INDIVIDUAL DIFFERENCES
IN SPEECH AND NON-SPEECH PERCEPTION
OF FREQUENCY AND DURATION

DISSERTATION

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

By

Matthew Joel Makashay, A.B., M.A.

* * * * *

The Ohio State University

2003

Dissertation Committee: Approved by

Dr. Keith Johnson, Adviser

Dr. Mary E. Beckman

Dr. Robert A. Fox

Adviser

Department of Linguistics
This dissertation investigates whether there are systematic individual differences in the perceptual weighting of frequency and duration speech cues for vowels and fricatives (and their non-speech analogues) among a dialectally homogeneous group of speakers. Many of the previous studies on individual differences have failed to control for the dialects of the subjects, which suggests that any individual differences that were found may be dialectal. Dialect production and perception tasks were included in this study to help ensure that subjects are not from dissimilar dialects. The main task for listeners was AX discrimination for four separate types of stimuli: sine wave vowels, narrowband fricatives, synthetic vowels, and synthetic fricatives. Vowel stimuli were based on the manipulation of duration and frequency of F1 for the vowels in “heed” and “hid”, while fricative stimuli were based on the manipulation of the fifth frequency centroid of the fricatives in “bath” and “bass”. Multidimensional scaling results indicate that there are subgroups within a dialect that attend to frequency and duration differently, and that not all listeners use these cues consistently across dissimilar phones. Results of this study will be relevant to the fields of perception, feature phonology, dialectology, and language change. If subgroups can have different perceptions of speech (but similar productions), this questions what is needed to classify dialect continua, and the ratios of these subgroups changing over time can explain some language mergers and shifts.
Dedicated to the memory of my grandmothers.
ACKNOWLEDGMENTS

I appreciate the support given by my advisor, Keith Johnson, and committee members, Mary Beckman and Rob Fox. Thank you for keeping my dissertation work on target. I also give thanks to Beverly Flanigan, Elizabeth Hume-O’Haire, and Ilse Lehiste for sharing their ideas and expertise with me, and to Bill Notz and Paul Cook for their generous assistance. And I do not forget the help I received from my fellow graduate students, Kiyoko Yoneyama, Pauline Welby, Georgios Tserdanelis, Martin Jansche, Tsan Huang, Janice Fon, Allison Blodgett, all the other labbies, and the members of the phonies and psychoids discussion groups.
VITA


1996 ........................................ M.A. Linguistics, Ohio State University, Columbus.

1995-present ................................ Graduate Teaching and Research Associate, Ohio State University.

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Linguistics

Studies in:
   Phonetics
   Speech Perception
# 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>iii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiv</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Theme</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Definitions</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Individual differences in perception</td>
<td>2</td>
</tr>
<tr>
<td>2. Previous research</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Somatosensation</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Gustation</td>
<td>10</td>
</tr>
<tr>
<td>2.3 Olfaction</td>
<td>11</td>
</tr>
<tr>
<td>2.4 Vision</td>
<td>12</td>
</tr>
<tr>
<td>2.4.1 Texture and color</td>
<td>13</td>
</tr>
<tr>
<td>2.4.2 Depth</td>
<td>14</td>
</tr>
<tr>
<td>2.4.3 Brightness</td>
<td>15</td>
</tr>
<tr>
<td>2.4.4 Spatial organization and visual search</td>
<td>16</td>
</tr>
<tr>
<td>2.4.5 Categorization</td>
<td>18</td>
</tr>
</tbody>
</table>
4.2.1 Talker ........................................... 92
4.2.2 Procedure for talker .......................... 93
4.3 Stimuli for perception experiments ............... 93
  4.3.1 Speech tokens ................................ 96
  4.3.2 Non-speech tokens ........................... 103
4.4 Listeners ......................................... 105
4.5 Procedure for listeners .......................... 105
  4.5.1 Questionnaires ................................ 106
  4.5.2 Familiarization ............................... 106
  4.5.3 Identification ............................... 106
  4.5.4 Discrimination ............................... 107
  4.5.5 Hearing test ................................ 108
  4.5.6 Forced choice identification ................. 108
  4.5.7 Commutation test ............................ 109

5. Results ............................................ 111

  5.1 Subject information ............................. 111
  5.2 Discrimination task ............................ 113
    5.2.1 Percent correct responses .................. 113
    5.2.2 Response RT ............................... 127
  5.3 Identification tasks ............................ 134
    5.3.1 Identification ............................ 134
    5.3.2 Forced choice identification ............... 138
  5.4 Group multidimensional scaling ................. 140
    5.4.1 Multidimensional scaling .................. 140
    5.4.2 Group MDS ................................ 141
    5.4.3 Sine wave vowels ........................... 144
    5.4.4 Narrowband fricatives ..................... 145
    5.4.5 Synthetic vowels ........................... 147
    5.4.6 Synthetic fricatives ....................... 148
    5.4.7 MDS summary ................................ 149
    5.4.8 Subject weights ........................... 149
  5.5 Subgroup multidimensional scaling ............... 150
    5.5.1 Sine wave vowels ........................... 158
    5.5.2 Narrowband fricatives ..................... 161
    5.5.3 Synthetic vowels ........................... 164
    5.5.4 Synthetic fricatives ....................... 167
  5.6 Possible explanations for perceptual differences 171
    5.6.1 Audition .................................... 171
    5.6.2 Commutation ............................... 173
    5.6.3 Production ................................. 176
5.6.4 Subdialects ........................................ 181
5.6.5 Music, language, and speech synthesis .......... 182
5.7 Language change .................................. 183

6. Discussion ........................................... 187
   6.1 Detection versus interpretation ..................... 187
   6.2 Identification .................................. 189
   6.3 Cue usage for segment and speech types ............ 190
   6.4 Cue usage consistency within subjects ............. 192
   6.5 Cue usage differences among subjects .............. 195
   6.6 Speech modality ................................ 197
   6.7 Feature phonology ................................ 203
   6.8 Sound change .................................. 204
       6.8.1 Sub-phonemic sound change in Columbus .... 205
       6.8.2 Prestigious listeners ......................... 206
   6.9 Conclusion .................................... 207

Appendices:
A. Questionnaires ..................................... 208
B. Vocoder trajectories ................................ 211
C. Ganong up-down solutions .......................... 238
D. Individual MDS solutions ......................... 248
Bibliography .......................................... 303
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Frequency and familiarity ratings for tokens</td>
<td>80</td>
</tr>
<tr>
<td>4.1</td>
<td>Cues involved in each continuum</td>
<td>94</td>
</tr>
<tr>
<td>4.2</td>
<td>Average vowel and fricative intensities in dB</td>
<td>98</td>
</tr>
<tr>
<td>4.3</td>
<td>Higher vowel formant frequencies in Hz</td>
<td>100</td>
</tr>
<tr>
<td>4.4</td>
<td>Lower fricative frequency poles in Hz</td>
<td>101</td>
</tr>
<tr>
<td>4.5</td>
<td>Commutation word groups</td>
<td>110</td>
</tr>
<tr>
<td>4.6</td>
<td>Vowel space and fricative tokens (with rhymes)</td>
<td>110</td>
</tr>
<tr>
<td>5.1</td>
<td>Subject background information</td>
<td>112</td>
</tr>
<tr>
<td>5.2</td>
<td>Subjects who had significant within-subject differences between their scores for frequency, duration, and both frequency and duration cues</td>
<td>125</td>
</tr>
<tr>
<td>5.3</td>
<td>Subjects who had significant within-subject differences between their scores for speech and non-speech</td>
<td>126</td>
</tr>
<tr>
<td>5.4</td>
<td>Subjects who had significant within-subject differences between their scores for vowels and fricatives</td>
<td>126</td>
</tr>
<tr>
<td>5.5</td>
<td>Subject description of non-speech stimuli</td>
<td>136</td>
</tr>
<tr>
<td>5.6</td>
<td>Subject description of speech stimuli</td>
<td>137</td>
</tr>
<tr>
<td>5.7</td>
<td>Duration values for vowels and fricatives</td>
<td>141</td>
</tr>
</tbody>
</table>
5.8 Average ERB values for different frequency measures of vowels and fricatives ........................................... 142
5.9 Average vowel and fricative loudness in sones ......................... 142
5.10 Correlations between factors for sine wave and synthetic vowels . . 143
5.11 Correlations between factors for narrowband and synthetic fricatives . 143
5.12 Stress and $R^2$ values for group MDS solutions ......................... 145
5.13 Ordering of individual MDS solution dimensions ....................... 154
5.14 Correlations of factors to individual MDS solution dimensions ........ 155
5.15 Duration importance for subjects ........................................ 157
5.16 Stress and $R^2$ values for subgroup MDS solutions on non-speech vowels 158
5.17 Stress and $R^2$ values for subgroup MDS solutions on non-speech fricatives161
5.18 Stress and $R^2$ values for subgroup MDS solutions on speech vowels . 165
5.19 Stress and $R^2$ values for subgroup MDS solutions on speech vowels . 168
5.20 Subject scores for the commutation test ................................ 174
5.21 Average F1 and duration values of the vowels in heed and hid for men and women ........................................ 176
5.22 Average F5 and duration values of the fricatives in bath and bass for men and women ........................................ 179
C.1 Subject 1’s (talker) Ganong up-down solutions .......................... 239
C.2 Subject 2’s Ganong up-down solutions ................................. 240
C.3 Subject 3’s Ganong up-down solutions ................................. 240
C.4 Subject 4’s Ganong up-down solutions ................................. 241
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Sample spectrograms of <em>heed</em> and <em>hid</em></td>
<td>80</td>
</tr>
<tr>
<td>3.2 Sample spectrograms of <em>bath</em> and <em>bass</em> at 44.1 kHz sampling rate</td>
<td>84</td>
</tr>
<tr>
<td>3.3 Sample spectrograms of <em>bath</em> and <em>bass</em> at 11.0 kHz sampling rate</td>
<td>84</td>
</tr>
<tr>
<td>3.4 Sample spectra at sampling rates of 22.1 and 11.0 kHz for [0] and [s]</td>
<td>85</td>
</tr>
<tr>
<td>3.5 Duration and average frequency values for <em>heed-hid</em> and <em>bath-bass</em> tokens</td>
<td>89</td>
</tr>
<tr>
<td>4.1 Average formant contours in Hz for [i] and [i]</td>
<td>95</td>
</tr>
<tr>
<td>4.2 Average formant contours in ERB for [i] and [i]</td>
<td>96</td>
</tr>
<tr>
<td>4.3 F1 contours for vowels in <em>heed-hid</em> tokens</td>
<td>99</td>
</tr>
<tr>
<td>4.4 Fifth noise band contours for fricatives in <em>bath-bass</em> tokens</td>
<td>100</td>
</tr>
<tr>
<td>4.5 Duration levels for the vowels of <em>heed-hid</em> and the fricatives of <em>bath-bass</em></td>
<td>102</td>
</tr>
<tr>
<td>4.6 Spectrograms of synthetic speech vowel tokens 41 (<em>hid</em>) and 13 (<em>heed</em>)</td>
<td>102</td>
</tr>
<tr>
<td>4.7 Spectrograms of synthetic speech fricative tokens 11 (<em>bath</em>) and 43 (<em>bass</em>)</td>
<td>103</td>
</tr>
<tr>
<td>4.8 Spectrograms of non-speech vowel tokens 41 (<em>hid</em>) and 13 (<em>heed</em>)</td>
<td>104</td>
</tr>
<tr>
<td>4.9 Spectrograms of non-speech fricative tokens 11 (<em>bath</em>) and 43 (<em>bass</em>)</td>
<td>104</td>
</tr>
<tr>
<td>5.1 Percent correct discrimination on pairs that differ only by frequency or duration</td>
<td>115</td>
</tr>
</tbody>
</table>
5.2 Percent correct values for the segment-type*order interaction . . . . 116
5.3 Percent correct values for different speech cues . . . . . . . . . . . 117
5.4 Percent correct values for non-speech and speech tokens . . . . . . 118
5.5 Percent correct values for vowels and fricatives . . . . . . . . . . . 118
5.6 Percent correct values for the segment-*speech-type interaction . . 119
5.7 Percent correct values for the cue-*speech-type interaction . . . . . 120
5.8 Percent correct values for the cue-*segment-type interaction . . . . 121
5.9 Percent correct values for the cue-*segment-*speech-type interaction . 122
5.10 Dendrogram showing subject clusters of similar responses . . . . . 125
5.11 Normalized RT for pairs that differ only by frequency or duration . . 128
5.12 Normalized RT for the speech-type*order interaction . . . . . . . . 129
5.13 Normalized RT for different speech cues . . . . . . . . . . . . . . . 130
5.14 Normalized RT for the cue-*speech-type interaction . . . . . . . . 131
5.15 Normalized RT for the cue-*segment-type interaction . . . . . . . 132
5.16 Normalized RT for the cue-*segment-*speech-type interaction . . . 133
5.17 MDS solution for sine wave vowels . . . . . . . . . . . . . . . . . 146
5.18 MDS solution for narrowband fricatives . . . . . . . . . . . . . . . 147
5.19 MDS solution for synthetic vowels . . . . . . . . . . . . . . . . . 148
5.20 MDS solution for synthetic fricatives . . . . . . . . . . . . . . . . . 149
5.21 Subject weights for frequency and duration dimensions . . . . . . 151
5.22 MDS solution for strong duration subgroup for sine wave vowels . . . 159
5.23 MDS solution for weak duration subgroup for sine wave vowels . . . 160
5.24 MDS solution for no duration subgroup for sine wave vowels . . . . 162
5.25 MDS solution for first dimension duration subgroup for narrowband fricatives .................................. 163
5.26 MDS solution for second dimension duration subgroup for narrowband fricatives .................................. 164
5.27 MDS solution for weak-to-no duration subgroup for narrowband fricatives 165
5.28 MDS solution for strong duration subgroup for synthetic vowels . . . 166
5.29 MDS solution for weak duration subgroup for synthetic vowels . . . . 167
5.30 MDS solution for no duration subgroup for synthetic vowels . . . . . 168
5.31 MDS solution for first dimension duration subgroup for synthetic fricatives ........................................ 169
5.32 MDS solution for second dimension duration subgroup for synthetic fricatives ........................................ 170
5.33 MDS solution for weak-to-no duration subgroup for synthetic fricatives 170
5.34 Audiograms for subjects 1–17 .................................................. 172
5.35 Subjects’ average F1 values of the vowels in heed and hid ............ 177
5.36 Subjects’ average duration of the vowels in heed and hid ............ 177
5.37 Subjects’ average F5 values of the fricatives in bath and bass ........ 178
5.38 Subjects’ average duration of the fricatives in bath and bass ........ 179
5.39 Michigan and Ohio values for duration of [i] and [i] ................. 184
5.40 Michigan and Ohio values for F1 frequency of [i] and [i] ............ 185
D.5 Subject 2’s sine wave vowels MDS solution .......................... 256
D.6 Subject 2’s narrowband fricatives MDS solution ..................... 257
D.7 Subject 2’s synthetic vowels MDS solution ............................ 257
D.8 Subject 2’s synthetic fricatives MDS solution .......................... 258
D.9 Subject 3’s sine wave vowels MDS solution ............................ 258
D.10 Subject 3’s narrowband fricatives MDS solution ...................... 259
D.11 Subject 3’s synthetic vowels MDS solution ............................. 259
D.12 Subject 3’s synthetic fricatives MDS solution .......................... 260
D.13 Subject 4’s sine wave vowels MDS solution ............................ 261
D.14 Subject 4’s narrowband fricatives MDS solution ...................... 262
D.15 Subject 4’s synthetic vowels MDS solution ............................. 262
D.16 Subject 4’s synthetic fricatives MDS solution .......................... 263
D.17 Subject 5’s sine wave vowels MDS solution ............................ 264
D.18 Subject 5’s narrowband fricatives MDS solution ...................... 265
D.19 Subject 5’s synthetic vowels MDS solution ............................. 265
D.20 Subject 5’s synthetic fricatives MDS solution .......................... 266
D.21 Subject 6’s sine wave vowels MDS solution ............................ 267
D.22 Subject 6’s narrowband fricatives MDS solution ...................... 268
D.23 Subject 6’s synthetic vowels MDS solution ............................. 268
D.24 Subject 6’s synthetic fricatives MDS solution .......................... 269
D.66 Subject 17’s narrowband fricatives MDS solution . . . . . . . . . . . . 301
D.67 Subject 17’s synthetic vowels MDS solution . . . . . . . . . . . . . . 302
D.68 Subject 17’s synthetic fricatives MDS solution . . . . . . . . . . . . . 302
CHAPTER 1

INTRODUCTION

1.1 Theme

This dissertation investigates individual differences in speech perception, how they relate to general auditory differences, what their possible causes are, and how the existence of perceptual differences can affect language. Whether or not the cause of a difference in perception is at the sensory level (e.g., one person has a poorer ability in detecting higher sound frequency cues than another) or at the cognitive level (e.g., one person depends less on high frequency sound cues in speech perception than another), the result can be the same (e.g., one person is poorer than another at distinguishing phones with differences in high frequency energy, such as [s] and [ʃ]). If a cue cannot be detected, it cannot be processed, and if a cue is detected but ignored, it still is not processed “correctly”. While my goal is to find individual or sub-dialectal differences in perceptual cognitive processing, the cause can be at the detection or interpretation level. Perhaps the individual who does not utilize higher frequency information can still perceive similar [s]-[ʃ] distinctions as another does if there is some sort of redundant cue substitution, such as relying more on lower frequency transitions to and from these fricatives. But what are the implications for
the definition of dialect (and language) if people have different perceptual processing of speech, yet still arrive at the same output?

1.2 Definitions

I will define some of the terminology I will be using, since the usage of these terms differs across disciplines. By *perception*, I mean the entire process that involves the receiving of a stimulus (including a complex stimulus which could possibly be broken into parts—e.g., a vowel has multiple formants that vary over time). This begins with the *detection* of a stimulus by one (or more) of the sense organs, and ends with the *interpretation* of the stimulus, which could be a simple identification of the stimulus, or incorporation of the information into some ongoing task such as understanding an utterance. Differences in *detection* ability are caused by differences in the *biology* (physiological and genetic) of the sense organs, while differences in *interpretation* are caused by differences in *cognition*. Of course, the brain is biological, but I make this distinction since cognitive processes can be learned or modified to a much greater extent than what adaptations can be made for one’s cochlea.

1.3 Individual differences in perception

Language researchers seek out “truths” just as researchers do in other fields. But in the perceptual world, people differ too greatly, even though many factors can be controlled (e.g., age, sex, dialect, and auditory profiles). It is difficult to find absolute values in speech perception that are the same for all listeners (e.g., a claim such as “all American English speakers judge a vowel with an F1 of 280 Hz and an F2 of 2250 Hz to be an [i]” would be disproved if one person heard it as [i]). So
subjects are assumed to have the same perceptual processing, and averages are taken, removing the variability that may have been caused by experimental error. But even if the standard deviation is given, how can one be sure that it is a random distribution without checking? And standard deviations of the group do not distinguish similar variance within each subject from variance among subjects.

Individual differences in perception have been found in all modalities (including somatosensation, gustation, olfaction, vision, and audition). The evidence suggests that both variation in physiology of the sense organs and differences in cognitive processing strategies are the sources. Dissimilarities in processing have been proposed to be based on individuals’ cue weightings (placing different amounts of importance on different components of the input signal) and their ability to suppress irrelevant input. The study of individual differences in signal detection theory (e.g., Ingham, 1970) indicates that both sensory and cognitive factors are required for explanation, since subjects may be different due to differences in sensitivity to the signal and due to differences in response bias.

One of the important consequences of individual differences is that the average response found for subjects is misleading if there are sub-groups within the population studied. Hirsh and Watson (1996) discuss how the “problem” of individual differences in auditory perception has been dealt with in the past:

When small numbers of subjects are used in a profile experiment, it appears that the only way to collect reasonably homogeneous data is to pretest larger groups of listeners and reject those who cannot, without lengthy training, discriminate on the basis of small differences in spectral shape. (p. 477)

And what occurs if the offending subjects are not weeded out:

In some cases, as in individual differences in the ability to hear out the details of complex auditory patterns, the range of thresholds in
discrimination performance can be as large as or even larger than the effects of the primary relevant stimulus parameters. (p. 476)

Baker, Hogan, and Rozsypal (1988) mention how subjects have been treated as “stimulus-driven organism[s]” (p. 402), and that:

[T]he possible causative or explanatory role of subject-based variables has been either largely ignored, or simply treated as “nuisance” parameters, to be controlled if possible (usually by a priori grouping), but rarely to be taken seriously as a basis for explaining the results. A noteworthy exception to this is found in studies of pathological groups as opposed to “normals”. (p. 402)

The issue of individual differences in perception (besides slight variation about a mean) has been brought up for many modalities. However, most of these effects are explained by different ages, genders, or genetic sub-populations instead of different modes of processing. Some fields explicitly acknowledge these sensory or genetic differences, such as X-chromosome color perception differences or heredity factors in taste, but do not usually go so far as to claim that individuals have slightly different psychological processes.

Some researchers directly state how ignoring individual differences will affect the validity of any generalizations made by a study. D. A. Stevens (1996) discusses how individual differences in taste can be hidden in seemingly unimodal populations. Other researchers support the purpose of their studies by citing real world differences that their field’s theories have not taken into account. Burdach, Köster, and Kroeze (1985) discuss how individual differences are usually dealt with as part of the error variance in olfactory research. But practical experience indicates there must be differences in odor acuity, otherwise there would not be screening tests to find sensitive individuals in the perfume and food industries.
Westerman and Cribbin (1998) noted that in previous research of depth cue integration, approximately half of the subjects’ performance would best be described by an additive model, while the others’ was multiplicative. But these differences did not generate much attention.

In some cases, these differences in processing input can have drastic consequences. Research on visual motion, such as Berthelon, Mestre, Pottier, and Pons’s (1998) study on the perception of motion in a vehicle approaching another car coming at a right angle at an intersection, has shown that one group of individuals is more likely to crash than another.

While individual differences in speech perception probably would not create results this drastic (except in cases of slips of the ear, or perhaps some perceptual shibboleths) they could provide insight into general linguistic theories. Most of these theories (as do many theories which involve humans) are not based on individuals but are based on the ideal individual as defined by the group average. Of course they acknowledge differences, but mostly treat the variance as an error factor.

Hodgson and Miller (1996) found that in a within-category trading relation between the F1 onset frequency and gap duration needed to hear a stop in say-stay-

s[...]/t/day continua, some subjects have a greater trading relation than average, and others no trading if the F1 value was too high for compensation. They state the importance of finding the basis of individual differences, such as long-term exposure to certain acoustic distributions. However, only a handful of hearing and linguistic studies have set out to study individual differences (e.g., Simon & Fourcin, 1976; Steffen-Batogowa & Katulska, 1984; Sharf, Ohde, & Lehman, 1988; Hazan & Rosen,
Individual differences in speech perception need to be kept distinct from dialectal differences. Like languages, dialects may have speech cues that are not contrastive for outsiders, or have different categorical boundary positions, causing differences in production and perception (e.g., Ladefoged & Broadbent, 1957; Labov & Ash, 1997; Thomas, 2000; Makashay, 2001). Unfortunately, only some studies that report individual differences control for dialect (e.g., Fox, 1983). Thus, most research on individual differences in speech perception may actually be reporting dialectal differences.

Other perceptual questions involve intra-individual differences, investigating whether cue weights are constant within an individual for different stimuli. For example, if an individual places twice as much importance on vowel formant frequency as compared to vowel duration, will there be the same weighting for analogous cues for fricatives at a higher frequency range? Are some people more constant across different contrasts in their weightings than others?

In this dissertation, first there will be a review of representative previous research that have found individual differences in perception, concentrating on hearing and language perception. Next there will be a description of a perceptual experiment involving discrimination of [i] and [i] on duration and frequency of F1, and of [θ] and [s] on duration and frequency of the fricative band closest to the fifth vowel formant. Both speech-like and non-speech-like tokens were used. The percentage correct and reaction time data will be analyzed for individual differences, including the use of
multidimensional scaling. Finally, the results will be discussed with regards to various issues in language processing, such as the role of perception in feature phonology and language change, whether speech is special or not, and if there is more support for prototype or exemplar theories.
Numerous studies have mentioned the wide range of subjects' perceptual abilities, but most have treated individual differences as a type of error to mention in passing, or as a justification to remove data from analysis. However, a few studies do investigate the possibility that these differences may have underlying perceptual or psychological significance. What follows are examples of research that mention individual differences, and the various ways it is dealt with. Non-auditory modalities will just include brief summaries of important findings, while auditory and speech will have more in-depth analysis on how individual differences are treated procedurally and theoretically.

2.1 Somatosensation

Some studies that found individual differences in the sense of touch involved whole-body vibration sensation, vibrotactile discrimination and scaling, and body orientation. Since the whole-body vibration experiment was a psychophysical study it can arguably be classified as a study resulting in physiological differences, while the remaining studies required the integration of information and are more psychological in nature.
Oborne and Humphreys (1976) were interested in how the human body responds to vibration, and found two sub-groups sensitive to two different frequency ranges. They relate this finding to previous studies they cite which found either the first (Ashley, 1970; Shoenberger & Harris, 1971) or the second (Miwa, 1967; Jones & Saunders, 1972; Oborne & Clarke, 1974) average range of frequency sensitivities, and link the cause of this discrepancy to the number of subjects from each sub-group that were used in these studies. Thus, this study demonstrates the importance of individual or sub-group differences in finding the “average” human sensitivities to stimuli.

Two studies that involved tactile objects are the Cholewiak and Collins (1997) vibrotactile perception and the (Hollins, Bensmaïa, Karlof, & Young, 2000) multidimensional scaling of textures. Cholewiak and Collins examined the sensitivity of subjects’ fingers to vibrotactile patterns. They found a wide range of ability in identification, with four groups having significantly different performance levels. These results are not due to differences in sensitivity since there was no significant correlation between vibrotactile thresholds and performance. Thus, individual differences in vibrotactile pattern recognition are due to higher-level processing tasks. Hollins et al. gathered dissimilarity and adjective scalings for surfaces. Most subjects have adjective scales that are two dimensional, while a couple require a third dimension. There were two sets of scales for which subjects would use one adjective scaling or the other to a significant degree, but not both. This provides evidence that people do not place the same emphasis of usage on different cues.
Nyborg (1971) researched how tactile cues can be used to judge orientation. The benefit in having specific over diffuse information depended on the overall accuracy of the subject. Increasing the salience of cues is of little benefit to some subjects if little use is made of them.

These studies found that there can be differences both in detection thresholds of touch, as well as the differences in the ability to use tactile cues in interpreting patterns with one’s fingers or in realizing the orientation of one’s body.

### 2.2 Gustation

Research on taste differences include chemical threshold studies and multidimensional scaling studies on food flavor and enjoyment.

Literature on individual differences in bitter taste was surveyed by Mattes and Beauchamp (2000). One well known individual difference in taste is the genetic sensitivity to phenylthiocarbamide (PTC), first discovered by A. L. Fox in 1931. Originally believed to be an example of a dominant/recessive gene trait for tasters/non-tasters, research by Olsen et al. (1989) suggested that more than one gene is needed (for an interaction of PTC and general taste sensitivity) since some tasters are “super-tasters” and there are also sub-groups of non-tasters. Genetics is not the only factor, however, since a study by Morton et al. (1981) on monozygotic twins found that only 55% of the variance of PTC thresholds can be explained by inheritance. Environmental factors do not endorse psychological effects, since the differences could be explained by pathological or developmental effects. Another bitter chemical, propylthiouracil (PROP), correlates with the number of taste buds on the anterior tongue (Mistretta, 1984) supporting a physiological explanation of differences in taste.
Stevens (1996) reports of a previous study (Stevens & Lawless, 1981) which looked at how age changes people's perception of fruits and vegetables. Multidimensional scaling of similarity ratings by 18–25, 36–45, and 56–65 year olds were given vectors by multiple regressions using attribute ratings of the flavors. Sweetness and hedonics were highly correlated for the younger group, but were split apart into separate dimensions for the middle group. The older group did not have either as principle dimensions, using mostly different types of intensity of flavor. Assuming these groups are all from the same population, and each will age similarly as the previous group did, this provides evidence that the importance one gives to cues can change over time.

