AN ACOUSTIC ANALYSIS OF VOICELESS OBSTRUENTS PRODUCED BY ADULTS AND TYPICALLY DEVELOPING CHILDREN

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Shawn L. Nissen, M.A.

The Ohio State University
2003

Dissertation Committee:

Professor Robert Allen Fox, Advisor
Professor Neal F. Johnson
Professor Michael D. Trudeau

Approved by

Advisor
Department of Speech and Hearing Science
ABSTRACT

A considerable amount of speech research conducted over the last five decades has attempted to explain how the human perceptual system so accurately and efficiently perceives speech. Although scientists have made great strides toward understanding the mechanisms of speech production and perception, the complex relationships between the acoustic structures of speech and the resulting psychological percepts have yet to be fully and adequately explained. In particular, there is a limited amount of knowledge on the acoustic nature of speech produced by younger children. Thus, this study examined the acoustic structure of voiceless obstruents produced by adults and typically developing children from 3 to 6 years of age. Forty speakers from four different age groups produced word initial voiceless obstruents /p, t, k, ŋ, s, ʃ/ in real word CV contexts, as well as a similar series of intervocalic obstruents drawn from non-words (VCV phonetic context). The acoustic structure of the speech tokens was described in terms of multiple acoustic parameters (durations, normalized amplitude, spectral peak location, spectral slope, and spectral moments). The results of this study indicated that multiple acoustic parameters of voiceless stops and fricatives vary systematically as a function of place of articulation, vowel context, speaker age, and gender. In particular, it was found that the spectral peak location, slope, and the first three spectral moments were able to distinguish between differing places of articulation. It was further shown that gender
differences for several acoustic characteristics can be found in children at a relatively young age, with the acquired sibilant contrast between /s/ and /ʃ/ found to be less distinguished in children than adults. It was found that acoustic separation between the two sibilant fricatives widened as the age of the speakers increased, thereby suggesting that the contrast continues to be fine tuned throughout young childhood toward a more adult-like stage. Discriminant analysis revealed evidence that classification models based on adult male data were sensitive to gender related differences even in the youngest age group.
Dedicated to my wife and children
ACKNOWLEDGMENTS

I wish to thank my advisor, Dr. Robert Fox, for the intellectual guidance, faith in my ability, and constant encouragement. He has been a tremendous professor, mentor, and friend. I am grateful for the opportunity to have worked with him.

I also want to thank the members of my doctoral committees – Dr. Neal Johnson, Dr. Michael Trudeau, Dr. Marios Fourakis, Dr. Keith Johnson, and Dr. Majorie Chan – for their constructive and insightful assistance during my program of study. I would also like to thank Dr. Yuan Zheng for serving as my graduate school representative.

I thank the staff (Ann, Vicki, Kathleen, Shannon, John, and Pete), clinical supervisors (Dr. Gail Whitelaw, JoAnn Donohue, Marcia Taber, Christy Goodman, and Nadine Ocock), and faculty in the department for their help in time of need.

Special thanks to the Department of Speech and Hearing Science and the graduate school at The Ohio State University for the financial assistance to conduct this research and to all the subjects who participated in this study.

I also wish to thank my colleagues in the lab (Chris Gibbons, Dr. Julie McGory, Dr. Elaina Frieda, Kim Rosenbauer, and Rachel Frens), thank you for your friendship.

My appreciation goes to Sam and Julia McCarter for their help and assistance.

I thank my parents and siblings for their unwavering love and support.

Finally, I would like to thank my loving wife and children, who have always sacrificed so that I might accomplish my dreams.
VITA

May 8, 1968 .......................... Born – Logan, Utah

1994 ................................. B.A. Humanities, Brigham Young University

1995 – 2002......................... Graduate Teaching and Research Associate
                                The Ohio State University

1998 ................................ M.A. Speech-Language Pathology,
                                The Ohio State University

2003 – present ..................... Instructor, Audiology & Speech-Language
                                Pathology, Brigham Young University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Speech and Hearing Science
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Literature Review</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Source-filter theory</td>
<td>7</td>
</tr>
<tr>
<td>2.1.1 Sound source for vowels and sonorants</td>
<td>8</td>
</tr>
<tr>
<td>2.1.2 Sound source for obstruent segments</td>
<td>9</td>
</tr>
<tr>
<td>2.1.2.1 Frication</td>
<td>9</td>
</tr>
<tr>
<td>2.1.2.2 Transient noise excitation</td>
<td>10</td>
</tr>
<tr>
<td>2.1.2.3 Combined transient and frication noise source</td>
<td>11</td>
</tr>
<tr>
<td>2.1.3 Filter transfer function</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Spectral moment analysis</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1 Obstruent discrimination</td>
<td>14</td>
</tr>
<tr>
<td>2.2.2 Spectral moments and phonological development</td>
<td>24</td>
</tr>
<tr>
<td>3. Methods</td>
<td>31</td>
</tr>
<tr>
<td>3.1 Subjects</td>
<td>31</td>
</tr>
<tr>
<td>3.2 Stimuli</td>
<td>33</td>
</tr>
<tr>
<td>3.3 Procedure and analysis</td>
<td>34</td>
</tr>
<tr>
<td>3.3.1 Recording</td>
<td>34</td>
</tr>
<tr>
<td>3.3.2 Segmentation</td>
<td>35</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>A summary of the variables included in the various classification models for CV stops</td>
<td>83</td>
</tr>
<tr>
<td>4.2</td>
<td>Results of discriminant classification using both cross-validation (jack-knife) and standard stepwise procedures for CV stops; all values represent % correct classification</td>
<td>84</td>
</tr>
<tr>
<td>4.3</td>
<td>Results of stepwise linear discriminant analysis when the classification model is trained on adult male data only for CV stops; all values represent % correct classification</td>
<td>85</td>
</tr>
<tr>
<td>4.4</td>
<td>Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for CV stops; all values represent % correct classification</td>
<td>86</td>
</tr>
<tr>
<td>5.1</td>
<td>Results of discriminant classification using both cross-validation (jack-knife) and standard stepwise procedures for VCV stops; all values represent % correct classification</td>
<td>120</td>
</tr>
<tr>
<td>5.2</td>
<td>Results of stepwise linear discriminant analysis when the classification model is trained on adult male data only for VCV stops; all values represent % correct classification</td>
<td>121</td>
</tr>
<tr>
<td>5.3</td>
<td>Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for VCV stops; all values represent % correct classification</td>
<td>122</td>
</tr>
<tr>
<td>6.1</td>
<td>A summary of the variables included in the various classification models for fricatives</td>
<td>165</td>
</tr>
<tr>
<td>6.2</td>
<td>Results of discriminant classification using both cross-validation (jack-knife) and standard stepwise procedures for CV fricatives; all values represent % correct classification</td>
<td>166</td>
</tr>
</tbody>
</table>
6.3 Results of stepwise linear discriminant analysis when the classification model is trained on adult male data only for CV fricatives; all values represent % correct classification.................................................. 167

6.4 Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for CV fricatives; all values represent % correct classification.................................................. 168

7.1 Results of discriminant classification using both cross-validation (jack-knife) and standard stepwise procedures for VCV fricatives; all values represent % correct classification.................................................. 209

7.2 Results of stepwise linear discriminant analysis when the classification model is trained on adult male data only for VCV fricatives; all values represent % correct classification.................................................. 210

7.3 Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for VCV fricatives; all values represent % correct classification.................................................. 211
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Segmentation of stop consonant</td>
<td>46</td>
</tr>
<tr>
<td>3.2</td>
<td>Segmentation of a fricative consonant</td>
<td>47</td>
</tr>
<tr>
<td>3.3</td>
<td>10 ms series of analysis windows for the stop consonants</td>
<td>48</td>
</tr>
<tr>
<td>3.4</td>
<td>20 ms series of analysis windows for the stop consonants</td>
<td>49</td>
</tr>
<tr>
<td>3.5</td>
<td>20 ms series of analysis windows for the fricative consonants</td>
<td>50</td>
</tr>
<tr>
<td>3.6</td>
<td>40 ms series of analysis windows for the fricative consonants</td>
<td>51</td>
</tr>
<tr>
<td>4.1</td>
<td>VOT by place of articulation and vowel context in CV stops</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Mean vowel duration by place of articulation in CV stops</td>
<td>55</td>
</tr>
<tr>
<td>4.3</td>
<td>Mean vowel duration by vowel context in CV stops</td>
<td>55</td>
</tr>
<tr>
<td>4.4</td>
<td>Mean vowel duration by place of articulation and vowel context in CV stops</td>
<td>56</td>
</tr>
<tr>
<td>4.5</td>
<td>Normalized amplitude as a function of place of articulation in CV stops</td>
<td>57</td>
</tr>
<tr>
<td>4.6</td>
<td>Normalized amplitude as a function of vowel context in CV stops</td>
<td>58</td>
</tr>
<tr>
<td>4.7</td>
<td>Mean spectral peak location as a function of place of articulation and window location for CV stops</td>
<td>60</td>
</tr>
<tr>
<td>4.8</td>
<td>Mean spectral peak location as a function of age group and window location for CV stops</td>
<td>61</td>
</tr>
<tr>
<td>4.9</td>
<td>Spectral peak location by place of articulation and vowel context in CV stops</td>
<td>62</td>
</tr>
</tbody>
</table>
4.10 Spectral peak location for male and female speakers as a function of place of articulation for CV stops…………………………………………………………..63

4.11 Spectral peak location for different age groups as a function of place of articulation for CV stops…………………………………………………………..64

4.12 Mean spectral slope as a function of place of articulation and window location for CV stops…………………………………………………………..65

4.13 Mean spectral slope as a function of age group and window location for CV stops…………………………………………………………..66

4.14 Mean spectral slope as a function of vowel context and place for CV stops…………………………………………………………..68

4.15 Mean spectral slope as a function of gender and age for CV stops………….68

4.16 Spectral mean as a function of place of articulation and window location for CV stops…………………………………………………………..69

4.17 Spectral mean as a function of age and window location for CV stops……….70

4.18 Spectral mean for vowel contexts as a function of place of articulation for CV stops…………………………………………………………..71

4.19 Spectral mean as a function of gender and age group for CV stops………….72

4.20 Spectral variance as a function of place of articulation and window location for CV stops…………………………………………………………..73

4.21 Spectral variance as a function of place of articulation for CV stops………….74

4.22 Spectral variance as a function of vowel context for CV stops………………..75

4.23 Spectral skewness as a function of place of articulation and window location for CV stops…………………………………………………………..76

4.24 Spectral skewness as a function of age and window location for CV stops…….76

4.25 Spectral skewness as a function of place of articulation and vowel context for CV stops…………………………………………………………..77

4.26 Spectral skewness as a function of gender and age group for CV stops……….78
4.27 Spectral kurtosis as a function of place of articulation for CV stops ..........79
5.1 Overall token duration as a function of place of articulation and age in VCV stops .................................................................90
5.2 Initial vowel duration as a function of place of articulation and age in VCV stops .................................................................91
5.3 VOT by place of articulation in VCV stops ........................................93
5.4 VOT by vowel context in VCV stops .............................................93
5.5 Final vowel duration as a function of vowel context in VCV stops ........94
5.6 Normalized amplitude as a function of place of articulation in VCV stops ....95
5.7 Normalized amplitude as a function of vowel context in VCV stops ..........96
5.8 Normalized amplitude as a function of place of articulation and vowel context in VCV stops .................................................................97
5.9 Spectral peak location as a function of place of articulation and window location for VCV stops .................................................................99
5.10 Spectral peak location as a function of age group and window location for VCV stops .................................................................99
5.11 Spectral peak location by place of articulation and vowel context in VCV stops .................................................................101
5.12 Spectral peak location for different age groups as a function of place of articulation for VCV stops .................................................................101
5.13 Spectral peak location for male and female speakers as a function of place of articulation for VCV stops .................................................................102
5.14 Spectral slope as a function of place of articulation and window location for VCV stops .................................................................103
5.15 Spectral slope as a function of age group and window location for VCV stops .................................................................104
5.16 Spectral slope as a function of place of articulation and vowel context for VCV stops .................................................................105
5.17 Spectral slope as a function of age group for VCV stops
5.18 Spectral mean as a function of place of articulation and window
    location for VCV stops
5.19 Spectral mean as a function of age and window location for VCV stops
5.20 Spectral mean for vowel contexts as a function of place of articulation
    for VCV stops
5.21 Spectral mean as a function of gender and age group for VCV stops
5.22 Spectral variance as a function of place of articulation and window
    location for VCV stops
5.23 Spectral variance as a function of place of articulation and vowel
    context for VCV stops
5.24 Mean spectral skewness as a function of place of articulation and
    window location for VCV stops
5.25 Spectral skewness as a function of age group and window location
    for VCV stops
5.26 Spectral skewness as a function of place of articulation for VCV stops
5.27 Spectral skewness as a function of place of articulation and vowel
    context for VCV stops
5.28 Spectral kurtosis as a function of place of articulation for VCV stops
6.1 Fricative duration as a function of place of articulation for CV fricatives
6.2 Fricative duration as a function of place of articulation and age for
    CV fricatives
6.3 Fricative duration as a function of vowel context for CV fricatives
6.4 Vowel duration as a function of place of articulation for CV fricatives
6.5 Vowel duration as a function of vowel context for CV fricatives
6.6 Mean vowel duration as a function of place of articulation and vowel
    context in CV fricatives
6.7 Normalized amplitude as a function of place of articulation in CV fricatives

6.8 Normalized amplitude as a function of vowel context in VCV fricatives

6.9 Normalized amplitude as a function of place of articulation and vowel context in VCV fricatives

6.10 Spectral peak location as a function of place of articulation and window location for CV fricatives

6.11 Spectral peak location as a function of age group and window location for CV fricatives

6.12 Spectral peak location by place of articulation in CV fricatives

6.13 Spectral peak location by place of articulation and age in CV fricatives

6.14 Spectral peak location by vowel context and gender in CV fricatives

6.15 Spectral peak location for male and female speakers as a function of place of articulation for CV fricatives

6.16 Spectral peak location as a function of different age for CV fricatives

6.17 Spectral slope as a function of place of articulation and window location for CV fricatives

6.18 Spectral slope as a function of age and window location for CV fricatives

6.19 Spectral slope as a function of gender and place for CV fricatives

6.20 Spectral slope as a function of vowel context and place of articulation for CV fricatives

6.21 Spectral slope as a function of gender and age group for CV fricatives

6.22 Spectral mean as a function of place of articulation and window location for CV fricatives

6.23 Spectral mean as a function of age and window location for CV fricatives

6.24 Spectral mean as a function of place of articulation for CV fricatives
6.25 Spectral mean as a function of place of articulation and gender for CV fricatives………………………………………………………………………………149
6.26 Spectral mean as a function of age group and gender for CV fricatives……..149
6.27 Spectral mean as a function of place of articulation and age for CV fricatives…………………………………………………………………………..150
6.28 Spectral variance as a function of place of articulation and window location for CV fricatives…………………………………………………………151
6.29 Spectral variance as a function of place of articulation for CV fricatives…..152
6.30 Spectral skewness as a function of place of articulation and window location for CV fricatives…………………………………………………………153
6.31 Spectral skewness as a function of age group and window location for CV fricatives…………………………………………………………………154
6.32 Spectral skewness as a function of place of articulation for CV fricatives…..155
6.33 Spectral skewness as a function of place of articulation and gender for CV fricatives………………………………………………………………….155
6.34 Spectral skewness as a function of age group and place of articulation for CV fricatives……………………………………………………………………157
6.35 Spectral kurtosis as a function of place of articulation and window location for CV fricatives……………………………………………………………157
6.36 Spectral kurtosis as a function of age and window location for CV fricatives………………………………………………………………………………158
6.37 Spectral kurtosis as a function of place of articulation for CV fricatives……..159
6.38 Spectral kurtosis as a function of gender and place for CV fricatives………..160
6.39 Spectral kurtosis as a function of age and place of articulation for CV fricatives……………………………………………………………………………..160
7.1 Overall token duration as a function of final vowel context in VCV fricatives..................................................................................172
7.2 Initial vowel duration as a function of final vowel context in VCV fricatives..................................................................................173

xvii
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.3</td>
<td>Fricative duration as function of place of articulation in VCV fricatives…174</td>
</tr>
<tr>
<td>7.4</td>
<td>Fricative duration as a function of final vowel context in VCV fricatives…175</td>
</tr>
<tr>
<td>7.5</td>
<td>Final vowel duration as a function of place of articulation in VCV fricatives…176</td>
</tr>
<tr>
<td>7.6</td>
<td>Final vowel duration as a function of final vowel context in VCV fricatives…176</td>
</tr>
<tr>
<td>7.7</td>
<td>Final vowel duration as a function of final vowel context in VCV fricatives…177</td>
</tr>
<tr>
<td>7.8</td>
<td>Normalized amplitude as a function of vowel context in VCV fricatives…178</td>
</tr>
<tr>
<td>7.9</td>
<td>Normalized amplitude as a function of place of articulation and vowel context in VCV fricatives…179</td>
</tr>
<tr>
<td>7.10</td>
<td>Spectral peak location as a function of place of articulation and window location for VCV fricatives…181</td>
</tr>
<tr>
<td>7.11</td>
<td>Spectral peak location as a function of age and window location for VCV fricatives…182</td>
</tr>
<tr>
<td>7.12</td>
<td>Spectral peak location as a function of place of articulation for VCV fricatives…183</td>
</tr>
<tr>
<td>7.13</td>
<td>Spectral peak location as a function of place of articulation and age group for VCV fricatives…183</td>
</tr>
<tr>
<td>7.14</td>
<td>Spectral peak location as a function of age group for VCV fricatives…184</td>
</tr>
<tr>
<td>7.15</td>
<td>Spectral peak location for male and female speakers as a function of place of articulation for VCV fricatives…185</td>
</tr>
<tr>
<td>7.16</td>
<td>Spectral slope as a function of place of articulation and window location for VCV fricatives…186</td>
</tr>
<tr>
<td>7.17</td>
<td>Spectral slope as a function of age group and window location for VCV fricatives…187</td>
</tr>
<tr>
<td>7.18</td>
<td>Spectral slope as a function of vowel context and place of articulation for VCV fricatives…188</td>
</tr>
</tbody>
</table>
7.19 Spectral slope as a function of age group and place of articulation for VCV fricatives...

7.20 Spectral mean as a function of place of articulation and window location for VCV fricatives...

7.21 Spectral mean as a function of age and window location for VCV fricatives...

7.22 Spectral mean as a function of place of articulation for VCV fricatives...

7.23 Spectral mean as a function of place of articulation and gender for VCV fricatives...

7.24 Spectral mean as a function of age group and gender for VCV fricatives...

7.25 Spectral mean as a function of gender and place of articulation for VCV fricatives...

7.26 Spectral mean as a function of place of articulation and age for VCV fricatives...

7.27 Spectral variance as a function of place of articulation and window location for VCV fricatives...

7.28 Spectral variance as a function of place of articulation for VCV fricatives...

7.29 Spectral skewness as a function of place of articulation and window location for VCV fricatives...

7.30 Spectral skewness as a function of age group and window location for VCV fricatives...

7.31 Mean spectral skewness as a function of place of articulation for VCV fricatives...

7.32 Spectral skewness as a function of place of articulation and gender for VCV fricatives...

7.33 Spectral skewness as a function of age group and place of articulation for VCV fricatives...

7.34 Spectral kurtosis as a function of place of articulation and window location for VCV fricatives...
7.35  Spectral kurtosis as a function of age and window location for VCV fricatives

7.36  Spectral kurtosis as a function of place of articulation for VCV fricatives

7.37  Spectral kurtosis as a function of age group and place of articulation
      for VCV fricatives
A considerable amount of speech research conducted over the last five decades has attempted to explain how the human perceptual system so accurately and efficiently perceives speech. Speech is transmitted primarily from speaker to receiver by means of acoustic energy. This is accomplished by encoding language into a physical acoustic waveform, which is then received and decoded by the perceptual system of the listener. Although the process of speech perception might appear simplistic to a naïve observer, in reality it is highly complex. Historically, scientists have postulated many theories to explain how humans are able to perceive speech at such a young age with relatively little effort. Although many of these endeavors have made great strides toward understanding the mechanisms of speech production and perception, the complex relationships between the acoustic structures of speech and the resulting psychological percepts have yet to be fully and adequately explained.

One problem encountered by individuals attempting to explain the process of speech perception is that the acoustic structures within the physical speech waveform are highly variable. In many instances, a direct and invariant connection between a particular phonetic percept and the resulting acoustic waveform has yet to be found. In
the field of speech science, this particular characteristic of the speech signal has been termed a “lack of invariance;” this characteristic causes problems for many theories of speech perception that rely upon the principle of invariance. As presented by Chomsky and Miller (1963), the principle of invariance claims that for every phonemic representation there must be a set of distinguishing features that uniquely identify that sound segment. The term “invariance” also has roots in the theories of Gibson (1979), who regarded invariants as “unique patterns in informational media that are invariably present when the property or event in the environment about which they provide information is present” (Fowler, 1994). Not only has the problem of noninvariance complicated a theoretical explanation of speech perception, but also from a practical standpoint, it continues to limit the progress of computer recognition of speech.

The acoustic realization of individual phonemes can vary considerably depending on multiple factors, some of which might be external to the production of the speech signal. One external source of variability may come from the ambient conditions of the speaking environment, such as room reverberation or background noise (Pisoni, 1997). These types of external noise sources often alter acoustic components of the speech signal prior to it reaching the ears of the listener.

There is also a considerable amount of acoustic variation between and within individual speakers. Between-speaker variation is due in large part to differing physical attributes of the speaker. Vocal tract size and shape, as well as vocal fold dimensions, are primary factors in determining the acoustic structure of a speech sound. Thus, the acoustic realization of speech sounds will differ according to the age, gender, and physical attributes of each individual speaker. This conclusion was supported by results
from a classic study conducted by Peterson and Barney (1952) and later replicated in Hillenbrand, Getty, Clark and Wheeler (1995). An acoustic analysis of speech samples elicited from 76 subjects of varying gender and age indicated that the acoustic variability of vowel tokens clustered according to the age and gender of the speaker. Female speakers produced formants that were higher in frequency than that of men, yet still somewhat lower than the vowel tokens spoken by the child subjects. Peterson and Barney estimated that on average the formant frequencies of the children were at least a half octave higher than the adult males. These results were not surprising considering the differences in vocal tract size and shape between the different speaker categories.

The acoustic parameters of the speech signal will change as a function of rate differences, changing voice quality, or context effects. In other words, a speaker will not always produce a particular phoneme exactly the same way. For example, the same acoustic cue can elicit differing phonetic percepts depending on the linguistic context. This phenomenon was illustrated in the classic study by Liberman and colleagues (1952). When a burst of acoustic energy centered around 1400 Hz was placed in front of a series of English vowels, listeners perceived different syllable types. If the burst preceded a high-front vowel, such as /i/, subjects identified the resulting syllable as /pi/. However, when the exact same burst was followed by a relatively lower vowel like /a/, the perceived syllable was consistent with /ka/. These early findings indicated that subjects perceived the same acoustic cue as different phonemes depending on the vowel context.

Conversely, it has been found that different acoustic cues elicit the same phoneme percept when placed in different linguistic contexts. The initial phoneme in the words “beat,” “bat,” “bought,” and “boot” is perceived by listeners as a /b/, even though the
acoustic cues formant transitions for each instance of /b/ is somewhat different depending on the following vowel context (Sussman, Fruchter, Hilbert, & Sirosh; 1998). Similar results have been found with other consonants as well (e.g., Liberman et al., 1954; 1967).

Assuming that the phonetic representation of a learned phoneme does not vary, the question then arises as to how listeners perceive speech from the midst of such large amounts of acoustic variability? The experimental findings illustrated previously do not verify the lack of invariance within the speech signal, but do suggest that the correspondence between the acoustic signal and the perceived phoneme representation is not simple in nature. Thus, researchers continue to search for more accurate models and descriptions of the complex process by which the humans perceive speech, a perceptual system that is able to decode language from the acoustic speech signal in a remarkably efficient and accurate manner.

Despite the lack of a simple and direct association between the acoustic signal and perceived phoneme, some researchers have found that when the appropriate analysis is applied to the proper acoustic parameters, the speech signal does yield unique and discrete patterns of acoustic energy. Fortunately, recent advancements in computer technology have allowed scientists to utilize more specific types of analysis on a variety of different acoustic parameters.

The continued search for acoustic invariance within the speech signal has been recently aided in part by a relatively new type of analytical tool, spectral moment analysis. This type of analysis examines the spectral characteristics of discrete time segments of the speech signal in terms of multiple statistical moments (i.e., mean, variance, skewness, and kurtosis). By combining these various “statistical snapshots” of
perceptually relevant segments of the physical waveform, researchers have aimed to identify unique and discrete patterns of acoustic energy within the speech signal.

Utilizing spectral moment analysis as well as more traditional types of acoustic analysis, several studies have found that the acoustic signal of obstruent sounds produced may contain discrete and stable spectral patterns (e.g., Forrest, Weismer, Melenkovic, & Dougall, 1988; Tomiak, 1990; Jongman, Wayland, & Jong, 2000). This research aimed to identify acoustic or spectral parameters that might serve to differentiate or discriminate adult’s productions of stop and fricative consonants in terms of place of articulation.

Another area of research has utilized spectral moment analysis to examine the speech of children. Several studies (e.g., Forrest, Weismer, Elbert, & Dinnsen, 1990; Miccio, 1996) have investigated the differences in acoustic structure of stop consonants produced by phonologically disordered and normally articulating children. Other researchers have also focused on changes in obstruent sounds as a function of normal aging and developmental maturation (Nittrouer, 1995; Fox, Nissen, McGory & Rosenbauer, 2001).

However, despite recent work in this area, there is still a limited amount of knowledge on the acoustic nature of voiceless obstruents produced by typically developing children younger than 6 years of age. Thus, the general aims of this dissertation are threefold:

1) Describe the acoustic structure of syllable initial and intervocalic voiceless obstruent productions of adults and typically developing children in terms of multiple acoustic parameters (duration, normalized amplitude, spectral peak location, spectral slope, and spectral moments).
2) Examine the systematic variation of obstruent production of both adult and typically developing children. More specifically, to what extent do the acoustic characteristics of voiceless stop and fricative productions change as a function of age, gender, place of articulation, and vowel context.

3) Determine how successful various types of discriminant functions classify stop and fricative productions in terms of place of articulation. Of further interest, is the extent to which classification models based on specific groups of speakers will successfully categorize the obstruents produced by other speaker groups of differing gender and age. This method of analysis provides insight into possible gender or age related changes in combinations of acoustic parameters and how such acoustic patterns may relate to potential cues to place of articulation.
CHAPTER 2

LITERATURE REVIEW

2.1 Source-Filter Theory

In a series of classic papers published in the early 1960s, the Swedish scientist Gunnar Fant outlined a theory of how human speech is produced. This theory is now referred to as the linear source-filter theory of speech production, or more commonly as Source-Filter Theory. According to this theory (Fant, 1960), speech sounds are the result of three mechanisms: the sound generating source, the filtering characteristics of the vocal tract, and the radiation characteristic into the external environment. The basic premise of source-filter theory is succinctly summarized by Fant (1968) in the following statement:

This basic principle stated in electrical engineering terms implies that somewhere in the vocal tract there originates a source, constituting the raw material of the sound whilst the wave propagation through the vocal tract provides a filtering, a shaping of the raw material resulting the speech wave as defined by the pressure-time wave at a certain distance from the speaker’s mouth. (p. 191)

Additional studies (e.g., Stevens & House, 1961; Stevens, 1989) have indicated that
Fant’s source-filter theory provides an accurate theoretical description of the basic underlying acoustic mechanisms of speech, regardless of the type of speech sound being produced.

2.1.1 Sound source for vowels and sonorant segments

According to source-filter theory, human speech is generated by three different types of source waveforms: 1) quasiperiodic oscillation, 2) frication, and 3) transient noise excitation. These sound sources may occur independently or in combination. In vowels or sonorant speech sounds, the primary acoustic source is the quasiperiodic oscillation of the vocal folds (Shadle, 1997). In American English, voicing is an important distinguishing feature among speech sounds. Sounds produced using this type of acoustic source are said to be voiced, whereas sounds produced without significant vocal fold vibration are considered voiceless in nature. Vowels and sonorants (i.e., liquids & glides) primarily utilize the voicing sound source, whereby other types of sound segments might combine voicing with an additional sound source (e.g., frication, transient noise burst).

In most languages, vocal fold vibration or voicing is created by eggesive pulmonic airflow passing through adducted vocal folds, generating a triangular pulse wave (Fry, 1979). The acoustic structure of this pulse wave is determined in part by the state of the glottis (vocal fold closure pattern), the physical nature of the folds during vibration (mass, length, and tension), as well as numerous aerodynamic factors. Spectrally, voiced sound sources are characterized by harmonic components, at multiples of the fundamental vibrating frequency that decrease in amplitude as the frequency of the harmonic increases.
2.1.2 Sound sources for obstruent segments

The additional sound sources of frication and transient noise excitation are also utilized in the production of human speech. These sources are generated when a pressurized flow of air encounters a relatively significant constriction or obstruction within the vocal tract. Speech sounds that are produced with a significant degree of frication or transient noise excitation are classified as obstruents. Obstruent speech segments may contain these sound sources in isolation or in combination with voicing. Fricatives, stops (plosives), and affricates all fall within this class of speech sounds, and are differentiated based upon the manner of the vocal tract constriction.

