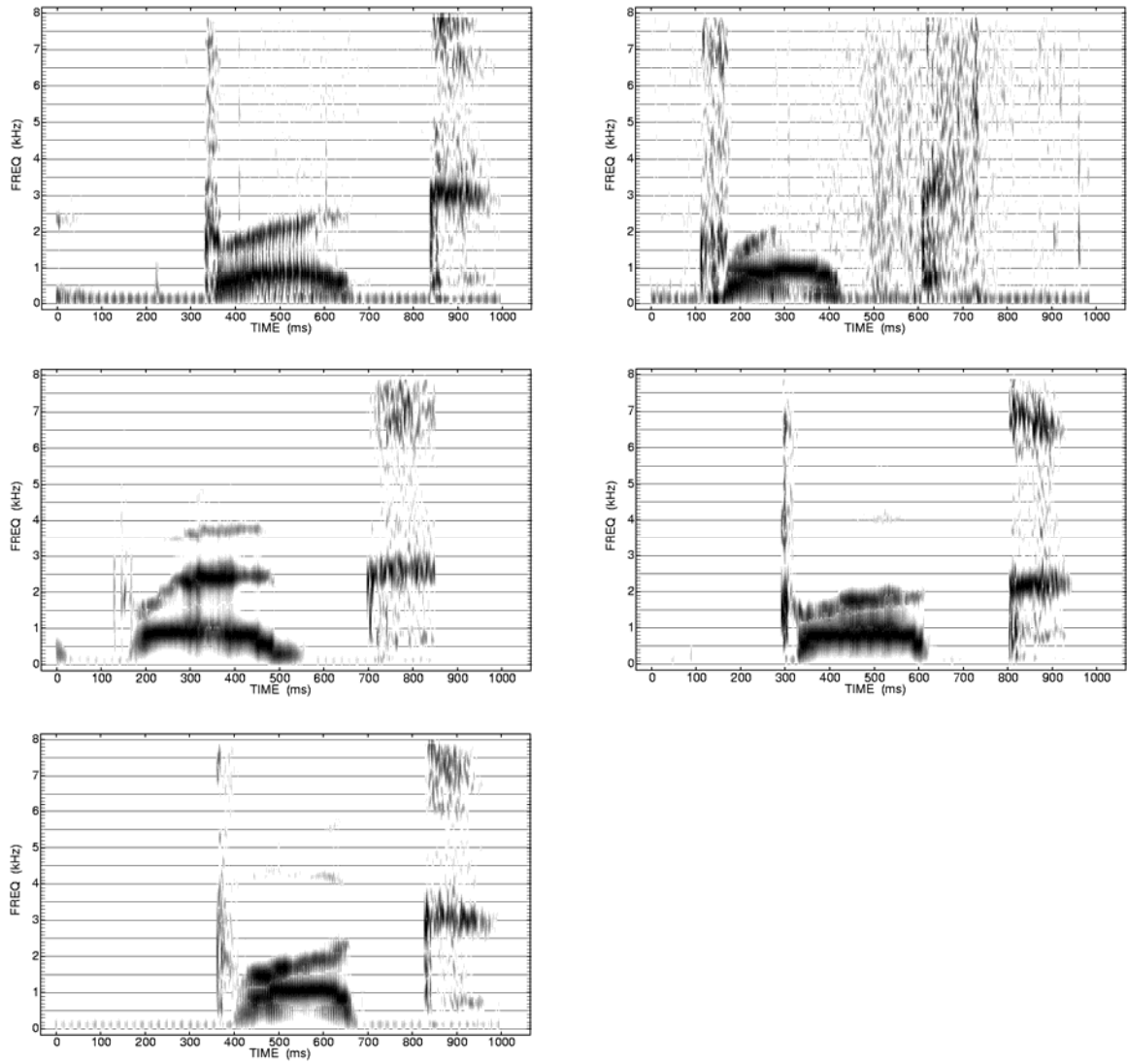


Figure 25. Five example spectrograms of the word “bug” spoken by C7. Note the long VOT following burst release.