These studies’ results indicate that individual differences in taste can be due to genetic factors (one may have a receptor for a molecule or not) or psychological make-up (pleasure may be associated with sweetness or not).

2.3 Olfaction

Individual differences in smell seem to be due to different thresholds caused by physiology, but an EEG study suggests there may be differences in processing as well.

Burdach et al. (1985) found a wide range of thresholds for odor (and taste) detection for subjects in a four-alternative forced choice procedure. Stevens and O’Connell (1996) divided subjects into three groups depending on how they classify pemenone (PEM)—those who call it strong and offensive are osmic, mild smell raters are allosmic, and undetectable smell raters are anosmic. Subjects then smelled other odorants and categorized them for use in group multidimensional scalings. The
allosmic group had a complex six cluster group of odorants, while the osmic and anosmic groups both had four clusters, but of different classifications and memberships, including one of the anosmics’ clusters being “odorless”.

Schwartz (2000) examined EEG activity of high, medium, and low accuracy subjects in detecting the presence of a subthreshold level of asoamyl acetate. “High” meaning slightly above chance, and “low” meaning slightly below chance. Interestingly, all subjects gave similar low smell intensity and confidence ratings overall. The high accuracy subjects did have higher EEG activity in a more anterior location (a shift toward the frontal cortex) than the others, but had no more awareness than the others of the odor. However, the high scorers gave better confidence scores (and higher anterior EEG activity) on hits, and worse confidence scores (and higher posterior EEG activity) on misses than the other subjects. So, either the high accuracy performers had a lower threshold, or they were better able to process weak information.

So different people have different abilities at detecting certain odorants, and possibly different success rates at processing weak cues.

2.4 Vision

Individual differences in detection ability in vision have been found for texture and color. Individual cognitive differences have been found for depth perception, spatial organization and visual scanning tasks, and visual categorization. For brightness, both cognitive and detection differences have been discovered.
2.4.1 Texture and color

Sutter and Hwang (1999) studied the ability of proposed visual filters in early visual processing to segregate simple and complex horizontal, vertical, and diagonal line patterns of one or two rectangular figures. The task was to indicate whether the innermost rectangle inside the pattern was oriented horizontally or vertically after only seeing the pattern for 50 ms. Simple patterns were faster than complex, but there was wide variation on the subjects’ ability to process the complex patterns. Subjects displayed different difficulties depending on the number of elements, whether the lines were vertical and horizontal or diagonal, and the particular combination of the number of elements and the line type. The explanation for these results is that there are different widths in orientation bandwidths for different angles across subjects, so some individuals may do poorer at two or one element detection because of stimulation of competing detectors or detectors not receiving as much of the pattern.

Color blindness is a well known difference in perception, with the most common types resulting in the confusion of red and green. Since medium and long wavelength genes are linked to the X-chromosome, more males are affected. Neitz and Jacobs (1990) studied differences among color normals, since in an earlier study (Neitz & Jacobs, 1986) females performed differently from males, who were themselves in different groups depending on the spectral positions of medium and long wavelength cones. The current study had color normal males comparing red-green mixtures to determine which was redder or greener. The majority fell into two discrete groups, with an outlier group to each side. Their prediction is that males may have up to two
medium and/or two long wavelength cones. Estévez (1986) also noticed these differences in red and green cone sensitivities, but any differences in blue were obscured by instrumental error.

Webster and MacLeod (1988) looked at other well-known causes of differences in color perception, such as different amounts of lens and macular pigments (which absorb some short wavelength light). Observers performed a color-matching task by adjusting three fixed primary wavelengths until they matched the color and brightness of the test light. Webster and MacLeod found individual differences for macular and lens pigment densities, amount of rod intrusion on the cones, and sensitivities and photopigment densities of all three cones.

2.4.2 Depth

Kumar and Glaser (1991) investigated how depth perception of an object can be influenced by the depths of surrounding objects. Subjects judged which of two stereoscopic dots was closer on a computer screen, while a surrounding rectangular frame (or array of dots) was set at different “angles” with the right or left edge “closer” to the viewer. One subject was not influenced at all by the surrounding context, but the others were to varying degrees, indicating that depth perception is influenced by how one integrates visual cues.

How surface slant is interpreted from line drawings was studied by Cowie (1998). Subjects changed the angle of a probe to match a specified angle on a surface, for drawings of various figures. The subjects were grouped into four types of perceptual strategy—some were bicategorical, setting angles close to either horizontal...
or vertical; some used the whole array of intended angles, only resorting to bicategories for figures with poor cues; some used a smaller range of angles; and some had mixed strategies. This demonstrates that there is a wide range of ability in extracting depth cues from flat representations of three dimensional objects.

Westerman and Cribbin (1998) investigated binocular and monocular cues for both computer (texture and luminance cues) and real world (disparity, accommodation, luminance, and texture cues) images. Subjects had the task of moving one probe to the same depth as another. Accuracy in individual cue conditions (such as for luminance or for texture) was not predictive of accuracy in combined cue conditions (such as for both luminance and texture) which does not support an additive model for all subjects. Therefore, subjects with greater cue integration skills performed better in the complex conditions.

2.4.3 Brightness

Bowen, Sekuler, Owsley, and Markell (1981) and Bowen (1986) examined individual differences of cognitive processing in pulse brightness perception. Usually there is a monotonic increase in brightness with duration, but for simultaneous pulses there may be a temporal brightness enhancement, in which a medium length pulse seems brighter than one of longer or shorter duration. The studies found three types of subjects with different results in comparing the brightness of two pulses. Roughly a third of the subjects found a shorter pulse brighter than a longer pulse if they had simultaneous onset or offset, a third found shorter brighter than longer only for simultaneous offset (otherwise the longer pulse is brighter as expected), and a third always found the longest pulse brightest. There were no differences between groups
for flicker or motion sensitivity, so the brightness effect is caused by different cognitive strategies in using temporal information in comparing pulses of light.

Yaguchi, Kawada, Shioiri, and Miyake (1993), however, found physiological differences in brightness perception. Subjects with normal color vision matched the brightness of different colors, and performed another brightness comparison task using flicker photometry (in which the subject increases the speed of alternation of two colors until they no longer flicker, which the theory predicts only occurs if the colors are of equal luminance, with brightness information only coming from the luminance channel (brightness information from the rods) and not the chromatic channel (color information from the cones)). Some subjects have the same levels of brightness for both tasks, meaning they only use the luminance channel to determine brightness, while other subjects’ results indicate that they use both luminance and chromatic channel information to determine brightness.

### 2.4.4 Spatial organization and visual search

Research on visual motion, such as Berthelon et al.’s (1998) study on the perception of motion in a vehicle approaching another car coming at a right angle at an intersection, has shown that one group of individuals is more likely to crash than another. Field-dependent subjects are those who use global information while interpreting visual input, but they may focus on irrelevant details in the environment. Field-independent subjects are more accurate than field-dependent subjects in these tasks since they can concentrate on the relevant cues and separate them from the surrounding context. Field-dependent drivers are more likely to have accidents
than field-independent drivers, and this cannot be attributed to age, sex, driving experience, or the mileage driven.

Jones and Kennedy (1995) found individual differences for a letter search within a list. They also noted psychophysical differences in thresholds for dynamic visual acuity of the letter “C” or its reverse image moving across the screen, for apparent movement of a box back and forth (as opposed to two boxes alternating) due to different ISIs, for bistable stroboscopic motion over three locations for two boxes each moving one position to the other side (as opposed to two pairs of boxes alternating with an overlap of one box from each pair) due to different ISIs, and for critical flicker frequency of a box due to the amount of contrast between its light and dark phases.

Buchsbaum, Pfefferbaum, and Stillman (1972) investigated strategies used in comparing objects. Subjects were eye-tracked while comparing a standard square with a test square to determine if the new square was smaller or larger. Subjects maintain their tendency as to which square to look at first. They also maintain a strategy of comparison in which they either alternate fixations between squares, or fixate at a few positions within one square before moving to the other. Hogeboom and Leeuwen (1997) also found individual preferences for a search strategy. Subjects were to compare two jigsaw-type objects to determine if they would fit together, or would they leave a gap between them. Nearly all used mostly a serial search, and the non-serial searchers were affected by the global shape of the pieces (e.g., if the joined shape would be symmetrical). A second test with a greater number of subjects and more difficult jigsaw pieces revealed that non-serial searchers may begin in the middle of the pieces. Interestingly, more than half of the subjects used serial or non-serial
searches depending on the type of jigsaw piece, and the usage was not the same across subjects. Serial searchers may be slower, but they are more accurate, focus on the local information, and had no global symmetry bias. Overall, subjects maintain one type of search strategy, but certain subjects may switch depending on the perceived complexity of the pieces in a speed/accuracy trade-off.

2.4.5 Categorization

Parks, Wall, and Bastian (1969) replicated Cross et al.’s (1965) identification and discrimination tasks of circles with angles cut out. Subjects were trained to label black circles with 42° or 46° deleted as “bub”, and 50° and 54° as “gug”.

One subject had better discrimination across the boundary than within categories (as expected), one had the opposite effect, and the third had a flat discrimination contour. A subsequent experiment with more subjects resulted in an even larger array of unusual discrimination contours. This study demonstrates that teaching (or failing to teach) people arbitrary, unnatural categories leads to uninterpretable individual differences.

2.5 Audition

Individual differences have been found that are both physiological and cognitive for frequency, cognitive for frequency patterns, physiological for intensity, and cognitive for amplitude and frequency envelopes. Different cognitive abilities are used in determining voice quality and sound localization, and there are physiological differences among individuals’ ear canals, cochleae, and brainstems.

146° is roughly the angle formed on a clock face with the hour and minute hands at 12h 07.7m, and 50° is about 12h 08.3m.
2.5.1 Frequency

Both cognitive and psychophysical differences have been found in frequency detection. Gilkey (1987), Berg and Green (1990), Doherty and Lutfi (1999), Kidd (1993), Neff and Dethlefs (1995), and Wright and Saberi (1999) have found different weightings across frequencies, different abilities in cross-spectra comparison, and different strategies for detection. Variation in the cochlea and its processes have been investigated by Greenberg and Larkin (1968), Wright (1995), and Hicks and Bacon (1999).

Gilkey (1987) investigated critical bands for a tone with fixed or random intensity masked by short narrow-band or long wide-band noise for three subjects. Increasing narrow-band masker duration decreases the randomized intensity effect for one subject more than the other two. Increasing the bandwidth of the noise allows one subject to overcome the effects of the randomizing level, aids another, but does nothing for the third. Gilkey proposed that some subjects can use information in the temporal or spectral fringe to help in detecting the signal. A following experiment used 25 reproducible noise samples to mask a tone, and used the results from four subjects in spectral and temporal weighting functions. The maximum positive spectral weight is at the signal frequency, and all subjects have negative weights at the highest frequencies, but some have negative weights for a band of low frequencies. All subjects have a negative weight before the onset of the signal, but are more variable at offset. His interpretation is that subjects can compare information from different spectral and temporal regions, which can aid some in overcoming masking.

Berg and Green (1990) were interested in how spectral weights would be affected by level perturbations of spectral components in a task to detect the increase
in level of the middle frequency. Three screened “good profile listeners” gave similar estimated weight patterns (in which the peak is at the target center frequency), but were very different at certain frequency components. One subject had a large positive low frequency weight at one component, another had a large negative higher frequency weight, and the third had overall lower weights for low frequencies than high. Subjects whose weights did not sum to zero had poorer performance in detecting the added tone to the middle component.

This procedure was also used by Doherty and Lutfi (1999) to test 15 normal (and 15 hearing impaired) subjects, with more target tones, and with more spectral components over a wider frequency range. They found wide variation in weight estimates, sometimes with the peak not being the target frequency. Kidd (1993) investigated how adding more spectral components centered around the signal affected its threshold. Out of 12 normal subjects, a third had no effect of number of components on threshold, but the rest showed a wide variety of improvement with an increase in the number of nonsignal tones.

Neff and Dethlefs (1995) review Neff et al.’s studies involving masking experiments with maskers composed of random frequency, amplitude and phase sinusoids composed of 2 up to 100 components. Forty-nine (over all experiments) normal subjects were picked if a pretest indicated they were affected by informational masking. Listeners were ordered from lowest to highest thresholds for the masked signal for each condition, and the lowest and highest quartiles were compared. The lowest quartile masked threshold listeners (i.e., the best detectors) had their thresholds increase monotonically with the number of masker components, while the highest quartile increased their thresholds for maskers from 2 to 10 components and then
decreased from 10 to 100 components. The highest quartile subjects stated their
difficulty in detecting the signal, and appeared to use differences in signal level in
an intensity-based strategy, instead of the cross-spectra comparison strategy of the
lowest quartile. There was not much difference between these two sub-groups for a
broadband noise, so the higher threshold group is not just unable to contend with
masking.

Since maskers in real life are not fixed, Wright and Saberi (1999) tested 10
random maskers with 10 tones. Four subjects were tested on signal detection in
three conditions in random order: fixed (same masker used throughout the block,
random across blocks), random-by-trial (same masker for both stimuli in a trial, but
random across trials), and random-by-interval (pair of random maskers for the stimuli
in a trial—only a subset of possible masker pairings was used). Uncertainty of the
masker does make the signal harder to detect. However, for the random-by-interval
condition, it appeared that subjects were choosing particular maskers as being more
likely to contain the signal, as thresholds were fluctuating wildly between maskers. A
following experiment was designed to determine the detectability of the signal and the
detectability of the masker, using the same subjects, stimuli, and conditions. Except
for one subject, at low signal levels listeners based responses on a particular masker
instead of the signal. The favored masker in a pair differed across listeners, but it
usually was the one more effective in masking the signal. So while one subject kept
the expected strategy for finding the signal, others who found the task too difficult
changed their strategy.

Greenberg and Larkin (1968) performed a series of masking experiments in
which they would insert random non-signal frequency probes (which would converge
to the signal frequency over time) to determine how the detectability of the signal would be affected. For groups of four subjects, usually one or two subjects would have a decline in performance due to the probes. None of the subjects reported being aware of the probes when informed afterwards, which also supports the conclusion that subjects can have different frequency filters or that off-frequencies are more attenuated for some people indicating differences in critical-bands.

Wright (1995) investigated delay of signal onset after masker onset for detection in notched and unnotched noise. There was a short delay between onsets of a short masker and the signal (burst condition), and a short or long delay between onsets of a long masker and the signal (short and long delay conditions). Seven normal subjects were prescreened to determine the temporal magnitudes of their burst temporal effects. Two had small effects, and the rest had large. The results of the first experiment using only notched noise found the two subjects expected to have small temporal effects (STE) did so, but the expected large temporal effects group had subgroup differences. Two subjects had large temporal effects (LTE), while three were termed as having mixed temporal effects (MTE). STE and MTE were similarly better than LTE for short onset delay, while MTE and LTE were similarly worse than STE in the burst condition. The second experiment used unnotched maskers, with similar results of the previous experiment, except that STE and MTE were similar for all conditions. The researchers’ explanation is that different subjects may have different types of channel frequency effects, as notched noise would affect the across-channel process since its frequency components were remote from the signal, and unnotched noise would affect across- and within-channel processes since it would have frequency components both at and remote from the signal.
Hicks and Bacon (1999) were interested in the nonlinearity of the cochlea, especially how it transfers from being linear for the lower frequencies to being nonlinear for the higher. Using symmetrical and asymmetrical (not centered around the signal frequency) notched noise maskers, six normal hearing subjects detected frequencies from 375 to 3000 Hz. Subject ERB values were similar for low frequencies but varied widely for the higher frequencies. The slopes of the frequency functions varied as well, being similar for subjects at low frequencies, but considerably different at high frequencies. This suggests that subjects can have different frequency nonlinearities, though Hicks and Bacon cited research by Fagelson and Champlin (1997) who proposed it may be due to localized regions of damage in the cochlea.

2.5.2 Temporal

Cognitive differences in temporal frequency patterns have been discovered by Ueda and Ohgushi (1987) and Deutsch (1995) while experimenting with Shepard’s (1964) complex octave tones that perceptually have endlessly rising (or falling) pitch. The tones in different pitch classes (e.g., C, C♯, D, D♯, E, F, F♯, G, G♯, A, A♯, and B) can have their fundamental frequencies doubled (or halved) and would still be considered to be in the same pitch class, but would be in the next higher (or lower) octave. For example, a tone at 440 Hz and a tone at 880 Hz are both considered A, but are separated by an octave. Shepard made complex tones (with a bell-shaped spectral envelope) that had frequencies from only one pitch class (e.g., C) from ten sequential octaves. Playing these in order (say, C up to B, and then repeating) would give the illusion that the pitch is always rising, since the frequencies in each complex
are higher than the previous, except at the spectral edges where low frequencies are added and high ones are dropped.

Ueda and Ohgushi (1987) use Shepard’s tones to test how subjects weight tone height (pitch) and chroma (octave). They found through multidimensional scaling that some subjects do weight one or the other more. Deutsch (1995) first gives examples of auditory illusions, such as a passage in Tchaikovsky’s *Sixth Symphony (The Pathétique)* in which two violin parts alternate the theme and also a second voice. But, it sounds as if one violin plays the theme and the other the accompaniment due to pitch proximity grouping. Deutsch investigated the tritone (half-octave) paradox, which is similar to Shepard’s method, but only uses two sequential tones “half-way across” an octave. She found that some subjects hear it as falling and others as rising. Using two sequences of tritones (i.e., two melodies), different subjects would hear one or the other melody as higher. Changing to different keys would cause the subjects to switch the order consistently. Deutsch found a correspondence between the octave band and a subject’s pitch range, but there were also dialect differences, and proposed that this may be due to different populations’ uses of F0 in gender identity).

2.5.3 Duration

Künnapas, Hallsten, and Söderberg (1973) investigated subjects’ estimation of duration (and intensity) of a 912 Hz sinusoid (and a vibrotactile stimulation). There appeared to be three main groups for duration estimation depending on the subjects’ power function relating stimulus and perception: high, middle, and low exponents.
2.5.4 Intensity

Differences in intensity judgments have been found as well. Popper, Parker, and Galanter (1986) and Schneider (1980, 1988) have found different loudness scales and contours across subjects, while McGuinness (1974) and J. H. Johnson, Turner, Zwislocki, and Margolis (1993) found different tolerance levels and JNDs.

Popper et al. (1986) had seven subjects perform three tasks on loudness estimation. Difference matching involved matching the level of a tone to indicate the difference in loudness between two other tones; difference estimation involved providing a positive real number to indicate this difference; and magnitude estimation involved providing a positive real number for the loudness of a tone. Three of the subjects used a loudness scale with a steeper slope for difference matching than for difference estimation.

Schneider investigated individual differences in loudness scales (Schneider, 1980), and in equal loudness contours of two frequencies (Schneider, 1988). Five subjects in Schneider (1980) compared the difference in intensity between two tones with that of another pair, indicating which pair has the greater loudness difference. One subject’s slope of perceived loudness was steeper than all others, and another less steep. Five subjects in Schneider (1988) compared the difference in intensity between two complex stimuli, having two frequency components across a critical band. Equal loudness contours across varied greatly across the subjects, indicating some judged the lower frequency as louder than the higher frequency at the same intensity, some the reverse, and some had more complex frequency relationships.
McGuinness (1974) tested 50 subjects for differences in loudness tolerances. Females have lower thresholds for higher frequencies, but have significantly lower tolerance levels than males for all frequencies. But more is needed than a sex effect to explain differences, since males were just as likely to have abnormally low loudness estimations as females were for high. J. H. Johnson et al. (1993) investigated the relationship between loudness and just noticeable differences for intensity differences over three subjects. Individual JND values were extremely different for certain intensity levels.

2.5.5 Complex sounds

By complex sounds, I mean stimuli in which more than one domain is being manipulated (such as frequency, intensity, and duration). Some cognitive research is done across domains (D. M. Johnson, Watson, & Jensen, 1987; Watson, 1987; Surprenant & Watson, 2001) or uses multidimensional scaling (Howard & Silverman, 1976; Christensen & Humes, 1996). Some investigate variance discrimination (Lutfi, Doherty, & Oh, 1996), pitch and timbre (Singh & Hirsh, 1992; Singh & Bregman, 1997), amplitude and frequency envelopes (Ueda & Akagi, 1990; Richards & Lentz, 1998), or the affects of age (Allen & Bond, 1997). Drennan and Watson (2001) examined how individual differences can be minimized through training.

D. M. Johnson et al. (1987) tested 14 subjects on a variety of auditory tests, including masked and quiet thresholds; tone intensity, frequency and duration discrimination; click train duration discrimination; gap duration discrimination; and tonal-sequence discrimination. They found different patterns of performance, including one subject within normal limits for intensity, frequency and order but below
average for duration, another subject normal for duration but below average on inten-
sity, frequency and order, and a third normal or better for all tests except frequency.
A second experiment on sixteen subjects was composed of the same tasks as before,
but added one for rhythm discrimination. One subject was within normal limits for
all tasks except for detecting quiet tones and intensity, another was normal except
for sequence discrimination, a third was poor only on intensity and frequency, and
a fourth was above average on all tasks but was poor for rhythm. High correlations
among similar tasks suggests independent dimensions of auditory capability.

A study which follows this multi-faceted perceptual testing methodology is
Surprenant and Watson’s (2001) testing of 93 normal-hearing subjects on non-speech
and speech tasks, including the discrimination of pitch, intensity, duration, rhythm,
embedded loudness, tone order, or nonsense syllable order; and the identification of
nonsense syllables; as well as the identification of stimuli presented in noise, including
consonants in /a/ syllables, words, and sentences. Subjects were grouped by their
overall discrimination performance into deciles, which showed large ranges of thresh-
olds for almost all of the individual tests. The speech identification tasks also showed
a wide range of performance, but a factor analysis separated speech from non-speech
performance. This cannot be used to support the claim that speech is distinct from
regular audition, since, as the authors point out, the non-speech tasks may be finding
differences that are too subtle to be linguistically relevant.

Watson (1987) summarizes experiments which investigated subjects’ abilities
to detect changes in frequency, intensity, or duration of complex tone patterns. There
were large differences in pattern discrimination abilities between the 10th and 90th
percentiles of 127 normal hearing subjects. The thresholds needed to detect a change
in pitch or intensity differed by a factor of 6, for duration a factor of 3, and for rhythm a factor of 4.

A multidimensional scaling analysis was performed by Howard and Silverman (1976) on 34 subjects for stimuli varied over fundamental frequency, waveform shape, and number and frequency of formants. A three dimensional solution with dimensions corresponding to fundamental, waveform, and having two high frequency formants or not. However, examining the subject weights reveals that while 21 subjects had greatest weights for the fundamental, four have greatest weights for the waveform, and the rest have low or balanced weights for all dimensions.

Christensen and Humes (1996) also used multidimensional scaling, varying between target and nontarget values of harmonicity, spectral shape, and amplitude envelope for 12 normal subjects. The weights of the subjects based on their similarity judgments indicated that six attended most to spectral shape, four to harmonics, one to amplitude, and one equally to spectral shape and harmonics. Ten of the subjects were then trained in distinguishing the all-target from the all-nontarget token for the following classification task on all tokens. Five subjects’ weights were highest for harmonicity, three for amplitude, and one for spectral shape. Two subjects use “majority rules” for the three cues, and the last had no decipherable pattern. Ten subjects were then trained on a small number of tokens to distinguish target from nontarget values of amplitude envelope. Testing on the remaining stimuli revealed that nine subjects had greater amplitude weights after training. These experiments demonstrate that people can have different cue weighting strategies, but are affected by experience.
Lutfi et al. (1996) examined how five normal subjects determined which complex tone was composed of octave frequencies with a greater variance in intensity of the components. All subjects had one or two different frequency weights zeroed out, with predominate weights on 1 to 3 different components. Two subjects may also base their decisions on the maximum tone level.

Singh and Hirsh (1992) tested the effect of changes in spectral locus and F0 of harmonic complex tones on the perception of pitch and timbre. Stimuli were created with four sequential harmonic tone components (out of a series of nine harmonics), and with one of 13 possible F0 values. Six listeners (including two musicians) indicated whether a sequence of two stimuli had a rise, fall or no change in pitch, a change in “something else” (timbre), or both. Most subjects reported a change in pitch and timbre when only the timbre had changed, but one listener (a non-musician) was consistently able to detect a change in timbre as such and not as a change in F0. Another subject (a musician) would report a change in timbre also as a change in pitch, even if the F0 had a moderate change in the opposite direction. Singh and Hirsh note that this goes against conventional wisdom, as musicians are expected to distinguish pitch from timbre. Since the non-musician separated pitch from timbre the most and one of the musicians did so the least, they conclude that restraint should be used in using “musician” as a group identifier.

How harmonic structure and temporal envelope influence segregation of complex tone sequences was investigated by Singh and Bregman (1997). An ABA-“gallop” format was created under four conditions in which the tones A and B had the same or different number of harmonics and same or different rise-decay envelope. During presentation, the F0 difference between the A and B tones would increase from
zero until the subject indicated the pattern segregated into two streams. Listeners had different absolute \( \Delta F_0 \) values required for segregation, and while most listeners were not affected by the number of harmonics, some were able to use this contrast to segregate at a lower \( F_0 \) difference. As found previously (Singh & Hirsh, 1992), musical experience is not a reliable group identifier, and the authors suggest the type of experience should be taken into account.

Individual differences in the perception of amplitude and frequency envelopes were found by Ueda and Akagi (1990) and Richards and Lentz (1998). Ueda and Akagi (1990) investigated how the perception of sharpness (a timbre attribute) in broadband noise was influenced by duration, intensity, and amplitude envelope shape (rectangular, triangular with linear rise and abrupt cutoff, and triangular with abrupt onset and linear fall) for 13 subjects. Two were eliminated after multidimensional scaling because of high stress values, low \( r^2 \) values, and “obscure stimulus configurations” (p. 816). Nine subjects perceived more sharpness with an increase in duration, while two heard more sharpness with a decrease. Seven judged an envelope with linear fall as sharper than one with linear rise, while the other four heard little difference in quality between the two.

Richards and Lentz (1998) examined sensitivity to changes in level and envelope patterns across frequency. Seven subjects had to discriminate a standard flat magnitude of all frequencies (for different numbers of components) from a 1-step increase in the level of low frequencies over high frequencies and an up-down raising and lowering of every other frequency. One subject’s best thresholds were shifted toward the higher number of components for the 1-step intensity difference. Three normal hearing subjects were used in an experiment to get psychometric functions of level,
phase, and differences in level and phase across frequency. For the phase task, an in-phase set of sinusoidally amplitude modulated (SAM) tones was the standard, to be distinguished from a 1-step shift of low frequency carrier SAM tones to lag behind the high frequency and an up-down lagging and leading of every other SAM tone. The third subject seems to use a Euclidean combination of the two cues in the joint condition, while the other two use summation of cues at low levels, and a function between summation and Euclidean combination for high levels. Another experiment involved a level task (with in-phase tones or fixed modulator phases across frequencies) and a phase task (with equal levels or fixed different levels across frequencies). The third subject used the two cues independently, while the others were affected by phase on sensitivity to level.

Allen and Bond (1997) investigated how categorization abilities differ with age. Three normal-hearing age groups (10 six to seven year olds, 11 ten to eleven year olds, and 11 twenty-two to twenty-nine year olds) gave similarity judgments on complex sounds that vary in spectral and temporal dimensions as found in speech and non-speech. Hierarchical clustering of weights revealed two groups: the older group with high weights for all dimensions included the adults, most of the 10-year-olds, and a couple of 7-year-olds; and the younger group with lower weights included most of the 7-year-olds and a few 10-year-olds. The younger listeners failed to make a distinction between tones and complex tones, and did not utilize temporal envelope distinctions. The older group gave greater weight to frequency, and had a more categorical organization of the sounds. So increased age or experience leads to increased categorization abilities.
One study that set out to investigate individual differences and whether they can be minimized through extended training is Drennan and Watson (2001). Forty-one listeners participated in a pure-tone intensity discrimination task, and one in which the sixth component of an eleven-component profile was modified. The discrimination performance (measured in $\Delta I$) of the subjects ranged from 0.3 to 3.0 dB for pure tones, and from 0.45 to 3.5 dB for profiles. A following experiment involved seven of the “good” and four of the “poor” profile listeners discriminating using the profiles of the previous experiment, but for 9000 trials to test for learning effects. Listeners appeared to approach different asymptotic levels (with differences between listeners as high as 15-20 dB), and although three of the “poor” listeners improved significantly, almost to the range of the “good” subjects, one subject did not seem to learn.