2.1.2.1 Frication.

Frication in speech is generated from a narrow constriction within the vocal tract. This constriction is formed by the close approximation of two articulators without direct tissue contact. As pressurized airflow passes through the constricted area, air particles transition from a generally laminar flow to a more turbulent pattern of movement. The turbulent flow of air particles following the constriction is characterized by the formation of complex patterns of air particle motion or “eddies” (Kent, 1997). These eddies within the airflow are created by rapid and irregular changes in air particle direction and velocity. Turbulence may also occur if the airflow encounters an abrupt obstacle within the vocal tract, such as the teeth (Shadle, 1990). This turbulent airflow then generates a sound wave of seemingly random pressure fluctuations (Kent & Read, 2002). Thus, the resulting waveform is aperiodic, without a discernable harmonic structure.

American English fricative consonants are divided into nine different phonemic categories, differentiated by the presence or absence of voicing coupled with five
different places of articulation. A list of these fricatives are as follows: labiodental - /f v/; dental - /θ ð/; alveolar - /s z/; palato-alveolar - /ʃ ʒ/; and glottal - /h/. The shape of the vocal tract for the fricative /h/ varies depending on the phonetic context, with its specific place of articulation determined by the articulation patterns of adjacent vowel sounds. Thus, the allophones of /h/ are sometimes referred to as voiceless vowels.

2.1.2.2 Transient noise excitation.

Stop consonants are characterized by an explosion or burst of air from the vocal tract. This burst is initiated when an active articulator such as the tongue or lips create a complete constriction within the vocal tract, thereby obstructing the outward flow of air through the oral cavity. If this oral obstruction is relatively complete and the nasal tract is also closed, as in oral stops, intraoral air pressure builds behind the stop closure. In the following discussion, the term “stop” will be used to refer to oral stop consonants and not include nasal stop consonants. During this period of constriction or stop closure, the acoustic result is relative silence. As the articulators separate, the pressurized air mass behind the closure is released causing a burst of acoustic noise energy.

Similar to the noise source found in fricatives, the stop burst is also composed of turbulent noise, except that it is transient in nature. Most stop bursts are between 10 and 50 ms in duration (Fry, 1979), whereas, the continuous noise source associated with fricative sounds is commonly over 150 ms in length (Jongman et al., 2000). If followed by a voiced speech sound, the stop burst is commonly followed by a transitional period, when the articulators are moving to approximate the vocal tract configuration of the following sound. The period of time between the transient stop burst and the beginning of voicing is termed the Voice Onset Time (VOT). Thus, the VOT may be defined as
the release of the stop closure relative to the onset of voicing. A stop’s VOT is usually positive, but in the case of some voiced stops the VOT is negative. In these instances, voicing begins prior to the release of the stop closure.

American English stop consonants are divided into six different phonemic categories, differentiated by the absence or presence of voicing coupled with three different places of articulation. A list of these stops are as follows: bilabial - /p b/; alveolar - /t d/; and velar - /k g/. A glottal stop also occurs in English, serving primarily as a type of boundary marker between two adjacent vowel segments. However, in American English the glottal stop is often not phonemic in nature (Kent, 1997).

2.1.2.3 Combined transient and frication noise source.

Affricate segments are produced by combining both a transient and frication noise source. Affricate phonemes in American English start with a complete obstruction of the oral cavity near the alveolar ridge, slightly posterior to the constriction for a /t/ or /d/. However, instead of separating the articulators widely to allow an explosive burst of pressurized air, the articulators are only slightly separated. As the pressure behind the closure escapes through the narrow opening between the two articulators, the production of frication noise occurs. Thus, the affricate is a generated by an initial stop closure followed immediately by frication noise. Thus, the voiceless and voiced English affricates are represented in the International Phonetic Alphabet as the symbols /tʃ/ and /dʒ/, respectively.

2.1.3 Filter transfer function

In human speech the acoustic structure of the waveform generated by a sound source or sources is modified by the filter transfer function of the vocal tract. In other
words, the acoustic energy is shaped or filtered by the shape and composition of the vocal tract in front of and behind the sound source. Historically, researchers (Fant, 1960, 1968; Stevens & House, 1961) have theorized that the vocal tract filter transfer function is derived from three components: radiation characteristic, poles, and zeros.

The radiation characteristic of the filter transfer function occurs as the acoustic waveform interfaces with the exterior atmosphere when it leaves the resonating cavities of the vocal tract. This characteristic suppresses the sound energy of lower frequencies more than higher frequencies, thereby acting as a high pass filter. The effect of the radiation characteristic is a relatively constant +6 dB/octave, regardless of the size of the mouth opening (Fant, 1968).

A pole is the summation of natural resonating frequencies within the cavities of the vocal tract. The term formant is often used when discussing the poles of a vowel spectrum. Zeros or antiformants are the inverse of poles, creating impedance to the transmission of particular sound frequencies. Thus, the zero component of a transfer function would eliminate a pole of identical frequency and limit the resonance of adjacent frequencies (Fant, 1968). Zeroes generally occur when the vocal tract is bifurcated, as is the case in the nasal cavity (nasalized vowels or consonants) or severally restricted (stops and fricatives).

In summation, the poles and zeros of the filter transfer function describe the resonating characteristics of the cavities within the vocal tract, and are therefore directly determined by the length and shape of those cavities. A primary constriction at the alveolar ridge would result in a relatively short front resonating cavity and a somewhat longer back cavity. In the case of an alveolar place of articulation with a transient noise
source [t or d], the filter transfer function allows an initial burst of high frequency sound energy at approximately 4 kHz to pass through the vocal tract, whereas, velar stops [k or g], with a more centralized articulation, exhibit a peak of acoustic energy in a lower frequency range around 2 kHz (Fry, 1979). Thus, the terminal nature of the sound structure is dramatically affected by the location of the constriction in the vocal tract of the constriction.

2.2 Spectral moment analysis

In the last decade researchers have begun to utilize spectral moment analysis in the search for invariant patterns of acoustic energy within the noise portion of obstruent speech sounds. This method of acoustic analysis uses a Fast Fourier transformed (FFT) spectra derived from a single window of analysis. Each FFT spectra is then considered a random distributed probability, from which four statistical moments are often calculated. Since spectral moments are in some sense a spectral “snap-shot” of the acoustic energy within a particular window of analysis, individually these statistics are static in nature. However, if these snap-shots are combined and examined in sequential order, spectral moment analysis can also provide information regarding the dynamic characteristics of a sound segment.

The first spectral moment describes the average energy distribution of a FFT power spectrum derived from a window of analysis, while the second moment describes the variability of frequencies over which the power spectrum is spread. The third spectral moment describes the skewness of the FFT distribution. Thus, a positive skewness would indicate that the median component of the FFT spectrum has a higher frequency than the mean. The spectral skewness is sometimes referred to as the “spectral tilt.” The
fourth moment describes the kurtosis or peakedness of the energy distribution. A low kurtosis value would indicate that the spectral distribution is relatively flat, while a high value reflects the presence of relatively well-defined spectral peaks (Jongman et al., 2000).

Historically, research involving the spectral moment analysis of obstruent speech segments might be divided into two divisions. One line of research examined if this type of statistical analysis could provide a high rate of obstruent classification. More specifically, these studies aimed to investigate if stop burst and fricative speech sounds could be statistically differentiated in terms of place of articulation based on patterns of spectral energy within their associated noise segments (e.g., Forrest et al., 1988; Jongman et al., 2000). A second associated area of research has utilized spectral moment analysis to examine how the acoustic properties of stop bursts and/or fricatives change as a function of normal speech development and aging (Nittrouer, 1995).

2.2.1 Obstruent discrimination

One of the earliest studies to utilize spectral moment analysis in examining the acoustic structure of speech segments was conducted by Forrest et al. (1988). In part, this study examined the spectral energy of three voiceless stops (p, t, and k) in word initial position. Obstruent segments were extracted from monosyllabic real words produced by 10 adult speakers of American English. Each target word was embedded in the carrier phrase “I can say ______, again.” However, as noted by the authors, the speech sample was not balanced by vowel type.

Following digitization at 20 kHz, the speech samples were then low and high pass filtered at 10 kHz and 70 Hz respectively. From this digitized waveform, the noise
portion of each obstruent was segmented for spectral analysis. A series of consecutive and overlapping portions of the waveform were analyzed using a 20 ms Hamming window. Both linear and Bark transformed power spectra were created for each analysis interval by pre-emphasizing the signal and thereafter applying a 512-point fast Fourier transform (FFT). These individual FFTs were then considered by Forrest et al. (1988) to be random distribution probabilities, from which they computed the first four statistical moments (mean, variance, skewness, and kurtosis).

Forrest et al. (1988) reported that the voiceless stops could be visually discriminated by spectral moments. Graphic representation of the first analysis window (to burst +10) indicated that the place of stop articulation could be visually differentiated if the spectral mean, skewness, and kurtosis were evaluated concurrently. Based on a visual analysis of the data, the investigators concluded that the second spectral moment of variance was not a contributing feature in the discrimination of the different voiceless obstruents and was therefore not included in subsequent statistical analysis.

Interestingly, more recent studies utilizing spectral moment analysis (e.g., Forrest et al., 1990; Nittrouer, 1995) have also not included spectral variance when examining the acoustic properties of obstruents. Jongman et al. (2000) concluded that the disregard for the second spectral moment in recent literature may be due in part to this original finding by Forrest et al..

Using the linear spectral moment data from the stop burst as input variables, the investigators conducted a stepwise discriminant analysis. This analysis further signified that the voiceless stops could be differentiated on the basis of the spectral patterns of energy within the burst of the stop. When utilizing only one window of analysis, all three
stops were classified with an average of 79.9% accuracy. However, when using dynamic spectral moment data from the first 40 ms of the burst, the classification rate increased to an average of 92%. More specifically, bilabial stops were classified at a rate of 95.4%, velar stops exhibited a slightly lower rate of 92.6%, and alveolar stops were correctly classified at a rate of 88%. Furthermore, a significant effect of speaker gender was not reported. When the discriminant function was “trained” on data from male speakers, 94% of all female productions of voiceless stops were classified successfully.

Interestingly, unlike the results of Kewley-Port (1983), a bark transformation of the power spectra actually decreased the classification accuracy of the stop burst spectra. Classification ambiguities were most commonly found between adjacent places of articulation (i.e., /p/ and /t/; /t/ and /k/). However, it is important to note that the study by Kewley-Port utilized trained judges classifying running LPC spectra rather than spectral moment analysis drawn from FFT spectra.

Historically, the spectral moment analysis of fricative consonants has proven less successful in discriminating between places of articulation than similar analysis conducted with stop consonants. As previously outlined in the earlier discussion concerning the spectral analysis of stop consonants, one of the earliest acoustic investigation of American English fricatives utilizing spectral moment analysis was conducted by Forrest et al. (1988). In this publication, the authors also reported on the acoustic characteristics of four voiceless fricatives (/f, θ, s, Ъ), which vary in terms of place of articulation. The methods by which the spectral moments were obtained and subsequently analyzed is similar to that utilized with stop consonants.
When all places of articulation were included in a discriminant analysis, it was found that voiceless fricatives were classified at a lower rate than voiceless stops. When using discriminant functions based on linear spectral moment from the initial 40 ms of the noise segment, the correct classification of the voiceless fricatives was less than 74.5%. For bark transformed moment data the rate was only slightly higher at approximately 77.7%.

When analyzing the initial 20 ms of frication, a relatively high proportion of the voiceless sibilant fricatives were correctly classified. A rate of 82.7% was reported for linear moments, while bark transformed spectra were classified at the high rate of 98.3%. Unlike the discrimination of voiceless stops which relied primarily on first spectral moment (mean), it was reported that the high rate of discrimination between the sibilant fricatives was based largely on the variable of spectral skewness.

As expected, examination of the classification ambiguities revealed that with both the linear and bark transformed data, most of the classification errors involved the non-sibilant contrast between /f/ and /θ/. Nearly half of the /θ/ tokens were incorrectly classified as /f/. Forrest et al. (1988) interpreted these results as possible evidence that the critical spectral cues for this particular contrast may occur later in the segment, in a transitional region approximately 20 ms prior to vowel onset. As noted by the authors, this conclusion is also supported by an earlier investigation of fricative spectral orientation (Bladon & Seitz, 1986).

In a partial replication of the methodology employed by Forrest et al. (1988), Tomiak (1990) further investigated the validity of the spectral moments invariant with voiceless fricative obstruents. A series of experiments analyzed a corpus of 35 different
CV syllables produced by six adult speakers (3 male/3 female). The syllables were a combination of a voiceless fricative /f, θ, s, ʃ, h/ followed by one of seven different monophthongal vowels /i, e, æ, a, u, ou, o/. In addition, each target word was embedded in the carrier phrase “please say _____ to me”. In a perceptual identification task, the recorded tokens were then categorized by a group of listeners as either “well identified” or “poorly identified” tokens.

In an initial experiment, Tomiak (1990) utilized a series of consecutive and overlapping 15 ms analysis windows, the first of which was centered over the start of the fricative segment. The number of analysis windows for each token varied depending on the overall duration of a particular fricative segment. Following the procedure outlined in Forrest et al. (1988), both linear and Bark transformed spectral moments were derived for each fricative token.

Tomiak found that when utilizing the first four spectral moments, a discriminant function was moderately successful in classifying the voiceless fricatives. Analysis applied to linear and Bark transformed data from all well identified tokens were classified at 78% and 74% respectively. Similar to the results reported by Forrest et al. (1988), it was reported that the classification of the sibilant fricative tokens (96% for both /s/ and /ʃ/) were significantly higher than the nonsibilant fricatives (67% for /f/; 44% for /θ/). When the two nonsibilant fricatives were excluded from the analysis, the overall classification rate increased to approximately 92%.

Interestingly, similar to the findings of Forrest et al. (1988), Tomiak (1990) did not find a significant classification advantage for Bark transformed spectral moment data over linear derived data. Assuming that the Bark transformed data is an accurate model.
of the actual processing of the peripheral auditory system, the author interpreted this lack of advantage as a “perceptual failing of the (spectral moments) metric” (Tomiak, 1990, p. 187). However, Tomiak further argued that since a Bark transformation is probably only an approximation of the functioning of the human auditory system, this seeming perceptual inconsistency was premature.

In addition, Tomiak (1990) designed several follow-up experiments to investigate whether the classification of voiceless fricatives could be improved by changing the location and length of the analysis windows. By utilizing windows of longer duration and differing locations, Tomiak aimed to more precisely target the location and duration of perceptually salient fricative information used during perception.

Results from these experiments indicated that spectral moment data from longer analysis windows (encompassing approximately 100 ms of sustained frication) may be more successful in correctly classifying voiceless fricatives. In addition, the poor classification of the nonsibilant fricatives (/f, θ/) was shown to improve when the analysis window included portions of the fricative offset. The author postulated that the salient perceptual information used to identify nonsibilant fricatives may be located in the segment of frication containing a transition to the following vowel. From these findings, Tomiak (1990) concluded that that classification based on spectral moment data could be improved if “token-appropriate analysis windows” were utilized during analysis.

The study conducted by Tomiak (1990) was comprehensive, considering that it utilized multiple window sizes and locations to provide spectral information on both sibilant and nonsibilant fricatives. However, it is important to note that since statistical tests of comparison (analysis of variance) were not applied to the spectral moment data,
reported differences between fricative spectral characteristics were only observational in nature.

In a more recent investigation, Jongman et al. (2000) further utilized spectral moment analysis to describe the acoustic characteristics of voiced and voiceless American English fricatives. This study was a comparative analysis of both static (spectral peak location, spectral moments, noise amplitude, noise duration, and F2 onset) and dynamic (locus equations and relative amplitude) acoustic parameters. The authors statistically compared their measurements with analysis of variance and discriminant analysis.

Similar to previous research (e.g., Heinz and Stevens, 1961; Shadle, 1990; Behrens and Blumstein, 1988a), Jongman et al. (2000) found that many acoustic parameters could be used to distinguish between sibilant and nonsibilant fricatives. However, contrary to earlier findings, the authors found that the spectral peak location and both normalized and relative amplitude could be used to distinguish among the four places of fricatives. In addition, Jongman et al. also found that place of fricative articulation could also be differentiated by spectral moments.

Jongman et al. (2000) obtained spectral peak locations using a 40 ms Hamming window centered over the mid-point of the fricative segment. Power spectra were then derived from each analysis window by preemphasizing the signal by 98% and subjecting it to a FFT. An accompanying LPC analysis was also conducted for purposes of comparison. Jongman et al. (2000) defined the spectral peak location as the frequency of the highest-amplitude peak of each FFT spectrum.
From a four-way analysis of variance (place x voicing x vowel x gender) conducted on the dependent measure of spectral peak location, the authors found a strongly significant main effect of place of articulation $[F(3,2876)=1083.72, p<0.0001, \eta^2 =0.512]$. Follow-up post-hoc tests (Bonferroni) indicated that all places of articulation were significantly different from each other. After collapsing the data across speaker, voicing, and vowel context, the average spectral peak location for the labiodental, dental, alveolar, and palato-alveolar places of fricative articulation were reported at 7733, 7470, 6839, and 3820 Hz, respectively. Thus, as the place of articulation moved to a more posterior position in the vocal tract, the overall spectral peak location decreased in frequency. Interestingly, this pattern did not remain the same across gender. Jongman et al. (2000) reported that female speakers in the study were found to display a slightly different pattern, with dental fricatives having a higher peak location than the more forward articulated labiodental fricatives.

Jongman et al. (2000) also calculated the acoustic measures of normalized and relative amplitude. Normalized amplitude was calculated by taking the difference in dB between the overall fricative amplitude and the rms amplitude of a small portion of the vowel (3 glottal cycles) centered over the location of its maximum amplitude. Relative amplitude was the difference between fricative amplitude and vowel amplitude at strategically selected frequency regions. For the fricatives /s, z, ñ, and ñ/, components within the F3 region (third vowel formant) were selected, whereas, components within the F5 region (fifth vowel formant) were used when analyzing the fricatives /f, v, ð, and ð/.
Supporting the previous findings of Behrens and Blumstein (1988a, b), Jongman et al. (2000) found that the acoustic measure of normalized amplitude could be used to distinguish between sibilant and nonsibilant fricatives. However, contrary to this earlier research, it was found that measures of both normalized and relative amplitude also differed significantly as a function of place of articulation. A four-way ANOVA \[F(3,2876)=1489.51, p<0001, \eta^2=0.591\] and subsequent post hoc tests indicated that all places of articulation were significantly different from each other with regard to normalized amplitude.

Measures of relative amplitude were also found to distinguish between all places of articulation. Interestingly, a main effect of gender was also found \[F(1,2876)=28.73, p<0001, \eta^2=0.009\]. The relative amplitude measures were decreased for female speakers (-9.8 dB) when compared to the male speakers (-8.1 dB). This difference was reportedly most distinct with dental fricatives (/θ, ð/).

Using analytical methods similar to those described in previous research (i.e., Forrest et al., 1988; Tomiak, 1990), this study (Jongman et al., 2000) also examined the first four spectral moments (including variance) of the fricative tokens. For each token, the authors derived FFT power spectra from 40 ms Hamming windows located at four different locations: the onset, mid-point, and end of frication, as well as centered over the fricative-vowel transition. The spectral moments were calculated from both linear and Bark transformed spectra. However, since no significant differences between the two types of spectra were found, only linear spectral moments were reported.

Jongman et al. (2000) found that overall the four spectral moments (mean, variance, skewness, and kurtosis) distinguished between the fricatives in terms of place of
articulation. These effects for fricative place remained quite robust even when comparing across window locations. It was reported that for all spectral moments at least three places of articulation were distinguished at all analysis windows, with an advantage to the first (onset) and fourth (transition) windows. These window locations were found to contain a small increase in distinctive information.

Unlike previous spectral research on fricatives (e.g., Forrest et al., 1988), Jongman et al. (2000) also examined the second spectral moment (variance). The authors reported that the spectral variance differed as function of place of fricative articulation. Post hoc tests (Bonferroni) revealed that most of this effect was due to strong differences in variance between sibilant and nonsibilant tokens. Sibilant fricatives exhibited low variance (3.15 MHz), while the nonsibilant fricatives were found to have almost twice the variance (6.28 MHz).

The authors also reported a main effect of gender for all four spectral moment measures. Female speakers were found to have a significantly higher overall spectral mean (5286 for females; 5018 for males), variance (4.9 MHz for females; 4.5 MHz for males), and kurtosis (1.64 for females; 1.44 for males). The spectral skewness for fricative tokens produced by female speakers was significantly lower (.084 for females; .145 for males). The authors summarized the spectra of the female speakers as having “clearer peaks and a concentration of energy towards higher frequencies” (p. 1257).

In addition, Jongman et al. (2000) also conducted a stepwise linear discriminant analysis on a subset of the acoustic data. This analysis was based on the following 21 predictor variables: spectral peak location, 4 spectral moments x 4 window locations, normalized and relative amplitude, duration, and F2 onset. Overall classification for this
analysis was reported at a rate of 77%. The rate of classification for the labiodental, dental, alveolar, and palato-alveolar places of fricative articulation were reported at 53%, 48%, 81%, and 88%, respectively. Similar to previous findings (e.g., Forrest et al., 1988; Tomiak, 1990), an examination of classification ambiguities revealed that most errors did not cross the “sibilant/nonsibilant distinction” (Jongman et al.). For example, under 5% of nonsibilant fricatives (/f, v, θ, δ/) were confused with sibilant fricatives (/s, z, ʃ, ʒ/), while over 26% of labiodentals (/f, v/) and dentals (/θ, δ/) were confused with each other. A subsequent analysis of the discriminant function coefficients indicated that the most useful variables in the fricative classification were spectral peak location, normalized amplitude, relative amplitude, and the first spectral moment (mean) at the onset and midpoint of frication. These five variables alone could account for a classification rate of 69%.

2.2.2 Spectral Moments and Phonological Development

Research involving the spectral analysis of obstruent speech segments has provided experimental evidence for the existence of distinct spectral patterns of acoustic energy. In many cases, these spectral patterns (spectral moments) have been found to facilitate the classification of obstruent sounds according to phonetic categories. In particular, spectral moments have been successfully utilized to distinguish between stop and fricative consonants in terms of place of articulation (e.g., Forrest et al., 1988; Jongman et al., 2000).

A second associated area of study has employed spectral moment analysis to investigate the changes within the acoustic properties of obstruent consonants (stops and fricatives) as a function of disordered and normal phonological development. In a study
in 1990, Forrest, Weismer, Hodge, Dinnsen, and Elbert examined the spectral characteristics of alveolar and velar stops in both normal developing and phonological disordered children. Perceptual assessment suggested that the disordered children did not have /k/ in their phonological inventories, substituting /t/ for /k/ in all word positions. The four children in both the normal and disordered subject groups were between 3.5 and 6.5 years of age.

The authors conducted a discriminant analysis based on the first, third, and fourth spectral moments derived from the first 40 ms of the stop burst. It was found that speech tokens drawn from the normally developing children were successfully classified with an overall rating of 82%. Examination of the canonical weights revealed that spectral mean and kurtosis were the primary variables utilized in the development of the discriminant function.

When applied to the spectral moment data obtained from the phonologically disordered children, the same discriminant model used with the normal children failed to discriminate between the alveolar and velar stops. Forrest et al. (1990) found that only one of the phonologically disordered children showed spectral differences (1st, 3rd, and 4th moments) between their /t/ and /k/ productions. Although this child’s alveolar and velar stops were not perceptually different, a discriminant analysis indicated they could be classified with 87% accuracy. However, the acoustic patterns utilized to differentiate between the /t/ and /k/ productions were markedly different than that of the normally developing children. Specifically, this child’s stops were being classified mainly on differences in kurtosis and skewness, rather than spectral mean. These results indicate that although a child’s stops may be successfully classified acoustically, the perceptual
salience of the stop contrast is dependent upon which spectral measures are found to be different. Interestingly, the phonologically disordered child that exhibited imperceptible acoustic differences between /t/ and /k/, eventually obtained the contrast without direct intervention. Whereas, the other members of the phonological disordered group did not acquire the /k/ sound.

More recent research conducted by Miccio (1996) utilized spectral moment analysis to investigate the acquisition of voiceless fricatives in normal and phonologically disordered children. This study was a longitudinal look at how the acoustic patterns of the /f, θ, s, and j/ fricatives might change as process of normal development or the function of clinical treatment. Three different groups of children participated in the study: a younger normal group, an older normal group, and a group of phonologically disordered children. Speech was elicited from the normal groups of children at the start of the experiment and five months later, while the fricative samples were collected from the disordered group prior to treatment, post treatment, and two months following the end of treatment. The spectral moment data (1st, 3rd, and 4th) were gathered from the initial 40 ms of each fricative token using an overlapping 20 ms Hamming window.

Similar to findings reported by Forrest et al. (1990), Miccio (1996) found that normally developing children exhibited acoustic differences between sounds prior to the emergence of a perceptual contrast. For example, the author found that acoustic differences in /s/ and /j/ occurred before a perceptual contrast between the two sounds was acquired. Miccio concluded that these acoustic distinctions may be predictors of phoneme-specific acquisition.
Although previous research by Kent (1976) indicated that variability in sound
production decreases with age, Miccio found no differences in variability across different
collection times. The author concluded that this lack of difference may be due to the
limited period of time between collections (5 months) or because of differences in
analysis methodology (temporal measures vs. spectral moment). It is important to note
that Miccio did not compute the second spectral moment, but rather compared the mean
variability coefficients between the second, third, and fourth analysis windows for the
relevant spectral moments. Interestingly, Miccio found that the variability of the first
spectral moment was higher in normal developing children than those classified as
phonologically disordered. A similar result was revealed by Forrest et al. (1994) when
investigating the contrast between alveolar and velar stops.

Miccio (1996) concluded that the acquisition of fricative contrasts followed a
developmental pattern. For example, all members of the youngest normal group first
acquired a distinction between sibilant and nonsibilant types of fricatives. These
individuals then developed contrasts between the two sibilant fricatives (/s/ and /ʃ/), after
which the oldest group of normal children was then able to exhibit a difference between
the two voiceless nonsibilant fricatives (/θ/ and /ʃ/). Across subjects in the treatment
group, the order of fricative attainment was reported to begin with /θ/ and /s/, after which
/ʃ/ and then /θ/ were then acquired and maintained.

One of the most influential studies on the developmental order of sound
acquisition utilizing spectral moment analysis was conducted by Nittrouer in 1995.
Through the examination of the acoustic patterns of several obstruents (/t, k, s, ʃ/) as
revealed by spectral moment analysis, Nittrouer aimed to describe any age-related
differences in the spectral properties of obstruent consonants. Specifically, the author investigated the theory that some of the articulatory gestures of children are not as precisely specified as those of adult speakers.

This study was a partial replication of an earlier experiment (Nittrouer et al., 1989), whereby, it was found that the spectral mean or centroid of sibilant fricatives (/s/ and /ʃ/) were more similar in children (3 to 7 years of age) than adults. The authors interpreted this finding as support for the conclusion that in terms of the first spectral moment the adult speakers exhibited a greater distinction or contrast between sibilant fricatives. In this study and subsequent research, Nittrouer and colleagues often refer to these distinctions as “consonant effects.”

In her 1995 study, Nittrouer collected speech samples from ten adults and 30 children ranging in age from 3 to 7 years. The corpus of recorded tokens included 12 consonant-vowel syllables created by combining the four obstruents /t, k, s, and ʃ/ with the monophthongal vowels /i, a, and u/. Each target syllable was produced in the carrier phrase “It’s a ____ Bob” and repeated five times. For the stop consonants, the 1st, 3rd, and 4th spectral moments were derived from the first 20 ms of the stop burst. Spectral moments from the fricatives was calculated from a 50 ms portion of frication located at least 75 ms prior to the voicing of the following vowel. The means of the spectral measures were then collapsed across repetition prior to statistical analysis.

In terms of the first spectral moment, Nittrouer (1995) found significant main effects of place of articulation for both the stop and fricative contrasts. This is not surprising considering the smaller front resonating cavity associated with the more anteriorly articulated sounds in each contrast (/t/ and /s/). However, of more interest was
the finding that the difference in spectral mean values between /s/ and /ʃ/ were larger for adult speakers than for children ages 3 through 7, yet the differences between /t/ and /k/ were relatively similar. From these findings, Nittrouer suggested that “. . . children do not initially differentiate articulatory gestures as well as adults do, and subsequently master some articulatory gestures sooner than others” (1995).

In addition, main effects for vowel context were also found to be significant in all age groups, indicating that even the youngest speakers were anticipating the articulation of the following vowel. Specifically, Nittrouer (1995) found that when collapsed across speaker and place, the obstruents preceding an /i/ vowel had the highest spectral mean. For the stop consonants, vowel context effects were reported to be greatest for velar stops. Nittrouer postulated that this place by vowel interaction was result of a forward movement of the tongue body in anticipation of the high front articulation of the /i/ vowel. This anticipatory action would then produce a shortening of the anterior resonating cavity, resulting in an increase in the spectral mean of the velar stop burst.

Interestingly, Nittrouer (1995) also found that the anticipatory coarticulation effects within the stops varied as a function of the age of speaker, with the children exhibiting a stronger effect of vowel context than the adults. Through subsequent analysis, the author reported that the increased vowel-context effect of the child speakers was derived from an increased sensitivity found only in velar stops, with no age differences found for alveolar stops.