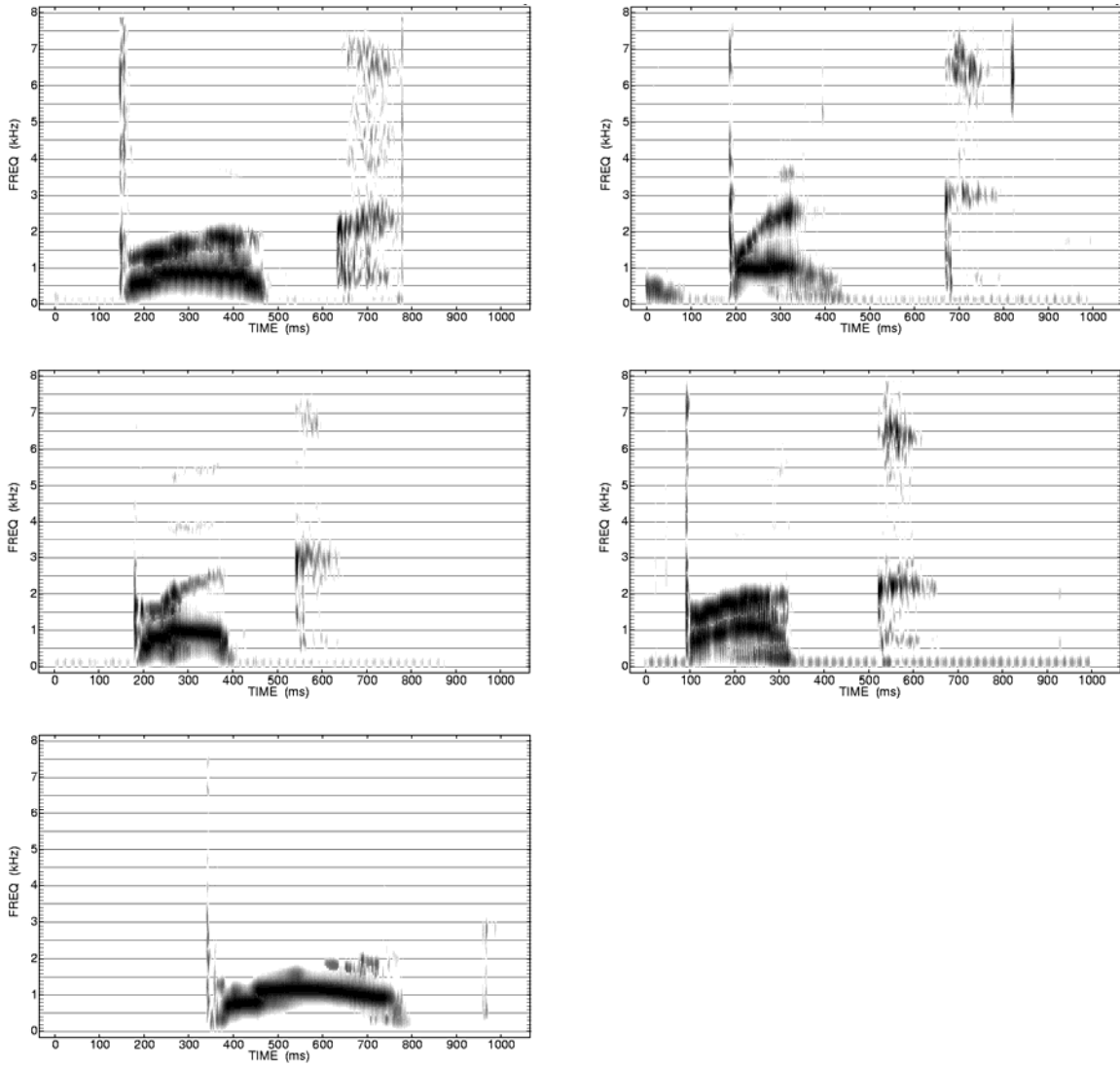


Figure 26. Five example spectrograms of the word “bug” spoken by C7. Note the short VOT following burst release.