2.5.6 Localization

Using two ears to find the source of sound is a task involving multiple cues that can be used by individuals. McFadden, Jeffress, and Russell (1973) and Simon, Collins, Jampolsky, Morledge, and Yu (1994) examined interaural time and intensity differences in localizing sound, while Bernstein, Trahiotis, and Hyde (1998) and Lorenzi, Gatehouse, and Lever (1999) also investigated effects of noise. Lutfi and Wang (1999) researched acoustic cues used in discriminating motion, while Wightman and Kistler (1999) examined front-back ambiguity resolution.

McFadden et al. (1973) investigated individual differences in the use of two binaural cues for localizing sound sources: interaural time differences (the ear closer to the sound source detects the wave first) and interaural intensity differences (the
ear closer to the sound source detects greater intensity). They hypothesize from previous research that since some individuals are sensitive to time while others are sensitive to intensity, and that since they were unable to train subjects to use the dispreferred cue, then auditory differences may be the cause. Bandnoise signals were created to range continuously and incrementally from having only intensity cues, to equal intensity and time cues, to only time cues, to conflicting equal intensity and time cues, and back to only intensity cues. Seventy-six subjects (three dropped for uninterpretable results) would indicate which side the auditory image is from. The cross-over point is the location in the range of stimuli at which a subject responds “left” and “right” each 50% of the time. This is in the conflicting cue portion of the range, and subjects more sensitive to time switch further along in the range. Subjects were classified as being time or level sensitive if the difference in performance between time-only and intensity-only cues was greater than 10%. This resulted in 32 unclassified subjects, 10 time sensitive subjects, and 31 intensity sensitive subjects. Lowering the criterion lowered the number of unclassified subjects to 17, but did not change the ratio of time:intensity subjects (1:3). The cross-over points for the time group were significantly further along the range than the intensity group.

Simon et al. (1994) also investigated interaural time and intensity differences of narrow-band noise at various frequencies on two old and three young subjects with normal hearing. Subjects varied the interaural intensity differences of a “pointer” to make it match the lateralization of a target with interaural temporal differences. The older subjects were more variable, and they and one of the young subjects had intensity and frequency effects. The other two young subjects might be more time sensitive, or adjust their judgment scales for the current stimuli. Bernstein et al.
(1998) found that the range of nineteen individual mean binaural masking level differences can be as large as 18 dB, and that five of the subjects needed twice as much delay as the rest to reach threshold in the interaural temporal differences condition.

Sound localization with compensation for noise was examined by Lorenzi et al. (1999). Pulse trains were presented at one of eleven speakers in a semi-circle around the front of the subject’s head from left ear to right. White noise could be present at the left, right, or center speaker. Four normal hearing subjects would indicate which speaker the target came from in a localization task, and whether or not they heard the signal in a detection task. Two of the subjects demonstrated pulling effects, as their location judgment for the signal moved toward the noise location. A third subject (and the fourth to some extent) displayed pushing effects, as the location for the signal moved away from the noise location.

Lutfi and Wang (1999) investigated the different cues that can be used for discriminating auditory movement. A standard tone with linear trajectory was created for headphones to move from left to right in front of the subject, while other tone linear trajectories were created with different starting points, velocity, or acceleration. Intensity, frequency (Doppler effect), and interaural time delay cues were manipulated so each would indicate a different velocity. Eleven normal subjects participated, but not for all conditions. The tasks were to indicate which of the two sounds was displaced farther to the right, which had greater velocity, or which had greater acceleration. In the displacement task, two subjects had greater weights for interaural time delay, while the other two had greater intensity weights. Three subjects preferred intensity cues for judging acceleration, while the other three depended on the Doppler effect. All six subjects used Doppler cues for velocity discrimination.
The front-back ambiguity in localization was examined by Wightman and Kistler (1999). White Gaussian noise was presented from 72 real sources, or 156 virtual sources via headphones using the subject’s own head-related transfer functions. Seven subjects judged the direction and distance of real and virtual sources with conditions for whether they were allowed to move their heads or not. Three listeners made large numbers of front-back confusions with no head movement, while two made very few. The tracking of subjects’ motions in the head movement condition showed that the trajectories were idiosyncratic, and that those who had the most front-back confusions made the largest head movements, which aided them in resolving the ambiguity. An examination of the head-related transfer functions could find no distinction between the good and poor locaters, suggesting that subject differences lie in ability and not in the acoustic cues shaped by their perceptual systems.

2.5.7 Head, pinna, ear canal, cochlea, and brainstem

Individual differences in hearing organs and brain auditory centers can also cause differences in perception. Hellström (1993) examined the relationship between the peak of sound transfer function (partially caused by ear canal length and cross-sectional area) and an individual’s temporary threshold shift due to exposure to noise, while Middlebrooks (1999) related directional transfer functions to pinna size and head width. Lauter and Loomis (1986, 1988) examined differences in auditory brainstem responses, and Don, Ponton, Eggermont, and Masuda (1994) examined sex differences for these responses, and how they might be caused by sex differences in the cochlea.
Hellström (1993) investigated if subjects with differences in the peak of sound transfer function (STF) have differences in temporary threshold shift (TTS). Previous research has noted differences in STF peak from free-field to tympanic membrane vary from 2 to 4 kHz, caused by differences in length or cross-sectional area of the ear canal, shape of its entrance, and acoustic impedance of the tympanic membrane. Ear canal length and volume has also been shown to influence frequency of peak TTS. Thirty-six normal hearing subjects were prescreened on STF—12 subjects with a higher STF value at 2 kHz composed the low-frequency group, 12 subjects with high values at 4 kHz made up the high-frequency group, and 12 with similar values at both frequency bands were the mid-frequency group. After exposure to narrow-bandpass white noise at the respective frequencies, audiograms were taken. TTS thresholds of low-frequency subjects were higher than mid-frequency which were in turn higher than high-frequency for low frequencies. The opposite pattern was found for high frequencies.

A study on directional transfer functions (DTFs) and individual differences in frequencies of peaks and notches was reported by Middlebrooks (1999). Thirty-three of the subjects in the study had physical dimensions measured in addition to having recordings taken from their ear canals of broadband sounds from various locations. Subjects’ DTFs have peaks and notches at different frequencies, and in certain instances one subject’s notch could be at another’s peak. The scaling of one subject’s DTF to another’s was substantially improved if pinna-cavity height and head width values were added to the equation. Middlebrooks calculates that, for example, at 8 kHz spectral feature differences from two random subjects would be perceptually distinct (greater than 1 ERB) half the time.
Lauter and Loomis (1986, 1988) measured the auditory brainstem responses (ABRs) to monaural right, monaural left, and binaural presentation of condensation clicks to seven normal hearing subjects. Within subject stability was greater than between for the five brainstem peaks they measured in terms of latency and amplitude. Individuals also had certain peaks that more stable in amplitude and latency across repetitions than other peaks, but these stable peaks were different for different subjects. Don et al. (1994) investigated whether sex differences in cochlear response times can explain sex differences in ABR. Rarefaction click stimuli with ipsilateral high-pass filtered noise at different frequencies were played to 23 females and 20 men with normal hearing. The females had larger amplitude and smaller I-V peak intervals in ABR than males. Compensating for cochlear response times significantly reduces the variability. Since female cochlear length is on average shorter than male, the smaller travel time for the wave may lead to better synchronization of cochlear output across all frequencies for women, and this may explain why their click-evoked auditory brainstem response waveforms are better defined with larger amplitudes.

2.5.8 Voice quality

Now a look into what features listeners find important in judging voice quality (but see Kreiman and Gerratt (1998) on the reliability of this type of rating). Multidimensional scaling studies have been performed for pathological voices heard by speech pathology students (Kempster, Kistler, & Hillenbrand, 1991) and for pathological and normal voices heard by speech pathologists and naïve listeners (Kreiman, Gerratt, Percoda, & Berke, 1992). Synthetic voice quality judged by experienced
listeners has been investigated as well (Klatt & Klatt, 1990), including MDS analysis (Hall, 2001).

Kempster et al. (1991) used 1 s of sustained /a/ from 30 women with dysphonia, and split them into two sets of 15 to lower the number of triadic comparisons. Twenty-five communication disorders graduate students listened to all triads from each set, choosing the voice least like the other two. Multidimensional scaling of the first set of voices revealed the first dimension had highest correlations with intensity and harmonics-to-noise ratio, the second dimension was correlated to F0, and the third was jitter. Nineteen listeners gave greatest weight to dimension 1 (D1), five to dimension 2 (D2), and one to dimension 3 (D3). Two subjects had low weights for all dimensions. For the second set of voices, D1 was intensity, D2 was log amplitude perturbation quotient and jitter, and D3 was F0. Twenty-one listeners had greatest weights for D1, two for D2, and two for D3. One subject had low weights for all dimensions. Listeners generally used the same dimensional ranking for both sets. So different people use different rankings of criteria in judging voice quality, but seem to use the same set of criteria.

Kreiman et al. (1992) used 1.67 s of /a/ spliced from 18 men with a variety of voice disorders as well as 18 normal male speakers. Eight speech pathologists and two otolaryngologists comprised the expert listening group and six untrained listeners comprised the naïve group. Multidimensional scaling was performed on dissimilarity scales of voice pairs for the four sets of data, and individual scaling results were compared to group scaling for each voice type. For experts listening to pathological voices, only three listeners approximated the group 3D solution (F0, partial period comparison, amplitude difference of the first two harmonics), and only one of the
three weighted them in this order. The rest of the experts only had two dimensions, and three of them had dimensions different from the group. The group solution for naïve listeners hearing pathological voices was 4D (F0, F0 and shimmer, F1 and jitter, amplitude difference of the first two harmonics and shimmer), but none of the individual solutions were like this (they were either three- or two-dimensional) and two of them had dimensions different from the group. For experts listening to normal voices, half had the same dimensions as the group (F0, shimmer and F3, uninterpreted), but one weighted them differently. The naïve group listening to normal voices had a 2D solution (F0, shimmer and F2/F1), but four individuals switched the dimensions, and two had different dimensions. The individual expert solutions were more varied than the naïve solutions, perhaps because the differing amounts of exposure to different types of pathologies. A further complication to the solutions is that even if two listeners have the same dimension, it does not have to be used in the same way. For example, some listeners used F0 linearly (high vs. low frequencies), while other used F0 abnormally (abnormal high or low frequency vs. normal range).

Klatt and Klatt (1990) varied the parameters of a synthetic female voice to improve the breathiness and make it more natural sounding. Five experienced listeners rated the breathiness, naturalness, and nasality of modified [a] tokens in comparison to a reference. An increase in the first harmonic amplitude induces sensation of breathiness for some, but others think it unnatural, and most think it nasal. An increase in tilt is breathy for some, but others think it unnatural or nasal. An increase in aspiration noise is breathy for most listeners. Klatt and Klatt used these different cue sensitivity results to set the parameters for breathiness to make it best
for all listeners: aspiration noise, tilt, longer open quotient, increased F1 and F2 bandwidths.

Another study on synthetic speech compared the output of different speech codecs in a multidimensional scaling task (Hall, 2001). Speech samples from a male and female talker were processed with different codecs and presented to sixteen experienced subjects for triadic comparison. For the male talker, older subjects were found to give less weight to the “naturalness” dimension and more weight to the “low-frequency” dimension than younger subjects.

2.6 Summary of individual differences across modalities

The common theme (besides finding physiological differences) is that different people use slightly different strategies in perceptual problem solving, but end up with similar answers. Differences in the amount of importance placed on certain types of information (cue weighting) and the ability to focus on the signal and suppress irrelevant noise (and the ability to distinguish the two) seem to be the major explanations of individual differences. Environmental factors (e.g., types of stimuli, frequency of occurrence) would play a large role in the development of a strategy. Besides using individual differences multidimensional scaling to compare people’s weights for dimensions, or hierarchical clustering to determine subgroups, little is done besides mentioning a few atypical subjects. People attend to different features for feeling textures (Hollins et al., 2000), have different levels of success in determining motion trajectories in their visual field (Berthelon et al., 1998) and even have different ways of hearing the same piece of music. What types of individual differences in speech perception are there to find?
2.7 Speech perception

First a sampling of differences in speech perception between dialects. This is to emphasize that some “individual” differences may be dialectal, as many studies do not report the subjects’ dialect(s). Next will follow studies that found individual differences in the perception of vowels, voicing, stop place, manner, fricatives, approximants, prosody, and sign language.

2.7.1 Dialectal differences

This is an overview of types of perceptual differences found by dialect. If speakers of a dialect can share certain perceptual properties due to their unique production, there may be sub-populations who share their own properties, instead of all differences being individual. Differences in vowel perception among listeners of different dialects have been investigated (Ladefoged & Broadbent, 1957; Fox, 1974; Janson, 1981; Janson & Schulman, 1983; Peeters, 1991; Miller & Grosjean, 1997; Niedzielski, 1998; Labov & Ash, 1997; Flanigan & Norris, 2000; Makashay, 2001), as well as consonant (Thomas, 2000) and accent (Sugito, 1978) differences.

One study that incorporated dialectal knowledge in vowel recognition using synthesized speech was Ladefoged and Broadbent (1957). Six synthetic versions of the sentence “Please say what this word is” were created by manipulating the first two formants, with results that sounded like different people with the same accent. What followed a repetition of each sentence carrier phrase was one of four words spoken by one talker—\textit{bit} /bɪt/, \textit{bet} /bɛt/, \textit{bat} /bæt/, and \textit{but} /bʌt/. It was found that the identification of the vowel in the target word was influenced by the vowel space constructed by the preceding carrier phrase. The authors also found an effect
of dialect, with different responses from subjects in separate socio-linguistic groups. After one carrier phrase, the R.P. English subjects were more likely to identify the test word as /bʌt/ and less likely to identify it as /bat/ than the Scottish subjects were. This was explained by the different vowel spaces of the dialects, as the token was based on the /ʌ/ of R.P., but is similar to the /a/ of the Scots vowel system.

Another study that explicitly looked at dialectal differences is one by Fox (1974) in which he conducted a cross-dialectal vowel perception experiment between Chicago and Norman, Oklahoma speakers. He finds differential trends in the perception of average formant values of /i, eɪ, a, u, ai, o, æ/ produced in a /t/ context produced by a speaker from each dialect, but nothing significant, leading him to suggest an experiment focusing on two groups with greater phonological differences.

Studies by Janson (1981) and Janson and Schulman (1983) find differences in perception for different dialects of Swedish. Janson (1981) played synthetic stimuli in a continuum from ett tag [etːɑːɡ] to ett tåg [etːoːɡ]. Stockholm subjects perceived more of the tokens as containing [oː] than Helsinki speakers of Swedish did. Janson and Schulman played synthetic stimuli in a continuum from sitt [sɪtː] to sett [setː] to sätt [sɛtː] to satt [satː]. Lycksele listeners, who produce a distinction between /e/ and /ɛ/, were slightly better in distinguishing a boundary between these tokens modeled on a Lycksele talker than Stockholm subjects, who do not produce that distinction.

Peeters (1991) examined subjects’ preferences for diphthongs in a cross-language (Dutch, English, and German) perception test. Synthetic continua that manipulated onset, transition, and offset duration were created for possible productions of each individual diphthong or long vowel, such as /ai/, /au/, /e/, and /o/. For each stimulus pair, subjects were to pick the diphthong they decide is better.
One subset of subjects included speakers of various dialects of East-Frisian, with the result that residents of different villages had different token preferences for /ai/ and /au/.

Miller and Grosjean (1997) investigated different dialects of French regarding whether vowel length was used for identification. Vowel continua were created based on standard French with synthetic endpoints côte /kot/ and cotte /kot/. Three continua were created with the same formant frequency values, but differentiated by short, medium, or long vowel length. The Swiss French listeners reported that the stimuli were good exemplars of the intended words, even though they were from another dialect. Swiss French listeners were found to be affected by duration in their answers, increasing the probability of choosing /o/ for longer durations, and /ɔ/ for shorter durations, while standard French listeners were not.

Niedzielski (1998) found that dialect identification could be manipulated by expectations. Subjects were to listen to a sentence spoken by a talker, and then pick from a set of resynthesized tokens the vowel which best matched the talker. When told that the talker was a fellow Detroiter, Detroit residents chose non-raised diphthongs that corresponded to General American English. When told the speaker was from Windsor, the subjects chose tokens with raised diphthongs that corresponded with the stereotypical Raised Canadian English. However, Detroit speakers were unaware that their English did not correspond to the General American English they claimed to speak. Not only does Detroit English employ Canadian Raising similar to that of their neighbors across the border, but Northern Cities Chain Shift has drastically shifted their vowels away from the system of General American English.
Two experiments on Birmingham speech with Birmingham, Chicago, and Philadelphia listeners were discussed by Labov and Ash (1997). In a forced choice identification task using a /k_d/ context elicited from Birmingham talkers in list format, Birmingham listeners usually performed better than their non-native counterparts. However, for certain words the natives did almost as poorly, such as for identifying /ɛ/ in ked [kɛd]. All three groups had about a 50% success rate for kide, but most of the correct judgments were on the diphthongal token, not the monophthongal ones. “The overwhelming majority of listeners, in Birmingham and elsewhere, failed to recognize the Birmingham production of kide as [kad] as the /kayd/ that it was intended for” (p. 532).² In the gating task with spontaneous speech, Birmingham subjects had a small advantage at the word level, or at the phrase level if the word was in a confusing phonetic context.

Flanigan and Norris (2000) tested Ohio students’ perception of southern Ohio pronunciation of words in isolation, phrases, and sentences. There are vowel mergers that have occurred or are occurring in some areas of central or southern Ohio. /ʌ/ has lowered to /a/ producing color~collar homophones in the south, and also color~collar~caller homophones for central Ohioans who have merged /a/ and /ɔ/. Monophthongization and backing have produced homophones such as fire~far and tower~tar in the south. /ɛ/ has tensed to /e/ in the south, producing special~spatial homophones, while /i/ has lax ed to /ɪ/ for some Ohioans, producing steel~still homophones. /u/ has lowered to /ʊ/ in many places, and to /o/ in the south, while /o/ is sometimes raised to /ʊ/ in the south, yielding homophones such as pool~pull,

²Although this can be explained by the fact that kide is not a word for most speakers of modern English, so there is no underlying /kayd/ for [kad] to map to as an alternative to /kad/ cod.
pool~pole, pole~pull, and pool~pull~pole. Southern Ohioans had a higher percentage of identification at the word and sentence level than northern Ohioans. Southern subjects with limited exposure to other dialects also perform significantly better at comprehending southern phrases than southerners who have had more exposure.

Diphthong perception across American English dialects was studied by Makashay (2001). Synthetic stimuli based on one talker from Binghamton, NY (northern US), and one from Birmingham, AL (southern US), were presented to subjects from both regions. The stimuli were 18 vowel continua in CVC context. These 10-step continua had initial tokens containing non-high vowels, such as hot /hɒt/, sad /sæd/, and bought /bɔt/ with formant values and durations manipulated to result in final tokens containing diphthongs, such as height /hɑɪt/, side /saɪd/, and bout /baʊt/. Diphthong production differs between these dialects. Binghamton has the Canadian raising of the North, whereby the nucleus of a diphthong is centralized (e.g., /ɑɪ/ > [ʌɪ]). Birmingham has the monophthongization of the South, in which (for this type of monophthongization) the nucleus of the diphthong is the entire vowel (e.g., /ɑɪ/ > [a]). In a direct boundary estimation task, northern subjects usually did not perceive any diphthongs along the continua that included southern monophthongized diphthongs as final tokens, while southern subjects did perceive the intended targets in these continua. However, northern subjects perceived southern non-monophthongized tokens as diphthongs earlier in the continua than southern subjects did. There was no significant difference between these groups for the continua that included northern diphthongs as endpoint targets. The differences in diphthong boundaries provide evidence that dialects are defined by perception as well as production, although the patterns of perception do not exactly mirror differences in production.
Thomas (2000) tested the use of offset spectral difference for determining the following consonant by central English Ohio and southern Spanish/English Texas dialects. The Ohio subjects produce Canadian Raising, while the Texas subjects do so to a lesser degree, if at all. Synthetic stimuli modeling the /ai/ offset values of the dialects were created in a series from tide to tight, modifying duration spectral offset, and duration of glottal pulsing during closure. Both groups were able to use the differences in diphthong duration and offset formant frequencies to distinguish a following /t/ from /d/, though the Texas subjects used the spectral difference to a lesser extent.

Sugito (1978) conducted identification tests of Japanese word accent types across dialects. An accent continuum using synthetic stimuli was created between two of the accent types used in the Osaka dialect (high/low, and low/high-falling) for /aka/. Subjects from Tokyo and Okayama dialects (which do not have low/high-falling accent, but treat it as low/high accent) did not perform significantly different from Osaka subjects in judging the boundary of high/low accent. Subjects from Nagasaki, which is losing accent contrast in production, made less of a distinction, as did those from Fukui and Yonezawa, which lack accent contrast.

2.7.2 Exemplars versus prototypes

One theoretically important conclusion that can be determined from the findings of one of the preceding dialect studies is that there is evidence to determine the winner of the exemplar-prototype debate. Prototype theories (e.g., Kuhl, 1991) propose that listeners store abstract, talker independent values of phonemes, and compare incoming values with them for classification. Exemplar theories (e.g., K.
Johnson, 1997) use the storage of previous examples for comparison with the new input. If there is a good fit to an exemplar, the stimulus is identified and the weights for that exemplar are strengthened. If there is no good match, but the stimulus is close enough to be considered part of a category, it is used as a new example of that category.

Prototypes cannot just be points in space, since even with normalization techniques (Miller, 1989, among others), the best shapes that can be achieved are irregularly shaped “clouds”, not spheres. So prototypes have to use a little bit of magnetic warping, pulling some areas of space in closer to the prototype, and pushing others further away. Instead of points in space, prototypes are actually composed of points and complex formulas.

Frequency of occurrence is an important aspect of exemplar theory, since experience shapes perception. But if prototype locations shift because of exposure to language throughout an individual’s life, then it is difficult to distinguish the two theories.

But how does the prototype theory handle allophonic and dialectal differences? Many southern dialects have monophthongization of the diphthong /aɪ/ to [a] (which is distinct from the monophthongal vowel /ɑ/, but can overlap its vowel space) in an open syllable or if the following consonant is voiced (Labov & Ash, 1997, among others). Both theories would need to be able to store dynamic information (e.g., duration and formant trajectory information of [aɪ] and [a] in the /ai/ space as opposed to the [ɑ] in the /a/ space). The exemplar theory can store examples of the [a] allophones along with the [aɪ] allophones in its /ai/ phoneme group, while the prototype theory
would need a prototype that includes dynamic [a'] and the more steady-state [a] while excluding [a], unless there is a prototype for each allophone.

Let us make this even more complicated by adding a northern dialect with Canadian Raising of /aI/ to [a'] (and sometimes to [i]) when preceding a voiceless consonant (Thomas, 2000; Makashay, 2001). For example, the northern /aI/ set would need to be able to accommodate words such as [ha'd] hide and [hɑ't] height. The exemplar model would act as before, and the prototype model would easily have a prototype for a fronting, rising diphthong because there is nothing similar. Now, what would happen if a northern talker were exposed to southern talkers? His perceptual system would change in comparison to unexposed northerners. Initially, he might experience confusion, as the southern [a] would probably overlap his /a/ space, and he would not know if he was hearing /aI/ or /a/—was the spoken southern word with a vowel that was either [a] or [ɑ] meant to be [had] /hɑd/ hide or [had] /hɑd/ hod? The exemplar theory would have little problem in incorporating the southern monophthongal /aI/—just add examples of [a] to the /aI/ group. Although there would be similar exemplars in both the /aI/ and /a/ groups, over time with greater exposure to monophthongal /aI/ the [a] exemplars in /aI/ would receive greater weights than the “incorrect” [a] exemplars in /a/. The prototype model has a more difficult time, since the northern dynamic /aI/ prototype would have to be radically changed to incorporate monophthongal [a].

An easier solution is to propose that there is a different set of prototypes for each phonetic system one learns, be it for a different language or a different dialect. If an individual already knows the talker, he can access the correct dialectal set of prototypes for perception. If the talker is unknown, all of the prototype sets are
accessed in parallel during the first few instances of speech in order to determine which has the greatest success of “prototypical hits”, and that set of prototypes is used for recognition. Similarly, the exemplar model can use different dialectal sets of exemplars, which would reduce the overlap between sets as previously mentioned.

But how do these theories handle the findings by Flanigan and Norris (2000), which found that southern Ohioans with limited exposure to other dialects did significantly better at comprehending southern phrases than those with exposure? Both theories can explain this by having a second set of representations for a different dialect, but only if the listener picked the wrong dialect. Flanigan and Norris raise the possibility that southern Ohio students who are aware of other dialects’ pronunciations may accommodate for them, and switch a southern interpretation of a vowel to a northern one. If a southerner exposed to another dialect has created a new set of prototypes (or exemplars) for the new dialect, then she would still have the same (or similar) set of prototypes (or exemplars) for her native dialect as an unexposed southerner who only has the southern set. She also should have the same perceptual ability for southern speech that she had before exposure to the new dialect. For example, if she hears enough tokens of northern [ha’d] for hide to create a separate northern /ai/ prototype or set of exemplars, then this should have no effect on her southern /ai/ prototype or exemplar set which may include the monophthongized [had] hide. But since she does worse in recognizing southern speech than subjects with less dialect exposure (and assuming she recognized the talker as southern, and was not switching between dialect representations at random), her exemplars or prototypes must have been modified by exposure to other dialects. For example, on hearing southern [had], she interpreted it as /hod/ hod instead of /haid/ hide. If she had two representations
she should have been able to map southern [had] onto her southern representation correctly. So she must have only one representation, originally southern, but modified by northern input.

The prototype model has a much more difficult task than the exemplar model in handling this situation. Under the exemplar model, the listener would gain examples of the northern [a'] in voiced and open syllable contexts. Previously, she had only a nearly unanimous set of [a] in this context, having [a'] examples only in voiceless contexts, to make up the southern /ai/. Although she would still hear southern [a] and gain those new exemplars, her northern [a'] exemplars would shift her representation away from that held by a southerner with less northern dialect experience. This would explain why she would do worse in identifying [a] as /ai/. She now has weaker weights for [a] and stronger weights for [a'] exemplars in /ai/ than an unexposed southerner has. So southern [had] would less likely be interpreted as /haid/ hide than nearby /had/ hod with her northern-influenced representation, in comparison to a southerner without that influence. The prototype model, on the other hand, cannot explain a worsening in the ability to correctly identify [a] as /ai/. Whether or not [a'] and [a] receive their own prototypes or there is only one prototype for /ai/ in the representation, what would cause the system to change its definition of a good [a] for /ai/? Southern talkers would not change their production of [a] as the result of a listener being exposed to northern [a'].