The findings of Nittrouer (1995), as well as the many others that have been previously discussed, indicate that spectral moment analysis is a valuable tool in examining the acoustic structure of obstruent sounds. However, despite recent work in
this area, there is still a limited amount of knowledge on the acoustic structure and developmental nature of young children’s speech. Thus, the general aim of this dissertation is to utilize the relatively new technique of spectral moment analysis to further study the obstruent productions of adults and typically developing children between 3 and 5 years of age.
CHAPTER 3

METHODOLOGY

The following chapter describes the methods by which a sample of speech from three groups of children and one comparative group of adults was elicited, recorded and subsequently analyzed. Acoustic and statistic analysis followed procedures outlined in previous research examining the acoustic properties of voiceless fricatives, namely Forrest et al. (1988) and Jongman et al. (2000). Methods used in this study were also draw from previous pilot studies completed by the author and additional colleagues (i.e., Dr. Robert A. Fox, Dr. Julie McGory, Dr. Elaina Freida, and Kimberly Rosenbauer). The elicitation, recording and analysis of these data were facilitated by several Matlab computer programs, created specifically for this study. These programs were created by the author, in cooperation and consultation with Dr. Robert A. Fox.

3.1 Subjects

Three groups of children between the ages of 3 and 6 (N = 30) and one comparison group of adults were recruited to participate in the study (N = 10). Subjects in the 3-year old group were between 3:0 and 3:11 years of age; the 4-year old group ranged between 4:0 and 4:11 years of age; and the 5-year old group contained children between 5:0 and 5:11 years of age. The adult subjects within the comparison group were between 18 and 40 years of age. Each group was
composed of an equal number of male and female subjects. Subjects were recruited from university and community preschool programs, local churches, community activity groups, and an already established database of former research subjects. Adult subjects were paid for their participation in the study. The parents or legal guardians were paid for their child’s participation.

All child and adult subjects were native speakers of American English. Parental report indicated that no children participating in the study had a diagnosed history of a speech, language, or hearing problem (excluding episodes of Otitis Media). Prior to participation all subjects were required to pass a hearing screening with pure tone air conduction thresholds of 25 DBHL for the frequencies 0.5, 1.0, 2.0, 4.0, and 6.0 kHz. An oral/motor screening was administered to the child subjects to ensure normal craniofacial structure and musculature. In addition, at the time of their participation in the study all subjects had visible front incisors.

Furthermore, all child speakers were required to pass a phonological screening (Goldman & Fristoe, 1986) prior to testing. This screening was conducted to determine if each child exhibited age-appropriate phonological development. Since many of the targeted phonemes are typically acquired between the ages of 3 and 6, no subject was excluded from the study based solely on the acquisition or quality of any particular phoneme, but only on the basis of a general standard of age appropriate phonological development. Thus, the GFTA was utilized as a screening tool and not for the purpose of a full phonological evaluation. If any subject failed any of the above-mentioned screening protocols, the party with legal consent was notified and given the appropriate referrals regarding follow-up evaluation.
3.2 Stimuli

The corpus of elicited productions consists of two different sets of stimulus items. The first set of productions contains 21 real words derived from 7 different voiceless obstruents /p, t, k, f, θ, s, j/ in initial position followed by 3 different monophthongal vowels /i, a, u/. This set of items were embedded within both monosyllabic and multisyllabic real words. To elicit relatively similar vocal emphasis across different productions, the targeted CV(C) syllable combinations were always in the initial and stressed position of each elicited word. Stimulus words in this set were produced in the carrier phrase “A _____.” In rare instances, an age appropriate word did not exist that exemplifies the targeted syllable. For example, since American English does not contain an age-appropriate word beginning with a /θu/ consonant-vowel combination, a proper name that follows the phonotactic rules of Standard English was substituted for the missing word. In this case, subjects were instructed to produce the proper name “thoot” in reference to a fictional character.

The second set of productions consisted of 21 nonsense words, created by preceding the above mentioned set of voiceless obstruents with the monophthongal vowel /o/ and followed by the vowels /i, a, u/ in a VCV syllable structure. Stimulus words in this set were produced in isolation. When producing this set of stimuli, speakers were instructed to place lexical stress on the second syllable, which contains the targeted consonant.

Since both sets of stimuli contained 21 words, the entire corpus of elicited productions for each subject totaled 42 different items. Stimulus items were repeated five times, yielding a total of 210 tokens per subject. By protocol, only the second, third, and fourth token of each stimulus item was segmented and subsequently analyzed. However, if one of these medial
tokens was found to be defective, the first or fifth token of that particular item was then used as an alternate. A token was considered defective if it contained peak-clipping or was only partially recorded.

3.3 Procedure and Analysis

3.3.1 Recording

Speech samples were recorded online to computer in a quiet room environment. More specifically, a high quality Shure SM10A-CN low impedance dynamic microphone and a Samson Mixpad-4 preamplifier were used to facilitate the recording of subject productions. The microphone was affixed to a head-set and placed approximately 4 cm from the speaker’s lips during recording.

The speech tokens were sampled at a rate of 44.1 kHz with a quantization of 16 bits by a Sound-blaster compatible sound card and subsequently saved directly to an internal computer disk. Following the recording of each stimulus token, a graphic presentation of the token was viewed to identify inappropriate recording levels (peak-clipping) or an insufficient recording window. If any of the above mentioned conditions occurred the token was rerecorded.

The real-word productions in stimulus set 1 were elicited through picture identification. Pictures representing the target words were presented on a 15 inch computer screen, positioned approximately 2 feet from each subject. All pictures utilized in the study were selected to be age appropriate for preschool aged children. The subjects were familiarized with the names of the pictures and the elicitation procedure prior to the recording session. If a subject incorrectly identified a picture as a different lexical item during the recording session, a prompt was then given and that particular item was rerecorded.
The nonsense words of stimulus set 2 were elicited by imitation. A pre-recorded sample of the nonsense words were used to provide the subjects with a consistent model to imitate. The elicitation order of both stimulus sets were randomized between presentations and across subjects. All elicitation and recording of stimuli was facilitated by a custom designed computer program written in Matlab programming language by the author and Dr. Robert A. Fox.

3.3.2 Segmentation

Segmentation of the target segments were conducted using waveform display assisted by spectrographic inspection (Cool Edit, 2000). To test for accuracy and reliability, 100 randomly chosen tokens were independently analyzed by a second party and subsequently correlated (r=.997, p<.001) to the original segmentation of these same tokens. Tokens were segmented at the following points:

1) Initial vowel onset: Initial moment of periodic energy
2) Initial vowel offset: The onset of the stop closure
3) Stop closure (intervocalic position): abrupt decrease in acoustic energy
4) Stop closure release: Sharp increase in diffuse noise energy
5) Frication onset: Rapid increase in zero-crossings and/or spectrographic identification of high frequency energy
6) Frication offset: Decrease in zero-crossings and/or absence of high frequency energy
7) Final vowel onset: Initial moment of periodic energy
8) Final vowel offset: End of periodic energy

Examples of these segmentation points applied to syllables containing a stop and fricative target are illustrated in figures 3.1 and 3.2, respectively.
3.3.3 Static Measurements and Analysis

3.3.3.1 Stop consonants.

Two series of analysis windows were used to analyze the acoustic structure of the stop productions. The first series was composed of five consecutive and overlapping 10 ms Hamming windows, with the first window in the series centered over the stop closure release. Subsequent windows were shifted forward in time by 5 ms increments, resulting in an overlap of 50%. Thus, the center of the last interval in the series was located 20 ms after the start of the stop burst. An example of this series of analysis windows is illustrated in figure 3.3. Each 10 ms window interval was then pre-emphasized by first-differencing. Though the need for pre-emphasis is minimalized when analyzing voiceless sounds, it was determined that such a procedure was necessary to more effectively compare subsequent results to findings previously published (e.g., Forrest et al., 1988; Nittrouer, S., 1992, 1995; Jongman et al., 2000).

Using a 512-point FFT, the spectral amplitudes of a series of frequency points were derived from the complex acoustic signal within each 10 ms window. In an FFT, the frequency resolution or spacing of the frequency points is determined by the window size and sampling rate. For example, this first series of analysis windows yields a 10 ms portion of digitized waveform sampled at a rate of 44.1 kHz per second, which contains a total of 441 numerical points of reference. However, the number of sample points in the FFT analysis must be of binary power (e.g., 512, 1024, 2048, etc.). Therefore, in order to conduct the 512-point FFT, the window of analysis was padded with 71 additional sample points set at zero (441 original reference points + 71 zero points = 512).
The second series of Hamming windows were 20 ms in duration and located at three
different locations. The first analysis interval was centered +10 ms from the release of the stop
burst, the second was centered at the mid-point, and the third window of analysis was centered -
10 ms from the end of the burst. The analysis windows in this series may or may not overlap
depending on the duration of the stop burst. An example of this series of analysis windows is
illustrated in figure 3.3. Similar to the previous series, each window in this series was also pre-
emphasized by first-differencing. However, the increased length of windows in the second series
(20 ms) required the application of a longer 1024-point FFT’s with zero padding.

The resulting FFT spectra from both series of analysis windows contained both real and
imaginary components. Therefore, only half of the frequency points were utilized in additional
calculations, with the 512-point and 1024-point FFT’s resulted in 256 and 512 real value
frequency samples, respectively. The additional frequency points with imaginary values were
disregarded in subsequent calculations. Thus, when recorded at a sampling rate of 44.1 kHz, the
512-point FFT applied to the 10 ms windows yielded an estimate of the spectral amplitude every
86.1 Hz (22,050/256=86.1). Whereas, the 1024-point FFT conducted on the 20 ms windows
produced an amplitude estimate every 43 Hz.

Prior to moment calculation the individual FFT spectra were converted to a normalized
power spectra. This calculation was accomplished by dividing the relative amplitude of each
frequency point by the sum amplitude of analysis points. Thus, following Forrest et al. (1988),
the normalized power spectra for the 512-point FFT’s were derived from the following
computation:

\[
P(k_j) = \frac{P(k_j)}{\sum(P(k_1) + \ldots + P(k_{256}))}
\]

\[P = \text{relative power}\]
\[k = \text{real-valued frequency point (j from 1 to 256)}\]
The normalized power spectra from the 1024-point FFT’s utilized a similar computation, with the exception that the real-valued frequency points (k_j) ranged from 1 to 512. The first frequency point (k_0) or dc component of each recorded sample does not provide useful information and is therefore not utilized when computing the normalized power spectra.

The normalized power spectra derived from the two series (10 ms, 20 ms) mentioned above were considered random distribution probabilities, from which the first four statistical moments were then computed. The spectral mean statistic was computed by taking the sum of each frequency point multiplied by the relative power of that point. Thus, the computation of the spectral mean statistic is as follows:

\[
\text{Spectral mean} = \sum [k_j (P_j)]
\]

\[
P = \text{relative power}
\]

\[
k = \text{real-valued frequency point}
\]

\[
j = \text{from 1 to 256 for 512-point FFT spectra}
\]

\[
\text{from 1 to 512 for 1024-point FFT spectra}
\]

The spectral variance statistic was calculated by taking the sum of each frequency point’s squared deviation from the mean. This calculation is summarized in the following notation:

\[
\text{Spectral variance} = \sum [(k_j - m_1)^2 P_j]
\]

\[
P = \text{relative power}
\]

\[
k = \text{real-valued frequency point}
\]

\[
j = \text{from 1 to 256 for 512-point FFT spectra}
\]

\[
\text{from 1 to 512 for 1024-point FFT spectra}
\]

\[
m_1 = \text{spectral mean}
\]

The spectral skewness statistic is a reflection of how the acoustic energy is distributed around the mean. This statistic is sometimes referred to as the “spectral tilt” because it conveys
the overall slant of acoustic energy. The $m_3$ was computed as follows:

$$\text{Spectral skewness} = \sum [(k_j - m_1)^3 P_j]$$

- $P$ = relative power
- $k$ = real-valued frequency point
- $j$ = from 1 to 256 for 512-point FFT spectra
  - from 1 to 512 for 1024-point FFT spectra
- $m_1$ = spectral mean

Since a direct comparison of skewness across different levels of variance is inappropriate (Forrest et al., 1988), the spectral skewness statistic was normalized and expressed as a coefficient. This computation may be expressed as the following:

$$\text{Spectral skewness}_{\text{normalized}} = \left[\frac{m_3}{(m_2^{3/2})}\right]$$

- $m_2$ = spectral variance
- $m_3$ = spectral skewness

The fourth spectral moment of kurtosis indicates the peakedness of the spectral distribution. A negative kurtic coefficient indicates a relatively flat spectral distribution, whereas a positive coefficient is characteristic of more prominent spectral peaks. The spectral kurtosis was computed as follows:

$$\text{Spectral kurtosis} = \sum [(k_j - m_1)^4 P_j]$$

- $P$ = relative power
- $k$ = real-valued frequency point
- $j$ = from 1 to 256 for 512-point FFT spectra
  - from 1 to 512 for 1024-point FFT spectra
- $m_1$ = spectral mean

To allow for direct comparisons, the fourth spectral moment was normalized for differences in spectral variance. In the fashion of Forrest et al. (1988), this normalization was conducted by the
following calculation:

\[
\text{Spectral kurtosis}_{\text{normalized}} = \left( \frac{m_4}{m_2^2} \right) - 3
\]

\( m_2 = \text{spectral variance} \)

\( m_4 = \text{spectral kurtosis} \)

All digital signal processing and acoustic analysis was conducted by custom designed computer programs written in Matlab programming language by the author and Dr. Robert Fox. A corpus of test tokens comprised of known acoustic components was utilized to evaluate the accuracy and reliability of these computer programs. For example, a test token composed of several sinusoidal frequencies (1 kHz, 3 kHz, and 5 kHz) of equal strength was analyzed by the computer programs and found to have the appropriate values for the various acoustic measures.

*Spectral peak loci* were estimated for each stop burst using the FFT interval spectra generated to compute the spectral moments. The overall spectral peak location of each stimulus token was based on the frequency component with the highest relative amplitude within a frequency range from 1.0 kHz to 11 kHz. To decrease the effects of lower frequency acoustic energy from possible voicing, frequencies below 1000 Hz were disregarded when figuring the spectral peak loci.

To ensure that the spectral peak locations were being calculated correctly, an addition measure of the overall spectral peak location was calculated using linear predictive coding (LPC) rather than FFT analysis. LPC is a type of acoustic analysis that predicts the resonating characteristics of the vocal tract from a small portion of the waveform. These characteristics are represented by a small number of combined coefficients, which form a “linear predictive formula” (Johnson, 1997). This formula is subsequently modified based on its ability to accurately predict subsequent portions of the waveform. The performance of a particular set of
coefficients is measured by a method of minimized error estimation, a measure of error based on the squared deviation between the predicted and actual value of sample point. LPC is helpful in determining the broad spectral peaks of the spectral envelope. However, unlike FFT analysis, it is ineffective in providing a more narrow spectral view of the harmonic frequencies.

3.3.3.2 Fricative consonants.

*Spectral moment* measures (mean, variance, skewness, and kurtosis) were computed for the fricative segments following the approach of Jongman et al. (2000). The spectral patterns of the fricative tokens were examined using two series of analysis intervals. The initial series employed three 20 ms Hamming windows, while the second series utilized a window length of 40 ms. The first analysis interval in each series was placed at the beginning of each fricative segment. Thus, for the 40 ms window the analysis interval was centered over a location in the waveform 20 ms after the start of frication. Subsequent intervals in each series were located at the mid-point and at the end of each fricative segment. An example of each series of analysis window is illustrated in figures 3.5 and 3.6, respectively.

Each analysis interval was pre-emphasized by first-differencing and then analyzed using the appropriate FFT with zero-padding. The initial series of 20 ms windows required the application of 1024-point FFT’s. Larger 2048-point FFT’s were used to derive spectra from the series of 40 ms windows. These individual FFTs were considered random distribution probabilities from which the first four spectral moments were computed. The computations of fricative spectral moments were accomplished according to the methods previously detailed in section 3.3.3.1.

*Spectral peak loci* were also estimated for each fricative using the FFT interval spectra generated to compute the spectral moments. The overall spectral peak location of each stimulus
token was based on the frequency component with the highest relative amplitude within a frequency range from 1.0 kHz to 15 kHz. To decrease the effects of lower frequency acoustic energy from possible voicing, frequencies below 1000 Hz were disregarded when figuring the spectral peak loci. The relative amplitude of the spectral peak is described in terms of dB SPL. Similar to the stop consonant analysis, an overall spectral peak location was also drawn from a separate linear predictive coding (LPC) analysis of the same interval windows. Likewise, a corpus of test tokens comprised of known frequency components was also used to measure the accuracy and reliability of these methods of analysis.

3.3.4 Dynamic Measurements and Analysis

3.3.4.1 Durations and ratios.

Segment durations and ratios will be computed from raw time points reported during the segmentation of the target segments. All duration measurements were reported in ms increments and calculated according to the following means:

Overall token duration (set 2): initial vowel onset to the final vowel offset

Initial vowel duration (set 1): vowel onset to offset

Initial vowel duration (set 2): vowel onset to stop closure

Stop closure duration (set 2): stop closure to stop release

Fricative duration (set 1 & 2): fricative onset to fricative offset

Final vowel duration (set 2): voice onset time to the final vowel offset

Voice onset time (set 1 & 2): periodic voicing relative to the stop closure release

3.3.4.2 Spectral slope.

A measure of spectral slope will be derived from the same power spectra generated during the spectral moment analysis. The slope will be derived from a linear regression line fit
to the relative amplitudes of FFT power spectra of each analysis window. More specifically, this spectral parameter will describe the amplitude slope of the FFT power spectra between 1 and 11 kHz.

3.3.4.3 Normalized amplitude.

The measure of normalized amplitude will be computed for each stop burst and fricative segment following the procedures described by previous researchers (Behrens & Blumstein, 1988b; Jongman, 2000). This measure will be a root-mean-square (rms) ratio of the entire noise segment (burst and frication) minus the rms amplitude of the strongest component within the initial 40 ms of the following vowel (measured in dB). This type of amplitude measure will serve to normalize for differences in speaker intensity (Jongman et al., 2000).

3.3.5 Statistical Analysis

All descriptive measures were collapsed across token repetition and grouped into 4 separate databases according to word (CV and VCV) and obstruent (stop and fricative) type prior to any statistical analysis. The general aims of this dissertation were addressed through the following types of statistics:

1) Descriptive statistics of the various acoustic parameters were used to illustrate the acoustic structure of syllable initial and intervocalic voiceless obstruent productions of the adult and child speakers.

2) Analysis of variance (ANOVA) was used to determine significant acoustic variation in the speakers stop and fricative productions as a function of place of articulation, vowel context, gender, or age group. Thus, the repeated measures ANOVAs were generally of mixed form, with between subject factors of gender and age group and within subject factors of place of articulation and vowel context. Dependent factors
included duration (vowel durations, voice onset time, fricative duration), amplitude (normalized amplitude), and spectral (peak, slope, moments) measures. Additional comparisons with one-way ANOVAs were conducted for each of the dependent measures to determine significant variation across window location and size. Partial eta squared ($\eta^2$) measures of effect size were computed with any significant ANOVA results. Furthermore, post hoc analyses consisted of pairwise comparisons, with Bonferroni adjustments for multiple comparisons.

3) Discriminant analyses is a statistical procedure used to classify individuals or cases into two or more mutually exclusive groups based on a set of predictor variables. One or more of these predictors or quantitative variables are combined to form a discriminant function or model based on how well it can maximize the differences among the groups being classified. In this study, discriminant analysis was used to examine how well a discriminant model based on the derived acoustic parameters could classify individual stop and fricative tokens by place of articulation. The accuracy of the classification is calculated on the percent of cases that are correctly classified into groups based on a particular classification function.

The classification functions will differ based on the acoustic parameters originally entered into the stepwise linear analysis and the particular individual or group of individuals upon which the function was developed. The discriminant functions will be developed using two different types of procedures, the cross-validation or “jack-knife” approach and an approach that utilizes a specific group of the data to train the classification model. In both approaches, variables are entered into the discriminant function in a stepwise manner.
In the cross-validation (jack-knife) approach, data from one subject was excluded and the remaining data used to estimate a discriminant function. This function was then used to classify the omitted subject’s data. Subsequently another subject’s data were omitted from the analysis and replaced by the data excluded originally. Again the discriminant function was estimated based on the larger data set and used to classify the excluded data set. This process was repeated until each set of individual subject data had been successfully or unsuccessfully classified. When the number of individual cases is relatively small, this type of discriminant procedure provides a more accurate classification (Forrest et al., 1990).

In the “specific group” approach, the overall data sample was split into two sub-samples. One sample was employed to develop or “train” a discriminant function and the other sub-sample was then used to validate the resulting function. Discriminant functions were trained on only the adult data, the adult male data, and the adult female data. These functions were then applied to specific speaker groups in the remaining data set. Both types of analysis contained input variables such as: duration measures, normalized amplitude, spectral peak location, slope, and moments (mean, variance, skewness, and kurtosis).
Figure 3.1  Segmentation of stop consonant
Figure 3.2  Segmentation of a fricative consonant
Figure 3.3 10 ms series of analysis windows for the stop consonants
Figure 3.4 20 ms series of analysis windows for the stop consonants
Figure 3.5 20 ms series of analysis windows for the fricative consonants
Figure 3.6 40 ms series of analysis windows for the fricative consonants
CHAPTER 4

ACOUSTIC ANALYSIS OF CV STOP CONSONANTS

The results found in the following section were derived from stops produced in a CV context. All data were collapsed across repetition for a given speaker prior to statistical analysis. Thus, these findings represent a by subject analysis. Unless otherwise noted, any main or interaction effects found to be significant at a .05 alpha level were discussed.

The term “place of articulation” refers to the place of consonant articulation only, while the term “vowel context” will be utilized when referring to differing vowel articulations. Prior to the presentation of findings for the spectral moment measures, comparisons of analysis window size and location will be presented. Immediately thereafter, the rationale for selecting the findings from one window size and location will be discussed, followed by the presentation of spectral results for only that particular analysis window size and location. Although specific statistics from additional analysis windows will not be reported, general patterns in the data that extend across these windows will also be described.
4.1 Duration and Amplitude Measures

4.1.1 Voice Onset Time

The voice onset time (VOT), defined as the initiation of periodic voicing relative to the stop closure release, was calculated for the stop consonants in the CV context. A four-way repeated measures ANOVA, with gender and speaker age as the between subject factors and place of articulation and vowel context as within subject factors revealed a significant main effect for place of articulation \( [F(2,64)=60.35, p<.001, \eta^2=.653] \). Subsequent pairwise comparisons indicated that all three places of articulation were significantly different in terms of VOT \( (p<.01) \). The overall mean VOT increased with posterior placement of articulation in the oral cavity \( (72.7 \text{ ms for } /p/; 81.4 \text{ ms for } /t/; \text{ and } 95.9 \text{ ms for } /k/) \).

A significant main effect of vowel context was also found \( [F(2,64)=23.58, p<.001, \eta^2=.424] \), with pairwise comparisons indicating that VOT differed as a function of all three vowel contexts \( (/i, \alpha, u/) \). These results indicate that the mean VOT decreased as the articulation of the following vowel moved posterior in the oral cavity \( (91.6 \text{ ms for } /i/, 82.2 \text{ ms for } /\alpha/, \text{ and } 76.0 \text{ ms for } /u/) \). However, as revealed by a significant place by vowel interaction effect \( [F(4,128)=13.07, p<.001, \eta^2=.290] \), this overall pattern was only maintained for velar stop consonants \( (/k/) \). The nature of these main effects (place and vowel), as well as the interaction effect (place by vowel), is illustrated in Figure 4.1. No other main effects (including gender and age group) or additional interaction effects were found to be significant.
4.1.2 Vowel Duration

There were significant differences in the duration of the vowel following the stop consonant (CV context) as a function of both place of articulation \([F(2,64)=86.17, p<.001, \eta^2=.729]\) and vowel context \([F(2,64)=20.86, p<.001, \eta^2=.395]\). As can be clearly seen in Figure 4.2, the main effect of place is due primarily to the increased duration of the vowel subsequent to a velar stop (/k/). The mean vowel durations following the /p, t, and k/ consonants were 142, 139, and 203 ms, respectively.

The main effect of vowel context was derived mainly from a significantly decreased duration for the high-back vowel (/u/), as illustrated in Figure 4.3. The durations for the /i, a, and u/ vowel contexts were 167, 172, and 144 ms, respectively.

These conclusions were statistically confirmed with pairwise comparisons (p<.01).
Figure 4.2  Mean vowel duration by place of articulation in CV stops.

Figure 4.3  Mean vowel duration by vowel context in CV stops.
In addition, the ANOVA revealed a significant place by vowel interaction \( [F(4,128)=56.53, p<.001, \eta^2=.639] \). As the articulation of the initial consonant moves posterior in the oral cavity (bilabial to velar), the duration of the two vowels (/i and /u/) increased dramatically. The nature of this interaction is further described in figure 4.4. All other main effects and interactions were not found to be significant.

Figure 4.4  Mean vowel duration by place of articulation and vowel context in CV stops.

4.1.3 Normalized Amplitude

Using a four-way ANOVA to examine the dependent measure of normalized amplitude (rms amplitude in dB of the entire stop burst relative to the strongest component in the following vowel), a main effect of place \( [F(2,64)=68.99, p<.001, \)
\(\eta^2 = .683\) was obtained. As illustrated in Figure 4.5, Bonferroni adjusted post hoc tests indicated that all three places of articulation were significantly (\(p < .01\)) different from each other in terms of normalized amplitude (-10.1 dB for /p/, -4.3 dB for /t/, and -6.1 dB for /k/). Collapsed across speaker and vowel context, the alveolar stops were found to have the highest normalized amplitude.

<table>
<thead>
<tr>
<th>Place of Articulation</th>
<th>Normalized Amplitude (in dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>-12</td>
</tr>
<tr>
<td>/t/</td>
<td>-10</td>
</tr>
<tr>
<td>/k/</td>
<td>-8</td>
</tr>
</tbody>
</table>

Figure 4.5 Normalized amplitude as a function of place of articulation in CV stops.

A main effect of vowel context was also found to be significant \(F(2,64) = 47.86, p < .001, \eta^2 = .599\), with differences in the normalized amplitude of the stop depending on the articulation of the following vowel. As shown in Figure 4.6, all three vowel contexts (/i, a, and u/) produced significantly (\(p < .001\)) different normalized amplitudes, with mean values of -5.0, -8.9, and -6.7 dB, respectively. Overall, the stop consonants had a higher
normalized amplitude when preceding an /i/ vowel. Although a place by vowel interaction was also found to be significant \[F(4,128)=5.42, p<.01\], the effect size was relatively small (\(\eta^2=.125\)). No other significant effects were revealed for the dependent variable of normalized amplitude.

![Normalized Amplitude as a Function of Vowel Context in CV Stops](image)

**Figure 4.6** Normalized amplitude as a function of vowel context in CV stops.

### 4.2 Spectral Measures

To assess the importance of the analysis window size, a one-way ANOVA was conducted on the spectral data. Utilizing the derived spectral measures as dependent variables (spectral peak, spectral slope, first four spectral moments), a main effect of window size was not found to be significant for any of the spectral measures. Considering this lack of significant statistical difference between the two series of
window sizes (10 and 20 ms), only data from the 20 ms series of analysis windows will be reported in the subsequent discussion of the stop consonants in the CV context. Furthermore, the initial 20 ms window of analysis was chosen to better compare any results to the findings of previous researchers who also derived spectral measures from the initial 20 ms of the stop burst (i.e., Forrest et al., 1988; 1990; 1994; Nittrouer, 1995).

A one-way ANOVA revealed significant differences in all spectral measures as a function of window location, except the fourth spectral moment of kurtosis. In addition, repeated measures, four-way ANOVAs and subsequent Bonferroni post hoc tests were conducted at each window location for all spectral measures. These ANOVA results, as well as other general patterns in the data that extend across window locations, are reported throughout this section (stops in a CV context) according to each type of measure. However, for purposes of consistency and the ability to compare these results to those in previous research (i.e., Forrest et al., 1988; 1990; 1994; Nittrouer, 1995), specific findings from only one window size and location will be discussed. Thus, only the analyses of spectral measurements derived from the first 20 ms analysis window (centered at +10 ms from the release of the stop burst) were reported in detail.

4.2.1 Spectral Peak

As previously mentioned, the results of a one-way ANOVA indicated that the overall measure of spectral peak differed significantly as a function of window location \([F(2,1079)=6.50, p<.01]\). As confirmed by Post hoc analysis \((p<.01)\), this effect was due primarily to the decreased overall frequencies derived from the first window of analysis. The overall spectral peaks for the first, second and third windows were 4690, 5108, and 5063 Hz, respectively.
Comparisons conducted at each analysis window revealed that the mean spectral peak locations of all three places of articulation were significantly distinguished (p<.01) at all but the second window location, as shown in Figure 4.7. In all window locations there were strong differences between the spectral peak loci of alveolar stops when compared to bilabial and velar stops. As shown in Figure 4.8, the patterns of age differences remained primarily intact across different windows of analysis.

![Figure 4.7 Mean spectral peak location as a function of place of articulation and window location for CV stops.](image)

As expected, when analyzing the spectral data from the first 20 ms analysis window, an ANOVA revealed a main effect for place of articulation [$F(2,64)=100.38$, $p<.001$, $\eta^2=.758$]. This ANOVA was of mixed design, with gender and speaker age as
the between subject factors and place of articulation and vowel context as within subject factors. Subsequent pairwise comparisons indicated that the spectral peak location of the three places of stop articulation (/p, t, k/) were significantly different (p<.001) from each other, with means of 4264, 6689, and 3116 Hz, respectively.