11.4. Appendix 4: Data review for each child subject

Child #1

Time 1

measure	mean	standard dev	low value	high value
number of words	9.3		5	10
voice onset time (voiced)	11.8	7.6	3	42
voice onset time (voiceless)	92.6	40.5	38	184
vowel duration	112.7	64.4	6	270
word duration	290.3	95.4	121	571
burst amplitude (overall)	40.0	6.3	29	55
vowel center amplitude	49.4	5.5	36	61
vowel center F1	1051	177	563	1440
vowel center F2	1938	282	1500	2828
vowel center F3	4569	268	4140	5367
F0 at voice onset + 20 ms	378	119	197	775

Time 2

measure	mean	standard dev	low value	high value
number of words	8.8		6	10
voice onset time (voiced)	10.1	5.2	4	20
voice onset time (voiceless)	72.0	46.2	9	205
vowel duration	134.6	186.8	27	1338
word duration	344.3	221.6	150	1601
burst amplitude (overall)	35.3	5.9	23	49
vowel center amplitude	39.4	6.9	26	54
vowel center F1	1161	162	789	1540
vowel center F2	1958	320	1356	2872
vowel center F3	4258	216	3592	4753
F0 at voice onset + 20 ms	323	115	123	730

Child #2

Time 1

measure	mean	standard dev	low value	high value
number of words	7.0		6	8
voice onset time (voiced)	25.4	12.6	6	55
voice onset time (voiceless)	90.4	53.0	21	231
vowel duration	326.8	179.3	63	668
word duration	627.9	232.7	180	985
burst amplitude (overall)	27.1	5.7	13	36
vowel center amplitude	28.7	4.3	19	37
vowel center F1	1027	168	591	1406
vowel center F2	2166	255	1460	2625
vowel center F3	4094	237	3703	4703
F0 at voice onset + 20 ms	290	28	222	375

Time 2

measure	mean	standard dev	low value	high value
number of words	8.8		6	10
voice onset time (voiced)	10.4	6.1	4	38
voice onset time (voiceless)	45.2	43.5	7	197
vowel duration	184.3	103.1	37	511
word duration	357.5	173.3	195	838
burst amplitude (overall)	31.9	10.6	19	49
vowel center amplitude	38.8	13.0	27	57
vowel center F1	994	319	750	1628
vowel center F2	1748	567	1386	2438
vowel center F3	3973	271	3105	4786
F0 at voice onset + 20 ms	386	148	233	800

Child #3

Time 1

measure	mean	standard dev	low value	high value
number of words	7.5		3	10
voice onset time (voiced)	17.8	15.9	4	77
voice onset time (voiceless)	81.9	35.3	31	159
vowel duration	176.6	77.4	36	403
word duration	368.0	109.1	180	585
burst amplitude (overall)	38.8	7.6	27	56
vowel center amplitude	47.2	4.8	36	56
vowel center F1	1237	160	931	1625
vowel center F2	2160	299	1636	2874
vowel center F3	4128	272	3690	4813
F0 at voice onset + 20 ms	413	128	226	855

Time 2

measure	mean	standard dev	low value	high value
number of words	8.7		6	10
voice onset time (voiced)	12.3	8.0	4	36
voice onset time (voiceless)	45.7	22.4	10	92
vowel duration	152.7	70.6	30	333
word duration	275.6	83.9	83	527
burst amplitude (overall)	39.6	4.9	29	50
vowel center amplitude	50.6	6.4	35	65
vowel center F1	1218	168	844	1550
vowel center F2	2056	290	1242	2693
vowel center F3	4014	374	3375	4837
F0 at voice onset + 20 ms	395	80	246	558

Child #4

Time 1

measure	mean	standard dev	low value	high value
number of words	8.5		5	10
voice onset time (voiced)	26.0	14.8	4	50
voice onset time (voiceless)	70.5	32.4	23	136
vowel duration	181.5	78.0	28	510
word duration	344.0	85.3	161	551
burst amplitude (overall)	34.4	5.4	24	46
vowel center amplitude	39.1	5.8	29	52
vowel center F1	935	152	572	1239
vowel center F2	1940	199	1313	2494
vowel center F3	4168	302	3469	4814
F0 at voice onset + 20 ms	252	45	79	373