However, it is possible that the assumption that the southern listener with northern experience correctly identified the talker is false. It is also possible that alternative dialect maps cannot be turned off, and a strong match to the wrong dialect map could outrank a weak match to the correct map.
2.7.3 Vowels

Vowels can be distinguished by formant frequency values (or ratios) at the “steady states” and in the transitions to adjacent segments, duration, intrinsic pitch, and amplitude, among others. Types of cue trading with individual differences found in vowels are formant transitions versus steady states (Lindblom & Studdert-Kennedy, 1967), consonantal context versus isolation (Rakerd & Verbrugge, 1985), Bark scale distance versus values of F1 and F2 (Zhukov & Chistovich, 1985), and formant values versus amplitude (Chistovich & Chernova, 1986). Other differences in vowel perception involve judgments of similarity or dissimilarity between vowels (Fox, 1982, 1983), formant values needed to distinguish one vowel from another (Kewley-Port & Atal, 1989; Kewley-Port & Watson, 1994; Aaltonen, Eerola, Hellström, Uusipaikka, & Lang, 1997), tenseness distinctions (Faber, 1992) and timbre affecting perceived duration (Pind, 1998).

Lindblom and Studdert-Kennedy (1967) examined the recognition of steady states [i] and [o] in isolation, and with transitions in the consonantal contexts [w, w] and [j, j]. All 10 subjects of “homogeneous dialectal backgrounds” have their average labeling crossover of narrow width within 3 stimuli of each other on the 20 point [u] to [i] in isolation continuum. But for the consonantal contexts, three subjects failed to show good crossovers. Most push their boundary to the consonant loci ([w] is past the [u] endpoint, and [j] is past the [i] endpoint), but one subject reverses the directions of the shifts, and two others push the boundary in [j, j] closer to [w] than they do in [w, w]. Lindblom and Studdert-Kennedy mention that the [j, j] is not a natural English sequence, which may explain its boundary’s high variance. They explain the shifting in terms of undershoot. This can be used to explain the individual differences
as people will have different exposures to undershoot depending on the rate of speech allowed in their social contexts (or if these subjects are from different dialects they may have centralized vowels to greater or lesser degrees). Results for the set of fast stimuli show that, most, but not all, significantly move the boundary in \([w\_w]\) even further towards \([w]\). There is no rate effect for \([j\_j]\) because no one hears sequences like that.

The effects of consonantal context were also shown by Rakerd and Verbrugge (1985) in a multidimensional scaling of /i, i, e, æ, a, ə, o, ʊ, u/ in /d.d/ and in isolation for 23 subjects. In isolation, the tense/lax third dimension had little or no perceptual salience for some, but was the most heavily weighted dimension for other subjects. This also could be due to dialectal differences. So perhaps lax vowels in isolation are different than in consonantal context (Rakerd, 1984), or perhaps lax vowels are not lax (and tense vowels are not tense) for some speakers (Raphael & Bell-Berti, 1975; Bell-Berti, Raphael, Pisoni, & Sawusch, 1979).

Zhukov and Chistovich (1985) cited and tested Fant’s (1959, 1983) claim that distance between \(F2'\) and \(F1\) along the Bark scale differentiates front vowels by height, as opposed to values of \(F1\). Using two-formant noise-excited stimuli, five Russian subjects were given a fixed values of \(F1\) and had to change \(F2\) to match the phonetic quality of a standard \([i]\) like sound. As \(F1\) increased, one subject did indeed match the Bark scale curve for chosen values of \(F2\). But two others just matched the \(F2\) of the stimulus no matter what the value of \(F1\). The remaining two subjects had flat \(F2\) values for lower values of \(F1\), but for higher values one approximated the Bark scale curve and the other was twice that. This indicates that some subjects are affected by \(F2\) in height judgment, as they have to raise \(F2\) as \(F1\) raises to maintain \([i]\) likeness,
while others can keep F2 relatively close to the standard. Vowel identification also
differed among subjects, as some had only [i] for low values of F1, others [i] and [i],
and one had [i] and [ø] when F2 was low. For high values of F1, most subjects heard
[ɛ], but some heard [e] and [ɛ]. Two subjects had different intrusive vowels for mid
values of F1—[e] and [i]. There was also a wide range of crossover points, but this
and the different identifications could be explained if the subjects are from different
dialects.

Chistovich and Chernova (1986) investigated identification of one- and two-
formant steady-state vowels based on formant and amplitude relations. Eleven Rus-
sian subjects identified trains of 16 formant pulses as one of [u, o, a, æ, e, ø, u, i, i]. For one of the one-formant signals, one subject heard [i] in only half the cases
as intensity increased, while the group average was 85%. The two-formant signals
resulted in wide identification, boundary location, and boundary width differences.
For example, in one continuum with constant formants but increasing intensity, vowel
sequences heard by three subjects were [u u i], [u i i], or [u ø]. Another continuum
resulted in [ø ø ø], [ø a æ], or [ø ø]. So even though there were boundary shifts due
to formant/intensity trade-off, some subjects’ boundaries corresponded to different
vowels. Some subjects did not use certain phone labels consistently, or at all. For
example, what the group would label [u], one subject would always identify as [i],
and another as [u]. A second experiment on four subjects involved frequency of the
harmonic with maximum amplitude, and spectral slope. There were also great dif-
fences among subjects in category boundary placement, category size, and in some
cases the order of vowels identified as peak harmonic increases. But some of these
vowel identification differences could be explained by dialect if the subjects were from different ones.

Fox (1983) performed a multidimensional scaling study for the vowels [i, i, ei, e, æ, a, ə, oʊ, u, ɔ, ai, au, ju] on 19 subjects from central Ohio. D1 was best represented by a front/back dimension, D2 was high/low, D3 was low-back onset, and D4 was mid/nonmid. Examining the subject weights that Fox provides reveals differences in subjects’ utilization of the dimensions. Nine subjects had highest weights for D1, seven for D2, and three for D3. This seems appropriate, since backness and height are two traditional vowel features (and perhaps the three D3 subjects are more sensitive to transitions than the others). However, only seven subjects had D1 and D2 as their two highest weighted dimensions, while six had D1 and D3, four had D2 and D3, and one had D2 and D4. But this is only an ordinal ranking of dimensional weights, as many values for subjects probably do not differ significantly from each other. In fact, six subjects had very similar weights for all four dimensions, indicating they attend to all of them. But it may be the case that different subjects make use of some vowel dimensions more than others. In a previous multidimensional monophthong scaling study that looked found a perception-production link, Fox (1982) found a main effect in the similarity judgments due to listener (as well as stimulus pair and the interaction). Looking at the INDSCAL solution, one notices that all the listener weights are about half the size of the speaker weights for the three dimensions (backness, height, and rounding). It may be that large individual differences of the listeners limits the model to a compromise fit of the group.

Although Kewley-Port and Atal (1989) found no individual differences in their results of a multidimensional analysis of small F1 and F2 differences in a small space,
there may be a pattern as to why subjects were dropped. Synthetic prototypes and
four variants around each prototype were created for the pairs such as /i-i/, /ɛ-æ/,
and /u-ʊ/ and tested on six to eight subjects. One criterion subjects needed to meet
was to not have a large asymmetry in dissimilarity scores for AB and BA orderings
of the dyads. Summing over two experiments, five subjects failed the symmetry test,
but they were all for /u-ʊ/ and /bu-ʊ/. Did these subjects have some sort of labial
masking effect? Are they from a dialect with little to no /u-ʊ/ contrast, or a contrast
at slightly different frequencies? Or does this just demonstrate the importance of
checking that individual differences are not caused by inability to perform the task?

Kewley-Port and Watson (1994) set out to determine the thresholds for for-
mant discrimination for synthetic /i, i, e, e, æ, a, æ, o, u, u/ by incrementing and
decrementing F1 and F2 separately. Subjects worked on subsets of vowels, but all
did the increment condition for /æ/. Thresholds for ∆F are highly variable for many
high F2 frequencies and /æ/’s F1, and more so for the increment than decrement
condition. Subjects with unusually high thresholds for a particular vowel formant did
not show poor performance for others, so it may be that people need higher thresholds
for formant change in certain vowels. Perhaps subjects were from different dialects?

Aaltonen et al. (1997) examined the perceptual magnet effect in Finnish. A
synthetic vowel continuum was created by raising F2 from /y/ to /i/. Seven sub-
jects from the Turku area were classified as “good categorizers” because their mean
boundary width was under one critical bandwidth, while six were “poor categoriz-
ers” because they needed more than a critical bandwidth to be certain of the vowel.
These subjects may also need to hear a lowered F3 for lip rounding. One of the
poor categorizers seemed to have no clear exemplars for /y/, another for /i/, and a
third was inconsistent in categorizing most of the stimuli. There was no clear relation between boundary locations and width. In the following experiment for /i/ goodness ratings, most of the poor categorizers (labeled the “high-P” group) picked a high F2 value for the best /i/, which would be classified as the prototype, and had ratings decreasing with F2 to the lowest ranked /i/, the nonprototype. The “Down” group picked stimuli close to the /y/ boundary as the best /i/ and the highest F2 as the worst, and the “Hill” group picked similar /i/ locations as the Down group but had a lower boundary. The Down and Hill groups (low-P group) had significant rating differences between prototype and nonprototype, but the high-P group did not. A discrimination experiment found the perceptual magnet effect for the low-P group, while the high-P group found it more difficult to detect differences from the nonprototype than the prototype. An experiment measuring mismatch negativity (MMN, an ERP component which reflects preattentive detection of deviations from standard sound) used a prototype or nonprototype as standard in comparison to a token that deviated in F2. The low-P group had lower MMNs for prototypes than nonprototypes, while there was no difference for the high-P group. So while there is evidence that the low-P group has a prototype, the high-P group may not if their null effect cannot be explained by unnaturalness of the stimuli or perhaps the hyperspace effect (K. Johnson, Flemming, & Wright, 1993).

Faber (1992) reports on previous work with Di Paolo on the apparent merger of cognate tense and lax vowels (/i/-/l/, /e/-/ɛ/, and /u/-/u/) before tautosyllabic /l/ in the Salt Lake Valley of Utah. An identification task involved two native speakers, the older distinguishing all tense/lax pairs in F1/F2 space (but /i/ was surrounded by /i/), while the younger appeared to have only the /eɪ/-/ɛɪ/ contrast. Nineteen
subjects could all identify the older speaker’s vowels and the younger’s /el/-/el/. But for native listeners hearing the younger subject’s other vowels, three could identify all of her tense/lax vowels, four could identify all except /ul/-/ul/, four could identify all but /il/-/il/, and three could not identify /ul/-/ul/ or /il/-/il/. An analysis of the younger speaker’s acoustics found that she has more prominent first harmonics (increased breathiness) for tense vowels over lax. Faber speculates that older Utah speakers distinguish tense/lax vowels by formants and (subtle) spectral profile. Younger speakers will have different types of exposure (such as hearing speakers with high F0 and large heads who might have F1 for /i u/ low enough to amplify F0). Vowel height and apparent breathiness may seem to covary, so some younger speakers will acquire a tense/lax distinction in formants and spectral profile, while others will just have formants.

One of Pind’s (1998) experiments investigated how vowel quality influenced the perception of vowel quantity in Icelandic. The vowel in sek [sĕ:k] is higher and fronter than the one in segg [sĕkː]. Three sets of synthesized continua were created that varied over duration—one for the F1/F2 values of [sĕ:k], one for [sĕk], and one with average values. Out of thirteen subjects in an identification task, four did not identify endpoints as different phonemes and two had only a marginal labeling crossover, leading Pind to the conclusion that some subjects give much greater weight to the spectral cue than the duration one.

### 2.7.4 Voicing

Cues for distinguishing voiced from voiceless sounds include VOT, F0, duration, aspiration, and delayed F1 onset. Liberman, Delattre, and Cooper (1958)
investigated voicing cue trading for word initial stops in synthetic stimuli, trading off cues such as presence of a voice bar or delayed F1 onset, aspiration, and place. Massaro and Cohen (1977) examined how duration, F0 frequency and amplitude, and VOT are used in the perception of voicing in synthetic fricatives.

Liberman et al. (1958) used hand-painted spectrographic patterns for the pattern playback to create CVs with onsets /b d g/ and vowels /e æ ə/. The stimuli ranged from having a voice bar, to not having a voice bar, to progressively delayed onsets of F1-though the end points are not typical voiceless stops since there is no aspiration. Up to half of the 28 subjects identified /b/ as /p/ as soon as it lost its voice bar, but this stop continua had the widest boundary of all stops. /d/ has the narrowest boundary width. Some listeners did not hear voiceless consonants at the end of the continua (and some did not hear /b/ at the beginning). A follow-up test was performed on eleven subjects supported the finding that individual differences are larger for the labial sequence than the others. Thirty-two subjects participated in an experiment (with no repetitions) that compared aspiration and delayed F1 versus only delayed F1. Adding the noise decreases boundary width, and more subjects identified the end of the continua as voiceless consonants than for delayed F1 alone. It would have been interesting to know the listeners’ dialects, as VOT for voiced stops can vary from fully voiced (mostly for the labial) to 0 VOT.

Massaro and Cohen (1977) first tested fricative duration versus voicing amplitude cues for the voicing distinction between /zi/ and /si/. They employed a multiplicative model to estimate the effects of these cues on the perception of /si/, fitting parameter values to seven subjects’ judgments by minimizing squared deviations between predicted and observed values. Two subjects had duration as their
more effective parameter, while the rest had voicing amplitude (a dialect difference?).

The next experiment traded F0, duration, and VOT (from fricative onset) against each other. One subject’s more effective parameter was VOT, but all the other’s were F0 (duration had little contribution). The final experiment flipped the usage of VOT, now measuring values from fricative offset. One subject had the best values for F0, the other five for VOT, while duration now comes in second.

2.7.5 Stop place

Manipulating formant transitions to the following vowel is one method used to change stop place identification (Liberman, Harris, Hoffman, & Griffith, 1957). The stop place cue trading relation that is most often investigated is release burst versus formant transitions to the following segment (Blumstein, Stevens, & Nigro, 1977; Hazan & Rosen, 1991; Hazan & Shi, 1995), while Cooper, Delattre, Liberman, Borst, and Gerstman (1952) examined formant transitions, and F1 rise (for voicing).

Using synthetic speech on the pattern playback, Liberman et al. (1957) created a /b d g/ continuum by changing the F2 transition to a following /e/. Five subjects unexperienced with synthetic speech and four experts performed an identification task and ABX discrimination. Two of the five naïve subjects were dropped for inconsistent categorization (perhaps because there were no stop place cues, just a transition to vowel, or there may be dialect or individual differences). Liberman et al. did not compare these two groups of subjects but combined them to determine how well discrimination can be predicted from identification. Examining the discrimination charts of a couple of subjects from each group finds that there is less variability in category boundary placement for the expert group than the naïve. One naïve subject
does not seem to have much discrimination between (or identification of) /d/ and
/g/, while an expert subject appears to have good discrimination within as well as
between categories.

Cooper et al. (1952) used pattern playback synthesis to survey four groups of
subjects (about 33 in each, from unspecified dialects) on their perception of burstless
tokens from a continuum of F2 transitions to a following steady-state vowel (/i, e,
ε, a, ə, o, u/). Two groups were told to identify the stops from one of /b, d, g/
but one group had an F1 rise and the other a straight F1 (“unvoiced”). The other
two groups were instructed to use /p, t, k/ during identification, and had the same
F1 distinction between them. Cue trading and vowel context effects can be found if
one compares across the groups. There were overlaps in judgment between subjects
in the /b, d, g/ with F1 rise group: /b/ or /d/ were heard by different subjects for
small F2 rises before /i/; /d/ or /g/ were heard for small F2 falls before /i, o, u/; and
/b/ or /g/ were heard for small F2 rises before /o/. For /p, t, k/ with F1 rise: /p/,
/t/ or /k/ responses were given by different subjects for small F2 falls before /i/, as
well as for the response overlaps mentioned previously for the /b, d, g/ group; /t/ or
/k/ responses overlapped for small to mid-sized falls in F2 before all vowels; and /p/
or /k/ responses by different subjects overlapped for small F2 rises before /ə, o, u/.
Subjects in the groups with no F1 rise had the same overlapping perceptions as their
respective F1 rise groups, but /t/ or /k/ responses by different subjects overlapped
for all vowels, to a greater extent.

Blumstein et al. (1977) examined how 24 subjects from unspecified dialects
perceived a synthetic [b d g] continuum with both bursts and transitions to a follow-
ing [a], just transitions, or with conflicting ([d g]) burst and transition cues. There
were three test groups for which each first identified all the stimuli, then underwent adaptation on one of labial, coronal, or velar stimuli, and then took a final identification task. The labial group listened to the adapting stimuli of [b] with burst and transition (the burst was more restricted in frequency than in nature), and with just transition. No conflicting place cue tokens were used since in pilot work they always sounded like [b]. The coronal group were adapted to [d] with burst and transition, [d] with transition, and the conflicting cue stimulus [d] with its transition but with a [g] burst. The velar group were adapted to [g] with burst and transition, [g] with transition, and the conflicting cue stimulus [g] with its transition but with a [d] burst. There were individual differences in the perception of the conflicting cue stimuli. Five of eight subjects in the coronal group consistently identified the conflicting cue stimulus as [g], and one of eight subjects in the velar group identified their conflicting cue stimulus as [d]. These subjects evidently weight burst cue information more highly than transitions. However, they still adapted to the transitions like the other subjects, since adaptation effects are due to acoustics, not labels. Adaptation to a stimulus shrinks its category size, as the boundaries to other categories move inwards. But some subjects almost lost the category entirely (e.g., less than 50% identification for [d]), and for both [d] and [g] adaptation, some subjects would hear [b] at the boundary between [d] and [g].

Hazan and Rosen (1991) were one of the few to purposefully set out to examine individual differences (but without controlling for dialect). Eighteen subjects (less two who were dropped for not reaching 90% accuracy on the endpoints) picked the identity of the initial stop in *bait-date* and *date-gate* continua. The continua
contained full-cues (both burst center frequency and following F2, F3 transitions varied between endpoints), formant-transition-cues (burst frequency is fixed while the transitions still varied) or burst-cues (burst frequency varied while the transitions are fixed). Variability was always greater in a reduced-cue condition than for full. All subjects had significant effects in the reduction of usefulness of either cue except for four subjects, but only in one continua each. Some subjects have similar effect for both reduced cues, while others have more effect for one over the other. This indicates that subjects utilize the burst and transition cues either proportionally or disproportionally. However, only the effect of burst-only cues correlates by subject for both word-pair continua. A follow up test on nine subjects (three dropped for not being able to label [d] endpoint at 90%, though all three were at 100% for [b]) was run on ba-da continua with and without the burst, and the same bait-date continua as before. All subjects had significant reduction of both cues in bait-date, but two do not for ba-da, which could be due to context effect from the following vowel.

2.7.6 Manner

Fitch, Halwes, Erickson, and Liberman (1980), Best, Morrongiello, and Robson (1981), and Hodgson and Miller (1996) all investigated the trading relationship between silent gap duration and formant onset values in the determination of whether a fricative or a fricative-stop sequence was heard.

Fitch et al. (1980) created synthetic continua of [s] followed by increasing silence duration and a flat formant onset of [lIt], and of [s] followed by increasing silence duration and the rising formant onset of [plIt], neutralizing other cues. Twelve subjects of unspecified dialects identified the tokens as slit or split. The researchers
reported that all subjects had a boundary shift at an average of 25 ms less silence needed to hear split with formant rises than without. But the shifts ranged from 8 to 40 ms by subject, and since 8 ms is the increment used in the duration difference between stimuli, one subject had a boundary shift between the two continua of one token. This subject may have only been relying on gap duration cues in determining whether the stop was there, and not on the presence or absence of spectral cues.

Best et al. (1981) used two synthetic say-stay continua that increased the silent gaps after the [s], with one continuum adding a poor [d] with a high F1 onset to the vowel, and the other adding a good [d] with a low F1 onset. Fifteen subjects from unspecified dialects performed identification tasks, and a three-way oddity discrimination task. This task included different comparison types: one cue, in which the three stimuli all had the same duration, but one differed from the others in F1 onset; two cooperating cues, in which there was a set longer gap duration before the good [d] low F1 onset (stay bias) than before the poor [d] high F1 onset (say bias); and two conflicting cues, in which there was a set longer gap duration before the poor [d] high F1 than before the good [d]. All subjects had the trading relation boundary shift (a good [d] low F1 needs less gap duration than a poor [d]), but had different performance levels on the oddity task. Ten subjects’ performance levels were ranked as cooperating cues > one cue > conflicting cues, while the other five had one cue performance equal to or lower than two conflicting cues. This makes one wonder whether these subjects actually have cue trading (the range on the identification task boundary shift went down to only 9 ms) or need both cues to distinguish fricative from fricative-stop. A following experiment used sine wave say-stay continua with oddity discrimination and AXB identification tasks (A and B were the continuum
endpoints). Sixteen subjects were told the stimuli were computer noise (unlimited number of possible categories), and six others were told they were distortions of say and stay (two possible categories). [The difference in response sets makes suspect the authors’ conclusion that since the phonetic listeners performed better than psychoacoustic listeners, speech must be special.] A post-test questionnaire weeded out four guessing subjects, and two computer noise subjects and one speech subject who heard other words than the ones being tested. The remaining subjects were grouped as: say-stay, for which three phonetic-biased and two computer noise subjects heard words, but sometimes tones and drips; temporal, of which three listened to gaps and two to the entire duration, with one being a phonetic-biased subject; and spectral, for which five listened for different types of frequency patterns, with one being phonetic-biased. Only the say-stay group had a trading relation. The temporal group’s results did not distinguish the two types of stimuli (i.e., frequency onset did not matter, just duration). And the spectral group distinguished between the two types of stimuli (near perfect discrimination), but had almost no trading relationship because each continuum was almost exclusively identified to a different category (i.e., a large top-down boundary shift instead of left-right). So duration did not completely compensate for the frequency onset difference.

Hodgson and Miller (1996) examined within category trading relations using a category goodness judgment task on 12 unspecified dialect subjects. Using a gap duration and F1 onset trading relation like in Best et al. (1981), the say-stay continua were expanded past stay to an exaggerated silent duration like s...t/day. (Another 10 subjects were used to replicate Best et al.’s results on the “normal” portion of the say-stay continua stimuli.) For the say-stay-s...t/day continua, nine of the subjects
have the expected trading relation. They gave low stay goodness ratings for short silences, rise to a peak, then lower again for the exaggerated silence durations, with the peak for high F1 (bad [d]) onset shifted to longer durations than the peak for low F1. The remaining three subjects had different results: one subject had a greater trading relation boundary shift, but the ratings for high F1 onset were consistently lower than for low F1, so perhaps F1 was almost too high to be compensated by increased duration; the other two subjects had low flat ratings for all high F1, as if there was no useful formant information. A following sine wave experiment, like Best et al.’s, involved 16 subjects being told they were hearing computerized say and stay words, all of whom later reported hearing the stimuli as words. Half the subjects had the expected trading relation boundary shift, but a greater duration was required for both peak ratings. For the other eight subjects: two had results similar to the remaining three subjects of the previous experiment; one subject had a wider rating peak for low F1; three increased their ratings with duration then flattened out (no internal category information); and two lowered their ratings as duration increased, as if they were focusing on silence duration and not phonetic information.

2.7.7 Fricatives

Weiner and Singh (1974) scaled discrimination reaction times of fricative pairs, while Whalen (1991), Hedrick and Ohde (1993), and Newman (1998) examined trading effects of fricative noise amplitude (or length) at certain frequency regions and transitions to an adjacent vowel. Since Newman was looking at perception and production links in individuals, her other experiments will be mentioned as well.
A multidimensional scaling analysis was performed by Weiner and Singh (1974) on the fricatives /f, v, θ, s, z, ñ, h/ before /ɑ/. The choice reaction time values of 10 subjects’ same-different responses were used as input to INDSCAL, resulting in a four-dimensional configuration (with dimensions interpreted as voicing, sibilant, front/back, and palatal). Subject weightings split the subjects into three groups (dialects?): three subjects weighed D2 the highest (and one of them has D1 the lowest), four subjects weighed D1 the highest and D2 the lowest, and three subjects had relatively equal weights that decreased with increasing dimensions. The first two groups have opposite weightings of importance for sibilant and voicing, each of which divide the fricatives into larger groups than the other dimensions can.

Whalen (1991) segmented the fricatives in *suit* and *shoot* to create hybrid fricatives (such as 1/8 of the beginning of [s] followed by 7/8 of the end of [ʃ], or 1/2 the beginning of [ʃ] and 1/2 the end of [s]). These would be followed by the transition (and vowel) corresponding to the second fricative, or the transition corresponding to the first fricative (with the second fricative between). Thirty-two subjects of unspecified dialects listened to the words and identified the fricative as either “s” or “sh”. The fricative portion with the longest duration was the one reported (though the second portion has more weight), but there was a trading relation between length and transition since the first fricative portion needs to be longer for identification if the transition corresponds not to it but to the second portion. There was also a context effect from the following [u], since an [ʃ] in first position needs to be longer for identification than an initial [s] would (i.e., listeners expect [s] to have low frequency noise if followed by [u]). A following experiment had an open set of responses, as four phonetically trained subjects transcribed the stimuli. There were varied responses.
Only two subjects heard non-speech for certain stimuli, as well as [p]-fricative clusters. One of the two also heard [sθ] when the fricative portions in [ʃs] had equal duration. Most subjects heard [tʃ] for short [s]-long [ʃ] stimuli, and [ks] for short [ʃ]-long [s], indicating that for these subjects the duration of the first fricative was too short to be any obstruent but a stop.

Hedrick and Ohde (1993) investigated the trading relation involving the amplitude of frication in relation to the amplitude of vowel onset in the F3 region for /s/-/ʃ/ and the F5, F6 region for /s/-/θ/. Synthetic fricatives for each pair were created that were place neutral for F1, F2, and F3 transitions to /i, a, u/. /s/-/ʃ/ continua were created in which the values for each token were place neutral, except that F3 frication noise varied from -20 to +20 dB relative to vowel F3 onset for each vowel. Similar /s/-/θ/ continua were created, except it was F5 and F6 that ranged from +20 to -20 dB relative to the vowel’s F5 onset. Continua were also created from these to vary the fricative duration. Ten normal hearing subjects of unknown dialects who passed a 70% accuracy pretest took the identification test. Not all of the subjects perceived an /s/-/ʃ/ distinction at the shortest (30 ms) fricative duration: three heard /s/ throughout the /a/ context, one heard /ʃ/ throughout /i/ and was inconsistent for /u/, and one was inconsistent for all 30 ms stimuli. So these subjects have minimum fricative duration (and vowel context effects) at higher values than others. A following experiment inserted silent gaps of different lengths between fricative and vowel for all the /sa/-/ʃa/ continua, and also had fricatives in isolation. Half of the subjects did not hear /s/ and /ʃ/ categories across all conditions—these subjects may need to hear a rise in amplitude between the fricative and vowel to hear /s/.
a gap follows the fricative (or if it is in isolation), the amplitude falls (to silence) and those subjects are more likely to hear /ʃ/.

Newman (1998) investigated individual differences to link perception and production in cue trading between fricative place and formant transitions to the following vowel. A synthetic /s/-/ʃ/-[x] continuum was created followed by a constant formant transition to /æ/, and a synthetic non-salient fricative selected by a pilot test preceded a continuum of /s/-/ʃ/-[x] fricative transitions to /æ/. Twenty-four subjects’ highest ranked /ʃ/ ranged 2739–2935 Hz for the fricative centroid, and 0.83–2.24 Bark for the formant transition differences. An unrelated task involved a VOT continuum (/bʌ/-/pʌ/-[phɑ]) which ranged from 8.25 to 291 ms. Twenty-seven subjects rated their best /p/ between 55 and 140 ms (though five subjects were dropped for choosing VOTs over 200 ms). Thirty-three subjects from multiple dialects participated in place adaptive testing on /bæ/-/dæ/, /dæ/-/ɡæ/, and /ɡæ/-/bæ/, but only 15 reached the criterion of having a high enough rating peak or being far enough from an endpoint to keep from being dropped from any continuum.

2.7.8 Approximants and gutturals

Individual differences have also been found for the perception of pharyngeal and uvular consonants (Alwan, 1989), and for approximants (which may or may not be categorical, depending on the subject), (Sharf et al., 1988; Iverson & Kuhl, 1996).