![Figure 4.8](image)

Figure 4.8  Mean spectral peak location as a function of age group and window location for CV stops.

A main effect of vowel context \(F(2,64)=17.19, \quad p<.001, \quad \eta^2=.349\) revealed that the spectral peak location was statistically different dependent upon the articulation of the following vowel (5255 Hz for /i/, 4294 Hz for /a/, and 4520 Hz for /u/). In particular, the peak location of the stop burst was significantly higher when preceding an /i/ vowel (p.<.001). However, there was also a place by vowel interaction \(F(4,128)=25.58,\)
p<.001, η²=.444], which indicated that this pattern was maintained across alveolar and velar places of stop articulation (/t/ and /k/), but was reversed for bilabial stops (/p/). These main effects (place and vowel), as well as the interaction effect (place by vowel), are shown in Figure 4.9.

![Figure 4.9 Spectral peak location by place of articulation and vowel context in CV stops.](image)

As expected, main effects of gender [F(1,32)=9.27, p<.01, η²=.225] and age group [F(3,32)=8.44, p<.001, η²=.442] were also found. As can be seen in Figure 4.10, across place of articulation the mean spectral peak location for female speakers (5045 Hz) was higher than males (4335 Hz). This pattern of difference is most prominent with alveolar stops (/t/).
In addition, the child speakers (5028 Hz) exhibited higher average peak locations than the adults (3677 Hz), as shown in Figure 4.11. Comparisons between the age groups indicated that the significant differences were primarily due to differences between the adult group and the three child groups (p<.01), with no differences between the child groups found to be significant. This result was not surprising considering the relatively larger size of the adult oral cavity and the lack of significant anatomic differences between the 3, 4, and 5 yr. old speakers. No other main or interaction effects were found when looking at the dependent measure of spectral peak location.
4.2.2 Spectral Slope

The spectral slope was calculated from a linear regression line fit to the relative amplitudes of the FFT power spectra of each analysis window. This measure has not been reported in previous spectral moment research (i.e., Forrest et al., 1988; Jongman et al., 2000; Nittrouer, 1995), however, in pilot research conducted by Fox and Nissen (2002) the spectral slope was found to be very effective in discriminating between voiceless English fricatives.

A repeated measures one-way ANOVA revealed that the mean spectral slope averaged across speakers, place, and vowel context differed significantly as a function of window location \( F(2,1079)=75.23, p<.001 \). Differences between all three window
locations were confirmed by subsequent Post hoc analyses (p<.01). As shown in Figure 4.12, comparisons conducted at each analysis window indicated that all three places of stop articulation were distinguished at each location (p<.01). In all window locations there were strong differences between the spectral slope of alveolar stops when compared to bilabial and velar stops. Age differences across window locations are displayed in Figure 4.13. As can be seen, the significant differences (p<.001) between the child speakers and the adults is maintained throughout the various window locations. These age group differences will be reported and discussed later in this section.

![Figure 4.12 Mean spectral slope as a function of place of articulation and window location for CV stops.](image-url)
A four-way ANOVA, with gender and speaker age as between subject factors and place of articulation and vowel context as within subject factors, was conducted on data from the first analysis window. A main effect of place of articulation \([F(2,64)=116.23, p<.001, \eta^2=.784]\) and subsequent pairwise comparisons \((p<.001)\) demonstrated that the mean spectral slope of all three stops were significantly different from each other (-0.165 for /p/, 0.790 for /t/, and -0.714 for /k/).

In addition, a significant effect of vowel context \([F(2,64)=53.56, p<.001, \eta^2=.626]\) indicated that the spectral slope of the three places of stop articulation were different from each other depending on the articulation of the following vowel (the mean slope for /i, a, u/ contexts were 0.389, -.241, and -0.237, respectively). The effect of vowel
context was mainly due to the significantly increased slope of stops preceding an /i/ vowel (p.<.001), as comparisons between /a/ and /u/ contexts were not found to be significant. The ANOVA also yielded a significant place by vowel interaction [F(4,128)=18.95, p<.001, \( \eta^2 = .372 \)]. As shown in Figure 4.14, the interaction revealed that the elevated slope found in the /i/ vowel context was negligible for bilabial stops, but highly significant for more posterior stop articulations (/t/ and /k/). This may be due to the fact that these two posterior stops involve the tongue as an active articulator, whereas the bilabial stop does not.

Interestingly, a main effect was obtained for both gender [F(1,32)=9.32, p<.01, \( \eta^2 = .226 \)] and age group [F(3,32)=9.19, p<.001, \( \eta^2 = .463 \)]. In addition, a significant gender by age group interaction was also noted [F(3,32)=3.10, p<.05, \( \eta^2 = .225 \)]. As can be seen in Figure 4.15, strong gender differences (p<.001) in spectral slope began with the 5 yr. old speakers and extended to the adults, with male speakers showing a dramatic decrease in slope at the initial 20 ms of the stop burst (window location 1). This result is interesting because anatomical differences in the oral cavity between male and female speakers are minimal in young children, which may suggest a difference in articulatory development in girls as opposed to boys starting at approximately 5 years of age.
Figure 4.14  Mean spectral slope as a function of vowel context and place for CV stops.

Figure 4.15  Mean spectral slope as a function of gender and age group for CV stops.
4.2.3 Spectral Mean

The first spectral moment (mean) was computed for the CV stop consonants at the different window locations. It was revealed through a one-way ANOVA that the dependent measure of spectral mean differed across window location \([F(2,1079)=22.67, p<.001]\). As shown in Figure 4.16, comparisons conducted at each analysis window revealed that all three places of stop articulation were distinguished at each location \((p<.001)\). Figure 4.17 illustrates age differences as a function of window location, as can be seen the significant differences \((p<.001)\) between the child speakers and the adults was maintained throughout different analysis windows.

![Figure 4.16 Spectral mean as a function of place of articulation and window location for CV stops.](image-url)
As expected, a four-way repeated measures ANOVA (place x vowel x gender x age group) conducted on the first 20 ms of the stop burst (analysis window 1) revealed a highly significant main effect for place of articulation \( [F(2,64)=165.91, p<.001, \eta^2=.838] \). Post-hoc analyses indicated significant differences \((p<.001)\) between all three places. Collapsed across speaker and vowel context the stop consonants (/p, t, k/) exhibited spectral means of 4542, 5772, and 3846 Hz, respectively. There was also a significant effect of vowel context \( [F(2,64)=33.02, p<.001, \eta^2=.508] \) and a significant place by vowel interaction effect \( [F(4,128)=23.15, p<.001, \eta^2=.420] \). Pairwise comparisons indicated that the main effect of vowel context was attributed primarily to the elevated spectral mean of the stop burst preceding an /i/ vowel \((p<0.001)\); whereas, the differences between /a/ and /u/ were not found to be statistically significant. As shown in Figure 4.17 Spectral mean as a function of age group and window location for CV stops.
4.18, the /i/ vowel context effect was pronounced in velar stops (/k/) and relatively reduced in bilabial (/p/) and alveolar stops (/t/). This pattern was also reported by Nittrouer (1995), who postulated that the effect was due to forward movement of the tongue body in anticipation of the high front articulation of the /i/ vowel. This anticipatory action would produce a shortening of the anterior resonating cavity and thereby result in an increase in the spectral mean of velar stops.

![Figure 4.18 Spectral mean for vowel contexts as a function of place of articulation for CV stops.](image)

Although not as strong as previous effects, significant differences in spectral mean were also found as a function of both gender \(F(1,32)=8.70, p<.01, \eta^2=.214\) and age group \(F(3,32)=6.13, p<.01, \eta^2=.365\), as well as a significant gender by age group
interaction \([F(3,32)=3.27 \ p<.05, \ \eta^2=.176]\). Interestingly, significant gender differences (p<.001) in spectral mean were found in stop consonants produced by the 5 yr. old speakers and extended to the adult age group, as shown in figure 4.19. Within these two age groups, male speakers show a dramatic decrease in spectral mean when compared to female speakers of the same age group (p<.001). Such differences in adult speakers can be anatomically explained by variation in vocal tract size across gender; on average the male vocal tract is about 20 cm longer than that of females (Fant, 1966). However, significant gender differences in vocal tract size are not usually considered to be present in younger children (i.e., 5 years of age). Thus, differences in spectral mean between 5 yr. old male and female speakers may be due to gender related patterns of articulatory development.

Figure 4.19  Spectral mean as a function of gender and age group for CV stops.
4.2.4 Spectral Variance

In an early study by Forrest et al. (1988), the second spectral moment of variance was found to contribute little to the discrimination of voiceless obstruents and was therefore not reported. Much of the subsequent research utilizing spectral moment analysis to examine stop consonants have also opted not to analyze the second moment. However, a more recent study examining voiceless fricatives (Jongman et al., 2000) found significant differences in the second moment across the place of articulation.

Individual four-way ANOVA’s (gender x age x place x vowel) were conducted on each of the three windows of analysis. As shown in Figure 4.20, these analyses revealed that at least two places of articulation were differentiated by spectral variance in all window locations, with all three places in the second and third analysis windows.

![Figure 4.20](image)

Figure 4.20  Spectral variance as a function of place of articulation and window location for CV stops.
For the first window location, a significant main effect of place \(F(2,64)=31.86, p<.001, \eta^2=.499\) was obtained for spectral variance. As shown in Figure 4.21, the spectral variance of the bilabial stops was significantly higher \(p<.001\) than the other two places of articulation (the variances for /p, t, and k/ were 6.7, 4.9, and 5.1 MHz, respectively).

A main effect for vowel context was also obtained \(F(2,64)=18.17, p<.001, \eta^2=.362\). The spectral variance of the first 20 ms of the stop burst was significantly lower \(p<.001\) when followed by an /i/ vowel, as is illustrated in Figure 4.22. However, this pattern was reversed for alveolar stops (/t/), which explains a small (an effect size of \(\eta^2=.124\)), yet significant place by vowel interaction effect \(F(4,128)=4.51, p<.01\).

![Figure 4.21](image-url)  Spectral variance as a function of place of articulation for CV stops.
4.2.5 Spectral Skewness

The CV stop consonant data were analyzed across window location through a series of four-way ANOVAs (gender x age x place x vowel). Subsequent post hoc analysis (p<.001) indicated that spectral skewness distinguished all three places of stop articulation at all but the second window of analysis, as can be seen in Figure 4.23. Furthermore, as shown in Figure 4.24, the patterns of age differences with regard to the spectral skewness remained primarily intact across the three different windows of analysis.

Figure 4.22  Spectral variance as a function of vowel context for CV stops.
Figure 4.23  Spectral skewness as a function of place of articulation and window location for CV stops.

Figure 4.24  Spectral skewness as a function of age and window location for CV stops.
As has been previously reported, a main effect of place \(F(2,64)=87.47, p<.001\) with a large effect size \(\eta^2=.732\) was obtained for spectral skewness at the first window of analysis. Post hoc analyses \(p<.001\) indicated that the mean spectral skewness of all three stops was significantly different from each other (-0.017 for /p/, -0.640 for /t/, and 0.997 for /k/). In addition, an effect of vowel context \(F(2,64)=15.42, p<.001, \eta^2=.325\), as well as a significant place by vowel interaction \(F(4,128)=14.97, p<.001, \eta^2=.319\) were obtained from the ANOVA. As can be seen in Figure 4.25, the effect of vowel context and associated interaction were largely due to the fact that the spectral skewness of the alveolar stop burst decreased significantly \(p<.001\) when preceding an /i/ vowel.

![Figure 4.25](image)

Figure 4.25  Spectral skewness as a function of place of articulation and vowel context for CV stops.
The ANOVA also revealed main effects for both gender \([F(1,32)=8.96, p<.01, \eta^2=.219]\) and age group \([F(3,32)=7.23, p<.001, \eta^2=.404]\), as well as a significant gender by age group interaction \([F(3,32)=3.59, p<.03, \eta^2=.252]\). Similar to the pattern of results obtained for spectral mean, when looking at skewness there were distinct gender differences that began with the 5 yr. old age group and was extended to the adults. Post hoc tests demonstrated that within these two age groups the gender differences were highly significant \((p<.001)\). Once again, since anatomical differences are minimal at this age, a possible explanation may be developmental in nature. These effects can be clearly seen in Figure 4.26.

In relation to spectral skewness, no other effects were found to be significant.

![Spectral Skewness](image)

Figure 4.26 Spectral skewness as a function of gender and age group for CV stops.
4.2.6 Spectral Kurtosis

Unlike the previously discussed spectral measures, the fourth spectral moment (kurtosis) did not vary significantly as a function of window location. A four-way mixed design ANOVA (gender x age group x place x vowel) conducted on the first window of analysis revealed only one significant effect for the dependent measure of kurtosis, that being place of articulation [F(2,64)=20.24, p<.001, η²=.388]. As shown in Figure 4.27, subsequent comparisons found that all three places of articulation were significantly different, with the spectral kurtosis increasing as the stop articulation moved posterior in the oral cavity.

![Figure 4.27 Spectral kurtosis as a function of place of articulation for CV stops.](image-url)
4.3 Discriminant Analysis for CV Stop Data

When classifying stops in the CV context, the acoustic parameters entered into the initial stepwise linear discriminant analysis were as follows: VOT, normalized stop amplitude, vowel duration, spectral peak, spectral slope, spectral mean, spectral variance, spectral skewness, and spectral kurtosis. The spectral moments measures were derived from only the first 20 ms window of analysis. The number and combination of acoustic variables included in the classification model changed as a function of the data set on which the model was trained. When utilizing a cross-validation procedure (jack-knife), each case in the analysis was classified based on a model developed from all other cases. A summary of the variables included in the various classification models can be found in Table 4.1.

As shown in Table 4.2, when classification results are based on a cross-validation (jack-knife) procedure, 75.6% of the entire data set (stops in CV context) was classified correctly in terms of place of articulation. The adults had a combined classification rating of 80.0%, while the 3, 4, and 5 year old speakers exhibited a rating of 72.2%, 71.1%, and 75.6 %, respectively. Overall, differences between male (72.2%) and female (75.0%) speakers were minimal, with the largest average differences found in the child speakers (71.1% for male and 79.2 % for female).

Classification results were also obtained by developing a discriminant function based on a specific group of speakers. Trained only on data from the adult speakers, a model of classification was composed of 4 of the original 9 variables input into the analysis. In order of their relative contribution to the classification, the variables were spectral skewness, vowel duration, spectral mean, and spectral slope. As shown in Table
4.1, the classification model based on these measures was able to correctly classify speech from the child speakers with an overall accuracy of 58.9%. The rate of success improved as the age of the speaker increased, with mean classification rates for the 3, 4, and 5 year old speakers being 53.3%, 58.9%, and 64.4%, respectively. These results may indicate that as children mature they develop more adult-like patterns of stop production.

To test if speaker gender effects could be found in the classification results, a discriminant function was trained on the adult male data. A comprehensive listing of classification scores for the discriminant model trained on only the adult male data is shown in Table 4.3. In a stepwise procedure involving the original 9 acoustic measures, only 3 were found to significantly (F>3.84) improve the classification of the training set. In order of contribution, the VOT, spectral variance, and spectral mean were combined to form the discriminant function. When applied to the training set, the discriminant function successfully classified 84.4% of the data by place of articulation, whereas the rest of the data set was found to have an overall rating of 58.7%. This difference in rate was due primarily to a less accurate classification of tokens produced by the child age groups (the mean rates for the 3, 4, and 5 year olds were 60%, 52.2%, and 55.6%, respectively). A breakdown by place of stop articulation revealed that for the child data the discriminant function was highly accurate in classifying alveolar stops (91.4%), but less successful with bilabial (33.3%) and velar stop consonants (51.4%). Excluding the adult male data, only minimal differences in gender were found (58.5% for male and 58.9% for female).
A discriminant model was also developed on data from the adult female speakers. This model was a combination of 4 acoustic measures (spectral mean, spectral skewness, vowel duration, and spectral slope; in order of contribution). As can be seen in Table 4.4, 91.1% of the stop tokens in the original training set were classified correctly, with a perfect classification of both bilabial and alveolar stops. When applied to the rest of the dataset, this discriminant model yields only a 60.3% overall classification rate (rates for /p t k/ were 81.0%, 41.0%, and 59.0%, respectively). Some of the lowest classification rates were found for alveolar stops produced by the 3 and 4 year old speakers, with means of 30.0% and 26.7%, respectively. Interestingly, stop data from the male children (63.7%) were classified by this function at a higher rate than from the female children (51.8%).
<table>
<thead>
<tr>
<th>Data Set Trained</th>
<th>Acoustic Variables in the Classification Model (by order of significance)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CV Context</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td>Spectral Peak</td>
</tr>
<tr>
<td></td>
<td>Final Vowel Duration</td>
</tr>
<tr>
<td></td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td>Spectral Peak</td>
</tr>
<tr>
<td></td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td></td>
<td>Final Vowel Duration</td>
</tr>
<tr>
<td></td>
<td>Spectral Kurtosis</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td></td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td></td>
<td>Final Vowel Duration</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult Speakers</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td></td>
<td>Final Vowel Duration</td>
</tr>
<tr>
<td></td>
<td>Spectral Mean</td>
</tr>
<tr>
<td></td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Entire Data Set</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td></td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td></td>
<td>Final Vowel Duration</td>
</tr>
<tr>
<td></td>
<td>Spectral Variance</td>
</tr>
<tr>
<td></td>
<td>VOT</td>
</tr>
<tr>
<td></td>
<td>Spectral Peak</td>
</tr>
<tr>
<td></td>
<td>Spectral Kurtosis</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1 A summary of the variables included in the various classification models for CV stops.
## Table 4.2 Results of discriminant classification using both cross-validation (jack-knife) and standard stepwise procedures for CV stops; all values represent % correct classification.

<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Training Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self Classification</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>—</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>72.2</td>
</tr>
<tr>
<td>- female only</td>
<td>66.7</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>71.1</td>
</tr>
<tr>
<td>- female only</td>
<td>80.0</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>75.6</td>
</tr>
<tr>
<td>- female only</td>
<td>82.2</td>
</tr>
<tr>
<td>Adult Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>80.0</td>
</tr>
<tr>
<td>- female only</td>
<td>84.4</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>72.2</td>
</tr>
<tr>
<td>- female only</td>
<td>75.0</td>
</tr>
<tr>
<td>Data Set Evaluated</td>
<td>Classification Model Trained on Adult Male Data</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>/p/</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>86.7</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>36.7</td>
</tr>
<tr>
<td>- female only</td>
<td>46.7</td>
</tr>
<tr>
<td></td>
<td>26.7</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>20.0</td>
</tr>
<tr>
<td>- female only</td>
<td>13.3</td>
</tr>
<tr>
<td></td>
<td>26.7</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>33.3</td>
</tr>
<tr>
<td>- female only</td>
<td>60.0</td>
</tr>
<tr>
<td></td>
<td>6.7</td>
</tr>
<tr>
<td>Adult Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>—</td>
</tr>
<tr>
<td>- female only</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>53.3</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>33.3</td>
</tr>
<tr>
<td>- female only</td>
<td>40.0</td>
</tr>
<tr>
<td></td>
<td>28.3</td>
</tr>
</tbody>
</table>

Table 4.3  Results of stepwise linear discriminant analysis when the classification model is trained on adult male data only for CV stops; all values represent % correct classification.
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Trained on Adult Female Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/p/</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>100.0</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>73.3</td>
</tr>
<tr>
<td>- female only</td>
<td>66.7</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>86.7</td>
</tr>
<tr>
<td>- female only</td>
<td>86.7</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>80.0</td>
</tr>
<tr>
<td>- female only</td>
<td>80.0</td>
</tr>
<tr>
<td>Adult Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>—</td>
</tr>
<tr>
<td>- female only</td>
<td>—</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>81.0</td>
</tr>
<tr>
<td>- female only</td>
<td>83.3</td>
</tr>
</tbody>
</table>

Table 4.4  Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for CV stops; all values represent % correct classification.
4.4 Summary of Significant Results

For stops produced in the CV context, ANOVAs indicated that the dependent measures of VOT, vowel duration, normalized amplitude, spectral peak, spectral slope, and all four spectral moments varied significantly as a function of place of articulation. Subsequent post hoc analysis revealed that similar to previous research (Forrest et al., 1988; Nittrouer, 1995), alveolar and velar stops were significantly different in terms of spectral mean and skewness. In addition, this study found that these two spectral parameters significantly separated bilabial stops. The measures of VOT, normalized amplitude, spectral peak, and spectral slope also varied significantly across all three places of stop articulation.

Main effects for vowel context were noted for vowel duration, normalized amplitude, spectral peak, spectral slope, and for all spectral moments, with the exception of kurtosis. Similar to research conducted by Nittrouer (1995), these vowel context effects usually varied as a function of the place of stop articulation, characterized by changes in the acoustic structure of alveolar and velar stops when followed by a high front vowel (/i/).

Significant effects of age and gender were found for the spectral measures of peak location, slope, mean, and skewness. Of particular interest, significant age by gender interaction effects were found for spectral slope, mean, and skewness. In general, strong gender differences for these measures began with the 5 year old speakers and extended to the adults. These results are interesting considering that anatomical vocal tract differences between male and female speakers are usually considered to be minimal in young children.
In terms of place of articulation, a discriminant analysis based on a cross-validation procedure was able to correctly classify 75.6% of the stop consonant data. The adult data were found to have classification rating of 80.0%, while the speech productions of the 3, 4, and 5 year old speakers exhibited ratings of 72.2%, 71.1%, and 75.6%, respectively. Across all speaker groups, gender differences were minimal (72.2% for male and 75.0% for female). When based on a discriminant function trained only on the adult data, speech from the child speakers was correctly classified with an overall accuracy of 58.9%, whereas the training set was classified at a rate of 84.4%.
CHAPTER 5

ACOUSTIC ANALYSIS OF VCV STOP CONSONANTS

The results found in the following section were derived from stops produced in a VCV context. The presentation of these findings will be similar to the format found in the previous chapter. For spectral measures, general comparisons of analysis window size and location will be presented, followed thereafter by more detailed findings from only the first 20 ms analysis window. Although specific statistics from additional analysis windows will not be reported, general patterns in the data that extend across these windows will also be described. All data were collapsed across repetition for a given speaker prior to statistical analysis. Unless otherwise noted, any main or interaction effects found to be significant at a .05 alpha level were discussed.

5.1 Duration and Amplitude Measures

For the stop consonant tokens in the VCV context, the phonemic category of the initial vowel (/o/) remained constant across all tokens, yet the following stop consonant (/p, t, k/) and final vowel (/i, a, u/) were varied to form 9 unique syllable combinations. A four-way repeated measures ANOVA, with gender and age as the between subject factors and place of articulation and vowel context as within subject factors was used to analyze a variety of duration and amplitude measures (overall token duration, initial
vowel duration, stop closure duration, VOT, final vowel duration, and normalized amplitude). All post hoc comparisons were adjusted for multiple comparisons (Bonferroni).

5.1.1 Overall Token Duration

No significant main effects or interactions were found for overall token duration. However, there was a slight trend toward longer overall tokens for the younger speakers, as shown in Figure 5.1.

![Overall Token Duration](image)

Figure 5.1 Overall token duration as a function of place of articulation and age in VCV stops.

5.1.2 Initial Vowel Duration

For the dependent measure of initial vowel duration, a main effect of final vowel context was the only significant finding revealed by the ANOVA $F(2,64)=8.59$, $p<.01$, 

90
Post hoc comparisons indicated that the main effect was the result of a significant (p<.01) decrease in the initial vowel duration when the final vowel was an /a/. These differences are illustrated in Figure 5.2.

**Figure 5.2** Initial vowel duration as a function of place of articulation and age in VCV stops.

### 5.1.3 Stop Closure Duration

A main effect of place of articulation \( [F(2,64)=18.15, p<.001, \eta^2=.362] \) was found to be significant for the dependent measure of stop closure duration. Subsequent pairwise comparisons indicate that the basis for this effect is derived from the significantly increased (p<.001) closure duration of bilabial stops. Collapsed across speaker and vowel context, the place of articulation means for /p, t, and k/ were 139, 126, and 123 ms, respectively.
The main effect of vowel context was also significant \( [F(2, 64) = 8.85, p < .001, \eta^2 = .217] \), characterized by a decrease in the stop closure duration when the following vowel was /a/. Post hoc tests indicated that only comparisons involving the /a/ vowel context were found to be significant \( (p < .01) \). The means for the three vowel contexts /i, a, u/ were 132, 124, and 132 ms, respectively. No other significant effects or interactions were noted for stop closure duration.

5.1.4 Voice Onset Time

The voice onset time (defined as the initiation of periodic voicing relative to the stop closure release), was calculated for the stop consonant tokens. There were significant differences in the VOT as a function of both place of articulation \( [F(2, 64) = 30.75, p < .001, \eta^2 = .490] \) and vowel context \( [F(2, 64) = 12.49, p < .001, \eta^2 = .281] \), as well as a significant place by vowel interaction effect \( [F(4, 128) = 8.41, p < .001, \eta^2 = .208] \). As can be clearly seen in figure 5.3 the main effect of place is due primarily to the decreased VOT of bilabial stops (86 ms for /p/; 101 ms for /t/; and 103 ms for /k/). The main effect of vowel context was derived mainly from an overall decrease in VOT when the stop precedes an /a/ vowel, as illustrated in Figure 5.4. The VOT means for the /i, a, and u/ vowel contexts were 103, 90, and 97 ms, respectively. These conclusions were statistically confirmed with post hoc comparisons \( (p < .01) \).
Figure 5.3 VOT by place of articulation in VCV stops.

Figure 5.4 VOT by vowel context in VCV stops.
5.1.5 Final Vowel Duration

As expected final vowel duration, differed significantly as a function of vowel type \( F(2,64) = 22.04, p < .001, \eta^2 = .408 \). Subsequent pairwise comparisons indicated that all three vowel types (/i, a, u/) were significantly (p < .01) different in terms of duration, with means of 280, 336, and 303 ms, respectively. Figure 5.5 illustrates these differences. No additional place or speaker effects were found to be significant for this particular measure of duration.

![Figure 5.5](image-url)

Figure 5.5 Final vowel duration as a function of vowel context in VCV stops.

5.1.6 Normalized Amplitude

Using a four-way ANOVA to examine the dependent measure of normalized amplitude (rms amplitude in dB of the entire stop burst relative to the strongest
component in the following vowel), a main effect of place \( F(2,64)=53.71, p<.001, \eta^2=.627 \) was obtained. As illustrated in Figure 5.6, Bonferroni adjusted post hoc tests indicated that all three places of articulation were significantly \((p<.001)\) different from each other in terms of normalized amplitude (-8.4 dB for /p/, -4.05 dB for /t/, and -6.1 dB for /k/). Collapsed across speaker and vowel context, the alveolar stops were found to have the highest normalized amplitude.

<table>
<thead>
<tr>
<th>Place of Articulation</th>
<th>Normalized Amplitude (in dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>-8</td>
</tr>
<tr>
<td>/t/</td>
<td>-4.05</td>
</tr>
<tr>
<td>/k/</td>
<td>-6.1</td>
</tr>
</tbody>
</table>

Figure 5.6 Normalized amplitude as a function of place of articulation in VCV stops.

A main effect of vowel context was also found to be significant \( F(2,64)=54.03, p<.001, \eta^2=.628 \), with differences in the normalized amplitude of the stop depending on the articulation of the following vowel. As shown in Figure 5.7, all three vowel contexts
(/i, a, and u/) produced significantly (p<.001) different normalized amplitudes, with mean values of -4.3, -8.2, and -6.0 dB, respectively. Overall, the stop consonants had higher normalized amplitude when preceding an /i/ vowel.

![Figure 5.7 Normalized amplitude as a function of vowel context in VCV stops.](image)

A place by vowel interaction was also found to be significant \([F(4,128)=7.65, p<.001, \eta^2=.193]\). As can be seen in Figure 5.8, the elevated normalized amplitude found in the /i/ vowel context is highly significant (p<.001) for bilabial and velar stop articulations (/p, k/), but alveolar stops (/t/) were found to have higher amplitude values in the /u/ vowel context. No other significant effects were revealed for the dependent variable of normalized amplitude.
5.2 Spectral Measures

For stop consonants in the VCV context, spectral moment data were derived from two series of analysis windows (10 and 20 ms). Windows within each series were located at subsequent and overlapping locations. A one-way ANOVA was conducted on the spectral data to evaluate differences in window size and location. Utilizing the derived spectral measures as dependent variables (spectral peak, slope, mean, variance, skewness, and kurtosis), no effects of window size were found to be significant. Considering the lack of significant statistical difference between the two window sizes (10 and 20 ms) and to better compare any results to findings presented in this study and throughout the
spectral moment literature (i.e., Forrest et al., 1988; 1990; 1994; Nittrouer, 1995), only data from the 20 ms series of analysis windows will be reported in the subsequent discussion.

Differences across window location were found to be significant in four of the six derived spectral measures (spectral slope, mean, variance, and skewness). These ANOVA results, as well as other general patterns in the data that extend across window locations, are reported throughout this section (stops in VCV context) according to each type of spectral measure. However, for purposes of consistency and the ability to compare these results to those previously reported (i.e., Forrest et al., 1988; 1990; 1994; Nittrouer, 1995), specific results from only the first 20 ms analysis window (centered at +10 ms from the release of the stop burst) will be reported in detail in this discussion.