Time 2

measure	mean	standard dev	low value	high value
number of words	9.7		9	10
voice onset time (voiced)	21.0	12.8	4	47
voice onset time (voiceless)	62.8	37.2	19	168
vowel duration	171.2	63.2	65	378
word duration	304.4	91.4	163	659
burst amplitude (overall)	35.4	5.1	22	45
vowel center amplitude	41.4	5.0	33	56
vowel center F1	930	124	703	1365
vowel center F2	2055	205	1571	2587
vowel center F3	3709	237	3188	4347
F0 at voice onset + 20 ms	242	31	179	396

Child #5

Time 1

measure	mean	standard dev	low value	high value
number of words	9.8		8	10
voice onset time (voiced)	14.5	7.6	4	40
voice onset time (voiceless)	61.8	23.5	21	129
vowel duration	135.4	54.6	48	293
word duration	279.1	97.6	123	540
burst amplitude (overall)	35.75	5.3	27	52
vowel center amplitude	43.2	6.4	31	59
vowel center F1	1136	190	785	1574
vowel center F2	2341	233	1781	2754
vowel center F3	4022	349	2359	4688
F0 at voice onset + 20 ms	368	142	170	844

Time 2

measure	mean	standard dev	low value	high value
number of words	9.2		8	10
voice onset time (voiced)	13.6	5.8	6	28
voice onset time (voiceless)	73.4	60.2	15	322
vowel duration	139.5	77.3	42	461
word duration	287.5	82.5	201	592
burst amplitude (overall)	36.4	5.4	17	47
vowel center amplitude	45.9	7.7	30	64
vowel center F1	1110	186	698	1659
vowel center F2	2315	274	1594	2776
vowel center F3	4074	272	3492	4758
F0 at voice onset + 20 ms	341	75	254	645

Child #6

Time 1

measure	mean	standard dev	low value	high value
number of words	9.3		7	10
voice onset time (voiced)	12.3	7.9	4	37
voice onset time (voiceless)	85.6	49.4	10	236
vowel duration	151.2	113.9	32	622
word duration	348.4	129.1	122	726
burst amplitude (overall)	21.1	5.9	11	39
vowel center amplitude	26.8	6.6	14	40
vowel center F1	1149	196	697	1622
vowel center F2	2235	217	1641	2813
vowel center F3	4460	326	3325	5156
F0 at voice onset + 20 ms	321	140	113	1042

Time 2

measure	mean	standard dev	low value	high value
number of words	9.8		8	10
voice onset time (voiced)	10.3	5.1	4	23
voice onset time (voiceless)	71.7	53.1	16	244
vowel duration	227.4	209.5	31	1015
word duration	413.4	259.2	89	1294
burst amplitude (overall)	39.0	6.2	20	52
vowel center amplitude	45.7	8.2	27	60
vowel center F1	1105	200	580	1645
vowel center F2	2231	288	1807	3155
vowel center F3	4211	270	3756	4788
F0 at voice onset + 20 ms	344	140	221	800

Child #7

Time 1

measure	mean	standard dev	low value	high value
number of words	9.5		7	10
voice onset time (voiced)	21.4	14.7	6	72
voice onset time (voiceless)	113.4	49.8	23	238
vowel duration	147.7	79.8	37	402
word duration	406.4	147.5	186	843
burst amplitude (overall)	25.9	5.7	5	36
vowel center amplitude	29.9	6.9	17	46
vowel center F1	943	117	721	1263
vowel center F2	2142	340	1487	2778
vowel center F3	3853	546	3497	4547
F0 at voice onset + 20 ms	254	39	178	351