Alwan (1989) examined F1 and F2 transitions in the perception of pharyngeal /ɣ/ and uvular /ʁ/ by six Arabic subjects (three from Baghdad, and one each from Kuwait, Beirut, and Khartoum). Synthetic continua varying on F1 transition to following /aa/ and F2 transition were made for each consonant individually. The
subjects’ tasks were to identify the consonant as /ʕ/, /v/ or /ʔ/ (/ʔ/ was included because all syllables must begin with a consonant), and give a goodness rating. For the pharyngeal F1 continuum, all subjects heard the categories in order of /v/, /ʔ/ and /ʕ/ as F1 transition onset increased. The Baghdad subjects had similar boundary crossover points, while the other subjects were more variable. For the uvular F1 continuum, all subjects have the /v/, /ʔ/ and /ʕ/ categories except for one of the Baghdad subjects (who had no /ʕ/, but a huge /ʔ/ category). The uvular/glottal boundary was variable across subjects, and strangely, the subjects from three diverse areas had boundaries that were more similar to each other than the Baghdad subjects had. The pharyngeal F2 continuum and the uvular F2 continuum stimuli were always heard as pharyngeal or uvular, respectively. The second experiment increased the F1 bandwidth of the uvular F1 continuum, and the F2 bandwidth of the pharyngeal F2 continuum to more natural values. The increased uvular F1 bandwidth had higher naturalness ratings than the uvular stimuli from the previous experiment. Now, none of the subjects hear /ʕ/ above chance. Only two of the Baghdad subjects have a /v/-/ʔ/ boundary, while the other Baghdad and the Kuwait subjects always hear /v/. There was no effect in increasing the pharyngeal F2 bandwidth. So bandwidth can be even more important than frequency for category discrimination for some subjects.

Sharf et al. (1988) previously found that some subjects have the ability to distinguish a child’s distorted /ɾ/ from /w/ while others cannot. They investigated whether training or feedback can allow clinicians to acquire this ability. Eight of nine speech pathology students passed an 85% identification criterion of child /w/-/ɾ/ continuum endpoints, and were split into two groups for counterbalanced feedback, training, and identification sessions. The stimuli were a synthetic /w/-/ɾ/ matrix of
F2 and F3 onset values that approximate child /w/, /r/, and distorted /r/ values followed by the vowel /eI/. Identification labels included the three consonants plus a “none” label to indicate none of the sounds were heard. Tokens with the greatest subject identification agreements were chosen as exemplary /w/, /r/, distorted /r/, or none. Having feedback was slightly better than having none, but there was little difference for training. However, half of the subjects can identify the exemplary distorted /r/ at 75% or higher with training. The second experiment was run to determine if the perception of a /wei/-/rei/ continuum of child values is categorical. Seven of the previous subjects and another who had taken the previous experiment underwent identification, and were split in two groups to do 1-step and 2-step ABX discrimination in one of two orders. There was no correlation with either ABX results with the identification of distorted /r/. A third experiment used 10 new subjects who had not been trained to replicate the second experiment. Distorted /r/ identification correlated with within-/r/ discrimination. So subjects who have less of a /w/-/r/ categorical relation are more able to identify distorted /r/.

Iverson and Kuhl (1996) used multidimensional scaling to examine twenty-eight subjects’ goodness judgments and identification of /r/ and /l/ before /a/ in synthetic stimuli matrix of F2 and F3 variations. The group was bimodal—18 subjects made /r/ and /l/ identifications, while nine made nearly all /l/ identifications. There was a significant difference in F3 of the best /l/ between groups. Two-dimensional scaling solutions were made for both groups, with dimension corresponding to F3 and F2. Subjects that had both liquids had a grouping for each, while those with mainly /l/ had a continuous /l/ group that flowed over to /r/. Perhaps there are dialect differences for liquids? These bad /r/ s for some subjects may have been caused by a
short F1 transition, so a following experiment used tokens with longer F1 transitions. The sixteen subjects performed as a single group. So an F1 transition length is important to some subjects as well as F3 in liquid identification.

2.7.9 Prosody

Large stress, intonation, and phrase boundary perceptual differences have been discovered across subjects, such as those found by Steffen-Batogowa and Katulska (1984) and Grant and Walden (1996).

Steffen-Batogowa and Katulska (1984) investigated the perception of main stress and stress group boundaries in Polish by 20 subjects. There were three types of stimuli: various utterances, including readings from novels, political speeches, lectures, poetry, and spontaneous speech by 10 talkers; a short story read by six talkers; and 24 isolated sentences, each with one neutral reading, and two readings each with a different word emphasized by two talkers. 73% of all subject pairs for the various utterances had significant differences of main stress placement, 39% for the short story, and 0% for the sentences (the laboratory speech). For subject pair differences in stress group boundaries, 89% were significant for the various utterances, 81% for the story, and 50% for the sentences. So there seem to be greater individual differences for boundary than stress placement, unless there is an interaction with dialect.

Grant and Walden (1996) investigated how prosody would be affected by having speech filtered with one of six different bandpass-filters covering different frequencies from 200 to 6100 Hz. There were three groups of 10 normal hearing subjects of unspecified dialects. One group identified the number of syllables and the stress patterns (using “X-X” to represent spondees, “o-X” for iamb, etc.) of one to three
syllable words. The duration of a word was unpredictable (e.g., some monosyllabic words were longer than trisyllabic words). A second group would identify rising versus falling sentence intonation as statements or questions, with duration being unpredictable. The third group identified the phrase boundaries for phrases like “A plus E times O,” “(A plus E) times O,” and “A plus (E times O),” with equivalent sentence duration. Three subjects’ performances for sentence intonation worsened as the bandpass-filters increased in frequency, but three others maintained relatively high scores for all filters, presumably because they can extract F0 from higher frequency harmonics better than the others can. Four subjects did worse on phrase boundaries as the filters went up, while four others kept equal performance. So some subjects are better able to use frequency, duration, and intensity cues at higher frequencies than others.

Howell (1993) studied cue trading effects of amplitude and F0 on the perception of vowel stress. Three subjects from southeast England and one from northwest England would read pairs of sentences that contained two syllable noun/verb stressed words with the same (underlying?) vowel (e.g., rebel). After each sentence, they would judge how well they conveyed the grammatical class of the word. Two weeks later, each would listen to the sentences of all the talkers and perform the same grammatical rating, and six months later, they would judge all the isolated words. Three of the four subjects had F0 as the most important cue to produce (and the fourth had F0 closely behind intensity). Two of the three subjects had centralization of both F2 and F1 as the next two most important cues (and one had intensity for the fourth). The third of the three subjects has duration for the second and third most important cues. For perception, each subject usually attends to the same properties
and in the same order as what they made in production. Other subjects would usually switch cue importance over to what the talker used (which is significant for both sentence and isolation contexts). There was negative cue trading on production and perception of F0 and intensity, but only on certain talkers. All subjects traded F0 and intensity for the first two talkers (and also duration and F0 on the first talker), but there were smaller effects on the others. So different subjects have different amounts of cue trading for different talkers.

2.7.10 Sign language

Spoken language is not the only mode of communication in which there are individual differences in perception. Poizner (1983) investigated how American Sign Language has shaped signers’ representation of visual movement. Five native ASL signers (congenitally deaf with deaf parents) and five non-signers (normally hearing) made triadic comparisons of signs made with point-lights, for lexical movements and for inflectional movements in two separate experiments. Hierarchical clustering and multidimensional scaling showed that while ASL and non-ASL subject groups were always significantly distinct, there were also individual differences in the weighting of the dimensions. For lexical movement, most ASL subjects had high weights for “repetition” and “arcness” dimensions. A couple had weights not as high, but had higher weights than the other ASL subjects for “plane” and “direction” dimensions, placing them closer to the hearing group. The weights of the groups are more similar for inflectional movement, and some ASL subjects had greater reliability of weights for the ASL preferred dimensions of “cyclicity” and “displacement” than for the hearer preferred “direction” and “iteration length”.

73
2.7.11 Non-speech

Research that examines non-speech with properties that mimic speech cues (Pisoni, 1977; Collins, 1984; Christensen & Humes, 1997), finds that the cues are treated in the same manner for both types of stimuli, but with differences between listeners.

Pisoni (1977) investigated non-speech temporal discrimination that mimics VOT. Eight subjects were extensively trained on the endpoints of two tones (500 Hz (T1) and 1500 Hz (T2)) that had the same offset time, but different onset lead and lag times of T1 (-50 ms to +50 ms). Two subjects had identification crossover points at the expected 0 ms onset lag time, but the other six had them shifted towards the lagging T1. A second experiment gave the same stimuli to 12 subjects who had no training. For percent correct of discrimination, one subject was at chance, four had a single peak for lagging T1 at +20 ms (i.e., two categories), and the rest had two peaks for leading T1 at -20 ms and lagging T1 at +20 ms (three categories). A third experiment had eight subjects trained on three stimuli (-50, 0, and +50 ms lead and lag times for T1). The expected identification crossover points were at -20 and +20 ms, but two subjects heard the negative onset category at the boundary between the other two above chance.

Collins (1984) investigated the discrimination of frequency-varying sinusoids which are analogous to F2 transitions and steady-states of CV and VC syllables. Four normal hearing subjects discriminated between glides preceding tones in one condition, and glides following tones in another. Three subjects had higher difference limens for glides preceding the tone. Seven normal hearing subjects participated in an experiment to find temporal masking on glides. For the four subjects tested on
the discrimination of tokens with low glides and tones, one subject had no temporal masking, and three had both forward masking on tone-glides and backward masking on glide-tones. For the four subjects tested on the discrimination of tokens with high glides and tones, all had backward masking, but only two had forward masking. For the four subjects who distinguished between glideless tones and tones with slightly high or low glides, all had backward masking, and none had forward. Collins relates this to Wang and Bilger’s (1973) finding that more errors are made on CV than VC syllables.

Christensen and Humes (1997) examined how training on non-speech cues affected people’s classification. Pairs of synthetic fricative stimuli were created that differed on three dimensions with three levels each: low, medium, and high values for center frequency, temporal gap duration between the fricatives, and slope of frequency transition of the second fricative. Nineteen subjects with normal hearing gave similarity judgments for all pairings of endpoint value stimuli with midpoint value stimuli for an MDS analysis. Eleven subjects had the greatest weights for center frequency (one was near 0 for temporal gap), four for temporal gap (one near 0 for frequency transition), and four for frequency transition. A following experiment was conducted on 18 of the subjects for which they were trained on three exemplars, tokens which had all low values (“circle”), all mid values (“triangle”) or all high values (“square”). GCM (generalized context model) weights were used with the MDS solution to predict performance. All subjects used frequency transition the most in identification, but the weights ranged from about .5 to 1, with center frequency usually being the second largest weight. Sixteen of the subjects were then trained on temporal gap duration, with short gaps as “circle”, medium gaps as “triangle”, and long gaps as
“square”. And again there was a range of weights of about .5 to 1 for the most important cue. This time it was temporal gap, a significant change in weight due to training. The second largest weight was usually for frequency transition. Ten subjects participated in the last experiment, with training for low, mid, and high frequency transition slopes. Two subjects did not reach the 80% accuracy criterion (which is odd since all subjects had the highest weights for transition previously) but were included anyway. This time the highest weighted cue was frequency transition for all but one subject (center frequency) with a range of about .4 to 1. These findings on non-speech suggest that listeners may focus on certain cues in speech.

2.7.12 Genetic

One interesting theory by Lieberman (1986) on linguistic variation is based on the premise that the biological substrate for speech is not uniform. Since many forms of dyslexia are genetically transmitted, there may be variations in the neural substrates used for speech sound identification. A genetically related group’s language should reflect what can and cannot be perceived well. But when they intermix with another group, the genes of the offspring will carry the perceptual variation into the speech community of the surviving language. The prediction is that if a large group joined the community, there could be language change, while if it were a small group, some of their children would have dyslexia with the standard language.

2.8 Weight changing

How successful are the attempts to switch a listener’s focus from one cue to another? Are subjects variable in their dependence on certain cues? Cue weighting can change due to experimental design and the stimuli used. When subjects were
trained on stimuli that differed on target and nontarget values of amplitude envelope, they had greater MDS weights for amplitude after training (Christensen & Humes, 1996). Many times, however, subjects seem unable to be trained to use a different cue. McFadden et al. (1973) failed to train subjects who used only one of interaural temporal differences or interaural intensity differences for sound localization to use the other cue.

Adaptation can cause a category to shrink. Blumstein et al. (1977) adapted subjects on a good stop token, with the result that boundaries to that category shifted inward.

There is also evidence that weights can change over time (not from a long-term study, but by studying different age groups of the same general population). Younger children were less likely to have MDS weights that result in organized sound categories than older children and adults (Allen & Bond, 1997). L2 learners narrow boundary width and increase monotonicity of identification function over time (Gass, 1984).

And, as this previous research has shown, individuals do tend to have differences in the way they attend to cues. But the differences in speech perception reported in the literature may have various causes. Some differences may be dialectal. Others may be the result of the type of stimuli used. Many studies test perception on impoverished stimuli, using the results to explain the cues and processes involved in real (complex) speech perception. (This raises the question for many of these types of studies if subjects would interpret these types of stimuli as speech if they were not told to do so.) The individual differences found might be due to increased error rates that would be hidden by ceiling effects in real speech. Or they may be true
individual differences in processing, shown, for example, when the removal of a cue greatly impairs a listener who weights it highly but not another listener who weights it lowly.

This dissertation examines how listeners are affected in a discrimination task when frequency and duration cues are manipulated. Subjects were not informed about the nature of the stimuli, and had to decide for themselves what the sounds were, and how they should be compared. Questions to be answered include whether a listener weights cues similarly across different phonetic contrasts (vowel vs. fricative), and how this can vary by individual.
CHAPTER 3

SPEECH CUES

In this study, trade-offs and sensitivity involving frequency and duration cues for vowel height and fricative place will be investigated because in these contrasts frequency and duration can be manipulated independently. The base tokens to be used will be heed [hid] and hid [hɪd], and bath [bæθ] and bass [bæs]. The effects of word frequency have been shown to affect perception (e.g., Broadbent, 1967) so these words were chosen based on their similar Kucera-Francis written frequencies, Brown verbal frequencies (Coltheart, 1981), and Hoosier mental lexicon familiarity ratings (Nusbaum, Pisoni, & Davis, 1984), shown in Table 3.1.

This chapter discusses previous studies that suggest that resonance frequency and duration both play a role in the perceptual identification and contrast of [i] versus [ɪ] (Section 3.1) and [θ] versus [s] (Section 3.2). The chapter ends with a description of the heed-hid and bath-bass continua used in this study.

\[3\]In shoe and sue, a pair that was originally considered for use, the talker had similar duration for [ʃ] and [s]. [ʃ] (\(M = 183 \text{ ms}, \ SD = 21\)) was not significantly shorter than [s] (\(M = 198 \text{ ms}, \ SD = 30\)), \(t(16) = 1.18, p = .25\) (two-tailed). However, [u] (\(M = 252 \text{ ms}, \ SD = 12\)) following [ʃ] was significantly longer than [u] (\(M = 223 \text{ ms}, \ SD = 27\)) following [s], \(t(16) = -2.887, p < .05\) (two-tailed). There was no post-vocalic [s]/[ʃ] pair with similar frequencies of occurrence that were easily synthesizable.
<table>
<thead>
<tr>
<th>Word</th>
<th>Kucera-Francis Written Frequency</th>
<th>Brown Verbal Frequency</th>
<th>Hoosier Mental Lexicon Familiarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>heed</td>
<td>8</td>
<td>0</td>
<td>6.1667</td>
</tr>
<tr>
<td>hid</td>
<td>6</td>
<td>0</td>
<td>6.5000</td>
</tr>
<tr>
<td>bath</td>
<td>26</td>
<td>3</td>
<td>7.0000</td>
</tr>
<tr>
<td>bass</td>
<td>16</td>
<td>0</td>
<td>7.0000</td>
</tr>
</tbody>
</table>

Table 3.1: Frequency and familiarity ratings for tokens

3.1 [i] and [ɪ]

Two of the acoustic differences between the vowels in [hid] and [hɪd] are F1 frequency values ([i] has higher F1) and duration ([i] is shorter). Example spectrograms of these vowels are shown in Figure 3.1. Stimuli based on these words would test the perception of the linguistic dimensions of vowel height and length (i.e., the tense/lax distinction).

Figure 3.1: Sample spectrograms of heed and hid, respectively.
Although many other cues can be used to distinguish \([i]\) from \([\ddot{i}]\) (e.g., \(F2\) or intensity) vowel duration and \(F1\) frequency contours were chosen because of corresponding cues in fricatives. Vowel duration cues can be compared with fricative duration cues, while \(F1\) frequencies in vowels can be contrasted with frequency differences used in fricative distinctions.

### 3.1.1 Tense/Lax

The tense/lax distinction (Jakobson, Fant, & Halle, 1963; Chomsky & Halle, 1968) has been the focus of much debate as to what the phonetic correlates are and whether the feature is needed phonologically (e.g., Delattre, 1962; Jakobson & Halle, 1964; Lass, 1976; Bauer, 1980). Features such as advanced tongue root (Stewart, 1967; Halle & Stevens, 1969), complex/simple (Trager & Smith, 1951), or peripheral/central also have been proposed to make a similar contrast.

One detail that should be mentioned in regards to the tense/lax issue is that for northern Columbus, OH residents (and for many other dialects), the closest tense vowel to \([\dddot{i}]\) is \([\dddot{e}]\) as shown by the vowel spaces in Appendix B. In any case, the main cues under focus are \(F1\) frequency and duration, which are in the following sections.

### 3.1.2 F1 frequency

The history of vowel formant research (e.g., Peterson & Barney, 1952; Potter & Peterson, 1948; Fairbanks & Grubb, 1961; Lehishe & Peterson, 1961; Hillenbrand & Gayvert, 1993; Hillenbrand, Getty, Clark, & Wheeler, 1995) is long, showing that differences in vowel height can distinguish \([i]\) from \([\dddot{i}]\). Formants may not be the only spectral information needed in vowel recognition (e.g., see Zahorian & Jagharghi,
1993) but since this study will be using formant synthesis, other vowel spectral cues will not be discussed.

How much of a frequency change in F1 is needed before it is noticed? Flanagan (1955) finds difference limens of steady state vowel formants to be roughly 5% of the formant frequency (about 14–20 Hz for men’s high front vowel F1 in Peterson and Barney (1952)), while Mermelstein (1978) finds DLs to be much larger in consonantal context. However, more recent research (Hawks, 1994; Kewley-Port & Watson, 1994; Kewley-Port, 1995) indicates that sensitivity can be even greater (<2% of the formant frequency).

Internal formant trajectory changes should be used, not just “steady-state” values, since research has shown that they contribute to intelligibility (Assmann & Katz, 2000). Much research has focused on the existence and perceptual importance of vowel dynamics (e.g., Potter & Steinberg, 1950; Tiffany, 1953; Strange, Verbrugge, Shankweiler, & Edman, 1976; Assmann, Nearey, & Hogan, 1982; Assmann & Katz, 2000; Di Benedetto, 1989; Makashay, 1999) finding that vowel formant movements convey categorical information.

3.1.3 Vowel duration

Although acoustic measurements show that [i] has greater duration than [ɪ] in similar environments (e.g., Peterson & Lehiste, 1960; House, 1961; Hillenbrand et al., 1995), whether this duration difference is needed perceptually is not as clear. Ainsworth (1972) found that [i] responses decreased (while [ɪ] responses increased) as duration increased, but this was for British English. Work by Massaro and Idson using stimuli with ambiguous formant values found that the identification of [i] improved
with increased duration, but for [i] there was either no difference due to duration (Massaro & Idson, 1978), or there was only a small decrease in accuracy with increased duration (Idson & Massaro, 1980). Hillenbrand, Clark, and Houde (2000) found that it was uncommon to have a duration induced shift for this contrast, and theorize that since the spectral properties are discrete, duration has a small role in identification.

Since these vowel formants differ on frequency level and transition rate to the following stop, it will not be possible to distinguish the effect of one from the other. Elliott, Hammer, Scholl, Carrell, and Wasowicz (1989) found that transitions can have lower frequency thresholds than steady sounds. Only the difference between an [i] F1 contour and an [i] F1 contour can be discussed, not whether it is due to the “steady-state” or due to the transitions. Previous research also has found that transitions aid in duration discrimination more than “steady-state” frequency (Dooley & Moore, 1988; van Wieringen & Pols, 1994), so the different Hz/ms rates of the different duration tokens would affect duration judgments.

3.2 [θ] and [s]

Unlike vowels, fricatives produced at different places of articulation usually do not have the same number of spectral peaks and do not have them in similar frequency regions. One method of referring to a certain fricative peak or centroid is to use the nearest vowel formant. In this instance, I am investigating frequency differences of [θ] and [s] in the F5 region, at about 4 kHz. This is based on Hedrick and Ohde (1993) (continuing the research of Stevens (1985)), who investigated relative amplitudes of these fricatives in the F5 and F6 regions. As I will be using formant synthesis (with five formants), I will refer to this frication region as the fifth frequency band, or F5.
The fricatives in [bæθ] and [bæs] contrast on frication centroid frequency in the F5 region of vowels ([s] has a higher centroid), and duration ([s] is longer). Example spectrograms of these fricatives are shown in Figure 3.2 and Figure 3.3. Figure 3.4 displays sample spectra with windows placed over the entire fricatives. These tokens test for perception of fricative place and sibilancy.

Figure 3.2: Sample spectrograms of bath and bass, respectively, at 44.1 kHz sampling rate. The F5 region is around 4 kHz.

Figure 3.3: Sample spectrograms of bath and bass, respectively, at 11.0 kHz sampling rate. The F5 region is around 4 kHz.
Manipulating F5 in the Klatt synthesizer changes the fricative spectral moments—one type of analysis of fricative acoustics (Forrest, Weismer, Milenkovic, & Dougall, 1988). Spectral moments are obtained from the power spectrum of a Fourier spectrum. Subsets of these four moments (mean, variance, skewness, and kurtosis) have been found to differentiate fricatives. Since formant synthesis requires frequency, bandwidth, and amplitude values to create a fricative spectrum, the fricatives will be modeled and described in terms of “F5” frequency instead of spectral moments. Figure 3.4 shows that much of the distinction between [θ] and [s] occurs above 5 kHz, which is the cutoff point for the formant synthesis being used. Although there is a difference in spectral peak frequency near F5 vowel frequency, it may be the peak bandwidth or other spectral information that is perceptually important.

3.2.1 Sibilant

Unlike the tense-lax distinction, the sibilant-nonsibilant distinction (similar to the strident feature of Jakobson et al. (1963) and Chomsky and Halle (1968)) has not
been expanded much past the basic characterization that sibilants are produced with an obstruction in the air stream to create a high degree of noisiness. [θ] has a lower intensity than [s] (by 10–15 dB) (Strevens, 1960; Jassem, 1965; Stevens, 1971, 1985; Behrens & Blumstein, 1988a, 1988b; Jongman, Wayland, & Wong, 2000). Stevens (1985) created a synthetic [sq]–[θq] continuum based on the amplitude of noise in the F5-F6 region, finding that [θ] is perceived if there is a rise in amplitude to the vowel, and [s] if a fall. However, there must be more than amplitude differences that distinguish sibilant [s] from nonsibilant [θ], because Behrens and Blumstein (1988b) found that while the amplitude of natural [s] (in CV syllables with [ia u]) can be reduced to make it sound [θ]-like, increasing the amplitude of natural [θ] does not make it sound like [s]. Distinctions between [s] and [θ] have been found in both frequency in the F5 region, and duration.

3.2.2 Fifth band frequency

Most studies on these fricatives contrast the overall spectral shapes, in which [θ] has diffuse energy over the full spectrum, while [s] has strong energy only above about 3.5 kHz (Strevens, 1960; Heinz & Stevens, 1961; Jassem, 1965; Behrens & Blumstein, 1988a). However, Abbs and Minifie (1969) found that [s] has major resonances around 4 kHz, while [θ] has major resonances around 3 kHz (but Jassem found that [θ]’s peak was not below [s]’s).

Other frequency cues found to distinguish fricatives (more for distinguishing [θ] from [f] than from [s]) are transitions from neighboring vowels (Harris, 1958; Heinz & Stevens, 1961; LaRiviere, Winitz, & Herriman, 1975).
3.2.3 Fricative duration

Previous studies have found that [θ] is significantly shorter than [s] (but sometimes only in certain environments, or only as a trend) (Abbs & Minifie, 1969; Larrivee et al., 1975; Behrens & Blumstein, 1988a; Hedrick & Ohde, 1993; Jongman et al., 2000). Perhaps, like tense vowels, sibilant fricatives require great articulatory effort than nonsibilants in order to force air through a narrow channel, resulting in a longer duration needed to secure the desired effect.

3.3 Intensity and loudness

Although weaker than in the [θ]-[s] contrast, one other measure that may distinguish [i] from [ɪ] is intensity. House (1961) claims that lax vowels have lower relative amplitudes than their tense pairs, which is contradictory to the source-filter theory that a more open vocal tract (higher F1) yields more airflow. However, Fairbanks, House, and Stevens (1950) found no significant difference in amplitude between [i] and [ɪ], while the measurements of Lehiste and Peterson (1959) show that the lax vowel has greater overall intensity, but the two vowels are approximately equal perceptually in loudness. The formant amplitudes reported by Peterson and Barney (1952) also have each of [i]’s formants lower than the corresponding [ɪ]’s. For the average overall vowel intensities produced by the talker in this study, [i] (M = 77 dB, SD = .9) was significantly higher by t test than [ɪ] (M = 71 dB, SD = 1.3).

Since the lax vowels are shorter than tense vowels, perceptually they may sound equally loud since loudness depends on frequency and duration in addition to amplitude.

For the talker’s fricative productions, [θ] (M = 58 dB, SD = 1.8) was significantly higher by t test than [s] (M = 54 dB, SD = .8), but this is probably due to the interdental being affricated.
3.4 Continua

As a preview to the discussion of the stimuli in Section 4.3, the organization of frequency and duration cues are shown in Figure 3.5, with four frequency levels and three duration levels. Each stimulus is coded by a two digit number, with the first digit for frequency level (1x, 2x, 3x, 4x) and second digit for duration level (x1, x2, x3). For example, the [i] in *hid* is stimulus 41 because it has the highest F1 value (the fourth level) and the lowest duration value (the first level). Stimulus 42, directly to the right of [i] is at the next higher duration level, while 31, directly below [i] is at the next lower F1 frequency level.

For the vowels in the [i]-[i] continuum, the [i] endpoint (41), has higher F1 and shorter duration than the [i] endpoint (13). While for the fricatives in the [θ]-[s] continuum, the [θ] (11) has lower F5 and shorter duration than the [s] (43) endpoint. There are also different orders of magnitude for the vowel and fricative stimuli as shown in Figure 3.5. Although fricative F5 occupies a larger acoustic spread than vowel F1 in Hz, it occupies a smaller perceptual range in ERB.

The prediction for the perception experiments reported in the following chapters is that all subjects will attend to frequency, but there will be differences in weighting. Many subjects may not attend to duration as much, with a wider range of differences.

Why would frequency and duration weights be constant across phones? Are there different weights for allophones of a phoneme? There can be disparate allophones. If vowels can be defined by frequency of their formants, might fricatives be better defined by spectral shape? Is longer duration needed only to reach the more
Figure 3.5: Duration and average frequency values for heed-hid and bath-bass tokens. Bath-bass tokens have greater frequency distribution in Hz (left), while heed-hid tokens have greater frequency distribution in ERB (right).
extreme values of a tense vowel? Is longer duration needed for strong intensity in fricatives?

Whether an individual uses one or more of these cues consistently across the two contexts, and how much individual variation exists are what will be studied in the following experiments.
CHAPTER 4

METHOD

To measure individual differences in speech perception and determine whether such differences persist consistently across different phonetic contrasts, the same group of dialectally homogeneous native northern Columbus, Ohio subjects participated in all experiments testing the stimuli discussed in Chapter 3 (following OSU research protocol #01B0222). The results of a hearing test will also be included in the analysis to determine if slight hearing loss is a factor in individual differences. Hearing loss in a certain frequency range might be detrimental, or the subject might be sensitive to any stimulus with frequency in that range (Doherty & Lutfi, 1999). A questionnaire was administered to help control for dialect effects, and auditory dialect screening found them to be similar.