5.2.1 Spectral Peak

Individual four-way ANOVA’s (gender x age group x place x vowel) conducted on each of the three windows of analysis revealed a significant effect of place of articulation for each window location. Post hoc comparisons of the mean spectral peak locations indicated that all three places of articulation were significantly distinguished (p<.001) at all but the second window location, as can be seen in Figure 5.9. Not surprisingly, the patterns of age differences remained primarily intact across different windows of analysis, as shown in Figure 5.10.
Figure 5.9  Spectral peak location as a function of place of articulation and window location for VCV stops.

Figure 5.10  Spectral peak location as a function of age group and window location for VCV stops.
As expected, when analyzing the spectral data from the first 20 ms analysis window, a four-way ANOVA (gender x age group x place x vowel context) revealed a main effect for place of articulation \(F(2,64)=103.49, p<.001, \eta^2=.764\). Subsequent pairwise comparisons indicated that the spectral peak location of the three places of stop articulation (/p, t, and k/) were significantly different (\(p<.001\)) from each other, with a mean of 4408, 7007, and 3179 Hz, respectively.

A main effect of vowel context \(F(2,64)=22.20, p<.001, \eta^2=.410\] revealed that the peak location of the stop burst was statistically different dependent upon the articulation of the following vowel (5509 Hz for /i/, 4571 Hz for /a/, and 4516 Hz for /u/). In general, the spectral peak location was significantly higher when preceding an /i/ vowel \(p<.001\). However, there was also a place by vowel interaction \(F(4,128)=19.88, p<.001, \eta^2=.383\], which indicated that this pattern was maintained across alveolar and velar places of stop articulation (/t, k/), but reversed for bilabial stops (/p/). These main and interaction effects are illustrated in Figure 5.11.

A significant effect of age group was also found \(F(3,32)=11.02, p<.001, \eta^2=.508\], which was not surprising considering the vocal tract size differences between children and adult speakers. As shown in Figure 5.12, post hoc comparisons indicated that the main effect of age group was the result of significant differences \(p<.01\) between the average spectral peak location of the adult group (3674 Hz) as compared to the three child groups (5262 Hz). No differences between the child groups were found to be significant.
Figure 5.11  Spectral peak location by place of articulation and vowel context in VCV stops.

Figure 5.12  Spectral peak location for different age groups as a function of place of articulation for VCV stops.
Although a main effect for gender [F(1,32)=11.02, p<.04] was found, the effect size was relatively small (η²=.132). Across place of articulation the mean spectral peak location for female speakers (5142 Hz) was higher than males (4588 Hz). As can be seen in Figure 5.13, this pattern of difference is most prominent with alveolar stops (/t/). In terms of spectral peak location, no other main or interaction effects were found.

![Figure 5.13 Spectral peak location for male and female speakers as a function of place of articulation for VCV stops.](image)

5.2.2 Spectral Slope

A one-way ANOVA revealed that the mean spectral slope averaged across speakers, place, and vowel context differed significantly as a function of window location
As shown in Figure 5.14, separate four-way ANOVA’s conducted on each of the three windows of analysis revealed that all three places of stop articulation were distinguished at each location (p<.001).

![Figure 5.14](image)

**Figure 5.14** Mean spectral slope as a function of place of articulation and window location for VCV stops.

Age differences across window locations are displayed in Figure 5.15. As can be seen, the significant differences (p<.001) between the child speakers and the adults is maintained throughout the various window locations. These age group differences will be reported and discussed later in this section.
Figure 5.15 Mean spectral slope as a function of age group and window location for VCV stops.

A four-way ANOVA, with gender and speaker age as the between subject factors and place of articulation and vowel context as within subject factors, was conducted on data from the first analysis window. A main effect of place of articulation \( F(2,64) = 90.79, p < .001, \eta^2 = .739 \) and subsequent pairwise comparisons \( p < .001 \) demonstrated that the mean spectral slope of all three stops were significantly different from each other \( -0.148 \) for /p/, \( 0.682 \) for /t/, and \( -0.661 \) for /k/.

In addition, a significant effect of vowel context \( F(2,64) = 60.26, p < .001, \eta^2 = .653 \) indicated that the spectral slope of the three places of stop articulation were different from each other depending on the articulation of the following vowel (the mean slope for /i, a, and u/ context was 0.420, -.237, and -.310, respectively). The effect of vowel
context is mainly due to the significantly increased slope of stops preceding an /i/ vowel (p<.001), as comparisons between /a/ and /u/ contexts were not found to be significant. The ANOVA also yielded a significant place by vowel interaction [F(4,128)=15.79, p<.001, η²=.330]. The elevated slope found in the /i/ vowel context was negligible for bilabial stops, but highly significant (p<.001) for more posterior stop articulations (/t, k/).

In the /i/ vowel context, bilabial stops were found to have the lowest mean spectral slope. However, when followed by an /a/ or /u/ vowel, velar stops exhibited the lowest slope. The nature of these effects can be seen in Figure 5.16 below.

![Figure 5.16 Spectral slope as a function of place of articulation and vowel context for VCV stops.](image)

A main effect was also obtained for age group [F(3,32)=15.58, p<.001, η²=.594] with the adult speakers having significantly (p<.001) higher slope values than the three
child age groups. No differences between the child groups were found to be significant.

The differences in spectral slope across the age groups is illustrated in Figure 5.17. Although there was a significant gender $[F(1,32)=4.62, p<.04, \eta^2=.126]$, the effect size was relatively small. Female speakers exhibited a slightly positive spectral slope (0.085), while male speakers were found to have a slightly negative slope (-0.170).

![Figure 5.17](image_url)

Figure 5.17 Spectral slope as a function of age group for VCV stops.

5.2.3 Spectral Mean

The first spectral moment (mean) was computed for the VCV stop consonants at three different window locations. It was revealed through a one-way ANOVA that the dependent measure of spectral mean differs across window location $[F(2,1079)=8.49, p<.001]$. As shown in Figure 5.18, a series of four-way ANOVAs revealed that each
window distinguishes all three places of stop articulation at a significant level (p<.01). Figure 5.19 illustrates age differences as a function of window location, as can be the significant differences (p<.001) between the child and adult speakers was maintained throughout different analysis windows.

A four-way repeated measures ANOVA (place x vowel x gender x age group) conducted on the first 20 ms of the stop burst (analysis window 1) revealed a highly significant main effect for place of articulation \[ F(2,64)=112.88, p<.001, \eta^2=.779 \]. Post-hoc analyses indicated significant differences (p<.001) between all three places of stop articulation. Collapsed across speaker and vowel context the stop consonants (/p, t, k/) exhibited spectral means of 4615, 5770, and 3893 Hz, respectively.

![Figure 5.18 Spectral mean as a function of place of articulation and window location for VCV stops.](image)
Figure 5.19  Spectral mean as a function of age and window location for VCV stops.

There was also a significant effect of vowel context \( [F(2,64)=30.48, p<.001, \eta^2=.488] \) and a significant place by vowel interaction effect \( [F(4,128)=23.11, p<.001, \eta^2=.419] \). Pairwise comparisons indicate that the main effect of vowel context is attributed primarily to the elevated spectral mean of the stop burst preceding an /i/ vowel (p<0.01); whereas, the differences between /a/ and /u/ were not found to be statistically significant. As shown in Figure 5.20, the spectral mean is increased in alveolar and velar stops when followed by an /i/ vowel, but decreases in bilabial stops. As previously mentioned, Nittrouer (1995) reported a similar vowel context effect for velar stops.
Significant differences in spectral mean were also found for gender
$[F(1,32)=4.43, p<.05, \eta^2=.122]$, however, unlike the differences reported with stops in the CV context, the effect size was relatively small. Overall, male speakers had lower mean values than did female speakers. As shown in figure 5.21, this gender difference was found in age groups other than the adults. In the 5 yr. old age group, the male speakers were found to have a significantly lower spectral mean than the female speakers of the same age. As previously discussed it is unlikely that this difference was due to any anatomical differences in vocal tract size. As expected, a main effect of age group $[F(3,32)=8.26, p<.001, \eta^2=.437]$ was also revealed by the ANOVA, with significant
differences (p<.002) between the adult group and the three younger groups of speakers.

No additional significant main or interaction effects were noted for the dependent measure of spectral mean.

Figure 5.21  Spectral mean as a function of gender and age group for VCV stops.

5.2.4 Spectral Variance

A one-way ANOVA indicated that the spectral variance varied as a function of window location \([F(2,1077)=7.73, p<.001]\). Post hoc tests revealed a significant (p<.001) difference between the first and third analysis windows, with mean spectral variances of 5.7 and 5.0 MHz, respectively.

In addition, individual four-way ANOVA’s (gender x age group x place x vowel) were conducted on each of three windows of analysis. As shown in Figure 5.22, these
analyses revealed that at least two places of articulation were differentiated by spectral variance at each window location, with all three places being distinguished in the third analysis window.

![Graph showing spectral variance as a function of place of articulation and window location for VCV stops.]

Figure 5.22 Spectral variance as a function of place of articulation and window location for VCV stops.

As previously mentioned, analyses of data derived from the first window location indicated a significant main effect of place \([F(2,64)=31.86, p<.001, \eta^2=.499]\) for spectral variance. As can be seen from Figure 5.23, the mean spectral variance of the bilabial stops was significantly higher \((p<.001)\) than the other two places of articulation (the variance measures for /p/, /t/, and /k/ were 6.9, 4.9, and 5.2 MHz, respectively).

A main effect for vowel context was also obtained \([F(2,64)=20.61, p<.001, \eta^2=.392]\). The spectral variance of the stops were significantly higher \((p<.001)\) when
followed by an /a/ vowel, as is illustrated in Figure 5.23. This figure also reflects a significant place by vowel interaction effect \( F(4,128)=6.41, p<.001, \eta^2=.167 \), bilabial and velar stops exhibited the lowest spectral variance in the /i/ vowel context, but alveolar stops were found to have lower variance when followed by an /u/ vowel.

Figure 5.23  Spectral variance as a function of place of articulation and vowel context for VCV stops.

5.2.5 Spectral Skewness

The results of a one-way ANOVA indicated that the overall measure of spectral skewness differed as a function of window location \( F(2,1077)=5.05, p<.01 \), with the first window having a significantly (\( p<.05 \)) more positive skewness (the mean skewness values for the three windows were .01, -.22, and -.25, respectively). A series of ANOVAs (gender x age x place x vowel) and subsequent post hoc analysis (\( p<.001 \))
indicated that spectral skewness distinguished all three places of stop articulation at all but the second window of analysis. These differences can be seen in Figure 5.24. Furthermore, as shown in Figure 5.25, the patterns of age differences with regard to spectral skewness remained intact across the different windows of analysis.

As has been previously reported, analyses (ANOVA) of the first window location indicated a main effect of place \([F(2,64)=63.60, p<.001, \eta^2=.665]\). Post hoc analyses (p<.001) indicated that the mean spectral skewness of all three stops were significantly different from each other ( -0.092 for /p/, -0.767 for /t/, and 0.916 for /k/). These differences are illustrated in Figure 5.26.

![Figure 5.24](image.png)

Figure 5.24  Mean spectral skewness as a function of place of articulation and window location for VCV stops.
Figure 5.25  Spectral skewness as a function of age and window location for VCV stops.

Figure 5.26  Spectral skewness as a function of place of articulation for VCV stops.
In addition, an effect of vowel context \([F(2,64)=20.05, p<.001, \eta^2=.385]\), as well as a significant place by vowel interaction \([F(4,128)=14.62, p<.001, \eta^2=.314]\) was obtained. As can be seen in Figure 5.27, the effect of vowel context and associated interaction are largely due to the fact that when followed by an \(/i/\) vowel bilabial stops exhibited the relatively positive skewness values, yet in a similar vowel context the alveolar and velar stops exhibited more negative values of skewness.

![Figure 5.27 Spectral skewness as a function of place of articulation and vowel context for VCV stops.](image)

As expected, the ANOVA also revealed a main effect for age group \([F(3,32)=7.23, p<.001, \eta^2=.404]\). The basis for this effect were significant differences (\(p<.001\)) between the three groups of younger speakers and the adult group, with the adults having a higher mean skewness. The mean values of skewness for the child (3, 4,
5 yr. old) and adult groups were -.12, -.23, -.26, and .68, respectively. No significant differences within the younger groups of speakers were found. All other effects were either not significant or had very small effect sizes.

5.2.6 Spectral Kurtosis

The fourth spectral moment (kurtosis) did not vary significantly as a function of window location. Furthermore, a four-way mixed design ANOVA (gender x age group x place x vowel) revealed only one significant effect for the dependent measure of kurtosis, that being place of articulation \( [F(2,64)=17.31, \ p<.001, \ \eta^2=.351] \). As shown in Figure 5.28, it was found that the main effect was primarily due to the significantly (\( p<.001 \)) decreased skewness associated with bilabial stops. Although not significant in every pairwise comparison, the overall trend is for the spectral kurtosis to increase as the stop articulation moved posterior in the oral cavity.

![Figure 5.28](image)

Figure 5.28 Spectral kurtosis as a function of place of articulation for VCV stops.
5.3 Discriminant Analysis for VCV Stop Data

A discriminant analysis was conducted on the acoustic data derived from the stop consonant productions in the VCV context. The acoustic parameters entered into the initial stepwise linear discriminant analysis were as follows: VOT, normalized stop amplitude, initial and final vowel duration, spectral peak, spectral slope, spectral mean, spectral variance, spectral skewness, and spectral kurtosis. The spectral moments measures were derived from the first 20 ms of the stop burst. The combination of acoustic variables included in each classification model was determined by the data set on which the model was trained and the specific discriminant procedure employed to develop the model. Both a cross-validation (jack-knife) and a standard stepwise procedure were utilized in this analysis. A summary of the variables included in the various classification models can be found in table 4.1.

As shown in Table 5.1, when classification results are based on a cross-validation (jack-knife) procedure, 71.4% of the entire data set (stops in VCV context) were classified correctly in terms of place of articulation. The adults had a combined classification rating of 81.1%, while the 3, 4, and 5 year old speakers exhibited a rating of 67.8%, 70.0%, and 66.7 %, respectively. Overall, differences between male (67.2%) and female (68.3%) speakers were minimal, with the largest average differences found in the 4 year old (57.8 % for male and 80.0 % for female) and adult speakers (88.9 % for male and 75.6 % for female).

Classification results were also obtained by developing a discriminant function based on a specific group of speakers. Trained only on data from the adult speakers, a model of classification was composed of 4 of the original 9 variables input into the
analysis. In order of their relative contribution to the classification, the variables were spectral skewness, final vowel duration, spectral mean, and spectral slope. Interestingly, this combination and ordering of acoustic variables is the same as that obtained when analyzing the stop consonants in the CV context, as shown in Table 4.1. These similarities indicate that the stops were being produced by the adults is a similar manner across the two linguistic contexts. However, classification models based on data from the child groups did not exhibit similar patterns of model composition across contexts. The most significant variable in the majority of classification models based on child speakers was the measure of spectral peak. However, subsequent variables included in the analyses varied widely as function of the age group and the linguistic context. This result may indicate some variability in the way the children produced their stops as a function of the linguistic environment.

As can be seen in Table 5.1, the classification model based on the adult data were able to correctly classify stops produced by the child speakers with an overall accuracy of 59.3 %. Unlike the classification rates found in the CV stops, the rate of success actually decreased slightly as the age of the speaker increased. The mean classification rates for the 3, 4, and 5 year old speakers were found to be 61.1%, 60.0%, and 56.7 %, respectively.

A discriminant function was trained on the adult male data. In a stepwise procedure involving the original 9 acoustic measures, 4 were found to significantly (F>3.84) improve the classification of the training set. In order of contribution, the spectral peak, spectral variance, spectral skewness, and the final vowel duration were combined to form the discriminant function. From this function, 93.3 % of the adult male
data were classified by place of articulation, whereas the rest of the data set was found to have an overall rating of 55.6%. This difference in rate was due primarily to a less accurate classification of tokens produced by the child age groups (the mean rates for the 3, 4, and 5 year olds were 54.4%, 53.3%, and 51.1%, respectively). A breakdown by place of stop articulation revealed that for the child data the discriminant function was highly accurate in classifying alveolar stops (95.5%), but less successful with bilabial (34.4%) and velar stop consonants (28.9%). Excluding the adult male data, only minimal differences in gender were found (58.5% for male and 58.9% for female). A comprehensive listing of classification scores for the discriminant model trained on only the adult male data is shown in Table 5.2.

A discriminant model was also developed on data from the adult female speakers. This model was a combination of 3 acoustic measures (spectral skewness, spectral slope, and spectral mean; in order of contribution). As can be seen in Table 5.3, 82.2% of the stop tokens in the original training set were classified correctly. Similar to results from the CV stops, a high rate of classification was found for both the bilabial (93.3%) and alveolar stops (93.3%). When applied to the rest of the dataset, this discriminant model yielded a 56.8% overall classification rate (rates for /p t k/ were 77.1%, 30.5%, and 62.9%, respectively). Some of the lowest classification rates were found for alveolar stops produced by the child speakers. Alveolar stop data from the 3, 4, and 5 year old speakers were found to be lower than chance levels at 23.3%, 30.0%, and 23.3%, respectively. Stops produced by the male children (63.7%) were classified by this function at a higher rate than the female children (51.8%).
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Training Set</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self Classification</td>
<td>All Adults</td>
<td>Male Adults</td>
<td>Female Adults</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>—</td>
<td>81.1</td>
<td>93.3</td>
<td>82.2</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td>67.8</td>
<td>61.1</td>
<td>54.4</td>
<td>53.3</td>
</tr>
<tr>
<td>- male only</td>
<td>73.3</td>
<td>—</td>
<td>55.6</td>
<td>53.3</td>
</tr>
<tr>
<td>- female only</td>
<td>77.8</td>
<td>—</td>
<td>53.3</td>
<td>53.3</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td>70.0</td>
<td>60.0</td>
<td>53.3</td>
<td>52.2</td>
</tr>
<tr>
<td>- male only</td>
<td>57.8</td>
<td>—</td>
<td>48.9</td>
<td>55.6</td>
</tr>
<tr>
<td>- female only</td>
<td>80.0</td>
<td>—</td>
<td>57.8</td>
<td>48.9</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td>66.7</td>
<td>56.7</td>
<td>51.1</td>
<td>53.3</td>
</tr>
<tr>
<td>- male only</td>
<td>68.9</td>
<td>—</td>
<td>60.0</td>
<td>57.8</td>
</tr>
<tr>
<td>- female only</td>
<td>64.4</td>
<td>—</td>
<td>42.2</td>
<td>48.9</td>
</tr>
<tr>
<td>Adult Speakers</td>
<td>81.1</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>- male only</td>
<td>88.9</td>
<td>—</td>
<td>—</td>
<td>80.0</td>
</tr>
<tr>
<td>- female only</td>
<td>75.6</td>
<td>—</td>
<td>71.1</td>
<td>—</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td>71.4</td>
<td>59.3</td>
<td>55.6</td>
<td>56.8</td>
</tr>
<tr>
<td>- male only</td>
<td>67.2</td>
<td>61.5</td>
<td>54.8</td>
<td>61.7</td>
</tr>
<tr>
<td>- female only</td>
<td>68.3</td>
<td>57.0</td>
<td>56.1</td>
<td>50.4</td>
</tr>
</tbody>
</table>

Table 5.1 Results of discriminant classification using both cross-validation (jack-knife) and standard stepwise procedures for VCV stops; all values represent % correct classification.
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Trained on Adult Male Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/p/</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>93.3</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>40.0</td>
</tr>
<tr>
<td>- female only</td>
<td>40.0</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>40.0</td>
</tr>
<tr>
<td>- female only</td>
<td>40.0</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>23.3</td>
</tr>
<tr>
<td>- female only</td>
<td>33.3</td>
</tr>
<tr>
<td>Adult Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>—</td>
</tr>
<tr>
<td>- female only</td>
<td>66.7</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>39.0</td>
</tr>
<tr>
<td>- female only</td>
<td>37.8</td>
</tr>
</tbody>
</table>

Table 5.2 Results of stepwise linear discriminant analysis when the classification model is trained on adult male data only for VCV stops; all values represent % correct classification.
### Table 5.3  Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for VCV stops; all values represent % correct classification.

<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Trained on Adult Female Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/p/</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>93.3</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>76.7</td>
</tr>
<tr>
<td>- female only</td>
<td>66.7</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>70.0</td>
</tr>
<tr>
<td>- female only</td>
<td>73.3</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>76.7</td>
</tr>
<tr>
<td>- female only</td>
<td>66.7</td>
</tr>
<tr>
<td>Adult Speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>93.3</td>
</tr>
<tr>
<td>- female only</td>
<td></td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>77.1</td>
</tr>
<tr>
<td>- female only</td>
<td>75.0</td>
</tr>
</tbody>
</table>

Table 5.3  Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for VCV stops; all values represent % correct classification.
5.4 Summary of Significant Results

Statistical analysis (ANOVA) conducted on stops produced in the VCV context determined that the dependent measures of stop closure duration, VOT, normalized amplitude, spectral peak, spectral slope, and all four spectral moments varied significantly as a function of place of articulation. Subsequent post hoc analysis revealed that all three places of stop articulation were significantly different in terms of normalized amplitude, peak location, spectral slope, spectral mean, and spectral skewness.

Main effects for vowel context were found for all of the obtained acoustic measures except overall token duration and spectral kurtosis. As mentioned in the previous chapter, these vowel context effects usually varied as a function of the place of stop articulation, characterized by changes in the acoustic structure of alveolar and velar stops when followed by the vowel /i/. Nittrouer (1995) attributed this effect to forward movement of the tongue body in anticipation of the high front /i/ vowel.

Significant effects of age were found for the spectral measures of peak location, slope, mean, and skewness. In general, age effects were due mainly to significant differences between the acoustic structure of the adult speaker group as compared to the three child groups. Differences between child groups were rarely found to be significant. Although significant gender differences were noted for the measures of peak location and spectral mean, the effect sizes were relatively small.

In terms of place of articulation, a discriminant analysis based on a cross-validation procedure was able to correctly classify 71.4% of the stop consonant data. The adult data were found to have classification rating of 81.1%, while the productions of the
3, 4, and 5 year old speakers exhibited ratings of 67.8%, 70.0%, and 66.7%, respectively. Across all speaker groups, gender differences were minimal (67.2% for male and 68.3% for female). When based on a discriminant function trained only on the adult data, speech from the child speakers was correctly classified with an overall accuracy of 59.3%, whereas the training set was classified at a rate of 81.1%. The model trained on the adult data was a combination of the variables spectral skewness, final vowel duration, spectral mean, and spectral slope. Interestingly, this combination and ordering of acoustic variables was the same for both the CV and VCV data sets, indicating that the stops consonants were being produced by the adults in a relatively consistent and similar manner across the two linguistic contexts. In contrast, classification models based on data from child speakers varied widely across phonetic contexts in terms of model composition. This result suggests that child speakers exhibit greater context dependent variability in the acoustic structure of their stop productions than adult speakers.
CHAPTER 6

ACOUSTIC ANALYSIS OF CV FRICATIVE CONSONANTS

The results found in the following section were derived from fricatives produced in a CV context. The presentation of these findings will be similar to the format found in the previous chapter. For spectral measures, general comparisons of analysis window size and location will be presented, followed thereafter by more detailed findings from only the middle 40 ms analysis window. Although specific statistics from additional analysis windows will not be reported, general patterns in the data that extend across these windows will also be described. All data were collapsed across repetition for a given speaker prior to statistical analysis. Unless otherwise noted, any main or interaction effects found to be significant at a .05 alpha level were discussed.

6.1 Duration and Amplitude Measures

These results were derived from a four-way repeated measures ANOVA, with gender and age as the between subject factors and place of articulation and vowel context as within subject factors. This analysis examined the dependent measures of fricative duration, final vowel duration, and normalized amplitude. Any post hoc tests conducted on significant effects were adjusted for multiple comparisons (Bonferroni).
6.1.1 Fricative Duration

For the measure of fricative duration, a significant main effect of place of articulation [$F(3,96)=14.16$, $p<.001$, $\eta^2=.307$] was found. As can be seen in figure 6.1, subsequent pairwise comparisons indicated that the basis for this effect was derived from the significantly increased ($p<.001$) duration of labiodental fricatives /f/. Collapsed across speaker and vowel context, the mean fricative durations for /f, θ, s and j/ were 182, 205, 205, and 199 ms, respectively.

There was also a significant place by age interaction effect [$F(9,96)=3.14$, $p<.003$, $\eta^2=.228$]. Adult speakers were found to have relatively shorter fricative durations when the primary constriction of the consonant was more anterior in the oral cavity (/f/ and /θ/). However, this trend changed for alveolar and palato-alveolar fricatives (/s/ and /j/),

![Figure 6.1](image)

Figure 6.1  Fricative duration as a function of place of articulation for CV fricatives.
with the 5 yr. old speakers exhibiting the shortest durations. This interaction is illustrated in Figure 6.2. Although some separation between age groups can be seen in this figure, a significant main effect was not found.

In addition, the ANOVA also revealed a significant difference in fricative duration as a function of vowel context \[F(2,64)=35.30, p<.001, \eta^2=.525\], as can be seen in Figure 6.3. Post hoc tests indicated that the most significant differences (p<.001) between contexts were characterized by a decrease in the fricative duration when followed by and /a/ vowel. Although not as great, a significant difference (p<.05) between the /i/ and /u/ vowel contexts was also found. The mean fricative durations for the /i/, /a/, and /u/ vowel contexts were 200, 186, and 208 ms, respectively. No other significant effects were found for the measure of fricative duration.

![Figure 6.2 Fricative duration as a function of place of articulation and age for CV fricatives.](image)
6.1.2 Vowel Duration

There were significant differences in the duration of the targeted vowel as a function of both place of fricative articulation \( F(3,96)=9.12, p<.001, \eta^2=.222 \) and vowel context \( F(2,64)=64.94, p<.001, \eta^2=.670 \), as well as a significant place by vowel interaction effect \( F(6,192)=68.15, p<.001, \eta^2=.680 \). As can be clearly seen in figure 6.4, the main effect of place was due to the decreased duration of the vowel following interdental and alveolar fricatives (171 ms for /i/; 189 ms for /θ/; 169 ms for /s/; and 186 ms for /ʃ/). As expected, significant differences (p<.001) were found between all three vowel contexts, with the duration increasing as the place of vowel articulation moved posterior in the oral cavity. This trend is further illustrated in Figure 6.5. Overall, the duration means for the /i/, /a/, and /u/ vowels were 153, 181, and 202 ms, respectively.
Figure 6.4 Vowel duration as a function of place of articulation for CV fricatives.

Figure 6.5 Vowel duration as a function of vowel context for CV fricatives.
In general, the trend of relative durations in the token final vowel remained similar when preceded by labiodental, dental, or alveolar fricatives. The /a/ vowel exhibited the longest mean duration, followed by the /u/ and /i/ vowels, respectively. However, when following a palato-alveolar fricative these trends in relative duration changed, with the /u/ vowel exhibiting the longest mean duration and the /a/ vowel having the shortest duration. These changes across place of fricative articulation and vowel context can be clearly seen in Figure 6.6. No other significant effects or interactions were noted.

![Figure 6.6](image-url)  
Figure 6.6  Mean vowel duration as a function of place of articulation and vowel context in CV fricatives.
6.1.3 Normalized Amplitude

Using a four-way ANOVA to examine the dependent measure of normalized amplitude (rms amplitude in dB of the entire fricative relative to the strongest component in the following vowel), a main effect of place \( F(3,96)=157.69, p<.001, \eta^2=.831 \) was obtained. As expected, Bonferroni adjusted post hoc tests indicated that this effect was derived primarily from significant differences \( p<.001 \) between the nonsibilant and sibilant fricatives. As is illustrated in Figure 6.7, the normalized amplitude of the nonsibilant fricatives (-13.7 dB for /f/, -11.9 dB for /θ/) was less than that of the sibilant fricatives (-3.6 dB for /s/, -3.0 dB for /ʃ/).

![Figure 6.7 Normalized amplitude as a function of place of articulation in CV fricatives.](image-url)
A main effect of vowel context was also found to be significant \([F(2,64)=16.08, p<.001, \eta^2=.334]\), with differences in the normalized amplitude of the fricative depending on the articulation of the following vowel. Overall, the normalized amplitude measures for the fricatives were significantly \((p<.001)\) decreased when followed by an /a/ vowel. As can be seen in Figure 6.8, the normalized values for /i, a, and u/ were –7.7, -9.3, and -7.2 dB, respectively.

![Figure 6.8](image)

Figure 6.8  Normalized amplitude as a function of vowel context in VCV fricatives.

A place by vowel interaction was also found to be significant \([F(6,192)=42.86, p<.001, \eta^2=.177]\). As can be seen in Figure 6.9, the decreased normalized amplitude
found in the /a/ vowel context was significant (p<.01) for /f, θ, and s/, but was actually slightly increased for /ʃ/. No other significant effects were revealed for the dependent variable of normalized amplitude.

![Figure 6.9 Normalized amplitude as a function of place of articulation and vowel context in VCV fricatives.](image)

6.2 Spectral Measures

For fricative consonants in the CV context, spectral moment data were derived at three window locations (onset, mid, and offset) for both a 20 ms and 40 ms series of analysis windows. To assess the importance of the analysis window size, a one-way ANOVA was conducted on the spectral data. Utilizing the derived spectral measures as dependent variables (spectral peak, spectral slope, first four spectral moments), a main
effect of window size was found to be significant for spectral slope [F(1,2878)=1635.72, p<.001], variance [F(1,2878)=10.39, p<.001], and skewness [F(1,2878)=12.59, p<.001].