Time 2

measure	mean	standard dev	low value	high value
number of words	9.7		8	10
voice onset time (voiced)	22.0	19.2	8	54
voice onset time (voiceless)	124.7	57.6	19	208
vowel duration	196.0	123.3	25	582
word duration	439.0	164.2	161	977
burst amplitude (overall)	35.6	4.4	25	51
vowel center amplitude	38.4	6.3	25	58
vowel center F1	969	151	586	1250
vowel center F2	1776	245	1219	2484
vowel center F3	3850	234	3414	4485
F0 at voice onset + 20 ms	260	55	97	437

Child #8

Time 1

measure	mean	standard dev	low value	high value
number of words	6.3		3	8
voice onset time (voiced)	18.2	11.8	3	43
voice onset time (voiceless)	100.9	35.0	59	169
vowel duration	164.7	102.5	36	472
word duration	370.7	127.4	148	636
burst amplitude (overall)	42.7	5.8	31	54
vowel center amplitude	42.4	5.0	32	51
vowel center F1	932	148	614	1290
vowel center F2	1943	313	1351	2742
vowel center F3	4498	280	3696	4969
F0 at voice onset + 20 ms	304	87	113	530

Time 2

measure	mean	standard dev	low value	high value
number of words	9.3		7	10
voice onset time (voiced)	17.6	12.2	3	53
voice onset time (voiceless)	102.9	43.0	19	215
vowel duration	149.6	92.3	66	664
word duration	355.6	102.8	132	772
burst amplitude (overall)	38.2	5.0	27	51
vowel center amplitude	40.3	8.7	25	60
vowel center F1	989	138	750	1272
vowel center F2	2071	231	1723	2813
vowel center F3	4550	322	3750	5344
F0 at voice onset + 20 ms	284	71	170	448

Child #9

Time 1

measure	mean	standard dev	low value	high value
number of words	6.67		2	9
voice onset time (voiced)	13.3	4.0	5	21
voice onset time (voiceless)	58.1	30.4	13	130
vowel duration	211.6	81.2	76	404
word duration	384.4	100.8	241	586
burst amplitude (overall)	38.2	7.7	3	46
vowel center amplitude	47.8	7.6	26	59
vowel center F1	1131	170	801	1480
vowel center F2	2070	240	1571	2531
vowel center F3	3962	253	3422	4442
F0 at voice onset + 20 ms	375	125	122	649

Time 2

measure	mean	standard dev	low value	high value
number of words	10.0		8	10
voice onset time (voiced)	11.4	6.4	6	30
voice onset time (voiceless)	54.0	26.5	21	145
vowel duration	196.1	126.6	46	793
word duration	379.2	151.8	147	969
burst amplitude (overall)	40.7	4.5	29	48
vowel center amplitude	50.8	6.6	32	61
vowel center F1	1155	120	873	1412
vowel center F2	2107	234	1455	2769
vowel center F3	4009	277	3497	4655
F0 at voice onset + 20 ms	361	80	154	600

Child #10

Time 1

measure	mean	standard dev	low value	high value
number of words	9.3		6	10
voice onset time (voiced)	16.3	10.7	3	42
voice onset time (voiceless)	105.7	102.7	15	561
vowel duration	195.6	122.9	34	527
word duration	422.3	208.9	154	1020
burst amplitude (overall)	42.4	5.7	29	53
vowel center amplitude	49.9	6.6	36	62
vowel center F1	1039	169	620	1307
vowel center F2	2154	290	1552	2672
vowel center F3	4227	310	3318	5056
F0 at voice onset + 20 ms	426	130	202	704

Time 2

measure	mean	standard dev	low value	high value
number of words	7.2		4	10
voice onset time (voiced)	12.2	4.8	4	23
voice onset time (voiceless)	109.1	79.6	31	308
vowel duration	235.5	126.5	69	546
word duration	427.7	134.1	203	716
burst amplitude (overall)	40.5	5.1	28	50
vowel center amplitude	55.1	8.1	37	67
vowel center F1	1142	192	713	1542
vowel center F2	1974	258	1458	2495
vowel center F3	4186	243	3598	4819
F0 at voice onset + 20 ms	432	161	224	926