4.1 Predictions

Previous studies have shown that members of different dialects have different perceptual strategies, but that does not apply in this case because these listeners are of the same dialect (at least as judged by their productions). If results for all these listeners pattern in the same way, then there is no evidence that members of a dialect can have different perceptual strategies. This also leads to the conclusion
that previous indications of individual differences in speech perception could well be
due to dialectal differences.

If there are differences, but each listener uses cues consistently for the vowels
and fricatives (for both speech sounds non-speech analogues), then there is no evidence
that these listeners have different strategies involving frequency and duration that
might be related to their experience with duration and frequency as speech cues for
these particular contrasts. If there are individual differences, then different possible
strategies on cue usage may be observed. Consistency in the treatment of vowels and
fricatives is part of a possible strategy.

Subjects’ environmental input will not be identical because they receive dif-
erent exemplars through life. But there may be subgroups of listeners with similar
experiences who have similar usage of frequency and duration cues for vowels and
fricatives. If all subjects are different, there may be a range of strategies among the
individuals. Or if differences are due to various levels of hearing ability, this should
be reflected in correlations of the hearing test results with individual differences in
fricatives or vowels.

4.2 Production

4.2.1 Talker

A male native northern Columbus, Ohio talker produced multiple tokens of
heed, hid, bath, and bass to be used as models for the values of the continua endpoints
in synthesized speech of the vowel and fricative tokens, and in sine wave analogues of
the vowels and narrow bandwidth noise analogues of the fricatives. The talker wore
a Shure SM10A head-mounted microphone, and was recorded onto a Sony DTC-790
DAT deck at a 44.1 kHz sampling rate with 16 bit quantization via a Symetrix SX202 mic preamp. The speech was digitally transferred to WAV format soundfiles with a Creative SBLive! soundcard. The sound files were down-sampled to 11.025 kHz for analysis because the Klatt synthesizer would be at this sampling rate. Since fricative noise and formants extend above the Nyquist frequency, certain measurements such as intensity would not be modeled correctly in synthesis if the same cut-off frequency was not used.

4.2.2 Procedure for talker

The word list was blocked by word, with fifteen repetitions of each word split over three blocks. Non-final tokens in each block were analyzed.

4.3 Stimuli for perception experiments

Entropic’s xwaves package was used for speech analysis. Ten equally spaced points for the vowels, starting at onset and ending at offset, and five equally spaced points for the fricatives were measured to map frequency contours. The first five formants of all vowels were measured, along with the first five energy band regions of the fricatives. Formant tracking with 1 ms frame step with a 25 ms Hamming window and LPC order of 18 for five formants was used for the vowels, with hand correction as necessary (occasionally for the higher formants). An autocorrelation method of DFT with 25 ms Hamming window with no liftering and an order of 14 was used to get the center frequencies of the fricatives. Hertz values were converted to ERB, averaged, and converted back to Hz. Averages were computed in ERB because this logarithmic auditory scale better models listeners’ frequency resolution than the linear acoustic Hz scale.
The Klatt synthesizer was used to produce synthetic versions of the words that are relatively speech-like, but with all values between tokens constant except for those under study. The sine wave analogs based on vowels and narrow bandwidth noise stimuli based on fricatives are not speech-like (although they can be interpreted as speech by some individuals), but use the same frequency and duration cues found in real speech. These non-speech tokens will be used to determine if the perception of what a subject considers to be speech is different from the perception of what a subject considers to be non-speech.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Cue continua:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel height [hi/id]</td>
<td>Frequency: F1</td>
</tr>
<tr>
<td>Fricative place [baeθ/s]</td>
<td>Duration: vowel length</td>
</tr>
<tr>
<td>Sine wave</td>
<td>Frequency: F5 noise band</td>
</tr>
<tr>
<td>Narrow bandwidth noise</td>
<td>Duration: fricative length</td>
</tr>
<tr>
<td></td>
<td>Frequency: lowest component</td>
</tr>
<tr>
<td></td>
<td>Duration: tone length</td>
</tr>
<tr>
<td></td>
<td>Frequency: highest component</td>
</tr>
<tr>
<td></td>
<td>Duration: noise length</td>
</tr>
</tbody>
</table>

Table 4.1: Cues involved in each continuum

Frequency was varied in four steps and duration in three steps in this two-dimensional stimulus continuum space, yielding 12 stimuli for each experiment. For the vowel height experiment, the first formant of the vowel ranged from its value in [i] to its value in [i] in four steps, and independently the duration of the vowel ranged in three steps from the duration of [i] to that of [i]. For the fricative place experiment, the fifth noise band ranged from the value of [θ] to the value of [s], and independently the duration ranged from [θ] to [s]. The non-speech vowel tokens were based on the vowel height stimuli, but the formants were replaced by time-varying
sinusoidal waves. Similarly, the narrow bandwidth noise stimuli were based on the fricative place stimuli.

ERB (Moore & Glasberg, 1983) average values are used to determine the average formants for the vowels and frequency bands for the fricatives at ten equally spaced time slices. As shown in the comparison of Figure 4.1 with Figure 4.2, although F2 varies in Hz more than F1, there is more of a perceptual change in F1 ERB values than in F2 between [i] and [ɪ].

![Graph](image1.png)

![Graph](image2.png)

Figure 4.1: Average formant contours in Hz (average ERB converted back to Hz) for [i] and [ɪ].

Obviously, frequency differences can be noticed at the onset of the second stimulus while duration differences might require hearing the entirety of the second stimulus, if changes in frequency slope are not noticed.
4.3.1 Speech tokens

The Delta system (Hertz, 1990) was used to create Klatt synthesizer parameters. The [h] portion for heed-hid was a 110 ms steady-state, devoiced, aspirated version of the first five formant values of the vowels, and could therefore be an aid in distinguishing the vowel tokens since F1 varies among tokens. Transition of F5 at the onset of the fricative might also indicate the place of the fricative in the bath-bass tokens. The Hz equivalent of the ERB average F5 transition from the vowel to [θ] and [s] was used until the onset of frication to make it constant for all the fricatives. The 10 ms transition from the final F5 value of the vowel to the different initial frequency values of the fricatives is short in relation to the duration of the fricative: 164, 228, or 296 ms. Vowel durations for the heed-hid stimuli were 162, 234, and 297 ms.

Although the talker devoiced his word-initial [b], it was voiced in the synthesizer parameters to facilitate its identification. The final [d] for the vowel tokens was voiced.
The overall average intensity of the vowels was measured from after the burst to before frication for the [æ] in *bath* and *bass*, and from after the end of aspiration to before closure for *heed* and *hid*. Intensity of the fricatives was measured from after the end of voicing (25 ms into the fricative) to the end of the fricative. Although small, subjects might hear a difference in intensity. Vowel intensity data was also used to find an F0 level that minimized intensity differences, as described below.

The frequencies of vowel formants F2 to F5 in *heed-hid* tokens (and F1 to F5 for the [æ] in the *bath-bass* tokens) were set to the Hz equivalents of the ERB average contours. The bandwidth of each vowel formant was constant across tokens. Thus, the only differences between the [i] and [i] endpoint tokens (and among the other tokens in between) were the F1 trajectories and the durations. Of course, changes in duration affected the rate of change of the formant and F0 contours.

The fricatives were given the same four lower noise bands, set to the Hz equivalents of the ERB average contours. This biases the perception toward [θ], since [s] typically does not have strong energy below 3500 Hz (Strevens, 1960). However, the amplitude was kept to [θ] values, because any value higher sounded too strong to be interdental. Looking at the talker’s spectrograms, the [θ] in *bath* is slightly affricated (i.e., there was not a complete closure, but there was a change in the intensity of frication). This was not modeled.

The F0 contour (H* L- L%) was modeled as a fall from 116 Hz to 70 Hz, with the fall starting 20% of the way into the vowel, and ending at 70 Hz for the entire final consonant. These values were chosen in order to space the harmonics evenly around the different F1 values to prevent differences in intensity. The voiceless fricatives would not have their intensities affected by F0, and had the same contour except...
that an F0 rise from 110 Hz was added from the release of the initial [b] of \textit{bath} and \textit{bass}, rising to 116 Hz at the vowel’s 20% point. Praat’s (Boersma & Weenik, 1996) intensity function was used to determine average intensity throughout the vowel or fricative segment, shown in Table 4.2. Loudness cannot be controlled for because it is a function dependent on frequency and duration. The differences in intensity are greater for the fricative portions, as F5 moves away from F4 into higher frequencies.

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Vowel</th>
<th>Fricative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speech</td>
<td>Non-speech</td>
</tr>
<tr>
<td>4</td>
<td>79.6</td>
<td>79.6</td>
</tr>
<tr>
<td>3</td>
<td>79.3</td>
<td>79.4</td>
</tr>
<tr>
<td>2</td>
<td>79.3</td>
<td>79.4</td>
</tr>
<tr>
<td>1</td>
<td>79.6</td>
<td>79.6</td>
</tr>
<tr>
<td>Dur.</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.2: Average vowel and fricative intensities in dB for tokens of different frequency and duration levels. These are not the intensities heard by listeners. For each of the four continua, word tokens were scaled to create a group average of 75 dB SPL.

Duration values were set as close as possible to the average values, limited by the fact that the number of milliseconds needed to be a multiple of nine (for the nine sections between the ten value points) for the vowels, and a multiple of four (between five value points) for the fricatives.

The first sets of continua were modified because three of the four endpoints did not sound like good examples of \textit{heed}, \textit{hid}, and \textit{bass}. This was not unexpected, as the F2 and higher formants were not exemplary of either [i] or [i], and [s] does not have lower frequency fricative energy. Phonetically trained listeners of various dialects, including one from Ohio, found that the \textit{hid} endpoint sounded like [hr̩d],
the *heed* endpoint was a poor example of [i] (too similar to [i]), and the *bass* endpoint was a poor example of [s] (too similar to [θ]). The continua were shifted one-fourth of the F1 frequency range toward the [i] side (i.e., vowel tokens with the highest F1 level were removed from the continuum, and a new level of tokens below the original lowest F1 level were added, using the same ERB distances as those between the other levels), and one-fourth of the F5 frequency range toward the [s] side (i.e., tokens with the lowest fifth noise band were removed, and a new level of tokens were added above the previous highest). Figure 4.3 displays the original and final F1 values for the vowels, while Figure 4.4 displays the F5 values for the fricatives.

![Figure 4.3: F1 contours for vowels in heed-hid tokens.](image-url)
Figure 4.4: Fifth noise band contours for fricatives in *bath-bass* tokens.

The remaining frequency values for the [i]-[i] vowel tokens (F2 through F5) are shown in Table 4.3, while the remaining frequency values for the [θ]-[s] fricative tokens (“F1” through “F4”) are in Table 4.4.

<table>
<thead>
<tr>
<th>Formant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>F5</td>
<td>3997</td>
<td>4055</td>
<td>4066</td>
<td>4055</td>
<td>4051</td>
<td>4047</td>
<td>4053</td>
<td>4038</td>
<td>4029</td>
<td>3982</td>
</tr>
<tr>
<td>F4</td>
<td>3495</td>
<td>3476</td>
<td>3454</td>
<td>3440</td>
<td>3413</td>
<td>3402</td>
<td>3406</td>
<td>3424</td>
<td>3475</td>
<td>3506</td>
</tr>
<tr>
<td>F3</td>
<td>2662</td>
<td>2647</td>
<td>2645</td>
<td>2645</td>
<td>2648</td>
<td>2628</td>
<td>2623</td>
<td>2607</td>
<td>2573</td>
<td>2447</td>
</tr>
<tr>
<td>F2</td>
<td>1971</td>
<td>1927</td>
<td>1924</td>
<td>1928</td>
<td>1916</td>
<td>1907</td>
<td>1893</td>
<td>1873</td>
<td>1826</td>
<td>1712</td>
</tr>
</tbody>
</table>

Table 4.3: Higher vowel formant frequencies in Hz.
The different duration values of the phones were also difficult to distinguish in pretests, resulting in many stimulus pairs that differ only in duration to always be classified as the same. To prevent this dimension from collapsing, the shorter tokens ([i] and [θ]) were shortened by one standard deviation from their respective means, and the longer tokens ([i] and [s]) were lengthened by one standard deviation from their respective means. The middle tokens were changed to the average lengths of the endpoints. The levels are shown in Figure 4.5.

The three duration values have linear increments because the psychophysical dimension of time has been difficult to determine. Some research suggests that time is a power function (Steiner, 1968; Abel, 1972), but others find no evidence that it is not linear (Bobko, Thompson, & Schiffman, 1977; Allan, 1983). In any case, the increments are above the JND at the duration levels involved (about 20-30 ms, but these studies were for single tones or noise, not embedded complex stimuli, although Bochner, Snell, and MacKenzie (1988) found no difference between the discrimination of speech or of tonal complex stimuli—10%–15% DL).

Spectrograms for the endpoints of the synthetic heed-hid continuum are shown in Figure 4.6, and the endpoints for bath-bath are shown in Figure 4.7.

Table 4.4: Lower fricative frequency poles in Hz.

<table>
<thead>
<tr>
<th>“Formant”</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>“F4”</td>
<td>3081</td>
<td>3064</td>
<td>3108</td>
<td>3137</td>
<td>3013</td>
</tr>
<tr>
<td>“F3”</td>
<td>2128</td>
<td>2109</td>
<td>2246</td>
<td>2143</td>
<td>2001</td>
</tr>
<tr>
<td>“F2”</td>
<td>1111</td>
<td>1319</td>
<td>1281</td>
<td>1047</td>
<td>1097</td>
</tr>
<tr>
<td>“F1”</td>
<td>392</td>
<td>426</td>
<td>492</td>
<td>434</td>
<td>548</td>
</tr>
</tbody>
</table>
Figure 4.5: Duration levels for the vowels of *heed-hid* and the fricatives of *bath-bass*.

Figure 4.6: Spectrograms of synthetic speech vowel tokens 41 (*hid*), left, and 13 (*heed*), right.
Figure 4.7: Spectrograms of synthetic speech fricative tokens 11 (bath), left, and 43 (bass), right. The fricative contrast is more easily seen in the narrowband stimuli shown in Figure 4.9.

4.3.2 Non-speech tokens

To create the sine wave stimuli for the heed-hid continuum, the Klatt parameters and the wavefile were input into a program written by Alex Francis and Howard Nusbaum, and modified by Keith Johnson. This takes the frequency information for the parameter file, and the amplitude information from the soundfile to create a sine wave for each of the five formants. For the bath-bass continuum, two Klatt parameter files were created—one that contained the fricative portion only, and one that contained [bæ] and the trailing voicebar into the onset of the removed fricative. The bandwidths of the fricatives were decreased to 10% of their former values. Only the [bæ] portion was replaced by sine waves. The F5 transition of the vowel was changed to the ERB average of the middle two (of the four) frequency transitions.

Spectrograms for the endpoints of the non-speech heed-hid continuum are shown in Figure 4.8, and the endpoints for bath-bath are shown in Figure 4.9. Sine wave speech (Remez, Rubin, Pisoni, & Carrell, 1981) sounds like electronic beeps, whistles, or tones, and narrow bandwidth fricatives sound like air flowing through
pipes or straws. Some people can hear the sine wave tokens as speech. So, to determine if speech is special, the synthetic speech tokens must be presented after the non-speech analogues in order not to prime subjects into hearing speech on the sine wave tokens that have the same frequency and duration values.

Figure 4.8: Spectrograms of non-speech vowel tokens 41 (hid), left, and 13 (heed), right.

Figure 4.9: Spectrograms of non-speech fricative tokens 11 (bath), left, and 43 (bass), right.
4.4 Listeners

Seventeen subjects from the northern Columbus area (including the talker who was reused as a listener months after his recordings) were recruited from introductory undergraduate linguistic classes, and were paid $40. Although students were informed that they would be listening to “ordinary” and “extraordinary” sounds, not “speech” and “non-speech”, it would not be surprising that they were biased towards speech. An individual would come at the same time on different days spanning one or two weeks.

4.5 Procedure for listeners

The main task for all experiments was AX discrimination, using a computer button box to record subjects’ same/different responses and reaction time. Subjects also identified all the tokens before and after the discrimination task to determine if they heard them as speech or not.

Subjects were tested in small groups wearing headphones. The experiments lasted for more than one session (four sessions for the experiments, and one for the hearing test). The order of the experiments were counterbalanced across subjects to control for learning effects. The sine wave sessions were always presented before the speech ones, to prevent the influence of the speech forms on the non-speech forms, but the order of vowel versus fricative tokens was counterbalanced.

Listeners wore Optimus Nova 71 headphones and underwent the experiment in a computer lab. The average background noise level penetrating the headphones was at 60 dB SPL ±2 dB, as measured by a RadioShack Sound Level Meter 33-2050, with fast C weighting. The average signal level was set to 75 dB SPL. The
Optimus headphones were chosen since they have a relatively flat response curve for the frequencies being tested, preventing a wide disparity in intensities due to the different frequency levels of F1 in the vowel tokens.

4.5.1 Questionnaires

Three questionnaires that were given to the listeners are reprinted in Appendix A. The first was filled out during recruitment, in order to select students from north Columbus who claim to have no speech or hearing impairments. The second was given at the first session to learn their caregivers’ places of residence. The post-test questionnaire was devised to determine if subjects are familiar with formant-based synthetic speech, as it has become more prevalent in society (e.g., weather reports), and if they had been exposed to sine wave speech.

4.5.2 Familiarization

The twelve tokens were played twice to listeners to familiarize them with the types of differences involved. For vowel and sine wave vowel tokens, listeners were informed that the differences were in the middle portion of the sounds, and at the end of the sounds for the fricative and narrow bandwidth fricatives. The orders of the two series were set to emphasize the frequency and the duration differences, with a 600 ms ISI.

4.5.3 Identification

Listeners then were presented all twelve tokens of one continuum (sine wave vowels, narrowband fricatives, synthetic vowels, or synthetic fricatives) in random order for the identification task, with a 1000 ms ISI. They were instructed to name the
tokens, if possible, or to describe them. They were informed it was acceptable to leave
the sheet blank, which about half did for their first session. A second identification
task was performed after discrimination, to determine if their perceptions had changed
during the course of the experiment.

4.5.4 Discrimination

The discrimination test was preceded by a practice set composed of ten pairs
from various parts of the continua (four identical, six off by one or more frequency
and/or duration steps).

Identical pairs were presented three times in each block of the test, to comprise
a total of 20% of the tokens. Non-identical pairs were presented once per block, and
were composed of three types. Thirty-six of the pairs differed only by frequency level
(e.g., 12 42), twenty-four of the pairs differed only by duration level (e.g., 21 23), and
seventy-two of the pairs differed by both frequency and duration (e.g., 11 43). There
were five blocks with token pairs presented randomly.

Subjects used an SRT response box, pressing the left-most button with the
index finger of their left hand to indicate tokens were the “same”, and pressing the
right-most button with the index finger of their right hand to indicate tokens were
“different.” All subjects were right handed. Tokens within each pair were presented
with a 300 ms ISI,\textsuperscript{5} and a 1200 ms inter-trial interval (with a maximum response
time set to 3 s). Subjects were given feedback on their accuracy and reaction time.
Subjects could rest after each block.

\textsuperscript{5}This value was chosen because Pisoni (1973) found highest $d'$ scores at a value of 250 ms ISI for
synthetic vowel discrimination, but at higher ISI for synthetic (stop) consonants (250, 500 ms, and
above).
4.5.5 Hearing test

A Maico MA-16 audiometer calibrated to the ANSI S3.6-1969 standard was used to obtain audiograms of the listeners in a sound attenuated booth. An unmasked presentation of tones from 125 to 8000 Hz were played by auditory conduction to the left and right ears. If one or more listeners has a high threshold at 125, 250, and 500 Hz tones (which encompass the range of F1 for [i]-[ɪ]) or at 3, 4, and 6 kHz test tones (which encompass the range of F5 energy region for [θ]-[s]), then the offsets can be used in a regression analysis to determine if hearing loss affects the results.

4.5.6 Forced choice identification

After the hearing test, listeners performed a Ganong up-down forced-choice identification task (Ganong & Zatorre, 1980). The four stimulus sets were played in the same order as in the previous sessions for the subject (non-speech vowel, non-speech fricative, vowel, fricative; or non-speech fricative, non-speech vowel, fricative, vowel). The labeled endpoints of a continuum were played in alternation, twice (e.g., 41, 13, 41, 13), and then the entire continuum was played from the first presented endpoint (e.g., 41) to the second (e.g., 13). The endpoints were played again in reverse order (e.g., 13, 41, 13, 41), and the entire continuum were played from the first presented endpoint (e.g., 13) to the second (e.g., 41). The order of stimuli had each duration level of a frequency played before advancing to the next frequency level. Listeners were forced to choose one label or the other (e.g., heed or hid). The average identification will be taken from the “up” and “down” sequence for each continuum. Since the stimuli are two-dimensional (with frequency and duration...
changes) a single cross-over point cannot be computed, but the divided spaces are provided in Appendix C.

### 4.5.7 Commutation test

A commutation test (Healey, 1975) was performed in the same session as the hearing test. This task involves recording the subject reading a list of words, starting the tape at random, and then playing back the tokens to the subject for identification. Subjects were recorded using the same procedure as the talker. This task tests the subject’s production and perception of minimal pairs.

In this instance, the (General American) distinctions of /u/-/u/-/o/ and /a/-/ʌ/ before /l/, and of /ɑ/-/ɔ/ were tested. These mergers were investigated by Thomas (1989, 1996, 2001) in Columbus and neighboring areas. From Thomas’s results, the expected findings of this study are that /ʌ/ should be merged with /w/, and /u/-/ɔ/ should be merged before /ʌ/, while /i/-/i/ before /l/ should not be merged. The words tested are in Table 4.5, but subjects also read /h_d/ and /b_t/ tokens in order to compare their vowel spaces. This wordlist is shown in Table 4.6.

Subjects read wordlists onto tape, beginning and ending the recording with the vowel space tokens. Words for the commutation task were repeated ten times, randomized within two blocks. Subjects had access to a visual metronome initially set to 40 bpm to keep them at a slow, steady pace. During practice, subjects were allowed to raise the rate to one more comfortable to them, and some did by up to 10 or 15 bpm.

---

6Post-vocalic /l/ vocalizes in most of the productions by these talkers.
### Table 4.5: Commutation word groups

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/-/ɪ/-/ɛ/</td>
<td>heed hid head</td>
</tr>
<tr>
<td>[ɪ]-[ɛ]</td>
<td>pin pen</td>
</tr>
<tr>
<td>/ɪl/-/ɪl/</td>
<td>feel fill</td>
</tr>
<tr>
<td>/ʌ/-/w/</td>
<td>which witch</td>
</tr>
<tr>
<td>/ɑ/-/ɔ/</td>
<td>cot caught</td>
</tr>
<tr>
<td>/ʌɪ/-/ɔɪ/</td>
<td>tour tore</td>
</tr>
<tr>
<td>/ɑl/-/ʌl/</td>
<td>doll dull</td>
</tr>
<tr>
<td>/ʌl/-/ʊl/-/ɔl/</td>
<td>pool pull pole</td>
</tr>
</tbody>
</table>

### Table 4.6: Vowel space and fricative tokens (with rhymes)

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>Words</th>
<th>Fillers</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>heed beat</td>
<td>he’d beet</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>hid bit</td>
<td>aide</td>
</tr>
<tr>
<td>/e/</td>
<td>aid bait</td>
<td></td>
</tr>
<tr>
<td>/ɛ/</td>
<td>head bet</td>
<td></td>
</tr>
<tr>
<td>/æ/</td>
<td>had bat</td>
<td></td>
</tr>
<tr>
<td>/ɑ/</td>
<td>odd bot (roBOT)</td>
<td></td>
</tr>
<tr>
<td>/ɔ/</td>
<td>awed bought</td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>owed boat</td>
<td></td>
</tr>
<tr>
<td>/u/</td>
<td>hood put</td>
<td></td>
</tr>
<tr>
<td>/ʊ/</td>
<td>who’d boot</td>
<td></td>
</tr>
<tr>
<td>/ʌ/</td>
<td>HUD (mud) but</td>
<td>heard</td>
</tr>
<tr>
<td>/ɜ/</td>
<td>herd Bert</td>
<td>Hyde byte</td>
</tr>
<tr>
<td>/aɪ/</td>
<td>hide bite</td>
<td></td>
</tr>
<tr>
<td>/aʊ/</td>
<td>how’d bout</td>
<td></td>
</tr>
<tr>
<td>/ɔɪ/</td>
<td>Hoid (void) Boyt (exPLOIT)</td>
<td></td>
</tr>
<tr>
<td>/θ/</td>
<td>bath</td>
<td></td>
</tr>
<tr>
<td>/s/</td>
<td>bass (gas)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6: Vowel space and fricative tokens (with rhymes)
CHAPTER 5

RESULTS

This chapter presents the results of the study. Section 5.1 provides background information for the subjects. Section 5.2 details the discrimination data (as count correct responses in 5.2.1, and response reaction time in 5.2.2), while in Section 5.3 the results of the identification and Ganong tasks are presented. The principal analyses in the study are reported in Sections 5.4 and 5.5, which report the results of multidimensional scaling (MDS) (see, e.g., Kruskal & Wish, 1978; Borg & Groenen, 1997; Cox & Cox, 2001) for groups and individual listeners. The individual differences found in these MDS analysis are the focus of this dissertation. Possible explanations for these perceptual differences will be investigated in Section 5.6, focusing on differences in audition, production, and subdialectal membership.

5.1 Subject information

All subjects reported no history of hearing disorders (although subject 10 reported that her father was partially deaf). Subjects ranged in age from 18 to 22, with an average of 20 years of age. All were White, except for subject 3 who has Black and White parents.
Table 5.1 lists subjects’ age, sex, main residence location, and childhood places of the parents. Fewer than half the subjects had moved once to a different part of the northern Columbus area. The main place of residence for a subject from 0–7 until about 12 years of age was listed, as this age range has been found to be the years for dialect acquisition (Chambers, 1992). Subjects 2 and 17 attended the same elementary and middle schools, while 9 and 15 attended the same middle school.

<table>
<thead>
<tr>
<th>S</th>
<th>Age</th>
<th>Sex</th>
<th>Location</th>
<th>Father</th>
<th>Mother</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>M</td>
<td>Worthington</td>
<td>downstate NY</td>
<td>OK; AR; Cincinnati</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>M</td>
<td>Worthington</td>
<td>Columbus</td>
<td>Columbus</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>F</td>
<td>Westerville</td>
<td>MO; IL</td>
<td>NW PA</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>M</td>
<td>Upper Arlington</td>
<td>Portsmouth; Cincinnati</td>
<td>Portsmouth</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>M</td>
<td>Clintonville</td>
<td>Toledo</td>
<td>Toledo</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>F</td>
<td>Hilliard</td>
<td>SE MI</td>
<td>W WV</td>
</tr>
<tr>
<td>7</td>
<td>19</td>
<td>F</td>
<td>Upper Arlington</td>
<td>GREECE</td>
<td>GREECE</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>M</td>
<td>Dublin</td>
<td>Columbus</td>
<td>Columbus; WI</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
<td>F</td>
<td>Clintonville</td>
<td>Marietta; Dayton; Toledo; Columbus; w WV</td>
<td>Mansfield; Shelby</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>F</td>
<td>NE Columbus</td>
<td>Toledo; Columbus</td>
<td>Columbus</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>M</td>
<td>Westerville</td>
<td>Columbus</td>
<td>Bellaire; Bethesda; Columbus</td>
</tr>
<tr>
<td>12</td>
<td>20</td>
<td>M</td>
<td>Worthington</td>
<td>ENGLAND; NW NY</td>
<td>Mansfield</td>
</tr>
<tr>
<td>13</td>
<td>18</td>
<td>F</td>
<td>N Columbus</td>
<td>Cleveland; Columbus</td>
<td>Cleveland</td>
</tr>
<tr>
<td>14</td>
<td>20</td>
<td>F</td>
<td>Clintonville</td>
<td>E PA; NE OH</td>
<td>Warren</td>
</tr>
<tr>
<td>15</td>
<td>21</td>
<td>F</td>
<td>Clintonville</td>
<td>N WV</td>
<td>E. Liverpool</td>
</tr>
<tr>
<td>16</td>
<td>19</td>
<td>M</td>
<td>Gahanna</td>
<td>Akron</td>
<td>Warren</td>
</tr>
<tr>
<td>17</td>
<td>22</td>
<td>M</td>
<td>Worthington</td>
<td>New Richmond</td>
<td>N KY</td>
</tr>
</tbody>
</table>

Table 5.1: Subject background information. For the areas in and around northern Columbus, Dublin is in the northwest, Worthington is in the north, Westerville is in the northeast, Hilliard is in the west, Clintonville is central, Upper Arlington is between Hilliard and Clintonville, and Gahanna is in the east. Parents’ childhood residences are listed as cities in Ohio, portions of other U.S. states, and other countries.
As discussed by Thomas (1989), dialect mixture is evident in Columbus, which is near the boundary between north and south midland dialects, and receives immigrants from all of Ohio, including the inland north dialect. This influx is demonstrated by the home towns of the subjects’ parents. Both parents were reported as having spent most or all of their childhoods outside of Columbus by all but four of the subjects. The fathers of subjects 8 and 11 lived their entire lives in Columbus, and subject 10’s mother did as well. Only subject 2 has both parents who have lived in Columbus their entire lives.