In addition, differences across window location were found to be significant in five of the six derived spectral measures (spectral slope, peak, mean, skewness, and kurtosis). These ANOVA results, as well as other general patterns in the data that extend across window locations, are reported throughout this section according to each type of spectral measure. However, for purposes of consistency and the ability to compare these results to those previously reported (e.g., Fox et al., 2001; Jongman et al., 2000), specific results from only the middle 40 ms analysis window (centered at the midpoint of the fricative) will be reported in detail in the following discussion.

6.2.1 Spectral Peak

Results of a one-way ANOVA indicated that overall the measure of spectral peak differed significantly as a function of window location [F(2,1437)=12.85, p<.001]. As confirmed by Post hoc analysis (p<.001), this effect was due primarily to the increased overall frequencies derived from the middle window of analysis. These results are not surprising considering that the initial and final windows of analysis would likely contain acoustic information regarding the transition from the vowels preceding and following the fricative target. The overall spectral peaks for the first, second and third windows were 7404, 8028, and 7308 Hz, respectively.

Individual four-way ANOVA’s (gender x age group x place x vowel) conducted on each of the three windows of analysis revealed that for all three window placements only the palato-velar fricatives (/ʃ/) were significantly distinguished (p<.001) from the other places of articulation in terms of mean spectral peak. These mean averages are
broken down by place of articulation in Figure 6.10. The patterns of age differences remained intact across different windows of analysis, as can be seen in Figure 6.11.

As expected, when analyzing the spectral data from the middle 40 ms of the fricative segment, an ANOVA revealed a main effect for place of articulation \(F(3,96)=128.23, p<.001, \eta^2=.800\). This ANOVA was of mixed design, with gender and speaker age as the between subject factors and place of articulation and vowel context as within subject factors. Subsequent post hoc tests indicated that the main effect of place was derived primarily from the significantly lower \(p<.001\) peak location of palato-velar fricatives \((/ʃ/)\), as well as a significantly \(p<.05\) higher location for labio-dental fricatives \((/f/)\). The mean spectral peak locations for all four places of fricative articulation (9397 for \(/f/\), 8721 for \(/θ/\), 8927 for \(/s/\), and 5068 for \(/ʃ/)\) are displayed in Figure 6.12.

![Figure 6.10 Spectral peak location as a function of place of articulation and window location for CV fricatives.](image-url)
Figure 6.11 Spectral peak location as a function of age group and window location for CV fricatives.

Figure 6.12 Spectral peak location by place of articulation in CV fricatives.
In addition, a significant place by age group interaction [$F(9,96)=3.17$, $p<.003$, $\eta^2=.229$] revealed that the spectral peak locations of the different places of fricative articulation are different dependent upon the age of the speaker. In particular, the difference in mean spectral peak between /ʃ/ and the other fricatives increased with the age of the speaker, this trend is illustrated in Figure 6.13. Furthermore, a change in the relative pattern of spectral peak locations across place of articulation is exhibited by the 4 year old speakers. Specifically, /θ/ fricatives show a higher spectral peak than /s/ fricatives only in 4 year old speakers. However, it is important to note that the mean differences between these two places were not significant.

![Figure 6.13 Spectral peak location by place of articulation and age in CV fricatives.](image-url)
Although no main effects of vowel context were revealed by the ANOVA, a significant vowel by gender effect was found [F(2,64)=6.52, p<.004, $\eta^2=.169$]. The relationship between gender and the mean spectral peak varies according to the vowel context. In the /i/ and /u/ vowel contexts, male speakers have a significantly higher (p<.01) mean spectral peak location. However, differences between genders is negligible for fricatives in the /a/ vowel context. This interaction is shown in Figure 6.14.

![Figure 6.14: Spectral peak location by vowel context and gender in CV fricatives.](image)

As expected, main effects of gender [F(1,32)=4.74, p<.05, $\eta^2=.129$] and age group [F(3,32)=14.05, p<.001, $\eta^2=.568$] were also found. As can be seen in Figure 6.15, across...
place of articulation the mean spectral peak location for female speakers (8286 Hz) was higher than males (7770 Hz). This pattern of difference is most prominent with sibilant fricatives (/s/ and /ʃ/).

![Mean Spectral Peak Location Chart](image)

Figure 6.15 Spectral peak location for male and female speakers as a function of place of articulation for CV fricatives.

As shown in Figure 6.16, the child speakers (8454 Hz) exhibited higher average peak locations than the adults (6749 Hz). An inverse relationship between the mean peak location and speaker age was noted. Specifically, the spectral peak decreased as a function of increased speaker age. Comparisons between the age groups indicated that the significant differences were primarily due to a differences between the adult group and the three child groups (p<.003), with no differences between the child groups found.
to be significant. This result is not surprising considering the relatively larger size of the adult oral cavity and the lack of significant anatomic differences between the child speakers. No other main or interaction effects were found when looking at the dependent measure of spectral peak location.

![Figure 6.16 Spectral peak location as a function of different age for CV fricatives.](image)

**6.2.2 Spectral Slope**

The measure of spectral slope was not reported in previous spectral moment research (e.g., Forrest et al., 1988; Jongman et al., 2000; Nittrouer, 1995), however, in pilot research conducted by Fox et al. (2001) the spectral slope was found to be an important cue in the classification of voiceless English fricatives. A one-way ANOVA revealed that the mean spectral slope averaged across speakers, place, and vowel context
differed significantly as a function of window location [F(2,1079)=75.23, p<.001]. Post hoc analyses indicated significant differences (p<.01) in pairwise comparisons that included the first window of analysis (onset). Fricatives analyzed at the onset window location were found to have significantly lower overall spectral slope values (4.3 for onset location, 5.3 for mid location, and 5.0 for final location). As shown in Figure 6.17, a series of ANOVAs (gender x age x place x vowel) and subsequent post hoc analysis revealed that three of the four places of fricative articulation were distinguished at each window location (p<.001), with all locations containing a lack of differentiation between nonsibilant fricatives (/f/ and /θ/).

<table>
<thead>
<tr>
<th>Window Location</th>
<th>Spectral Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset</td>
<td>3</td>
</tr>
<tr>
<td>Mid</td>
<td>4</td>
</tr>
<tr>
<td>Final</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 6.17  Spectral slope as a function of place of articulation and window location for CV fricatives.
In addition, age differences across window locations are displayed in Figure 6.18. As can be seen, the general pattern of slope values between the age groups was maintained throughout the various window locations, with the exception of the middle window. In the first and third window locations the slope values increased with an increase in speaker age and then were relatively lower for the adult speakers. However, for data analyzed in the middle portion of the fricative, the slope values of the adult speakers were found to be relatively higher than all three groups of child speakers. These age group differences will be reported and discussed in more detail later in this section.

![Figure 6.18 Spectral slope as a function of age and window location for CV fricatives.](image)

A four-way ANOVA, with gender and speaker age as the between subject factors and place of articulation and vowel context as within subject factors, was conducted on

142
data from the middle analysis window. A strongly significant main effect of place of articulation [$F(6,36) = 266.16$, $p < .001$, $\eta^2 = .893$] was derived from the ANOVA. As previously discussed, subsequent pairwise comparisons ($p < .001$) demonstrated that three of the four places of fricative articulation were statistically different in terms of mean spectral slope, with no significant difference noted between nonsibilant fricatives (/f/ and /θ/). The mean spectral slope values for the different fricative types (/f, θ, s, ñ/) were 3.42, 3.40, 5.46, and 9.08, respectively.

A significant place by gender interaction [$F(6,36) = 10.31$, $p < .001$, $\eta^2 = .244$] indicated that the mean spectral slope of the four fricative types varied as a function of the gender of the speaker. As illustrated in Figure 6.19, female speakers exhibited higher slope values for all places of fricative articulation, except for the alveolar fricatives (/s/).

In addition, the ANOVA also yielded a significant place by vowel interaction [$F(4,24) = 4.77$, $p < .001$, $\eta^2 = .130$]. The interaction reveals that when followed by an /i/ vowel the (/f/, /θ/, and /ñ/) fricatives show a slight elevation in slope, whereas the /s/ fricatives actually decrease in slope for this particular vowel context. This interaction is displayed in Figure 6.20 below.

Interestingly, a main effect was obtained for age group [$F(3,32) = 4.84$, $p < .01$, $\eta^2 = .312$], as well as a significant gender by age group interaction [$F(3,32) = 3.08$, $p < .05$, $\eta^2 = .224$]. Spectral slope values rise as the age of speaker increases (the mean slope values for 3 year old, 4 year old, 5 year old, and adult speakers were 4.61, 5.18, 5.70, and 5.86, respectively). Significant differences ($p < .05$) were found between the youngest group and the two older age groups, as well as between the 4 year old speakers the adult speakers. However, the extent of these differences between age groups varies widely as a
Figure 6.19 Spectral slope as a function of gender and place for CV fricatives.

Figure 6.20. Spectral slope as a function of vowel context and place of articulation for CV fricatives.
function of gender. This interaction effect can be clearly seen in Figure 6.21 below. No other main or interaction effects were found to be significant for the dependent measure of spectral slope.

![Figure 6.21 Spectral slope as a function of gender and age group for CV fricatives.](image)

**6.2.3 Spectral Mean**

The first spectral moment (mean) was computed for the CV fricative consonants at all three window locations (onset, mid, and offset). A one-way ANOVA revealed that the overall spectral mean differed across window location [$F(2,1437)=8.05$, $p<.001$]. Individual four-way ANOVA’s (gender x age x place x vowel) were conducted on each of the three windows of analysis. As shown in Figure 6.22, each window distinguished three places of fricative articulation, with the greatest differences found between /s/ and the other three fricatives. Although no significant differences in age group were found
for any analysis window in terms of spectral mean, Figure 6.23 does illustrate a surprising change in the relative pattern of data for the middle analysis window. As can be seen, the youngest age group exhibited a relatively high overall spectral mean for the onset and offset windows, but a relatively low mean for the middle window. This result is surprising considering the anatomical differences in vocal tract size between the youngest speakers and the adults. Assuming that the fricatives are produced in a similar manner, speakers with a smaller vocal tract should exhibit higher spectral mean values, which was the case in the onset and offset portions of frication. Thus, one possible explanation for the 3 year old speakers lowered spectral means in the middle window may be developmental, rather than anatomical in nature. The results for this window (mid location) will be reported in more detail in the following discussion.

![Graph showing spectral mean as a function of place of articulation and window location for CV fricatives.](image)

Figure 6.22  Spectral mean as a function of place of articulation and window location for CV fricatives.
A four-way repeated measures ANOVA (place x vowel x gender x age group) conducted on the middle 40 ms of the fricative revealed a significant main effect for place of articulation [$F(3, 96) = 30.32$, $p < .001$, $\eta^2 = .487$]. As shown in Figure 6.24, post-hoc analyses indicated significant differences ($p < .001$) between three of the four places of fricative articulation. No statistical difference in spectral mean was found between /θ/ and /ʃ/ fricative tokens. Collapsed across speaker and vowel context, the fricative consonants (/f, θ, s, ʃ/) exhibited spectral means of 4987, 5497, 6602, and 5713 Hz, respectively.

Figure 6.23 Spectral mean as a function of age and window location for CV fricatives.
Figure 6.24  Spectral mean as a function of place of articulation for CV fricatives.

There was also significant place by gender interaction effect [$F(3,96)=5.11$, $p<.004$, $\eta^2=.138$]. As shown in Figure 6.25, this interaction is characterized by higher spectral means for male speakers on the three more anterior produced fricatives (/f, θ, s/) and a lower spectral mean for the more posterior /ʃ/ fricative. It is somewhat surprising to find that the male speakers exhibited elevated spectral means in the majority of fricative types. However, this result is further explained by a closer investigation of gender differences within specific age groups. Figure 6.26 further illustrates that the elevated means for male speakers was found primarily in the child speakers, with no significant difference between genders in the adult group. A significant main effect of gender was not found to be significant when data from all age groups were included in the ANOVA. However, if the adult speakers are excluded from the analysis, a one-way
Figure 6.25 Spectral mean as a function of place of articulation and gender for CV fricatives.

Figure 6.26 Spectral mean as a function of age group and gender for CV fricatives.
ANOVA reveals a significant effect of gender \(F(1,358)=7.68, p<.01\) within the child speakers. Since anatomical differences in vocal tract size are negligible for the child speakers, any gender differences are most likely developmental in nature.

In addition, the four-way ANOVA (place x vowel x gender x age group) also indicated a significant place by age group \(F(3,96)=11.31, p<.001, \eta^2=.515\) interaction. As can be seen in Figure 6.27, a strong contrast between /s/ and /ʃ/ was exhibited by the 5 year old speakers and further differentiated by the adults. The spectral mean of the /ʃ/ fricative was significantly lower than /s/. This within-sibilant contrast is not evident in the younger groups of speakers (3 and 4 yr. old). Previous results by Nittrouer (1995) found a similar place by age group interaction. Her results indicated that the difference in spectral mean values for /s/ and /ʃ/ are larger for adult speakers than for those of children.

![Figure 6.27 Spectral mean as a function of place of articulation and age for CV fricatives.](image)
ages 3 through 7. Nittrouer concluded that these differences were the product of
development differences, indicating that even children at 7 years of age had not yet
acquired adult-like gestures for /s/ and /ʃ/. These results support Nittrouer’s conclusion,
however, they also indicate that at 5 years of age children are probably starting to
approximate an adult-like contrast between sibilant fricatives.

6.2.4 Spectral Variance

A one-way ANOVA revealed that the overall spectral variance did not differ significantly
as a function of window location. Individual four-way ANOVA’s conducted at each
window location indicated that all four places of articulation can be differentiated from
each other in terms of spectral variance, as can be seen in Figure 6.28.

Figure 6.28 Spectral variance as a function of place of articulation and window location
for CV fricatives.
As mentioned previously, a significant main effect of place \([F(3,96)=50.48, p<.001, \eta^2=.612]\) was obtained for spectral variance. Subsequent pairwise comparisons indicated that all four fricative types (/f, θ, s, j/) were significantly (\(p<.001\)) different in terms of variance, with means of 6.26, 5.38, 2.39, and 3.30 MHz, respectively. Interestingly, when collapsed into two groups according to sibilance, the measure of spectral variance was found to significantly (\(p<.001\)) differentiate between sibilant and nonsibilant fricatives. These differences are further illustrated in Figure 6.29. No other effects or interactions were found to be significant for spectral variance.

Figure 6.29  Spectral variance as a function of place of articulation for CV fricatives.
6.2.5 Spectral Skewness

A one-way ANOVA revealed that the mean spectral skewness averaged across speakers, place, and vowel context differed significantly as a function of window location \[F(2,1437)=4.69, p<.01\]. Through a series of four-way ANOVAs (gender x age x place x vowel), it was found that the pattern of spectral skewness within different places of articulation remained similar across window locations. As can be seen in Figure 6.30, the only fricative to have a significant (p<.001) difference in skewness was the palato-velar fricative /\gamma/.

Age differences across window locations are displayed in Figure 6.31. As can be seen, the significant differences (p<.001) between the child speakers and the adults is maintained throughout the various window locations. These age group differences will be reported and discussed later in this section.

![Figure 6.30 Spectral skewness as a function of place of articulation and window location for CV fricatives.](image-url)
As previously mentioned, analyses of data derived from the middle window location indicated a strongly significant main effect of place \([F(3,96)=242.97, p<.001, \eta^2=.884]\) for spectral skewness. As can be seen from Figure 6.32, the main effect of place was due to the significantly (\(p<.001\)) elevated skewness of palato-velar fricatives. This conclusion was confirmed through subsequent pairwise comparisons. The mean skewness for /f, θ, s, j/ was -2.23, -2.18, -1.88, and 0.21, respectively.

There was also a significant place by gender interaction effect \([F(3,96)=4.86, p<.004, \eta^2=.132]\). As shown in Figure 6.33, this interaction was characterized by a gender difference for alveolar fricatives (-1.56 for male speakers and -2.21 for female speakers), but not for any other places of fricative articulation. However, these differences were not large enough to cause an overall main effect of gender.
Figure 6.32  Spectral skewness as a function of place of articulation for CV fricatives.

Figure 6.33  Spectral skewness as a function of place of articulation and gender for CV fricatives.
In addition, the four-way ANOVA (place x vowel x gender x age group) also indicated a strongly significant main effect of age group \[F(3,32)=24.96, \, p<.001, \, \eta^2=.701\] and a significant place by age group \[F(9,96)=12.37, \, p<.001, \, \eta^2=.537\] interaction effect. Post hoc tests indicated significant \((p<.001)\) differences between the child speakers and the adults, with no differences between the child groups found to be significant. This main effect of age group is further explained by examining the place by age group interaction. As can be seen in Figure 6.34, the differences in age group are found primarily in comparisons of the fricative /ʃ/. The contrast in spectral skewness between /ʃ/ and the other fricative types widened as the age of the speaker increased, possibly indicating the developing acquisition of the /ʃ/ fricative toward an adult-like stage of articulation. These results provide some evidence indicating this development begins at approximately four years of age and continues to be fine tuned throughout young childhood (at least through age six). This same general trend across age groups was also reported by Nittrouer (1995), who concluded that similar age-related acoustic differences were the product of continuing articulatory development.

6.2.6 Spectral Kurtosis

A one-way ANOVA revealed that the mean spectral kurtosis averaged across speakers, place, and vowel context differed significantly as a function of window location \[F(2,1437)=58.35, \, p<.003\]. A series of four-way ANOVAs (gender x age x place x vowel) indicated that all four places of fricative articulation were significantly distinguished \((p<.001)\) at both the onset and offset window locations, as is shown in Figure 6.35. However, for the middle analysis window only two places of articulation were differentiated by the measure of spectral kurtosis. Age differences across window
Figure 6.34 Spectral skewness as a function of age group and place of articulation for CV fricatives.

Figure 6.35 Spectral kurtosis as a function of place of articulation and window location for CV fricatives.
locations are displayed in Figure 6.36. As can be seen, the relative pattern of kurtosis values varied widely as a function of window location and age group. Any differences in place or age group derived from the middle window location will be reported and discussed in more detail later in this section.

A four-way ANOVA, with gender and speaker age as the between subject factors and place of articulation and vowel context as within subject factors, revealed significant main effect of place of articulation \( [F(3,96)=266.16, p<.001, \eta^2=.893] \). As previously discussed, subsequent pairwise comparisons demonstrated that only two of the four places of fricative articulation were statistically different \( (p<.001) \) in terms of mean spectral kurtosis. These differences were due mainly to the decreased kurtic value of palato-velar fricatives. As shown in Figure 5.37, the mean kurtic values for the four fricative types (/f, θ, s, j/) were found to be 3.78, 3.77, 3.54, and 1.46, respectively.

![Figure 6.36 Spectral kurtosis as a function of age and window location for CV fricatives.](image-url)
A significant place by gender interaction \([F(3,96)=10.31, p<.001, \eta^2=.244]\) indicated that the mean spectral kurtosis of the four fricative types varied as a function of the gender of the speaker. As illustrated in Figure 6.38, the values of kurtosis for male speakers (3.96) were higher than female speakers (3.58) for nonsibilant fricatives, yet lower for sibilant fricatives (1.76 for males and 3.23 for females). As can be clearly seen, the overall gender differences for sibilant fricatives are derived mainly from differences within fricatives articulated near the alveolar ridge (/s/).

As can be seen in Figure 6.39, the ANOVA also yielded a significant place by age group interaction \([F(9,96)=4.64, p<.001, \eta^2=.305]\). Although an overall main effect of age group was not found to be significant, the interaction reveals a highly significant age difference \((p<.001)\) for /ʃ/ fricatives. This finding may indicate that with regard to
Place of Articulation

Male
Female

Figure 6.38 Spectral kurtosis as a function of gender and place for CV fricatives.

Figure 6.39 Spectral kurtosis as a function of age and place of articulation for CV fricatives.
kurtosis, the child speakers have not yet acquired an adult-like palato-velar fricative. Similar age differences concerning the measure of spectral kurtosis have not been reported in previous research (i.e., Nittouer, 1995).

6.3 Discriminant Analysis for CV Fricative Data

Discriminant analysis was conducted to determine how accurately the derived acoustic measures could categorize fricative tokens in the CV context by place of articulation. The acoustic parameters entered into the initial stepwise linear discriminant analysis were as follows: fricative duration, normalized fricative amplitude, final vowel duration, spectral peak, spectral slope, spectral mean, spectral variance, spectral skewness, and spectral kurtosis. The spectral moments measures were derived from the middle 40 ms of the fricative. Several different discriminant functions or models were utilized to classify the fricative data based on the procedure (jack-knife or standard stepwise) and the data set on which the function was trained. A summary of the variables included in the various classification models can be found in table 6.1.

As shown in Table 6.2, when classification results are based on a cross-validation (jack-knife) procedure, 69.2% of the entire data set (fricatives in CV context) were classified correctly in terms of place of articulation. The adults had a combined classification rating of 65.0%, while the 3, 4, and 5 year old speakers exhibited a rating of 70.0%, 70.0%, and 77.5%, respectively. Overall, differences between male (67.5%) and female (68.8%) speakers were minimal. An examination of classification rates across age groups indicated that gender differences within individual ages ranged from only 3.3% for the adult speakers to 10.0% for the 5 year old speakers.
Utilizing a standard stepwise procedure, a model of classification was trained on acoustic data derived from the adult speakers. As shown in Table 6.2, the spectral measures of skewness, mean, variance, slope, and kurtosis were combined to form the discriminant model. This classification model was able to correctly classify fricatives produced by the child speakers with an overall accuracy of 42.2%. The rate of successful classification improved as the age of the child speakers increased. The mean classification rates for the 3, 4, and 5 year old age groups were found to be 34.2%, 38.3%, and 54.2%, respectively. Interestingly, the classification model categorized the fricative productions of the male speakers (48.9%) at a higher rate than the female children (35.6%).

Another classification model was created by using only the adult male data as the training set. As can be seen in Table 6.3, using the derived measures of spectral skewness, mean, slope, variance, and kurtosis (listed in order of contribution), this model successfully classified 85.0% of the training data by place of fricative articulation. Classification rates for the sibilant fricatives in the training set was 100%, while the nonsibilant fricatives were correctly identified 70% of time. Errors in classification did not cross the sibilant/nonsibilant distinction. In other words, nonsibilant fricatives were only confused with each other. When applied to the remaining data sets, the classification model categorized fricatives with an overall accuracy of only 44.3%. This difference in rate was due primarily to a large decrease in accurate classification for fricatives produced by the child age groups. The mean rates for the 3, 4, and 5 year olds were 28.3%, 40.0%, and 55.8 %, respectively. Further examination of the different places of fricative articulation revealed that the discriminant function was more accurate in
classifying the children’s labiodental fricatives (75.6%), compared to the decreased rates found for the other fricative types (37.8% for /θ/, 33.3% for /s/ and 18.9% for /ʃ/). The classification model accurately identified only 6.7% of the 3 year old female productions of /s/ and /ʃ/. However, the classification rates for the sibilant fricatives improved dramatically as the age of the speaker increased (the mean rates for the 3, 4, and 5 year old data were 11.7%, 18.3%, and 48.4 %, respectively). As discussed previously, these results support the theory that young children’s (up to age 7) sibilant fricative gestures are not differentiated in an adult-like manner and are therefore continuing to be “fine-tuned” as the child matures (Nittrouer, 1995). Excluding the adult male data, gender differences of 11.0% were found (50.6% for males and 39.6% for females). Interestingly, fricative data from males was classified at a higher rate than the females in every age group.

A standard stepwise discriminant analysis of data produced by the adult female speakers resulted in a classification model that contained 4 acoustic parameters. Listed in the order of their contribution to the classification, these parameters include the following: spectral slope, spectral mean, normalized fricative amplitude, and spectral skewness. As can be seen in Table 6.4, 83.3% of the fricative tokens in the original training set were classified correctly. Perfect classification (100%) was found for the sibilant fricatives, while the nonsibilant fricatives /f/ and /θ/ were categorized at the lower rate of 60.0% and 73.3%, respectively. Similar to the previous classification model, nonsibilant fricatives were not misidentified as sibilant fricatives, but only with each other. When applied to the rest of the dataset, the model yielded an overall classification rate of 51.0%. Similar to the model developed on adult male data, the lowered rate of this classification model was also due in part to the decreased rates found for the child
productions of the sibilant fricative /ʃ/ (17.6%), yet the classification of the /s/ fricative was nearly perfect at 96.7%. The classification accuracy for the child productions of the nonsibilant fricatives /f/ and /θ/ were 24.4% and 27.8%, respectively. Interestingly, fricative data from the male children (48.3%) were classified by this function at a higher rate than from the female children (38.9%). Mean rates for /f, θ, s, and ʃ/ were found to be 21.9%, 49.5%, 70.5%, and 61.9%, respectively. Some of the lowest classification rates were found for labiodental fricatives produced by the child speakers, with an average accuracy of 17.7%. Overall, differences between male (54.2%) and female (55.8%) speakers were minimal, with the largest average differences found in the adult speakers (60.0% for male and 83.3% for female).
<table>
<thead>
<tr>
<th>Data Set Trained</th>
<th>Acoustic Variables in the Classification Model (by order of significance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV Context</td>
<td>VCV Context</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
<tr>
<td>Spectral Skewness</td>
<td>Spectral Skewness</td>
</tr>
<tr>
<td>Spectral Mean</td>
<td>Spectral Mean</td>
</tr>
<tr>
<td>Spectral Variance</td>
<td>Spectral Variance</td>
</tr>
<tr>
<td>Spectral Slope</td>
<td>Spectral Slope</td>
</tr>
<tr>
<td>Normalized Amplitude</td>
<td>Normalized Amplitude</td>
</tr>
</tbody>
</table>
| Spectral Skewness        | Spectral Skewess...
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Training Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self Classification</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>—</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td>70.0</td>
</tr>
<tr>
<td>- male only</td>
<td>61.7</td>
</tr>
<tr>
<td>- female only</td>
<td>68.3</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td>70.0</td>
</tr>
<tr>
<td>- male only</td>
<td>73.3</td>
</tr>
<tr>
<td>- female only</td>
<td>70.0</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td>77.5</td>
</tr>
<tr>
<td>- male only</td>
<td>85.0</td>
</tr>
<tr>
<td>- female only</td>
<td>75.0</td>
</tr>
<tr>
<td>Adult Speakers</td>
<td>65.0</td>
</tr>
<tr>
<td>- male only</td>
<td>75.0</td>
</tr>
<tr>
<td>- female only</td>
<td>78.3</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td>69.2</td>
</tr>
<tr>
<td>- male only</td>
<td>67.5</td>
</tr>
<tr>
<td>- female only</td>
<td>68.8</td>
</tr>
</tbody>
</table>

Table 6.2 Results of discriminant classification using both cross-validation (jack-knife) and standard stepwise procedures for CV fricatives; all values represent % correct classification.
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>/f/</th>
<th>/θ/</th>
<th>/s/</th>
<th>/ʃ/</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Training Set</td>
<td>66.7</td>
<td>73.3</td>
<td>100</td>
<td>100</td>
<td>85.0</td>
</tr>
<tr>
<td>3 year old speakers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>66.7</td>
<td>23.3</td>
<td>16.7</td>
<td>6.7</td>
<td>28.3</td>
</tr>
<tr>
<td>- female only</td>
<td>80.0</td>
<td>33.3</td>
<td>26.7</td>
<td>6.7</td>
<td>36.7</td>
</tr>
<tr>
<td>4 year old speakers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>83.3</td>
<td>40.0</td>
<td>23.3</td>
<td>13.3</td>
<td>40.0</td>
</tr>
<tr>
<td>- female only</td>
<td>80.0</td>
<td>46.7</td>
<td>40.0</td>
<td>6.7</td>
<td>43.3</td>
</tr>
<tr>
<td>5 year old speakers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>76.7</td>
<td>50.0</td>
<td>60.0</td>
<td>36.7</td>
<td>55.8</td>
</tr>
<tr>
<td>- female only</td>
<td>93.3</td>
<td>40.0</td>
<td>93.3</td>
<td>60.0</td>
<td>71.7</td>
</tr>
<tr>
<td>Adult speakers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- female only</td>
<td>60.0</td>
<td>60.0</td>
<td>26.7</td>
<td>13.3</td>
<td>40.0</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>75.2</td>
<td>36.2</td>
<td>35.2</td>
<td>30.5</td>
<td>44.3</td>
</tr>
<tr>
<td>- female only</td>
<td>84.4</td>
<td>40.0</td>
<td>53.3</td>
<td>24.4</td>
<td>50.6</td>
</tr>
</tbody>
</table>

Table 6.3  Results of stepwise linear discriminant analysis when the classification model is trained on adult male data only for CV fricatives; all values represent % correct classification.
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Trained on Adult Female Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/f/</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>60.0</td>
</tr>
<tr>
<td>3 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>23.3</td>
</tr>
<tr>
<td>- female only</td>
<td>33.3</td>
</tr>
<tr>
<td>4 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>16.7</td>
</tr>
<tr>
<td>- female only</td>
<td>20.0</td>
</tr>
<tr>
<td>5 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>13.3</td>
</tr>
<tr>
<td>- female only</td>
<td>13.3</td>
</tr>
<tr>
<td>Adult speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td></td>
</tr>
<tr>
<td>- female only</td>
<td></td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td>21.9</td>
</tr>
</tbody>
</table>

Table 6.4 Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for CV fricatives; all values represent % correct classification.
6.4 Summary of Significant Results

For fricatives produced in the CV context, statistical analysis (ANOVA) indicated that the dependent measures of fricative duration, vowel duration, normalized amplitude, spectral peak, spectral slope, and all four spectral moments varied significantly as a function of place of articulation. However, only spectral variance was found to significantly distinguish all four places of fricative articulation. Interestingly, with the exception of Jongman et al. (2000) and Tomiak (1990), the measure of spectral variance has often not been analyzed or reported in the previous literature (e.g., Forrest et al., 1988; Miccio, 1996; Nittrouer, 1995). The measures of spectral peak, slope, mean, skewness, and kurtosis were able to separate non-sibilant from sibilant fricatives, as well as between the two sibilant fricatives (/s/ and /ʃ/). However, these measures were not able to differentiate between the non-sibilant fricatives (/θ/ and /ð/).