12. Bibliography

- Assman, P. F. and Katz, W. F. (2000). "Time-varying spectral change in the vowels of children and adults." J. Acoust. Soc. Am. 4(108): 1856-66.
- Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. Invariance and variability in speech processes. D. H. K. J. S. Perkell. Hillsdale, N. J., Lawrence Earlbaum Associates: 36-51.
- Ferguson, C. A. and Farwell, C. B. (1975). "Words and sounds in early language acquisition." Language 51: 419-439.
- Gibson, T. and Ohde, R. N. (2003). "Development of coarticulation in 17-21 month old children." The convention of the American Speech-Language-Hearing Association in Atlanta, Georgia, November 23.
- Goldman, R. and Fristoe, M. (2000). Goldman Fristoe Test of Articulation 2. Circle Pines, MN, American Guidance Service, Inc.
- Goldstein, U. G. (1980). An articulatory model for the vocal tracts of growing children. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- Goodell, E. W. and Studdert-Kennedy, M. (1993). "Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: a longitudinal study." J. Speech Hear. Res. 36(4): 707-27.
- Hodge, M. M. (1989). A comparison of spectral-temporal measures across speaker age: Implications for an acoustic characterization of speech maturation, University of Wisconsin-Madison.
- Katz, W. F., Kripke, C. and Tallal, P. (1991). "Anticipatory coarticulation in the speech of adults and young children: acoustic, perceptual, and video data." J. Speech Hear. Res. 34(6): 1222-32.
- Kent, R. D. (1983). The segmental organization of speech. The production of speech. P. MacNeilage. New York, Springer-Verlag: 57-89.

- Kewley-Port, D. and Preston, M. S. (1974). "Early apical stop production: a voice onset time analysis." J. Phon. 2: 195-210.
- Koenig, L. L. (2000). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds." J. Speech, Lang., Hear. Res. 43: 1211-1228.
- Kuijpers, C. T. L. (1993). "Temporal aspects of the voiced-voiceless distinction in speech development of young Dutch children." J. Phon. 21: 313-327.
- Lee, S., Potamianos, A. and Narayanan, S. (1999). "Acoustics of children's speech: developmental changes of temporal and spectral parameters." J. Acoust. Soc. Am. 105(3): 1455-68.
- Leonard, L. B., Newhoff, M. and Mesalam, L. (1980). "Individual differences in early child phonology." Applied Psycholinguistics 1: 7-30.
- Locke, J. L. (1983). Phonological acquisition and change. New York, Academic Press.
- Macken, M. A. and Barton, D. (1979). "The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants." J. Child Language 7(1): 41074.
- Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. Language Production. B. Butterworth. London, Academic Press. 2: 3-50.
- Menyuk, P. and Klatt, D. H. (1975). "Voice onset time in consonant cluster production in children and adults." J. Child Language 2: 223-231.
- Netsell, R., Lotz, W. K., Peters, J. E. and Schulte, L. (1994). "Developmental patterns of laryngeal and respiratory function for speech production." J. Voice 8(2): 123-131.
- Nittrouer, S. (1993). "The emergence of mature gestural patterns is not uniform: evidence from an acoustic study." J. Speech Hear. Res. 36(5): 959-72.
- Nittrouer, S. (1995). "Children learn separate aspects of speech production at different rates: evidence from spectral moments." J. Acoust. Soc. Am. 97(1): 520-30.
- Nittrouer, S., Studdert-Kennedy, M. and McGowan, R. S. (1989). "The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults." J. Speech Hear. Res. 32(1): 120-32.