5.2 Discrimination task

5.2.1 Percent correct responses

Percent and count correct values were used instead of $d'$ for duration and fricative tokens because it is unclear what the “false alarm” corresponds to. If a subject correctly discriminates a pair that differs on frequency as being different, it can presumably be classified as a “hit”. But if the subject mistakenly decides that an identical pair is different, should it be classified as a false alarm if it is unknown why it was not judged to be the same? Did the subject decide that there was a difference in frequency, or a difference in duration? The corresponding problem occurs for finding the $d'$ of duration. Percent and correct values do not depend on these types of relations.

Figure 5.1 displays the percent correct responses for token pairs that differed only by frequency or duration for the (sine wave) non-speech vowels, (narrowband) non-speech fricatives, (formant synthesis) speech vowels, and (formant synthesis) speech fricatives. Subjects as a whole appear to have equal abilities in frequency
and duration discrimination for the fricative tokens, but make more duration mistakes for the vowels, as illustrated by the relationships to the \( x = y \) line. There also appears to be a wider range of abilities for duration discrimination than for frequency discrimination in the vowels.

Since there were two presentation orders (vowel tokens presented on first and third sessions, or on second and fourth), a univariate GLM analysis was computed to determine if there were any differences between the two groups on their sensitivity to duration and frequency (using count correct values, with results displayed as percentages). Effects of speech/non-speech tokens, and vowel/fricative were analyzed as well, and the estimated marginal means will be shown for the relevant tests.

There was no main effect of presentation order on sensitivity between the groups \((F(1, 15) = .410, p = .531)\). The only interaction involving presentation order was whether the stimuli were vowel-like or fricative-like \((F(1, 15) = 6.104, p < .05)\), with results shown in Figure 5.2. This is explained by a learning effect on whether a group received vowels or fricatives first. A post-hoc test using Tukey HSD\((N = 4, df = 15)\) on the estimated marginal means was performed to compare levels in the interaction. Subjects that received fricatives first (with a mean score of 82.5% correct discrimination) had significantly lower scores on them than for vowels (90.7%), and their fricative scores were significantly lower than those of subjects who received vowels first (lower than fricatives, 85.0%, and vowels, 89.9%).

All other factors had main effects and significant interactions. First the main effects will be discussed, then the two- and three-way interactions, and then finally interactions involving subjects.
Figure 5.1: Percent correct discrimination on pairs that differ only by frequency (x-axis) or duration (y-axis) by subject. (a) Sine wave vowels. (b) Narrowband fricatives. (c) Synthetic vowels. (d) Synthetic fricatives.
Figure 5.2: Percent correct values for the segment-type*order interaction.

The type of cue that differed within a pair (frequency, duration, or both) was significant \(F(2, 30) = 101.496, p < .05\), and each comparison was significant in Tukey HSD\(N = 3, df = 30\). Results are displayed in Figure 5.3. Scores for duration difference alone (with mean 72.6%) are significantly lower than for frequency alone (84.4%), which in turn are significantly lower than for pairs that differ in both frequency and duration (93.2%).

There are significant effects for speech-type and for segment-type. Speech pairs are discriminated correctly significantly more often than non-speech pairs (88.5% vs. 85.6%, \(F(1, 15) = 39.058, p < .05\), shown in Figure 5.4, and vowel pairs are significantly higher than fricative pairs (90.3% vs. 83.8%, \(F(1, 15) = 23.541, p < .05\), shown in Figure 5.5. The interaction between these factors is also significant \(F(1, 15) = 23.838, p < .05\), with values displayed in Figure 5.6. Tukey HSD\(N = \)
Figure 5.3: Percent correct values for different speech cues: duration, frequency, and both duration and frequency.

4, \( df = 15 \) reveals that vowel speech pairs (89.0%) are not significantly different from vowel non-speech pairs (91.6%), and vowel speech pairs are not significantly higher than fricative speech pairs (88.0%). Only non-speech fricatives (79.6%) are significantly different from the others, so it appears that the poor discrimination of these stimuli causes the speech/non-speech and vowel/fricative effects.

Cue-type interacts with both speech-type \( (F(2, 30) = 21.532, p < .05) \) and segment-type \( (F(2, 30) = 25.332, p < .05) \). The only comparisons between cue- and speech-type that are not significant with Tukey HSD\( (N = 6, df = 30) \) are speech and non-speech tokens that differ by both frequency and duration (93.9% and 92.4%, respectively), and speech and non-speech tokens that differ by frequency (84.5% and 84.4%, respectively), illustrated in Figure 5.7. Duration discrimination,
Figure 5.4: Percent correct values for non-speech and speech tokens.

Figure 5.5: Percent correct values for vowels and fricatives.
however, is distinct between speech (78.1%) and non-speech (67.2%). Speech and non-speech tokens have significantly lower scores with duration than with frequency, and significantly lower scores with either cue singularly than both together.

For cue- and segment-type comparisons, shown in Figure 5.8, the comparisons that are not significantly different with Tukey HSD\( (N = 6, \, df = 30) \) include vowels which differ by both frequency and duration (96.9%) that are not distinct from vowels with frequency cues (91.4%), and the duration of fricatives (76.4%) with the frequency of fricatives (77.4%). One comparison that goes against the main effect that vowels are better discriminated than fricatives involves duration. Correct discrimination by duration for vowels (68.9%) is significantly lower than for fricatives (76.4%). Otherwise, vowels are distinguished better than fricatives for frequency (91.4% vs. 77.4%) and both frequency and duration (96.9% vs. 89.5%). These two
interactions indicate that duration is the cue that differs in discriminability between speech and non-speech tokens, that adding duration to frequency does not aid in the discrimination of vowels, but that duration is similar to frequency in distinguishing fricatives.

A three-way interaction between cue-, speech-, and segment-type was significant ($F(2, 30) = 3.950, p < .05$), shown in Figure 5.9. There is no distinction in the discrimination accuracy of pairs that differ by duration between non-speech vowels (64.6%) and non-speech fricatives (69.8%), but there is for the corresponding speech tokens (83.0% vs. 73.2%). For speech tokens, there are no distinctions between speech vowels and fricatives for frequency (87.5% vs. 81.5%) or for frequency and duration (95.0% vs. 92.9%). There is a distinction between speech and non-speech tokens for fricatives that differ by frequency and duration (92.9% vs. 86.0%), and by frequency
Figure 5.8: Percent correct values for the cue-*segment-type interaction.

only (81.5% vs. 73.4%). Speech vowels that differ by frequency are significantly worse than non-speech vowels (87.5% vs. 95.4%), but speech is significantly better than non-speech for vowels by duration (73.2% vs. 64.6%). Duration and frequency are not significantly different for non-speech fricatives (69.8% vs. 73.4%) or for speech fricatives (83.0% vs. 81.5%). Speech vowels benefit from adding duration to frequency (87.5% for frequency alone vs. 95.0% for frequency and duration). But having both frequency and duration (98.8%) was not significantly better than frequency alone (95.4%) for non-speech vowels.

Three three-way interactions included subjects. Using Figure 5.1 as a guide will aid in interpreting the differences among subjects for frequency or duration discrimination accuracy. The cue-type, segment-type and subject interaction had only
one significant difference between subjects with Tukey HSD ($N = 102, df = 30$). Subject 3 was significantly worse than all subjects except 6 and 15 in discriminating vowel duration. However, examining within-subject comparisons finds that many subjects had different levels of accuracy for the different types of stimuli. As expected, subject 3 was significantly worse in discriminating vowel duration than all other segment and cue combinations. Frequency and duration cues for vowels led to significantly higher scores than duration alone for subjects 1, 2, 6, 9, and 11–15. Vowel frequency cues were also higher than duration for subjects 6, 9, 11, 12, and 15. Frequency and duration cues for fricatives scored higher than duration alone for subjects 11 and 16. Subject 4 scored significantly higher on fricatives that had duration and frequency
cues than just frequency, while subject 12 was significantly better on fricative frequency than vowel frequency. Subjects 5, 7, 8, and 17 had no significant differences for any cue and segment combinations.

The speech-type, segment-type, and subject interaction had a couple of subjects significantly different from others with Tukey HSD($N = 68$, $df = 31.682$). Subject 3 was significantly worse on speech vowels than all other subjects except 9 and 15. Subjects 3 and 12 were both significantly worse than all other subjects except 1, 2, 6, 7, 11, 13 on non-speech fricatives. Very few subjects had within-subject interactions. Subjects 12 and 16 both had non-speech fricatives with significantly smaller scores than all other combinations. Subject 3’s score for speech fricatives was significantly higher than for non-speech fricatives and speech vowels.

There were numerous significant differences for the cue-type, speech-type, segment-type, and subject interaction using Tukey HSD($N = 8772$, $df = 30$). Stimulus pairs that had both frequency and duration differences had only one significant difference: Subject 3 scored lower than all subjects except 12 and 15 on the non-speech fricatives. There were no significant differences on both cue pairs for non-speech vowels, speech vowels, or speech fricatives. There were also no significant differences for tokens that differed in frequency for speech and non-speech vowels. For speech fricatives, subject 10 had significantly lower frequency scores than 1, 3, 5, 14, and 17. For non-speech fricative frequency, subject 12 scored lower than all subjects except 2, 3, 4, 9, and 10; subjects 3 and 4 scored lower than 1, 7, 8, 11, and 15; subject 9 had the same results as 3 and 4 but scored lower than 14 and 17 as well; and subject 7 scored higher than 2–4, 9, 10, and 12. Duration accuracy had significant subject differences.

---

7The non-integral degrees of freedom is due to computation via the Satterthwaite method.
for all four types of stimuli. For speech vowels, subject 3 scored lower on duration than all subjects, 9 scored lower than 4, 5, and 8, while 15 scored lower than 8. For non-speech vowels: 3 scored lower on duration than all subject except 6, 12, and 15; 6 scored lower than all except 3, 9, 11, 12, and 15; 15 scored lower than 4, 5, 7, 8, 10, 13, 16, and 17; 12 scored lower than 10, 16, and 17; and both 9 and 11 scored lower than 16. Duration scores for speech fricatives were significantly lower for subject 11 than for 1, 2, 4–8, and 17, and lower for subject 14 than 4, 7, and 17. In regards to duration scores for non-speech fricatives: Subject 2 scored significantly higher than 3, 8, 12, and 14–17; 14 scored lower than 2 and 7; 12 has the same results as 14 but also scored lower than 4 and 5; 15 has the same results as 12 but also scored lower than 1, 6, 9, and 13; and 16 is the same as 15 but also lower than 10 and 11.

A hierarchical cluster analysis using the nearest neighbor method over squared Euclidean distance was performed on subjects' arcsine transformed percent correct responses for all cues, segment types, and speech types. The dendrogram is shown in Figure 5.10. Two groups of subjects that were most similar to each other were one group of subjects 5 and 8, and another group of 2 and 4. Subject 3’s scores were the most different from all others.

Many within-subject interactions were significant as well, and are shown in Tables 5.2–5.4. Examining how a particular subject is more adept at using certain cues in one environment over another will take place in the discussion chapter.

Individual differences will be explored more thoroughly in the multidimensional scaling section. The next section investigates RT differences.
Figure 5.10: Dendrogram showing subject clusters of similar responses. Subjects that are more similar to each other join further to the left.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-speech</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowels</td>
<td>1–6, 8, 9, 11–15</td>
<td>-</td>
<td>1–15, 17</td>
<td>-</td>
</tr>
<tr>
<td>Fricatives</td>
<td>15, 16</td>
<td>2</td>
<td>8, 11-17</td>
<td>2, 4, 9, 12</td>
</tr>
<tr>
<td>Speech</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowels</td>
<td>3, 9, 11, 13</td>
<td>-</td>
<td>2, 3, 6, 9, 11–13, 15, 17</td>
<td>-</td>
</tr>
<tr>
<td>Fricatives</td>
<td>11, 14</td>
<td>4</td>
<td>11, 14</td>
<td>2, 4, 6, 10, 13</td>
</tr>
</tbody>
</table>

Table 5.2: Subjects who had significant within-subject differences between their scores for frequency, duration, and both frequency and duration cues
<table>
<thead>
<tr>
<th>Levels</th>
<th>Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-speech &lt; Speech</td>
</tr>
<tr>
<td>Vowels</td>
<td></td>
</tr>
<tr>
<td>Dur.</td>
<td>6, 12</td>
</tr>
<tr>
<td>Freq.</td>
<td>-</td>
</tr>
<tr>
<td>Both</td>
<td>-</td>
</tr>
<tr>
<td>Fricatives</td>
<td></td>
</tr>
<tr>
<td>Dur.</td>
<td>8, 12, 15, 16, 17</td>
</tr>
<tr>
<td>Freq.</td>
<td>3, 9, 12</td>
</tr>
<tr>
<td>Both</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.3: Subjects who had significant within-subject differences between their scores for speech and non-speech

<table>
<thead>
<tr>
<th>Levels</th>
<th>Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowels &gt; Fricatives</td>
</tr>
<tr>
<td>Non-speech</td>
<td></td>
</tr>
<tr>
<td>Dur.</td>
<td>16</td>
</tr>
<tr>
<td>Freq.</td>
<td>2–6, 9, 12–14, 16</td>
</tr>
<tr>
<td>Both</td>
<td>3</td>
</tr>
<tr>
<td>Speech</td>
<td></td>
</tr>
<tr>
<td>Dur.</td>
<td>-</td>
</tr>
<tr>
<td>Freq.</td>
<td>2</td>
</tr>
<tr>
<td>Both</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.4: Subjects who had significant within-subject differences between their scores for vowels and fricatives
5.2.2 Response RT

For each of the four stimuli types (sine wave vowel, narrowband fricative, synthetic vowel, synthetic fricative), median RT values were taken for each pair of identical stimuli, and were averaged to get a subject’s base reaction time. Since subjects performed the four tasks over many days, there may have been differences in speed that need to be controlled. Median RT values were taken for each duration, frequency, and frequency and duration pair, and subtracted from the average same pair RT. Similarly to the count correct data, possible effects of presentation order, speech/non-speech tokens, and vowel/fricative were analyzed as well. A second analysis using RT starting at the offset of the second stimulus achieved the same levels of significance, and will not be presented.

Figure 5.11 displays individual reaction times for each of the four stimulus types (after subtracting each subject’s own average RT value for the “same” pairs). Theoretically, subjects should have the most difficulty in discriminating tokens that are identical, with the task becoming easier with increasing amounts of differences between tokens (see, e.g., literature reviews by Welford, 1960; Shepard, Kilpatric, & Cunningham, 1975; Nosofsky, 1992). So, token pairs that subjects believe to differ only slightly should have reaction times slightly faster than identical pairs (here set to 0 ms, and indicated by dotted lines), and pairs that are perceived to differ greatly should have much faster reaction times. However, Figure 5.11 shows that some subjects take longer to discriminate pairs that differ only by duration than those that are identical. Subjects were almost always faster on distinguishing stimuli by frequency than by duration, shown by the reaction times being above the $x = y$ line.
Figure 5.11: Normalized RT for pairs that differ only by frequency (x-axis) or duration (y-axis) by subject. (a) Sine wave vowels. (b) Narrowband fricatives. (c) Synthetic vowels. (d) Synthetic fricatives.
As was found for the percent correct data, there was no main effect of group presentation order \( (F(1, 15) = .653, p = .432) \). There was an interaction of presentation order with speech type \( (F(1, 15) = 5.448, p < .05) \), shown in Figure 5.12. Subjects who received fricative tokens after vowel tokens had slower normalized RTs (-106 ms) on average than subjects who received fricatives first (-164 ms). This may be due to training, because subjects who received non-speech fricatives first were also exposed to the sine wave portion preceding each narrowband fricative, while subjects who received the non-speech vowels first were not exposed to any type of fricative stimuli.

![Figure 5.12: Normalized RT for the speech-type*order interaction.](image)
Only the type of cue that differed within a pair had a main effect \(F(2, 30.01) = 55.185, p < .05\). Pairs that differed by duration (-58 ms) yielded RTs that were significantly slower (using Tukey HSD\(N = 3, df = 30.01\)) than those that differed by frequency (-154 ms) or differed by frequency and duration (-165 ms), as shown in Figure 5.13. This differs from the percent correct data (cf. Figure 5.3) in that there is no difference in RT between pairs that differ by frequency and pairs that differ by frequency and duration, but there may be a ceiling effect on speed.

![Figure 5.13: Normalized RT for different speech cues: duration, frequency, and both duration and frequency.](image)

Unlike the count correct data, there was no overall significant difference between speech and non-speech \(F(1, 15) = 2.012, p = .177\) or vowels and fricatives.
\((F(1,15) = 1.964, \ p = .181)\), but they each had an interaction with cue-type, and there was an interaction involving all three factors.

The significant interaction \((F(2, 30.006) = 11.887, \ p < .05)\) between cue-type and speech-type is shown in Figure 5.14. Tukey HSD\((N = 6, \ df = 30.006)\) shows that the only comparisons that are not significant are between non-speech pairs that differ by frequency and duration (-155 ms), non-speech pairs that differ by frequency (-151 ms) and speech pairs that differ by frequency (-156 ms). Comparing these results with the count correct data (cf. Figure 5.7) reveals again that there is no speed benefit for combined duration and frequency cues over frequency alone (but this is true only for non-speech in this case) that corresponds to the improvement in performance. Also, there is a difference in speed between speech and non-speech for the combined cues that does not show up in performance accuracy.

Figure 5.14: Normalized RT for the cue-\(\text{*speech-type}\) interaction.
Figure 5.15 displays the interaction between cue-type and segment-type \((F(2, 30.002) = 38.105, p < .05)\). All comparisons with Tukey HSD\((N = 6, df = 30.002)\) were significant except that there was no difference between vowels (-171 ms) and fricatives (-158 ms) for pairs that differed by both frequency and duration, and no difference between frequency pairs and duration and frequency pairs for vowels (-169 ms and -171 ms, respectively) or fricatives (-139 ms and -158 ms, respectively). Besides the differences between RT and performance discussed previously that may be due to speed ceiling effects, another difference is that there is a faster RT for fricatives that differ by frequency than by duration, but no corresponding performance increase (cf. Figure 5.8).

![Figure 5.15: Normalized RT for the cue-segment-type interaction.](image-url)
The three-way interaction with cue-type, segment-type, and speech-type was significant ($F(2, 30.005) = 6.765, p < .05$). The results are shown in Figure 5.16. RTs between -150 ms and -180 ms are not significantly different from each other, and RTs between -120 ms and -155 ms are not significantly different from each other (i.e., frequency and duration is not faster than frequency alone, the main effect). However, fricatives and vowels always have significant differences in RTs for non-speech, while that is only true for duration pairs in speech. Comparing these results with the percent correct data shows many differences, but the overall pattern of increased performance with faster reaction time still exists (cf. Figure 5.9).

![Figure 5.16: Normalized RT for the cue-*segment-*speech-type interaction.](image-url)
There were interactions involving subjects and cue-type \((F(30, 22.403) = 2.913, p < .05)\), subjects, cue-type, and segment-type \((F(30, 30) = 2.990, p < .05)\), and subjects, speech-type, and segment-type \((F(15, 34.164) = 34.035, p < .05)\). Individual RT differences will not be discussed, but individual differences resulting from multidimensional scaling will be presented in Sections 5.4 and 5.5.

5.3 Identification tasks

Although there were significant differences in discrimination, whether there were differences in identification of the stimuli is an important follow-up question. Many perceptual studies assume the endpoints are good examples of phones, but do not test if their listeners agree. If there is no agreement by listeners on the identity of the endpoints, then there is no reason to expect them to have similar discrimination patterns. There were two types of identification tasks in this experiment. One had an open response set in which the subjects described or tried to identify the tokens before and after each discrimination trial. The other was a Ganong up-down forced-choice identification task that occurred after the subjects had finished their discrimination trials.

5.3.1 Identification

Before and after each discrimination trial, subjects identified each of the twelve tokens presented in random order using an open set, or described them if identification was not possible. Tables 5.5 and 5.6 show the general description they had of the stimuli, what words they thought they heard, and whether they described any frequency or duration differences among the stimuli. The words “high,” “higher,” “low,” “lower,” “rising,” and “falling” were interpreted as reference to differences in
frequency, while “short,” “shorter,” “long,” and “longer” were interpreted as reference to differences in length. Subjects 4, 6, and 7 indicated they could hear the sine wave vowel stimuli as words, most of which contained a high front vowel or palatal glide. Subject 1 could interpret the narrowband fricatives as word-like as well, but did not hear a fricative. Subject 13 failed to hear words for the synthetic vowels, and she along with subject 9 also failed to hear the synthetic fricative tokens as words. Subject 14 identified the synthetic vowel stimuli as vocal, but nothing more definite than humming.

All subjects who identified synthetic vowel tokens as words indicated they heard [i] except for 4 and 17, and all heard [i] except for 8, 16, and 17. However, most subjects did not hear [s] in the synthetic fricative words except for 1, 2, 3, 8, and 14, but all heard [θ] except for 4, 6, 7, and 10.

Comparing the number of subjects who described stimuli with frequency or duration terms while listening to the presentation of the 12 tokens yields a significant result for vowels. As shown in Tables 5.5 and 5.6, twelve subjects report hearing frequency differences for the non-speech vowel stimuli, but only two did so for the speech vowel stimuli ($\chi^2(1, N = 34) = 12.143, p < .05$). Counts of duration descriptions were not significantly different between non-speech vowels and vowels ($\chi^2(1, N = 34) = .472, p = .492$). Neither counts of frequency descriptives ($\chi^2(1, N = 34) = .515, p = .473$) nor duration descriptives ($\chi^2(1, N = 34) = .118, p = .732$) were different between non-speech fricatives and speech fricatives. Of course, a subject may have heard frequency differences in speech vowels but neglected to mention them, but this explanation fails to account for the fact that the number of
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>speech-like</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>what</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>computer beep</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>cymbals/air through straw/monkey hiss</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>3</td>
<td>synthesizer/horn</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>air</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>here</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>electronic beep</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>air through straw</td>
<td>-</td>
<td>-</td>
<td>dur</td>
</tr>
<tr>
<td>6</td>
<td>horn/computer voice</td>
<td>you</td>
<td>freq</td>
<td>-</td>
<td>hissing/air through straw</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>word-like</td>
<td>eee, uh, duh, ew</td>
<td>freq</td>
<td>-</td>
<td>ventilator/labored breathing</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>8</td>
<td>beep</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>puff of air</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>cymbal</td>
<td>-</td>
<td>-</td>
<td>dur</td>
</tr>
<tr>
<td>10</td>
<td>mechanical beeping</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>wheezing</td>
<td>-</td>
<td>-</td>
<td>dur</td>
</tr>
<tr>
<td>11</td>
<td>beep</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>whistle</td>
<td>-</td>
<td>freq</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>horn/bird</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>-</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>14</td>
<td>electronic music/chirps</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>dog</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>air through straw</td>
<td>-</td>
<td>freq</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>computer sound</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>computer whoosh</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.5: Subject description of non-speech stimuli, with identified words and whether stimuli were described in terms of frequency and duration
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>heed, heat, hid, hit</td>
<td>-</td>
<td>-</td>
<td>bath, bass, bash, bat</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>human speech</td>
<td>heat, heap, he, heath, hid, hey</td>
<td>-</td>
<td>-</td>
<td>human speech &amp; static/air</td>
<td>path, pass, paft, pafls</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>male voice</td>
<td>heat, hih-eat, hay-eat, hate</td>
<td>-</td>
<td>dur</td>
<td>static</td>
<td>bath, bass, bassth, baff, bafth</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>hid, person</td>
<td>-</td>
<td>-</td>
<td>computer word with ending</td>
<td>bat</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>synthesized voice</td>
<td>he’d, he, hid, hit</td>
<td>freq</td>
<td>dur</td>
<td>computer &amp; ending noise</td>
<td>bath</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>6</td>
<td>deep voice</td>
<td>heed, heat, heaved, hid</td>
<td>-</td>
<td>-</td>
<td>human/frog &amp; air through tube</td>
<td>bath</td>
<td>-</td>
<td>dur</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>heed, heat, hid, hit, head, het</td>
<td>-</td>
<td>-</td>
<td>word with scratching</td>
<td>baff, bat</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>peed, peat, peaved, paid, pate, ped, pew’d</td>
<td>-</td>
<td>-</td>
<td>word with hissing</td>
<td>bath, bass, baths, bats, mat</td>
<td>-</td>
<td>dur</td>
</tr>
<tr>
<td>9</td>
<td>electronic voice</td>
<td>heed, hit, hayed, hate, head</td>
<td>-</td>
<td>dur</td>
<td>-</td>
<td>freq</td>
<td>dur</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Russian accent</td>
<td>heed, peat, pit, head</td>
<td>-</td>
<td>dur</td>
<td>voice with static</td>
<td>baa, minute</td>
<td>-</td>
<td>dur</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>sheep, ship, shaved, head</td>
<td>freq</td>
<td>dur</td>
<td>-</td>
<td>math, brush</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>12</td>
<td>man</td>
<td>cheers, shit</td>
<td>-</td>
<td>-</td>
<td>human/computer</td>
<td>bath, death</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>bath, math, bass, mass</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>voice</td>
<td>huh, hum, humm</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>bath, math, bass, mass</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>heed, keyed, kid</td>
<td>-</td>
<td>dur</td>
<td>voice &amp; wind</td>
<td>bath</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>16</td>
<td>computer voice</td>
<td>sheet</td>
<td>-</td>
<td>dur</td>
<td>computer voice</td>
<td>math</td>
<td>freq</td>
<td>dur</td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>head</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>bath</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.6: Subject description of speech stimuli, with identified words and whether stimuli were described in terms of frequency and duration
subjects who reported duration differences is nearly the same for both sets of vowel stimuli.

5.3.2 Forced choice identification

How well do listeners’ responses in the open set identifications correspond to forced-choice identifications in the subsequent Ganong up-down task? Subjects 1–3, 5–7, 9–12, and 15 used both [i] and [i] words in their descriptions of the speech vowel stimuli. Their Ganong up-down solutions (shown in Appendix C) show clear distinctions of [i] and [i] by frequency or by frequency and duration. Some subjects show larger confusion boundaries than others, and some have only clustering of vowel identity at one or both endpoints (and subject 3 seems to have an area of confusion between long and short [i]), but overall their solutions are interpretable and expected.