Fricatives produced in a CV context were found to be less sensitive than the stop consonants to vowel context, with main effects being found for fricative duration, vowel duration, normalized amplitude. These results are similar to previous research by Jongman et al. (2000) and unlike research presented by Nittrouer (1995).

Significant effects of age were found for the spectral measures of peak location, slope, and skewness. Similar to the results of Nittrouer (1995), significant place by age interaction effects were found for spectral peak, slope, and skewness. In general, the interaction effects were due mainly to a widening in the acoustic distinction between /s/ and /ʃ/ as the age of speaker increased. Supporting the conclusions of Nittrouer, these age related differences may be the product of continuing articulatory development by the
child speakers toward an adult-like pattern of production. Gender differences were only noted for the measure of spectral peak location.

In terms of place of articulation, a discriminant analysis based on a cross-validation procedure was able to correctly classify 69.2% of the fricative data. Productions from the adults were found to have classification rating of 65.0%, while the speech productions of the 3, 4, and 5 year old speakers exhibited ratings of 70.0%, 70.0%, and 77.5%, respectively. Across all speaker groups, gender differences were minimal (67.5% for male and 68.8% for female). When based on a discriminant function trained only on the adult male data, the classification accuracy differed significantly as a function of both the fricative place of articulation and the age of the speaker. Speech data from the training set (adult male) was classified at an overall rate of 85.0%, while speech from the child speakers was correctly classified at the lower rate of 41.3%. This decrease in classification was primarily due to the inability of the discriminant function to categorize the sibilant fricatives of the child speakers (33.3% for /s/ and 18.9% for /ʃ/).

Interestingly, the classification rates for the child speakers’ data increased dramatically as the age of the speaker increased. Thus, these data support the theory that young children’s sibilant fricative articulations are not differentiated in an adult-like manner, but are continuing to be “fine-tuned” as the child matures (Nittrouer, 1995). Of further interest, is the finding that the discriminant function developed on adult male fricative data was better at classifying the fricative productions of male children relative to the female children. This gender difference was more pronounced with the older groups of child speakers.
CHAPTER 7

ACOUSTIC ANALYSIS OF VCV FRICATIVE CONSONANTS

The results found in the following section were derived from fricatives produced in a VCV context. The presentation of these findings will be similar to the format found in the previous chapter. For spectral measures, general comparisons of analysis window size and location will be presented, followed thereafter by more detailed findings from only the middle 40 ms analysis window. Although specific statistics from additional analysis windows will not be reported, general patterns in the data that extend across these windows will also be described. All data were collapsed across repetition for a given speaker prior to statistical analysis. Unless otherwise noted, any main or interaction effects found to be significant at a .05 alpha level were discussed.

7.1 Duration and Amplitude Measures

For speech tokens in the VCV context, the phonemic category of the initial vowel /ɑ/ remaining constant across all tokens, yet the following fricative consonant (/ð, θ, s, ʃ/) and final vowel (/i/, /a/, and /u/) varied as function of place of articulation. A four-way repeated measures ANOVA, with gender and age as the between subject factors and place of articulation and vowel context as within subject factors was used to analyze a variety of duration and amplitude measures (overall token duration, initial vowel
duration, fricative duration, final vowel duration, and normalized amplitude). When indicated all post hoc tests were adjusted for multiple comparisons (Bonferroni).

7.1.1 Overall Token Duration

The ANOVA revealed a significant main effect of final vowel context \([F(2,64)=15.01, p<.001, \eta^2=.321]\). Pairwise comparisons indicating that the overall token duration decreased significantly \((p<.01)\) when the final vowel was an /i/ (653 ms) when compared to the other vowel contexts (701 ms for /a/ and 685 ms for /u/). These differences are illustrated in Figure 7.1. No other main effects or interaction effects were found to be significant.

Figure 7.1  Overall token duration as a function of final vowel context in VCV fricatives.
7.1.2 Initial Vowel Duration

For the dependent measure of initial vowel duration, a main effect of final vowel context was revealed by the ANOVA \[ F(2,64)=6.29, p<.004, \eta^2=.212 \]. Post hoc comparisons indicated that the main effect was the result of a significant \( p<.003 \) difference in the initial vowel duration when the final vowel context contained an /a/ as compared to an /u/ vowel context. This difference is illustrated in Figure 7.2. No other paired comparisons were found to be significant.

![Initial Vowel Duration as a function of final vowel context in VCV fricatives.](image)

**Figure 7.2** Initial vowel duration as a function of final vowel context in VCV fricatives.

7.1.3 Fricative Duration

There were significant differences in fricative duration as a function of both place of articulation \[ F(3,96)=13.03, p<.001, \eta^2=.289 \] and vowel context \[ F(2,64)=29.30, \eta^2=.\]
As can be seen in figure 7.3, the most significant differences in fricative duration were found in comparisons between nonsibilant fricatives (206 ms for /f/ and 202 ms for /θ/) and fricatives articulated near the alveolar ridge (226 ms for /s/). The main effect of vowel context was derived mainly from an overall decrease in fricative duration when followed by an /a/ vowel, as illustrated in Figure 7.4. The mean fricative durations for the /i/, /a/, and /u/ final vowel contexts were 216, 202, and 220 ms, respectively. These conclusions were statistically confirmed with post hoc comparisons (p<.001). Although the interaction effect was significant, the effect size was relatively small (η²=.091).

![Figure 7.3 Fricative duration as function of place of articulation in VCV fricatives.](image)
7.1.4 Final Vowel Duration

A main effect of place \([F(3,96)=5.72, p<.002, \eta^2=.152]\) revealed that the final vowel duration was statistically different dependent upon the articulation of the preceding fricative (328 ms for /f/, 324 ms for /θ/, 320 ms for /s/, and 306 ms for /ʃ/). As shown in Figure 7.5, the final vowel duration increased as the preceding fricative was articulated at a more anterior position in the oral cavity. Subsequent post hoc analysis indicated significant (p<.05) differences between /ʃ/ and the other places of fricative articulation. As expected, final vowel duration differed significantly as a function of vowel type \([F(2,64)=39.72, p<.001, \eta^2=.554]\). As shown in Figure 7.6, subsequent pairwise comparisons indicated that all three vowel types (/i, a, u/) were significantly (p<.01) different in terms of duration, with means of 290, 356, and 313 ms, respectively.
Figure 7.5  Final vowel duration as a function of place of articulation in VCV fricatives.

Figure 7.6  Final vowel duration as a function of final vowel context in VCV fricatives.
7.1.5 Normalized Amplitude

Using a four-way ANOVA to examine the dependent measure of normalized amplitude, a main effect of place \[F(3,96)=96.26, p<.001, \eta^2=.751\] was obtained. As expected, Bonferroni adjusted post hoc tests indicated that this effect was derived primarily from significant differences (p<.001) between nonsibilant and sibilant fricatives. As is illustrated in Figure 7.7, the normalized amplitudes of the nonsibilant fricatives (-11.5 dB for /f/, -10.7 dB for /θ/) were less than that of the sibilant fricatives (-3.7 dB for /s/, -3.0 dB for /ʃ/). These results are similar to previous research with adults (Jongman et al., 2000) and pilot research with children (Fox et al., 2001).

Figure 7.7  Final vowel duration as a function of final vowel context in VCV fricatives.
A main effect of vowel context was also found to be significant \( F(2,64)=28.05, p<.001, \eta^2=.467 \), with differences in the normalized amplitude of the fricative depending on the articulation of the following vowel. The effect of vowel context was mainly due to the significantly decreased amplitude of fricatives preceding an /a/ vowel (\( p<.001 \)), as comparisons between /i/ and /u/ contexts were not found to be significant. As shown in Figure 7.8, the mean normalized amplitude of fricatives preceding the /i, a, and u/ vowel contexts were -6.87, -8.55, and -6.21, respectively.

![Figure 7.8](image)

**Figure 7.8** Normalized amplitude as a function of vowel context in VCV fricatives.

A place by vowel interaction was also found to be significant \( F(6,192)=11.38, p<.001, \eta^2=.262 \). As can be seen in Figure 7.9, the decreased normalized amplitude
found in the /ɑ/ vowel context is highly significant (p<.001) for nonsibilant fricatives, yet negligible for sibilant fricatives. No other significant effects were revealed for the dependent variable of normalized amplitude.

![Normalized amplitude as a function of place of articulation and vowel context in VCV fricatives.](image)

**Figure 7.9** Normalized amplitude as a function of place of articulation and vowel context in VCV fricatives.

### 7.2 Spectral Measures

For fricative consonants in the VCV context, spectral moment data were derived at three window locations (onset, mid, and offset) for both a 20 ms and 40 ms series of analysis windows. To assess the importance of the analysis window size and location, a one-way ANOVA was conducted on the spectral data. A main effect of window size was found to be significant for spectral slope \[ F(1,2878)=1645.84, p<.001 \], spectral variance \[ F(1,2878)=8.9, p<.004 \], and skewness \[ F(1,2878)=11.63, p<.002 \].
In addition, differences across window location were found to be significant in five of the six derived spectral measures (spectral slope, peak, variance, skewness, and kurtosis). These ANOVA results, as well as other general patterns in the data that extend across window locations, are reported throughout this section (fricatives in VCV context) according to each type of spectral measure. However, for purposes of consistency and the ability to compare these results to those previously reported (e.g., Fox et al., 2001; Jongman et al., 2000), specific results from only the middle 40 ms analysis window (centered at the fricative midpoint) was reported in detail in the following discussion.

7.2.1 Spectral Peak

Results of a one-way ANOVA indicated that the overall measure of spectral peak differed significantly as a function of window location \([F(2,1437)=20.37, p<.001]\). As confirmed by Post hoc analysis \((p<.001)\), this effect was due primarily to the increased overall frequencies derived from the middle window of analysis. These findings are not surprising considering that the initial and final windows of analysis probably contain acoustic information regarding the transitions to and from the surrounding vowels. The overall spectral peaks for the first, second and third windows were 7084, 8326, and 7198 Hz, respectively.

Individual four-way ANOVA’s were conducted on each of the three windows of analysis. The mean averages are broken down by place of articulation and age group in Figures 7.10 and 7.11, respectively. Subsequent pairwise comparisons indicated that for all three window placements only the palato-velar fricatives (//f//) were significantly distinguished \((p<.001)\) from the other places of articulation in terms of mean spectral
peak. Furthermore, the patterns of age differences remained primarily intact across different windows of analysis. Not surprisingly, the adult speakers exhibited lower peak values in all three window locations.

![Figure 7.10 Spectral peak location as a function of place of articulation and window location for VCV fricatives.](image)

As expected, a main effect for place of articulation \( F(3,96)=167.44, p<.001, \eta^2=.840 \) was revealed by an ANOVA when analyzing the spectral data from the middle 40 ms of the fricative segment. The ANOVA was of mixed design, with gender and speaker age as the between subject factors and place of articulation and vowel context as within subject factors. Subsequent post hoc tests indicated that the main effect of place was derived primarily from the significantly lower (\( p<.001 \)) peak location of palato-velar
Figure 7.11 Spectral peak location as a function of age and window location for VCV fricatives.

The mean spectral peak locations for all four places of fricative articulation (9275 Hz for /f/, 9071 Hz for /θ/, 8872 Hz for /s/, and 5110 Hz for /ʃ/) are displayed in Figure 7.12.

The ANOVA also revealed a significant place by age group interaction $[F(9,96)=5.21, p<.001, \eta^2=.328]$ indicating that the spectral peak locations of the different places of fricative articulation were different as a function of the age of the speaker. In particular, the difference in mean spectral peak between /ʃ/ and the other fricatives increased with the age of the speaker, this trend is illustrated in Figure 7.13. Similar to the findings of Nittrouer (1995), these results support the position that adults differentiate their sibilant fricatives more than children.
Figure 7.12 Spectral peak location as a function of place of articulation for VCV fricatives.

Figure 7.13 Spectral peak location as a function of place of articulation and age group for VCV fricatives.
A significant effect of age group was also found \(F(3,32)=19.59, p<.001, \eta^2=.648\], which is not surprising considering the vocal tract size differences between child and adult speakers. As shown in Figure 7.14, post hoc comparisons indicated that the main effect of age group was the result of significant differences \(p<.001\) between the average spectral peak location of the adult group (6787 Hz) as compared to the three child groups (8513 Hz). No differences between the child groups were found to be significant.

![Figure 7.14](image)

Figure 7.14  Spectral peak location as a function of age group for VCV fricatives.

In addition, a main effect for gender \(F(1,32)=6.07, p<.02, \eta^2=.160\] was also found. Collapsed across age group, place of articulation, and vowel context, the mean spectral peak location for female speakers (8328 Hz) was higher than males (7836 Hz).
Although the effect size was relatively small ($\eta^2=.085$), a significant place by gender interaction effect [$F(3,96)=2.95$, $p<.05$] indicated that gender differences were more prominent in sibilant fricatives. This interaction is illustrated Figure 7.15. In terms of spectral peak location, no other main or interaction effects were found.

![Figure 7.15 Spectral peak location for male and female speakers as a function of place of articulation for VCV fricatives.](image)

7.2.2 Spectral Slope

A one-way ANOVA revealed that the mean spectral slope averaged across speakers, place, and vowel context differed significantly as a function of window location [$F(2,1437)=26.84$, $p<.001$]. As shown in Figure 7.16, a series of 4-way ANOVAs indicated that three of the four places of fricative articulation were distinguished at each
window location \((p<.001)\), with all locations containing a lack of differentiation between nonsibilant fricatives. Age differences across window locations are displayed in Figure 7.17. As can be seen, the general pattern of slope values between the age groups is maintained throughout the various window locations.

<table>
<thead>
<tr>
<th>Window Location</th>
<th>Spectral Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset</td>
<td>/c4/</td>
</tr>
<tr>
<td>Mid</td>
<td>/c4/c5</td>
</tr>
<tr>
<td>Offset</td>
<td>/c4/c6</td>
</tr>
</tbody>
</table>

Figure 7.16  Spectral slope as a function of place of articulation and window location for VCV fricatives.

A four-way ANOVA, with gender and speaker age as the between subject factors and place of articulation and vowel context as within subject factors, revealed a highly significant main effect of place of articulation \([F(3,96)=266.16, p<.001, \eta^2=.893]\) for the middle window of analysis. As previously discussed, subsequent post hoc analysis with a
Bernoulli adjustment for multiple comparisons demonstrated that three of the four places of fricative articulation were statistically different ($p<.001$) in terms of mean spectral slope, with no significant differences noted between nonsibilant fricatives. The mean spectral slope values for the different fricative types (/f, θ, s, ʃ/) were 3.41, 3.53, 5.53, and 9.10, respectively. A significant place by gender [$F(3,96)=10.68$, $p<.001$, $\eta^2=.250$] interaction indicated that the differences in mean spectral slope across place of articulation varied as a function of speaker gender. As illustrated in Figure 7.18, female speakers exhibited higher slope values for all places of fricative articulation, except for alveolar fricatives (/s/).
Furthermore, the ANOVA also indicated a main effect of age group \([F(3,32)=6.44, p<.003, \eta^2=.377]\), with the oldest two age groups having the highest overall slope values (4.55 for 3 yr. old, 5.39 for 4 yr. old, 5.97 for 5 yr. old, and 5.67 for adults). A significant place by age group interaction \([F(9,96)=2.99, p<.004, \eta^2=.219]\) revealed that the two oldest age groups exhibited the highest relative slope values for all places of fricative articulation, except inter-dentals (\(/\theta/\)). These main and interaction effects are displayed in Figure 7.19.
In addition, a significant effect of vowel context \( F(2,64)=5.89, p<.005, \eta^2=.155 \) indicated that the overall spectral slope of the fricative tokens varied as a function of the articulation of the following vowel (the mean slope for the /i, a, and u/ contexts were 5.25, 5.38, and 5.55, respectively). Although the overall main effect of vowel context was significant, post hoc tests indicated that no pairwise comparisons yielded significant differences. For the dependent measure of spectral slope, no other main or interaction effects were found to be significant.

### 7.2.3 Spectral Mean

A one-way ANOVA revealed that the overall spectral mean did not differ significantly as a function of window location. Individual four-way ANOVA’s (gender x
age group x place x vowel) revealed that each window location distinguished only two
places of articulation in terms of spectral mean, as shown in Figure 7.20. The only
significant differences in spectral mean were between /s/ and the other fricative types.

As can be seen in Figure 7.21, significant differences (p<.001) in age group were
found for the onset and offset analysis windows, but not for the window analyzing the
middle 40 ms of the fricative. The lack of statistical difference between age groups in the
middle window is somewhat surprising, considering the anatomical differences in vocal
tract size between the youngest speakers and the adults. However, one possible
explanation for the lack of significant difference in spectral mean for the middle window
may be developmental, rather than anatomical in nature. The results for this window
(mid location) will be reported in more detail in the following discussion.

![Figure 7.20 Spectral mean as a function of place of articulation and window location for
VCV fricatives.](image)
A four-way repeated measures ANOVA (place x vowel x gender x age group) conducted on the middle 40 ms of the fricative revealed a significant main effect for place of articulation \([F(3,96)=31.45, p<.001, \eta^2=.496]\). As shown in Figure 7.22, post-hoc analyses indicated significant differences \((p<.001)\) in spectral mean between /s/ and the other fricative types, but no statistical difference was found between the other fricatives. Collapsed across speaker and vowel context, the fricative consonants (/f, θ, s, ʃ/) exhibited spectral means of 5176, 5447, 6710, and 5722 Hz, respectively.
There was also significant place by gender interaction effect \([F(3,96)=7.85, p<.001, \eta^2=.197]\). As shown in Figure 7.23, this interaction was characterized by higher spectral means for male speakers on the three more anterior produced fricatives (/f, θ, s/) and a lower spectral mean for the more posterior /ʃ/ fricative. Similar to what was found with fricatives in the CV context, it is somewhat surprising to find that the male speakers were found to have elevated spectral means in the majority of the fricative types.

Overall, gender differences were found to be significant \([F(1,32)=4.66, p<.05]\), yet the effect size was relatively small \((\eta^2=.127)\). An examination of the gender differences across age groups indicated that most of the differences were within the 3 yr. old and 5 yr. old speakers, as can be seen in Figure 7.24. Although the gender differences in the adult group were not significant, it is still surprising to find adult male
Figure 7.23 Spectral mean as a function of place of articulation and gender for VCV fricatives.

Figure 7.24 Spectral mean as a function of age group and gender for VCV fricatives.
speakers with higher overall spectral means than the adult female speakers. Considering the differences in vocal tract size (male larger than female), elevated male spectral means are likely related to the manner of articulation, rather than any anatomical differences.

Examination of the adult gender differences across place of articulation revealed a possible explanation for the elevated male spectral mean values. As can be seen in Figure 7.25, in more posterior articulated fricatives (/s/ and /ʃ/), where the anatomical size of the vocal tract would have a greater effect on the spectral means, the adult male subjects were found to have lower mean values. The overall elevation of the mean values for the adult male speakers is derived primarily from a significant (p<.001) gender differences for the inter-dental fricatives. This fricative is produced with a primary constriction in the anterior portion of the vocal tract near a speaker’s dentition and would thereby be less sensitive to size differences in vocal tract anatomy.

Figure 7.25 Spectral mean as a function of gender and place of articulation for VCV fricatives.
In addition, the four-way ANOVA (place x vowel x gender x age group) also indicated a significant place by age group \([F(9,96)=8.95, p<.001, \eta^2=.456]\) interaction. As can be seen in Figure 7.26, an adult-like contrast between sibilant fricatives was characterized by the /ʃ/ fricative having a significantly lower spectral mean than that of /s/. This within-sibilant contrast was not evident in the youngest group of speakers (3 yr. old), and was only minimally present in 4 year old speakers. Although the 5 yr. old speakers exhibited a greater within-sibilant contrast than the younger speakers, their /s/ and /ʃ/ productions were still not as differentiated as the adults. These findings are similar to those found in previous research (Nittrouer, 1995), who concluded that even children at 7 years of age had not yet acquired adult-like gestures for sibilant fricatives.

![Figure 7.26 Spectral mean as a function of place of articulation and age for VCV fricatives.](image)

Figure 7.26 Spectral mean as a function of place of articulation and age for VCV fricatives.
7.2.4 Spectral Variance

A one-way ANOVA revealed that the overall spectral variance differed significantly as a function of window location \([F(2,1437)=7.35, \ p<.002]\), with the middle window of analysis having lower overall spectral variance than both the onset and offset window locations. However, as can be seen in Figure 7.27, the relative pattern of differences across differing places of fricative articulation remained intact throughout the window locations. Individual four-way ANOVA’s (gender x age group x place x vowel) indicated that three of four places of articulation can be significantly (\(p<.01\)) differentiated at all windows locations, and four places at the middle window location.

<table>
<thead>
<tr>
<th>Window Location</th>
<th>Spectral Variance (in MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset</td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td></td>
</tr>
<tr>
<td>Offset</td>
<td></td>
</tr>
</tbody>
</table>

Figure 7.27  Spectral variance as a function of place of articulation and window location for VCV fricatives.
As mentioned previously, a significant main effect of place \(F(3,96)=69.41, \ p<.001, \eta^2=.684\) was obtained for spectral variance. Subsequent pairwise comparisons indicated that all four fricative types (/f, θ, s, j/) were significantly (p<.001) different in terms of variance, with means of 6.26, 5.11, 2.35, and 3.13 MHz, respectively. Interestingly, when collapsed into two groups according to sibilance, the measure of spectral variance was found to significantly (p<.001) differentiate between sibilant and nonsibilant fricatives. These differences are further illustrated in Figure 7.28. No other effects or interactions were found to be significant for spectral variance.

![Figure 7.28](image)

Figure 7.28  Spectral variance as a function of place of articulation for VCV fricatives.
7.2.5 Spectral Skewness

A one-way ANOVA revealed that the mean spectral skewness averaged across speakers, place, and vowel context differed significantly as a function of window location \[ F(2,1437)=7.01, p<.002 \]. Through a series of four-way ANOVAs (gender x age x place x vowel), it was found that the pattern of spectral skewness within different places of articulation remained similar across window locations. As can be seen in Figure 7.29, the only fricative to have a significant \( p<.001 \) difference in skewness was the palato-velar fricative. Age differences across window locations are displayed in Figure 7.30. As can be seen, the significant differences \( p<.001 \) between the child speakers and the adults is maintained throughout the various window locations. These age group differences will be reported and discussed later in this section.

![Figure 7.29](image-url)  
Figure 7.29  Spectral skewness as a function of place of articulation and window location for VCV fricatives.
As previously mentioned, analyses of data derived from the middle window location indicated a strongly significant main effect of place \([F(3,96)=304.03, p<.001, \eta^2=.905]\) for spectral skewness. As can be seen from Figure 7.31, the main effect of place was due to the significantly \((p<.001)\) elevated skewness of the palato-velar fricatives. This conclusion was confirmed through subsequent pairwise comparisons.

The mean skewness values for /f, θ, s, ʃ/ were -2.16, -2.21, -1.88, and 0.22, respectively. An overall main effect of gender was not found to be significant. However, there was a significant place by gender interaction effect \([F(3,96)=3.46, p<.05, \eta^2=.098]\). As shown in Figure 7.32, this interaction is characterized by a gender difference for alveolar fricatives (-1.64 for male speakers and -2.34 for female speakers), but not for any other places of fricative articulation.
Figure 7.31 Mean spectral skewness as a function of place of articulation for VCV fricatives.

Figure 7.32 Spectral skewness as a function of place of articulation and gender for VCV fricatives.
In addition, the four-way ANOVA (place x vowel x gender x age group) also indicated a strongly significant main effect of age group \([F(3,32)=25.86, p<.001, \eta^2=.708]\) and a significant place by age group \([F(9,96)=11.45, p<.001, \eta^2=.518]\) interaction effect. Post hoc tests indicated significant (\(p<.001\)) overall differences in spectral skewness between the child speakers and the adults, with no differences between the child groups found to be significant. This main effect of age group is further explained by examining the place by age group interaction. As can be seen in Figure 7.33, the differences in age group are found primarily in comparisons of the fricative /ʃ/.

The contrast in spectral skewness between /ʃ/ and the other fricative types became stronger as the age of the speaker increased. These results are similar to findings reported by Nittrouer (1995), who concluded that the development of the /ʃ/ fricative

![Figure 7.33](image-url)  
**Figure 7.33** Spectral skewness as a function of age group and place of articulation for VCV fricatives.
production continues to be modified toward an adult-like articulation pattern at least through age seven. These findings provide additional evidence to support Nittrouer’s conclusions regarding the spectral development of /ʃ/ fricatives. No other main or interaction effects were found to be significant for the dependent measure of spectral skewness.

7.2.6 Spectral Kurtosis

A one-way ANOVA revealed that the mean spectral kurtosis averaged across speakers, place, and vowel context differed significantly as a function of window location [F(2,1437)=9.47, p<.001]. A series of four-way ANOVAs (gender x age x place x vowel) indicated that two of four places of fricative articulation were significantly (p.<.05) differentiated at the onset and middle window locations. The measure of spectral kurtosis did not differentiate between fricatives in the offset analysis window, as shown in Figure 7.34. Age differences across window locations are displayed in Figure 7.35. As can be seen, the relative pattern of kurtosis values remained constant across age groups in the first two analysis windows, but varied widely as a function of age group in the offset window. In particular the adult speakers changed from having the lowest overall kurtosis values at the onset and middle of the fricative, to having the highest kurtic values for the last 40 ms of the fricative. Any differences in place or age group derived from the middle window location will be reported and discussed in more detail later in this section.
Figure 7.34 Spectral kurtosis as a function of place of articulation and window location for VCV fricatives.

Figure 7.35 Spectral kurtosis as a function of age and window location for VCV fricatives.
A four-way ANOVA, with gender and speaker age as the between subject factors and place of articulation and vowel context as within subject factors, revealed significant main effect of place of articulation \( [F(3,96)=19.04, p<.001, \eta^2=.373] \). As previously discussed, subsequent pairwise comparisons demonstrated that only two of the four places of fricative articulation were statistically different \( (p<.001) \) in terms of mean spectral kurtosis. These differences were due mainly to the decreased kurtic value of palato-velar fricatives. As shown in Figure 7.36, the mean spectral kurtosis values for the different fricative types (/f, θ, s, ʃ/) were 3.43, 3.88, 3.63, and 1.26, respectively.

![Spectral Kurtosis as a function of place of articulation for VCV fricatives.](image)

Figure 7.36 Spectral kurtosis as a function of place of articulation for VCV fricatives.
The ANOVA also yielded a significant place by age group interaction
\[ F(9,96)=8.11, p<.001, \eta^2=.432 \]. Although an overall main effect of age group was not
found to be significant, the interaction reveals a highly significant age difference (p<.001)
for /ʃ/ fricatives, as shown in Figure 7.37. For this place of fricative articulation, the
adult speakers have a higher value of kurtosis than the three child age groups. This
finding may indicate that with regard to kurtosis, the child speakers have not yet acquired
an adult-like /ʃ/ fricative, even at age five. Age differences for the fourth spectral
moment have not been reported in previous research (i.e., Nittrouer, 1995).

![Figure 7.37](image_url)  
Figure 7.37  Spectral kurtosis as a function of age group and place of articulation for
VCV fricatives.
7.3 Discriminant Analysis for VCV Fricative Data

When classifying fricatives in the VCV context, the acoustic parameters entered into the initial stepwise linear discriminant analysis were as follows: fricative duration, normalized fricative amplitude, initial and final vowel duration, spectral peak, spectral slope, spectral mean, spectral variance, spectral skewness, and spectral kurtosis. The spectral moments measures were derived from the middle 40 ms of the fricative. Several different discriminant functions or models were utilized to classify the fricative data based on the procedure (jack-knife or standard stepwise) and the data set on which the function was trained. A summary of the variables included in the various classification models can be found in table 6.1.

As shown in Table 7.1, when classification results are based on a cross-validation (jack-knife) procedure, 69.8% of the entire data set (fricatives in VCV context) were classified correctly in terms of place of articulation. The adults had a combined classification rating of 76.7%, while the 3, 4, and 5 year old speakers exhibited a rating of 69.2%, 69.2%, and 80.0%, respectively. Overall, differences between male (70.8%) and female (67.9%) speakers were minimal. However, an examination of classification rates across age groups indicated that gender differences within individual ages was more substantial, ranging from 11.7% in the 5 year old speakers to 16.7% in the youngest age group.

Trained only on data from the adult speakers, a model of classification was composed of the 4 spectral measures. In order of their relative contribution, the variables included in the classification model were spectral skewness, variance, mean, and slope. As can be seen in Table 7.1, the classification model based on the adult data were able to
correctly classify fricatives produced by the child speakers with an overall accuracy of 46.9%. Collapsed across fricatives and gender the rate of successful classification improved as the age of the child speakers increased. The mean classification rates for the 3, 4, and 5 year old age groups were found to be 38.3%, 48.3%, and 54.2%, respectively.