- Nittrouer, S., Studdert-Kennedy, M. and Neely, S. T. (1996). "How children learn to organize their speech gestures: further evidence from fricative-vowel syllables." J. Speech Hear. Res. 39(2): 379-89.
- Ohde, R. N. (1984). "Fundamental frequency as an acoustic correlate of stop consonant voicing." J. Acoust. Soc. Am. 75: 224-230.
- Ohde, R. N. (1985). "Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children." J. Acoust. Soc. Am. 78(5): 1554-1561.
- Prather, E. M. and Hedrick, D. L. (1975). "Articulation development in children aged two to four years." J. Speech Hear. Disord. 40(2): 179-91.
- Repp, B. H. (1986). "Some observations on the development of anticipatory coarticulation." J. Acoust. Soc. Am. 79(5): 1616-19.
- Robb, M. P. and Saxman, J. H. (1985). "Developmental trends in vocal fundamental frequency of young children." J. Speech Hear. Res. 28(3): 421-7.
- Robb, M. P. and Saxman, J. H. (1990). "Syllable duration of preword and early word vocalizations." J. Speech Hear. Res. 33: 583-93.
- Robb, M. P., Saxman, J. H. and Grant, A. A. (1989). "Vocal fundamental frequency characteristics during the first two years of life." J. Acoust. Soc. Am. 85(4): 1708-1717.
- Robb, M. P. and Smith, A. B. (2002). "Fundamental frequency onset and offset behavior: a comparative study of children and adults." J. Speech, Lang., Hear. Res. 45: 446-456.
- Sereno, J. A., Baum, S. R., Mearan, G. C. and Lieberman, P. (1987). "Acoustic analysis and perceptual data on anticipatory labial coarticulation in adults and children." J. Acoust. Soc. Am. 81(2): 512-19.
- Sereno, J. A. and Lieberman, P. (1987). "Developmental aspects of lingual coarticulation." J. Phon. 15(3): 247-57.

- Siren, K. A. and Wilcox, K. A. (1995). "Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech." J. Speech Hear. Res. 38(2): 351-59.
- Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E. and Bird, A. (1990). "Iowa articulation norms project and its Nebraska replication." J. Speech Hear. Disord. 55: 779-798.
- Smith, B. L. (1978). "Temporal aspects of English speech production: A developmental perspective." J. Phon. 6: 37-67.
- Smith, B. L. and Kenney, M. K. (1994). "Variability control in speech production tasks performed by adults and children." J. Acoust. Soc. Am. 96(2 Pt. 1): 699-705.
- Smith, B. L., Kenney, M. K. and Hussain, S. (1996). "A longitudinal investigation of duration and temporal variability in children's speech production." J. Acoust. Soc. Am. 99(4 Pt 1): 2344-49.
- Snow, D. (1994). "Phrase-final syllable lengthening and intonation in early child speech." J. Speech Hear. Res. 37(4): 831-40.
- Stathopoulos, E. T. and Sapienza, C. (1993). "Respiratory and laryngeal measures of children during vocal intensity variation." J. Acoust. Soc. Am. 94(5): 2531-41.
- Stathopoulos, E. T. and Weismer, G. (1985). "Oral airflow and air pressure during speech production: A comparative study of children, youths, and adults." Folia Phoniatr. 37: 152-159.
- Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. Language Perception and Production. A. Allport, D. Mackay, D. Prinz and E. Scheerer. London, Academic: 67-84.
- Studdert-Kennedy, M. (1990). Language development from an evolutionary perspective. Biological and behavioral determinants of language development. N. Krasnegor, D. Runbaugh, R. Schiefelbusch and M. Studdert-Kennedy. Hillsdale, N.J., Lawrence Erlbaum Associates: 5-28.
- Suchato, A. (2004). Classification of stop consonant place of articulation. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

- Sussman, H. M., Duder, C., Dalston, E. and Cacciatore, A. (1999). "An acoustic analysis of the development of CV coarticulation: a case study." J. Speech Lang. Hear. Res. 42(5): 1080-96.
- Sussman, H. M., Minifie, F. D., Buder, E. H., Stoel-Gammon, C. and Smith, J. (1996). "Consonant-vowel interdependencies in babbling and early words: preliminary examination of a locus equation approach." J. Speech Hear. Res. 39(2): 424-33.
- Turnbaugh, K. R., Hoffman, P. R., Daniloff, R. G. and Ashber, R. (1985). "Stop-vowel coarticulation in 3-year-old, 5-year-old, and adult speakers." J. Acoust. Soc. Am. 77(3): 1256-57.