When trained on the adult male fricative productions, the measures of spectral slope, mean, skewness, and variance were found to significantly (F>3.84) improve the classification of the training data set (listed in order of contribution). A comprehensive listing of classification scores for the discriminant model trained on only the adult male data is shown in Table 7.2. This model was found to successfully categorize 71.7% of the training data by place of fricative articulation. When applied to the remaining data sets, it classified the fricatives with an overall accuracy of only 45.0%. This difference in rate was due primarily to a large decrease in accurate classification for tokens produced by the child age groups (the mean rates for the 3, 4, and 5 year olds were 33.3%, 43.3%, and 50.0 %, respectively). A breakdown by place of fricative articulation revealed that the discriminant function was more accurate in classifying the children’s nonsibilant fricatives (42.2% for /f/ and 75.6% for /θ/), compared to the decreased rates found for the sibilant fricatives (32.2% for /s/ and 18.8% for /ʃ/). The classification model failed to discriminate any of the 3 year old female productions of /s/ and /ʃ/. However, it is important to note that the classification rates for the sibilant fricatives increases with the age of the speaker. This trend in classification rates can be associated with the significant acoustic differences discussed previously in this section (i.e., spectral mean and skewness). These results support the theory that children’s sibilant fricative gestures are not as differentiated as adults and that a child’s articulation of these fricatives continues
to be modified toward an adult-like form as the child matures (Nittrouer, 1995). Excluding the adult male data, gender differences of 6.8% were found (48.9% for males and 42.1% for females).

A model was also developed on data from the adult female speakers. This model utilized only 3 acoustic measures to classify the training data (spectral skewness, spectral variance, and spectral mean; in order of contribution). As can be seen in Table 7.3, 85.0% of the fricative tokens in the original training set were classified correctly. Perfect classification (100%) was found for the sibilant fricatives, while the nonsibilant fricatives /f/ and /θ/ were categorized at the rate of 73.3% and 66.7%, respectively. When applied to the rest of the dataset, the model yielded an overall classification rate of 44.3%.

Similar to the model developed on adult male data, the lowered rate of this classification model was also due in part to the decreased rates found for the child productions of the sibilant fricative /ʃ/ (17.6%), yet the classification of the /s/ fricative was nearly perfect at 96.7%. The classification accuracy for the child productions of the nonsibilant fricatives /f/ and /θ/ were 24.4% and 27.8%, respectively. Interestingly, fricative data from the male children (48.3%) were classified by this function at a higher rate than from the female children (38.9%).
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Training Set</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self Classification</td>
<td>All Adults</td>
<td>Male Adults</td>
<td>Female Adults</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>—</td>
<td>79.2</td>
<td>71.7</td>
<td>85.0</td>
</tr>
<tr>
<td>3 Year Old Speakers</td>
<td>69.2</td>
<td>38.3</td>
<td>33.3</td>
<td>40.0</td>
</tr>
<tr>
<td>- male only</td>
<td>60.0</td>
<td>—</td>
<td>35.0</td>
<td>45.0</td>
</tr>
<tr>
<td>- female only</td>
<td>76.7</td>
<td>—</td>
<td>31.7</td>
<td>35.0</td>
</tr>
<tr>
<td>4 Year Old Speakers</td>
<td>69.2</td>
<td>48.3</td>
<td>43.3</td>
<td>43.3</td>
</tr>
<tr>
<td>- male only</td>
<td>78.3</td>
<td>—</td>
<td>48.3</td>
<td>41.7</td>
</tr>
<tr>
<td>- female only</td>
<td>66.7</td>
<td>—</td>
<td>38.3</td>
<td>45.0</td>
</tr>
<tr>
<td>5 Year Old Speakers</td>
<td>80.0</td>
<td>54.2</td>
<td>50.0</td>
<td>41.7</td>
</tr>
<tr>
<td>- male only</td>
<td>75.0</td>
<td>—</td>
<td>63.3</td>
<td>46.7</td>
</tr>
<tr>
<td>- female only</td>
<td>86.7</td>
<td>—</td>
<td>36.7</td>
<td>36.7</td>
</tr>
<tr>
<td>Adult Speakers</td>
<td>76.7</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>- male only</td>
<td>68.3</td>
<td>—</td>
<td>—</td>
<td>60.0</td>
</tr>
<tr>
<td>- female only</td>
<td>81.7</td>
<td>—</td>
<td>61.7</td>
<td>—</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td>69.8</td>
<td>46.9</td>
<td>45.0</td>
<td>44.3</td>
</tr>
<tr>
<td>- male only</td>
<td>70.8</td>
<td>53.9</td>
<td>48.9</td>
<td>48.3</td>
</tr>
<tr>
<td>- female only</td>
<td>67.9</td>
<td>40.0</td>
<td>42.1</td>
<td>38.9</td>
</tr>
</tbody>
</table>

Table 7.1 Results of discriminant classification using both cross-validation (jack-knife) and standard stepwise procedures for VCV fricatives; all values represent % correct classification.
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Trained on Adult Male Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/f/</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>73.3</td>
</tr>
<tr>
<td>3 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>46.7</td>
</tr>
<tr>
<td>- female only</td>
<td>53.3</td>
</tr>
<tr>
<td>4 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>53.3</td>
</tr>
<tr>
<td>- female only</td>
<td>40.0</td>
</tr>
<tr>
<td>5 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>46.7</td>
</tr>
<tr>
<td>- female only</td>
<td>50.0</td>
</tr>
<tr>
<td>Adult speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>30.0</td>
</tr>
<tr>
<td>- female only</td>
<td>33.3</td>
</tr>
<tr>
<td>5 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>26.7</td>
</tr>
<tr>
<td>- female only</td>
<td>66.7</td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td>45.7</td>
</tr>
<tr>
<td>- male only</td>
<td>44.4</td>
</tr>
<tr>
<td>- female only</td>
<td>46.7</td>
</tr>
</tbody>
</table>

Table 7.2 Results of stepwise linear discriminant analysis when the classification model is trained on adult male data only for VCV fricatives; all values represent % correct classification.
<table>
<thead>
<tr>
<th>Data Set Evaluated</th>
<th>Classification Model Trained on Adult Female Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/f/</td>
</tr>
<tr>
<td>Original Training Set</td>
<td>73.3</td>
</tr>
<tr>
<td>3 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>30.0</td>
</tr>
<tr>
<td>- female only</td>
<td>33.3</td>
</tr>
<tr>
<td>4 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>20.0</td>
</tr>
<tr>
<td>- female only</td>
<td>13.3</td>
</tr>
<tr>
<td>5 year old speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>23.3</td>
</tr>
<tr>
<td>- female only</td>
<td>40.0</td>
</tr>
<tr>
<td>Adult speakers</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td></td>
</tr>
<tr>
<td>- female only</td>
<td></td>
</tr>
<tr>
<td>Entire Data Set*</td>
<td></td>
</tr>
<tr>
<td>- male only</td>
<td>30.5</td>
</tr>
<tr>
<td>- female only</td>
<td>38.3</td>
</tr>
</tbody>
</table>

Table 7.3 Results of stepwise linear discriminant analysis when the classification model is trained on adult female data only for VCV fricatives; all values represent % correct classification.
7.4 Summary of Significant Results

Statistical analysis (ANOVA) conducted on fricatives produced in the VCV context determined that all of the acoustic measures except overall token duration varied significantly as a function of place of articulation. Similar to the fricatives produced in a CV context, only spectral variance distinguished all four places of articulation in the VCV fricative tokens. Significant differences between nonsibilant and sibilant fricatives were obtained for most of the remaining acoustic parameters.

Main effects for vowel context were found for the measures of overall token duration, vowel duration, fricative duration, normalized amplitude, and slope. When compared to the stop data, the acoustic structure of the fricatives was found to be less sensitive to the vowel context. As mentioned in the previous chapter, these vowel context effects are similar to previous research conducted with adult speakers (Jongman et al., 2000).

Significant effects of age were found for the spectral measures of peak location, slope, and skewness. Similar to results found with CV fricatives, significant place by age interaction effects were found for spectral peak, slope, mean, skewness, and kurtosis. These interaction effects were due mainly to an increased acoustic distinction between the sibilant fricatives /s/ and /ʃ/. In terms of these acoustic parameters, the distinction between the /s/ and /ʃ/ fricative contrast increased with age. These data support Nittrouer’s (1995) conclusion that for this particular sound contrast children often continue to develop their articulation toward an adult-like pattern of production. Gender differences were only noted for the measure of spectral peak location and mean.
In terms of place of articulation, a discriminant analysis based on a cross-validation procedure was able to correctly classify 69.8% of the stop consonant data. The adult data were found to have classification rating of 76.7%, while speech from the 3, 4, and 5 year old speakers exhibited ratings of 69.2%, 69.2%, and 80.0%, respectively. Across all speaker groups, gender differences were minimal (70.8% for male and 67.9% for female). Similar to results found in the CV context, when a discriminant function was trained only on the adult male data, the classification accuracy differed significantly as a function of both the fricative place of articulation and the age of the speaker. Speech data from the training set (adult male) was classified at an overall rate of 71.7%, while speech from the child speakers was correctly classified at the lower rate of 42.2%. This decrease in classification was primarily due to the inability of the discriminant function to categorize the sibilant fricatives of the child speakers (32.2% for /s/ and 18.8% for /ʃ/). As with the speech data from the CV context, the classification rates for the child speakers’ sibilant productions in the VCV context increased dramatically as the age of the speaker increased. Interestingly, the discriminant function developed on adult male fricative data was better at classifying the fricatives of male speakers than those produced by female speakers. This pattern of gender differences extended throughout all age groups and both phonetic contexts.
CHAPTER 8

GENERAL SUMMARY AND DISCUSSION

Using spectral moment analysis as well as more traditional types of analysis (i.e., amplitude and duration), this study examined the acoustic structure of voiceless obstruents produced by adults and typically developing children from 3 to 6 years of age. This research aimed to describe not only the acoustic structure of the obstruent tokens, but also determine if the measured acoustic parameters (durations, normalized amplitude, spectral peak, spectral slope, and the first four spectral moments) varied as a function of speaker or phonetic context. Furthermore, this study investigated how the combinations of derived acoustic parameters might be utilized to classify the obstruent tokens by place of articulation. This method of analysis may provide insight into possible gender or age related changes in patterns of obstruent production and how such patterns may relate to potential cues to place of articulation. These purposes were addressed through three research questions raised at the outset of this investigation. In the following discussion, the summarized results of each experimental question are presented and discussed according to obstruent type (stops and fricatives). Specific results for the stop and fricative tokens were obtained from the initial 20 ms of the stop burst and the middle 40 ms of the fricative, respectively.
1) Describe the acoustic structure of syllable initial and intervocalic voiceless obstruent productions of adults and typically developing children in terms of multiple acoustic parameters (durations, normalized amplitude, spectral peak location, spectral slope, and spectral moments).

Stop Consonants

In general, the acoustic parameters of the stop consonants were found to be similar to results presented in previous research (i.e., Forrest et al, 1988; 1990; 1994; Nittrouer, 1995). As expected, when compared to the other places of stop articulation, bilabial stops were found to have a relatively short VOT (79 ms for /p/; 91 ms for /t/; and 100 ms for /k/). In the VCV context, stop duration was also found to differ as a function of the stop articulation with /p, t, and k/ exhibiting durations of 139, 126, and 123 ms, respectively.

Collapsed across both phonetic contexts (CV and VCV), the mean spectral peak locations for the different stop articulations were found to be significantly different. Alveolar stops exhibited the highest mean peak values (6848 Hz), followed by bilabial (4336 Hz) and velar stops (3148 Hz), respectively. Bilabial and velar stops exhibited a falling spectral slope (-0.156 and -0.687), whereas, alveolar stops were found to have a rising spectrum (0.736).

In terms of spectral moment measures, bilabial stops were found to have a relatively low spectral mean (4578 Hz), a low skewness (-0.054), and a slightly decreasing value of kurtosis (-0.105). Alveolar stops were characterized by a relatively high spectral mean (5771 Hz), a strongly negative value of skewness (-0.703), and a
positive kurtosis (1.418). These findings of skewness and kurtosis are somewhat contrary to previous research, which found alveolar stops to exhibit relatively low values for each. Similar to results reported in previous research (i.e., Forrest et al., 1988; Nittrouer, 1995), the mean spectral energy of velar stops (3869 Hz) was found to be lower than other places of stop articulation. In addition, velar stops exhibited relatively higher values of skewness (0.956) and kurtosis (2.661). The spectral variance, a measure often not reported by other researchers, was found to be significantly higher in bilabial stops (6.8 MHz), followed by velar (5.1 MHz) and alveolar stops (4.9 MHz), respectively.

**Fricative Consonants**

Differences in duration between sibilant (224 ms) and nonsibilant fricatives (204 ms) were found to be significant in the VCV context, but not in the CV context, where differences in duration were due primarily to the decreased length of labiodental fricatives (182 ms). Fricative duration differences were sensitive to the articulation of the following vowel in both phonetic contexts. When collapsed across all places of articulation, the fricative duration decreased an average of 16 ms when preceding an /a/ vowel.

As expected, examination of the dependent measure of normalized amplitude (rms amplitude in dB of the entire fricative relative to the strongest component in the following vowel) revealed significant differences between the nonsibilant and sibilant fricatives. In both phonetic contexts, the normalized amplitude of the nonsibilant fricatives was significantly less than that of the sibilant fricatives. This finding is similar to previous research with adults (Jongman et al., 2000) and pilot research with children (Fox et al., 2002).
Collapsed across phonetic context, the mean spectral peak locations for (/f, θ, s, j/) were found to be 9336, 8896, 8899, and 5089 Hz, respectively. In general, the spectral peaks in this study were somewhat higher than findings of previous research (Jongman et al., 2000; Miccio, 1996; Tomiak, 1990) conducted with adults. However, this difference is understandable considering the young age and small vocal tract size of the majority of the speakers evaluated in this study. When only the adults are included in the analysis these average peak locations drop significantly. The fact that the spectral peaks in the youngest group of speakers often approached and sometimes exceeded 10 kHz raised an important issue with regard to methodology. Most of the previous spectral moment research (e.g., Forrest et al., 1988; Jongman et al., 2000; Tomiak, 1990) utilized a sampling rate of 20 to 25 kHz, often followed by low pass filtering at 10 or 11 kHz. As also noted by Miccio (1996) and Nittrouer (1995), it is possible that sampling and filtering at these rates may not account for a significant amount of higher spectral energy found in the speech of younger children. Thus, particular fricative contrasts (e.g., the /s/ vs. /ʃ/) may be artificially attenuated in children. By reevaluating one adult and one child at a higher sampling rate, Nittrouer determined that the possible effects a lowered sampling rate might have on her results were minimal. However, since only one child was reevaluating, it remains unclear whether a higher sampling rate for all subjects would have resulted in a statistically significant change in the study’s findings. To avoid these concerns, speech samples in the current study were sampled at a rate of 44.1 kHz.

The measure of spectral slope has not been reported in the majority of previous spectral moment literature (i.e., Jongman et al., 2000; Miccio, 1996; Nittrouer, 1995; Tomiak, 1990). Collapsed across phonetic context, the slope values in this study were
positive in nature and increased as the fricative articulation moved posterior in the oral cavity (3.41 for /f/, 3.46 for /θ/, 5.49 for /s/, and 9.09 for /ʃ/). Although the relative order of values were different than was reported in Fox & Nissen (2002), these results are similar in that the measure of spectral slope clearly differentiates non-sibilant and sibilant fricatives.

Alveolar fricatives were found to have the highest spectral mean value, with no significant differences found between the other fricative types. Collapsed across phonetic context, the fricative consonants (/f, θ, s, ʃ/) were characterized by spectral means of 5081, 5472, 6656, and 5717 Hz, respectively. In general, the values of these spectral means are similar to the findings reported by Jongman et al. (2000), with the exception of the elevated means found for the /ʃ/ fricative. This difference is due primarily to age differences, which will be discussed further in the following section. In both the CV and VCV contexts, the second spectral moment of variance was higher for non-sibilant fricatives (6.26 MHz for /f/ and 5.25 MHz for /θ/) than for sibilant fricatives (2.37 MHz for /s/ and 3.21 MHz for /ʃ/). These measures of variance are very similar to results previously reported by Jongman et al.

The spectral measures of skewness were almost identical across phonetic contexts, with the fricatives /f, θ, and s/ exhibiting negatively skewed energy distributions of -2.19, -2.20, and -1.88, respectively. The spectral distribution for the palato-velar fricative /ʃ/ was found to be skewed in a slightly positive direction (0.21). In terms of the fourth spectral moment (kurtosis), all fricative types (/f, θ, s, ʃ/) were found to have positive values of 3.60, 3.82, 3.58, and 1.36, respectively. Thus, the average
energy distributions were flattest for /ʃ/ fricatives. The range of these kurtosis values is higher than results obtained by Fox & Nissen (2002) and Nittouer (1995), yet very similar to those reported by Jongman et al. (2000).

2) Examine the systematic variation of obstruent production of both adult and typically developing children. More specifically, to what extent do the acoustic characteristics of voiceless stop and fricative productions change as a function of place of articulation, vowel context, age, and gender.

Stop Consonants

In both the CV and VCV contexts, the dependent measures of VOT, normalized amplitude, spectral peak, spectral slope, and all four spectral moments (including variance) were found to vary significantly across stop place of articulation. These results are similar to previous findings (i.e., Forrest et al., 1988; Nittouer, 1995), which indicate that the spectral moment measures of mean and skewness differentiate alveolar and velar stops. Results of this study not only support this finding, but further indicate that the first and third spectral moments also significantly separate bilabial stop consonants. In addition, subsequent post hoc analysis demonstrated that all three stops were significantly different from each other in terms of normalized amplitude, spectral peak, and spectral slope; measures not reported by Forrest et al. and Nittouer. Unlike Forrest et al., this study did find significant differences between stops in terms of the second spectral moment (variance). Subsequent post hoc analysis demonstrated that the measure of spectral variance was useful in distinguishing /p/ from /t, k/ across all age groups.

Results of this study further indicated that the acoustic structure of the stop burst varied as a function of the following vowel. Significant vowel context effects were found
for the measures of VOT, normalized amplitude, spectral peak, slope, and the first three spectral moments. Similar to findings presented by Nittrouer (1995), these vowel context effects were primarily the result of changes in the acoustic parameters of alveolar (/t/) and velar (/k/) stops when followed by a high front vowel (/i/). Nittrouer postulated that this effect was due to forward movement of the tongue body in anticipation of the high front articulation for the /i/ vowel. In theory, this anticipatory action would produce a shortening of the anterior resonating cavity and thereby result in an increase in the spectral mean of velar stops. Results from this study support Nittrouer’s conclusion in that the stops (/t, k/), which were most strongly affected by the /i/ vowel context, utilize the tongue as an active articulator. Likewise, a reduced effect was found for bilabial stops, which do not actively use the tongue as an articulator. This pattern of coarticulation was found across all age groups.

Further comparisons between the age groups indicated that the acoustic parameters of peak location and spectral mean differed significantly between the adult speakers and the three groups of child speakers, with no differences among the child groups found to be significant. This result was not surprising considering the relatively larger size of the adult oral cavity and the lack of significant anatomic differences between the 3, 4, and 5 year old speakers.

Interestingly, both age and gender differences were also found for the measures of spectral slope, mean, and skewness. Age by gender interaction effects indicated that strong gender differences in spectral slope, mean, and skewness began with the 5 year old speakers and extended to the adults, with male speakers showing a dramatic decrease in slope and mean, as well as an associated increase in skewness. These results are
interesting because anatomical differences in the oral cavity between male and female speakers are usually considered to be minimal in young children, which may suggest a possible difference in articulatory development in girls as opposed to boys starting at approximately 5 years of age. Similar patterns of age and gender differences were also reported in previous research (Fox et al., 2001) involving fricative tokens.

**Fricative Consonants**

In both phonetic contexts, main effects of place were found for the dependent measures of normalized amplitude, spectral peak, spectral slope, and all four spectral moments. However, only spectral variance was found to significantly distinguish all four places of fricative articulation. Interestingly, with the exception of Jongman et al. (2000) and Tomiak (1990), the measure of spectral variance has often not been analyzed or reported in the previous literature (e.g., Forrest et al., 1988; Miccio, 1996; Nittrouer, 1995). In general, these results are in agreement with the research of Jongman et al., who also found that spectral variance could be used to acoustically distinguish four places of fricative articulation.

Although only spectral variance distinguished all four fricative types, many of the derived acoustic parameters were found to successfully distinguish three out of four places of fricative articulation, in the CV context. The measures of spectral peak, slope, mean, and skewness were able to separate non-sibilant from sibilant fricatives, as well as between the two sibilant fricatives (/s/ and /ʃ/). However, these measures were not able to differentiate between the non-sibilant fricatives (/f/ and /θ/). Across age groups, these results are contrary to research indicating that the spectral peak location is able to distinguish between non-sibilant fricatives (Jongman et al., 2000). However, it was
found that the spectral mean values distinguished /s/ from the other fricative types, regardless of speaker age. As expected, strong differences between non-sibilant and sibilant fricatives were also found for the measures of fricative duration and normalized amplitude.

Across both phonetic contexts, results of this study indicate that the spectral structure of the fricative consonants was less sensitive than the stops to the articulation of the following vowel. Significant differences across vowel context were only found with the measures of fricative duration and normalized amplitude. These findings are similar to results presented by Jongman et al., 2000, but unlike the findings of Nittrouer (1995), which suggested that the magnitude of vowel context effects differed as a function of speaker age.

Of particular interest, significant place by age interaction effects were noted for the measures of spectral peak location, mean, skewness, and kurtosis. In each of these measures, the interaction effect was due mainly to a widening in the acoustic distinction between /s/ and /ʃ/ as the age of speaker increased. In terms of spectral mean, a contrast between /s/ and /ʃ/ was exhibited by the 5 year old speakers and further differentiated by the adults, however, this within-sibilant contrast was not evident in the younger groups of speakers (3 and 4 yr. old). Likewise, a greater distinction in spectral skewness was found between /s/ and /ʃ/ as the speakers increased in age. These findings suggest that the development of the sibilant contrast continues to be fine tuned throughout young childhood toward an adult-like stage of articulation.

This same general trend across age groups was also reported by Nittrouer (1995), who concluded that age-related acoustic differences were the product of
continuing articulatory development. Her results indicated that the difference in spectral mean values for /s/ and /ʃ/ were larger for adult speakers than for those of children ages 3 through 7. Nittrouer concluded that these differences were the product of development differences, indicating that even children at 7 years of age had not yet acquired adult-like gestures for /s/ and /ʃ/. In terms of spectral peak, mean, skewness, and kurtosis, the results of this study support these conclusions.

Significant main effects of gender were only found for the measure of spectral peak location. As expected, female speakers exhibited a slightly higher mean peak location than male speakers. Interestingly, analysis indicated that the gender differences previously found in adults (Jongman et al., 2000) and in older children and adolescents (Fox et al., 2001) were not present in the 3-5 year old speakers that participated in this study.

3) **Determine how successful various types of discriminant functions classify stop and fricative productions in terms of place of articulation.** Of further interest, is the extent to which classification models based on specific groups of speakers will successfully categorize the obstruents produced by other speaker groups of differing gender and age.

**Stop Consonants**

Discriminant analysis was utilized to classify stop consonants in terms of place of articulation. Collapsed across phonetic context, a discriminant function based on a cross-validation (jack-knife) procedure was able to correctly classify the entire set of stop consonants with an accuracy of 73.5%. A breakdown by place of stop articulation revealed that the discriminant function was moderately accurate in classifying alveolar
stops (77.9%), and somewhat less successful with bilabial (74.2%) and velar stop consonants (68.3%). Stops produced by the adults had a combined classification rating of 80.5%, while data from the 3, 4, and 5 year old speakers exhibited a lower rating of 70.0%, 70.5%, and 71.1%, respectively. Overall, differences between tokens produced by male (69.7%) and female (71.6%) speakers were minimal.

A discriminant function was also trained only on data from the adult speakers. This model of classification was developed on four acoustic variables. In order of their relative contribution to the classification, the variables were spectral skewness, final vowel duration, spectral mean, and spectral slope. Interestingly, this combination and ordering of acoustic variables was the same for both the CV and VCV data sets, indicating that the stops consonants were being produced by the adults in a relatively similar manner across the two linguistic contexts. In contrast, classification models based on data from child speakers varied widely across phonetic contexts in terms of model composition. This result suggests that child speakers exhibit greater context dependent variability in the acoustic structure of their stop productions than adult speakers.

Collapsed across phonetic context, classification models based only on adult data were able to discriminate 82.7% of the original training set and only 59.1% of the stop tokens produced by the child speakers. This difference in classification rate provides evidence for age-related differences in the acoustic characteristics of stops. However, unlike research conducted on the fricatives of older children (Fox & Nissen, 2002), discriminant functions trained on only male adult and only female adult data did not reveal systematic gender differences in the stop consonants of 3-5 year old child speakers.
Fricative Consonants

Discriminant analysis was also conducted on the fricative consonants to determine how well combinations of derived acoustic parameters could classify the fricatives in terms of place of articulation. When collapsed across the CV and VCV contexts, discriminant functions based on a cross-validation (jack-knife) procedure classified the entire set of fricative consonants with an accuracy of 69.5%. A breakdown by place of fricative articulation revealed that the discriminant functions were much more accurate in classifying sibilant fricatives (87.1%) than non-sibilant fricatives (53.5%). The mean classification rates for the four fricative types (/f, θ, s, ʃ/) were 59.2%, 47.9%, 84.6%, and 89.6%, respectively. Overall, fricatives produced by the adults had a combined classification rating of 70.8%, while data from the 3, 4, and 5 year old speakers exhibited an accuracy of 69.6%, 69.6%, and 78.7%, respectively. Differences between tokens produced by male (69.1%) and female (68.3%) speakers were minimal.

Results based on discriminant functions trained only on data from the adult male speakers indicated that the classification accuracy differed significantly as a function of both the fricative place of articulation and the age of the speaker. When collapsed across context, the combined classification rate for the sibilant fricatives in the training sets was 91.6%, while the nonsibilant fricatives were correctly identified 60.0% of time. Interestingly, errors in classification did not cross the sibilant/nonsibilant distinction. In other words, nonsibilant fricatives were only confused with each other. However, when applied to fricative data from the child speakers, the classification model categorized fricatives with an overall accuracy of only 41.8%. The rate of successful classification
improved as the age of the child speakers increased, with mean rates for the 3, 4, and 5 year olds being 30.8%, 41.6%, and 52.9%, respectively.

A breakdown by place of fricative articulation revealed that the discriminant functions were more accurate in classifying the children’s nonsibilant fricatives (58.9% for /f/ and 56.7% for /θ/), when compared to the low rates found for the sibilant fricatives (32.7% for /s/ and 18.8% for /ʃ/). In particular, the classification models were found to discriminate only 3.3% of the 3 year old female productions of /s/ and /ʃ/. However, it is important to note that the classification rates increased with the age of the speaker. The classification models correctly categorizing 45.8% of the 5 year old speaker’s sibilant fricatives. This trend in classification rates can be associated with the significant acoustic differences discussed previously in this section (i.e., spectral mean and skewness). These results support the theory that children’s sibilant fricative gestures are not as differentiated as adults and that a child’s articulation of these fricatives continues to be modified toward an adult-like form as the child matures (Nittrouer, 1995). Discriminant functions trained only on adult female data yielded similar patterns of results.

Excluding the adult male data on which the discriminant functions were developed, combined gender differences of 8.9% were found (49.7% for males and 40.8% for females). Interestingly, similar to results found in Fox & Nissen (2002), discriminant functions developed on adult male fricative data were better at classifying the fricatives of male speakers than those produced by female speakers. This pattern of gender differences extended throughout all age groups and across both phonetic contexts.
CHAPTER 9

CONCLUSION

The results of this study indicate that multiple acoustic parameters of voiceless stops and fricatives vary systematically as a function of place of articulation, vowel context, speaker age, and gender. In particular, it was found that the spectral peak location, slope, and the first three spectral moments were able to distinguish between differing places of articulation. It was further shown that gender differences for several acoustic characteristics can be found in children at a relatively young age. In addition, these data show that similar to previous research (i.e., Nittrouer, 1995), the acquired sibilant contrast between /s/ and /ʃ/ is less distinguished in children than adults. It was found that acoustic separation between the two sibilant fricatives widened as the age of the speakers increased, thereby suggesting that the contrast continues to be fine tuned throughout young childhood toward a more adult-like stage. Discriminant analysis revealed evidence that classification models based on adult male data were sensitive to gender related differences even in the youngest age group.
LIST OF REFERENCES


APPENDIX A

Experimental Stimuli

A. Initial position - Real words in the sentential context “a ______ ” (3 repetitions)

| i | peanut  | teapot  | key   | field | thief  | seal  | sheep |
| a | pocket  | Thomas  | car   | fox   | thought| sock  | shark |
| u | Pooh    | toothbrush | cougar | food-dish | “Thoot”| soup  | shoe  |

B. Intervocalic position - Nonsense words in citation form (3 repetitions)

| i | æpi | æti | æki | æfi | æθi | æṣi | æʃi |
| a | æpa | æta | æka | æfα | æθa | æsæ | æʃa |
| u | æpu | ætu | æku | æfu | æθu | æsɛ | æʃu |