On the other hand, for the open set identification, subject 4 never named a token with a word that has [i], subjects 8 and 16 never had [i], subject 17 did not use either [i] or [i], and subjects 13 and 14 did not hear words. In their forced-choice Ganong identifications, subjects 4, 16, and 17 do not seem to have vowel spaces that are different from subjects that heard both vowels in the open set labeling, though 16’s boundary seems wider than usual. However, subject 8 has no reliable [i] category, and 13 and 14 have a boundary in the middle of their [i] tokens.

From these findings, it appears that the majority of subjects were using [i] and [i] labels in their speech vowel discrimination task, although there are a few extraneous [e] and [ɛ] vowels for many of them. There were no significant effects on whether subjects identified vowels as [i] and [i] or not based on their arcsine
transformed percentages correct for frequency \((F(1, 15) = 1.087, p = .314)\), duration \((F(1, 15) = 2.999, p = .104)\), or both cues \((F(1, 15) = 1.321, p = .268)\).

For fricatives, only subjects 1–3, 8, and 14 used both \(\theta\) and \(s\) words in their open set identification. In their Ganong solutions, all but subject 8 have a good frequency or frequency and duration boundary, while 8 seems to have a duration boundary. Subjects 5, 11, 12, 15–17 never identified a word as having an \([s]\), 4, 6, 7, and 10 never heard \([\theta]\) or \([s]\), and 9 and 13 never heard speech.

For their Ganong solutions, 12 has a frequency division between the fricatives, as does 5 but with a small \([s]\) region, while 16 has a duration boundary but with a very poorly defined \([s]\) region. But 11 has no \([s]\) region, 15 and 17 have no reliable \([s]\) or \([\theta]\) regions. Although they heard neither fricative independently, subject 4 could divide them mostly by frequency, 6 and 7 by duration (although the \([s]\) category for 7 was poor). Subject 10 on the other hand had unclear categories. Subject 9 previously never heard the tokens has speech, but gave a good frequency and duration boundary, while the fricative space for 13 was uninterpretable.

It would seem to be a problem if most of the subjects did not identify the endpoints in the same manner for the items they are discriminating, but there was no significant difference between those who spontaneously identified fricatives as \([\theta]\) and \([s]\) and those who did not for frequency \((F(1, 15) = 3.221, p = .093)\), duration \((F(1, 15) = .261, p = .617)\), or both \((F(1, 15) = .003, p = .959)\)
5.4 Group multidimensional scaling

5.4.1 Multidimensional scaling

Reaction time data and counts of correct rejections (i.e., stimulus pairs that subjects labeled as different) were used to create matrices for individual differences multidimensional scaling (INDSCAL) (Carroll & Chang, 1970). Ratio scaling was used so that the resulting dimensional values could be correlated with the ERB frequency, duration, and loudness values of the stimuli. The reaction time data (inverted to make dissimilarity matrices) failed to produce a duration dimension in many of the subjects' individual solutions, as well as for the group solutions (the highest correlation between duration and a dimension was .741 for the synthetic fricatives). This was due to the fact that many subjects took an extremely long time to determine that a pair differed only by duration, as shown in Figure 5.11. Thus, only the MDS solutions based on the counts of the correct rejections will be discussed.

Stress values that best matched Spence’s (1979) stress formula were used in determining the optimal number of dimensions, as well as finding the stress “elbow” and deciding whether the solution was interpretable. How well a perceptual dimension correlated with stimulus values (e.g., F1 frequency or duration) was used in interpreting the solutions. In certain instances, more dimensions were found than what can be supported statistically. However, since duration was usually a weak dimension, and group and subject solutions had similar configurations, higher dimensions were included in the solutions to prevent duration information from being lost.
5.4.2 Group MDS

Asymmetric dissimilarity matrices on correct rejection counts were created separately for all four data types. The dimensions are labeled with the stimulus values that correlate with them best. Duration values are shown in Table 5.7.

<table>
<thead>
<tr>
<th>Vowel duration level</th>
<th>Fricative duration level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>ms</td>
<td>162</td>
</tr>
</tbody>
</table>

Table 5.7: Duration values for vowels and fricatives

The frequency contours of the vowels and fricatives are not level, so the average frequency in ERB were used as the frequency factor. Subjects could hear these contours, so the sum of the total absolute frequency change over time (total frequency change), and the main portion of frequency change (frequency offglide) were used as factors as well. Frequency offglide was the change from the fourth to ninth points in the vowels (from the extreme value to the point just before the transition to the following consonant) shown in Figure 4.3, and the change from the second to fourth points in the fricatives shown in Figure 4.4. These three frequency factor values are shown in Table 5.8.

Another factor that could not be completely controlled for is loudness, since it varies with frequency and duration. Average loudness values were obtained in Praat over the vowel or fricative portions, shown in Table 5.9 (intensity values are shown in Table 4.2). These values are in relation to the average human perception of
Table 5.8: Average ERB values for different frequency measures of vowels and fricatives

Table 5.9: Average vowel and fricative loudness in sones for tokens of different frequency and duration levels
Since MDS solutions are standardized, the frequency and duration values for the tokens were standardized. Correlations between the factors are shown in Table 5.10 for the non-speech and speech vowels, and Table 5.11 for the non-speech and speech fricatives. Loudness correlates highly with duration for sine wave vowels and with frequency for synthetic vowels so it is not an independent factor. Two of the frequency measures (frequency offglide and total frequency change) are highly correlated with each other as well. For the fricatives, loudness correlates highly with frequency for both the non-speech and speech tokens.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Duration</th>
<th>Freq.</th>
<th>Glide</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency offglide</td>
<td>0.000</td>
<td>0.215</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total frequency change</td>
<td>0.000</td>
<td>-0.458</td>
<td>0.751</td>
<td></td>
</tr>
<tr>
<td>Sine wave loudness</td>
<td>0.836</td>
<td>0.471</td>
<td>0.294</td>
<td>-0.017</td>
</tr>
<tr>
<td>Synthetic loudness</td>
<td>0.205</td>
<td>0.975</td>
<td>0.259</td>
<td>-0.399</td>
</tr>
</tbody>
</table>

Table 5.10: Correlations between factors for sine wave and synthetic vowels

<table>
<thead>
<tr>
<th>Factors</th>
<th>Duration</th>
<th>Freq.</th>
<th>Glide</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency offglide</td>
<td>0.000</td>
<td>-0.648</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total frequency change</td>
<td>0.000</td>
<td>0.045</td>
<td>0.685</td>
<td></td>
</tr>
<tr>
<td>Narrowband loudness</td>
<td>-0.444</td>
<td>0.849</td>
<td>-0.655</td>
<td>-0.069</td>
</tr>
<tr>
<td>Synthetic loudness</td>
<td>-0.477</td>
<td>0.835</td>
<td>-0.522</td>
<td>0.061</td>
</tr>
</tbody>
</table>

Table 5.11: Correlations between factors for narrowband and synthetic fricatives
INDSCAL MDS was performed on the individual data from all subjects to create a solution for each of the four sets of stimuli. For each solution, this yields a group space, as well as the weights for individual subjects on each dimension. MDS takes distances as input, such as the number of times a subject correctly distinguishes the tokens in a pair—dissimilarity is used as a distance. The output is a coordinate space (matrix) of the stimuli in N dimensions. The first dimension is the most “important” since its variance distributes the stimuli over the widest region of MDS space, while each succeeding dimension accounts for less and less variance.

In a few cases, two dimensions with similar stress values will switch rank position as additional dimensions are added. Ordering was kept strict (e.g., the first dimension in a one-dimensional solution will always be kept the first dimension even if in the two-dimensional solution the new dimension has a slightly better stress value).

The next four sections contain the group MDS solutions for each of the four data types. Phones are superimposed on the figures next to their “prototype” tokens (e.g., ‘i’ next to vowel token 13, and ‘“i”’ next to its non-vowel token). All MDS solutions were scaled with the slope from a regression model on the disparities with the original count correct data, although there was little difference in the resulting scales. Stress and R² values are shown in Table 5.12. Weights will be discussed afterwards.

5.4.3 Sine wave vowels

The four-dimensional solution for the sine wave vowel stimuli is shown in Figure 5.17. Frequency values have a correlation of .963 with the first dimension, total frequency change have a correlation of .960 with the second, and duration has
<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Vowels Stress</th>
<th>Vowels $R^2$</th>
<th>Fricatives Stress</th>
<th>Fricatives $R^2$</th>
<th>Speech Vowels Stress</th>
<th>Speech Vowels $R^2$</th>
<th>Speech Fricatives Stress</th>
<th>Speech Fricatives $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.358</td>
<td>0.538</td>
<td>0.334</td>
<td>0.487</td>
<td>0.331</td>
<td>0.516</td>
<td>0.327</td>
<td>0.510</td>
</tr>
<tr>
<td>3</td>
<td>0.323</td>
<td>0.590</td>
<td>0.266</td>
<td>0.590</td>
<td>0.236</td>
<td>0.673</td>
<td>0.260</td>
<td>0.596</td>
</tr>
<tr>
<td>4</td>
<td>0.163</td>
<td>0.821</td>
<td>0.245</td>
<td>0.626</td>
<td>0.189</td>
<td>0.752</td>
<td>0.216</td>
<td>0.676</td>
</tr>
<tr>
<td>5</td>
<td>0.138</td>
<td>0.858</td>
<td>0.236</td>
<td>0.643</td>
<td>0.173</td>
<td>0.778</td>
<td>0.190</td>
<td>0.725</td>
</tr>
<tr>
<td>6</td>
<td>0.125</td>
<td>0.878</td>
<td>0.227</td>
<td>0.666</td>
<td>0.161</td>
<td>0.802</td>
<td>0.181</td>
<td>0.743</td>
</tr>
</tbody>
</table>

Table 5.12: Stress and $R^2$ values for group MDS solutions

a correlation of .981 with the fourth. Nothing had a high correlation with the third dimension, but the solution was not lowered to three dimensions because the third dimension was still uninterpretable, and the fourth duration dimension would be lost.

Although it appears that frequency, total frequency change, and an uninterpretable frequency dimension make up the first three dimensions, it may just be a tetrahedron of frequency. A tetrahedron maximally disperses four points in three dimensions. Duration makes up the fourth dimension, but the mid and long values are closer together, and therefore less perceptually distinct, than the short and mid values. Although it appears that the lower two frequency levels are close together in panel 5.17.c, panels a and b show that the levels are distinct. It is only the frequency/duration viewpoint that makes them appear close.

5.4.4 Narrowband fricatives

The three-dimensional solution for the narrowband fricative stimuli is shown in Figure 5.18. In this solution, duration makes up the second dimension, with total frequency change the third, and frequency the first as it was previously. Frequency
Figure 5.17: MDS solution for sine wave vowels. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by uninterpretable third dimension (which is probably a function of frequency). (c) First dimension frequency by fourth dimension duration.
has a correlation of .928 with its dimension, duration has .855, and total frequency change has .676. As in the previous solution, although it appears that two of the frequency levels are collapsed in panel 5.18.a, the next panel shows that they are distinct.

Figure 5.18: MDS solution for narrowband fricatives. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension total frequency change.

5.4.5 Synthetic vowels

The solution for the synthetic vowels is three-dimensional, as shown in Figure 5.19. The first dimension correlates to frequency as in the previous solutions, but the second dimension is frequency offglide, not total frequency change. Duration makes up the third dimension, and the mid and long duration values are closer
together than the short and mid values like they were for the sine wave vowels. Frequency has a correlation of .897, frequency offglide has .887, and duration has .934. And, as discussed previously, the highest two frequency levels are not as close as they seem in the second panel.

Figure 5.19: MDS solution for synthetic vowels. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension duration.

5.4.6 Synthetic fricatives

The MDS solution for synthetic fricatives is shown to be two-dimensional in Figure 5.20, with frequency and duration as the first and second dimensions, respectively. Frequency has a correlation of .910 with the first dimension, and duration has a correlation of .965 with the second dimension.
Figure 5.20: MDS solution for synthetic fricatives. First dimension frequency by second dimension duration.

5.4.7 MDS summary

Notice that the duration dimension is more robust for the fricative scalings than for the vowels. The second and third levels of duration for the vowels are closer together. However, the frequency levels seem to have wider dispersion for the vowels than the fricatives. Speech stimuli have fewer dimensions than non-speech stimuli.

5.4.8 Subject weights

MDS solutions indicate the degree of importance of a dimension to a particular subject by the size of its weight value. The higher the value, the more weight a subject holds for that group space dimension. However, subject weights only provide information for dimensions that are included in the group solution. A subgroup of subjects may place importance on a different dimension not used by a larger group of subjects, and therefore it is not included in the group solution.
Examining the weights in Figure 5.21 shows that the values are somewhat similar to the percent correct discrimination data shown in Figure 5.1, although a direct comparison cannot be made because the percent correct frequency versus duration discrimination data does not include pairs that differ by frequency and duration, while the MDS solution does. Frequency weights for vowels are higher than duration weights, which matches the discrimination results, but duration weights for fricatives are too low in comparison to how well subjects discriminated duration differences. The wide range of duration weights for vowels mirrors the large differences in duration discrimination ability. Many subjects cluster together in both sets of data, such as 7, 10, and 17 for the sine wave vowels, and subject 3 has the lowest weights and scores for vowel duration.

5.5 Subgroup multidimensional scaling

One can use the weights of the group solution to differentiate subjects by performing cluster or profile analyses, but these all assume that the group solution is correct for everyone. If a subject has a low weight for a particular dimension, how low does it need to be before the subject is assumed to not have that dimension? What if a minority of subjects have a dimension that did not make the group solution? Subjects' MDS weirdness ratings will indicate how poorly they fit the group solution, but not why they do not fit.

To determine if some subjects do not fit the group solution, MDS solutions were made for each subject individually, using the same counts of correct rejections

---

8Although Kruskal and Wish (1978) suggest that an additional dimension should be added if it receives high weights by a few subjects even if it does not account for much variation, what about the cases in which a dimension receives low or medium weights by a few subjects?
Figure 5.21: Subject weights for frequency (x-axis) and duration (y-axis) dimensions. (a) Sine wave vowels. (b) Narrowband fricatives. (c) Synthetic vowels. (d) Synthetic fricatives.
that were used in the group solutions. Procrustes rotation\(^9\) analysis (see, e.g., Hurley & Cattell, 1962; Borg & Lingoes, 1987; Borg & Groenen, 1997; Dryden & Mardia, 1998; Cox & Cox, 2001) was used to orient each subject’s solution to a set of target matrices, choosing the one with the best fit. Subjects with similar solutions would be combined to create a new INDSCAL solution for this subgroup. This method is similar to PINDIS (see, e.g., Lingoes & Borg, 1978), but with more than one group solution.

For the Procrustes rotation, target matrices were created for all possible solutions, using the frequency, duration, frequency offglide, total frequency change, and loudness values of the stimuli as factors for possible dimensions. Since MDS solutions are standardized, each factor was centered to 0 and its elements were divided by its (new) root-mean-square. All combinations of factors were used, from two to five dimensions (i.e., frequency by duration, frequency by frequency offglide, frequency by total frequency change, \..., frequency by duration by frequency offglide, \..., frequency by duration by frequency offglide by total frequency change by loudness). The justification for using these acoustic values from the stimuli is that they were found to correlate highly with the unrotated INDSCAL solutions. Loudness was added because many solutions had uninterpretable dimensions, and although only a few subjects described stimuli as “louder” and “softer”, loudness seemed to be the next perceptual measure to use after ERB frequency and duration.

Correlations between a target matrix and the subject’s solution rotated to that target matrix were used to pick the target matrix which best matched the subject’s solution. Dimension order was determined by the ranking of each factor’s correlation

\(^9\)The MDS solutions for an individual were rotated, but the INDSCAL solutions for a group cannot be rotated.
strength starting from the subject’s two-dimensional solution on up to the “correct” N-dimensional solution determined by the stress values and interpretability.

Table 5.13 displays the ordering of each subject’s MDS solution for each of the four stimulus types. A question mark indicates that the correlation is low, and the correlations of the factors with the subjects’ MDS solutions are shown in Table 5.14. All the subjects have a frequency dimension for each set of data (usually the first or second dimension), except for subject 7 on synthetic fricatives, but she still has a frequency offglide dimension. More important is the fact that many subjects do not have duration dimensions. Even more interestingly, the lack of a duration dimension does not always match up with a low duration weight in the group solution. Subject 16 has a high duration weight in the group solution for sine wave vowels in relation to most of the other subjects, but he does not have a duration dimension in his individual solution. On the other hand subject 3 usually has the lowest or one of the lowest duration weights for all group solutions, but even she has a duration dimension for her synthetic vowel solution, while higher weighted 2 and 14 do not.

This provides justification that these subjects do not compose the same perceptual population. Not only are the subject weights from the group MDS solutions incomparable for determining whether a subject weights one dimension more than another (cf. the treatment of color-normal and color-blind subjects in section 20.2 of Borg & Lingoes, 1987), but there is the greater problem is determining what these perceptual spaces actually represent. The underlying assumption of INDSCAL is that the input data are from the same group. One would not combine the data of subjects from two languages to form a single group space (unless, for example, the subjects were all bilinguals) because this space might not be representative of either language.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2?</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2?</td>
<td>3</td>
<td>4?</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.13: Ordering of individual MDS solution dimensions including *frequency* (Freq.), *duration* (Dur.), *frequency offglide* (Glide), *total frequency change* (Total), and *loudness* (Loud.).
<table>
<thead>
<tr>
<th>Sub.</th>
<th>Sine wave vowels</th>
<th>Narrowband fricatives</th>
<th>Synthetic vowels</th>
<th>Synthetic fricatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.887</td>
<td>0.836</td>
<td>0.793</td>
<td>0.539</td>
</tr>
<tr>
<td>2</td>
<td>0.883</td>
<td>0.791</td>
<td>0.839</td>
<td>0.795</td>
</tr>
<tr>
<td>3</td>
<td>0.887</td>
<td>0.989</td>
<td>0.725</td>
<td>0.374</td>
</tr>
<tr>
<td>4</td>
<td>0.925</td>
<td>0.986</td>
<td>0.834</td>
<td>0.784</td>
</tr>
<tr>
<td>5</td>
<td>0.912</td>
<td>0.870</td>
<td>0.855</td>
<td>0.775</td>
</tr>
<tr>
<td>6</td>
<td>0.919</td>
<td>0.803</td>
<td>0.969</td>
<td>0.794</td>
</tr>
<tr>
<td>7</td>
<td>0.950</td>
<td>0.976</td>
<td>0.933</td>
<td>0.741</td>
</tr>
<tr>
<td>8</td>
<td>0.925</td>
<td>0.931</td>
<td>0.840</td>
<td>0.794</td>
</tr>
<tr>
<td>9</td>
<td>0.875</td>
<td>0.845</td>
<td>0.810</td>
<td>0.528</td>
</tr>
<tr>
<td>10</td>
<td>0.942</td>
<td>0.962</td>
<td>0.768</td>
<td>0.942</td>
</tr>
<tr>
<td>11</td>
<td>0.904</td>
<td>0.433</td>
<td>0.837</td>
<td>0.389</td>
</tr>
<tr>
<td>12</td>
<td>0.924</td>
<td>0.652</td>
<td>0.823</td>
<td>0.785</td>
</tr>
<tr>
<td>13</td>
<td>0.873</td>
<td>0.800</td>
<td>0.789</td>
<td>0.662</td>
</tr>
<tr>
<td>14</td>
<td>0.900</td>
<td>0.663</td>
<td>0.822</td>
<td>0.649</td>
</tr>
<tr>
<td>15</td>
<td>0.908</td>
<td>0.987</td>
<td>0.892</td>
<td>0.791</td>
</tr>
<tr>
<td>16</td>
<td>0.622</td>
<td>0.821</td>
<td>0.749</td>
<td>0.961</td>
</tr>
<tr>
<td>17</td>
<td>0.916</td>
<td>0.946</td>
<td>0.815</td>
<td>0.940</td>
</tr>
</tbody>
</table>

Table 5.14: Correlations of factors to individual MDS solution dimensions for frequency (Freq.), duration (Dur.), frequency offglide (Glide), total frequency change (Total), and loudness, (Loud.).
Furthermore, this should be applied to dialects, and to perceptual groups (which may represent sub-dialects, or may represent cognitive distinctions not related to dialect background).

To compare these perceptual subgroups, the subjects needed to be divided in some fashion. Since there is greater disparity in the treatment of duration than frequency (some subjects had no duration dimension, some had a strong low dimension, and others had a weak higher dimension), subjects were placed into three groups according to how their MDS solutions treated duration for each of the four stimuli sets. MDS solutions for the fricatives had many subjects who had duration as their first dimension, so they could be divided into groups that had duration as their first dimension, as their second dimension, or as third and higher dimension (or not at all). Vowel MDS solutions did not have as many subjects with duration as their first or second dimension, so they were divided into groups that had duration as their first or second dimension, as their third or higher dimension, or had no duration dimension. These subject groupings are illustrated in Table 5.15.

Although these subgroups were created based on their duration dimension rank, duration may not be the only factor playing a role. A GLM on subjects’ arcsine transformed percent correct discriminations on frequency or duration with duration subgroupings (which will be discussed in the next few sections) as the factors shows that for sine wave vowels frequency scores ($F(2, 8) = 15.637, p < .05$) but not duration scores ($F(2, 8) = 3.921, p = .065$) are better segregated by the subgroupings. The same holds true for narrowband fricatives ($F(2, 8) = 18.118, p < .05$ for frequency, and $F(2, 8) = .571, p = .586$ for duration). However, for synthetic vowels, both frequency ($F(2, 8) = 7.062, p < .05$) and duration ($F(2, 8) = 6.052, p < .05$)
### Table 5.15: Duration importance for subjects.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>High</th>
<th>Mid</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-speech</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowels</td>
<td>7, 10, 17</td>
<td>2, 4, 5, 8, 12, 14</td>
<td>1, 3, 6, 9, 11, 13, 15, 16</td>
</tr>
<tr>
<td>Fricatives</td>
<td>2, 4, 9, 10, 12</td>
<td>5, 6, 8, 11, 13, 16, 17</td>
<td>1, 3, 7, 14, 15</td>
</tr>
<tr>
<td>Speech</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowels</td>
<td>1, 4, 5, 6, 8, 10, 12, 13, 16</td>
<td>3, 7, 9, 11, 15, 17</td>
<td>2, 14</td>
</tr>
<tr>
<td>Fricatives</td>
<td>2, 4, 6, 7, 10, 13, 17</td>
<td>1, 5, 8, 9, 11, 15, 16</td>
<td>3, 12, 14</td>
</tr>
</tbody>
</table>

Table 5.15: Duration importance for subjects. *High* corresponds to duration as a first or second dimension for vowels, and as a first dimension for fricatives. *Mid* corresponds to duration as a third or higher dimension for vowels, and as a second dimension for fricatives. *Low* corresponds to duration as not being represented by a dimension for vowels, and as third or higher dimension or not being represented by a dimension for fricatives.

Differences are significantly represented in the groupings, while for synthetic fricatives neither are \( F(2, 8) = 3.736, p = .071 \) for frequency, and \( F(2, 8) = 3.727, p = .072 \) for duration). This indicates that it is not only duration discrimination ability that decides the subgroupings.

For example, one subject may have a strong duration dimension but an even stronger frequency dimension, while another subject may have have a weak duration dimension but an even weaker frequency dimension. The former subject would have her strong duration results displayed in the second dimension, while the latter subject would have his weak duration in the first dimension. Therefore, while the following analyses describe duration as being “strong” or “weak”, it is in comparison with the subjects’ performance on frequency.
5.5.1 Sine wave vowels

Only one subject (17) had duration as his first dimension, so subjects were grouped as follows: subjects with duration as their first or second dimension were classified as the strong duration subgroup (subjects 7, 10, and 17), subjects with duration as their third or higher dimension were classified as the weak duration subgroup (subjects 2, 4, 5, 8, 12, 14), and subjects with no duration dimension were classified as the no duration subgroup (subjects 1, 3, 6, 9, 11, 13, 15, 16). Stress and R\textsuperscript{2} values are shown in Table 5.16.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>High Stress</th>
<th>High R\textsuperscript{2}</th>
<th>Mid Stress</th>
<th>Mid R\textsuperscript{2}</th>
<th>Low Stress</th>
<th>Low R\textsuperscript{2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.318</td>
<td>0.541</td>
<td>0.352</td>
<td>0.548</td>
<td>0.335</td>
<td>0.596</td>
</tr>
<tr>
<td>3</td>
<td>0.215</td>
<td>0.710</td>
<td>0.309</td>
<td>0.611</td>
<td>0.294</td>
<td>0.661</td>
</tr>
<tr>
<td>4</td>
<td>0.179</td>
<td>0.768</td>
<td>0.154</td>
<td>0.838</td>
<td>0.155</td>
<td>0.847</td>
</tr>
<tr>
<td>5</td>
<td>0.158</td>
<td>0.805</td>
<td>0.122</td>
<td>0.884</td>
<td>0.128</td>
<td>0.885</td>
</tr>
</tbody>
</table>

Table 5.16: Stress and R\textsuperscript{2} values for subgroup MDS solutions on non-speech vowels

Figure 5.22 shows the MDS solution for the strong duration subgroup, with frequency, duration, and frequency offglide composing the dimensions. Comparing this with the group MDS solution for sine wave vowels in Figure 5.17 reveals that duration has increased its explanatory power of the solution by moving from the fourth to the second dimension. Frequency values have a correlation of .971 with the first dimension, duration has a correlation of .854 with the second, and frequency offglide has a correlation of .710 with the third.
The MDS solution for the weak duration subgroup is shown in Figure 5.23, with total frequency change, frequency, frequency offglide, and duration composing the dimensions. This is nearly identical with the group MDS solution in Figure 5.17. Although in panel c it appears that frequency has some ordering difficulty with its middle two levels, panels a and b show that frequency is still tetrahedral, so the problem is with the frequency by duration vantage point and not the dimension. Total frequency change has a correlation of .714 with the first dimension, frequency has a correlation of .803 with the second, frequency offglide has a correlation of .942 with the third, and duration has a correlation of .942 with the fourth.

The no duration subgroup has its MDS solution shown in Figure 5.24, with frequency, total frequency change, frequency offglide, and an uninterpretable factor composing the dimensions. This fourth dimension appears to be very close to being
Figure 5.23: MDS solution for weak duration subgroup for sine wave vowels. (a) First dimension total frequency change by second dimension frequency. (b) First dimension total frequency change by third dimension frequency offglide. (c) Second dimension frequency by fourth dimension duration.
duration, but the second frequency level is reversed in relation to the others. This indicates that the subjects are able to distinguish tokens by duration, but are not consistent with their own responses or with each other. Frequency has a correlation of .996 with the first dimension, total frequency change has a correlation of .773 with the second, and frequency offglide has a correlation of .707 with the third, while duration has only a correlation of .403 with the fourth dimension.

5.5.2 Narrowband fricatives

The first dimension duration subgroup was composed of subjects 2, 4, 9, 10, and 12. The second dimension duration subgroup was composed of subjects 5, 6, 8, 11, 13, 16, and 17. The weak-to-no duration subgroup was composed of subjects 1, 3, 7, 14, and 15. Stress and $R^2$ values are shown in Table 5.17.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td>0.308</td>
<td>0.256</td>
<td>0.235</td>
<td>0.227</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.560</td>
<td>0.642</td>
<td>0.684</td>
<td>0.701</td>
</tr>
<tr>
<td>Stress</td>
<td>0.321</td>
<td>0.243</td>
<td>0.218</td>
<td>0.207</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.524</td>
<td>0.648</td>
<td>0.690</td>
<td>0.715</td>
</tr>
<tr>
<td>Stress</td>
<td>0.356</td>
<td>0.297</td>
<td>0.252</td>
<td>0.228</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.436</td>
<td>0.515</td>
<td>0.593</td>
<td>0.642</td>
</tr>
</tbody>
</table>

Table 5.17: Stress and $R^2$ values for subgroup MDS solutions on non-speech fricatives

Figure 5.25 shows the MDS solution for the strong duration subgroup, with duration, and frequency for the dimensions. Comparing this with the group MDS solution for narrowband fricatives in Figure 5.18 reveals that duration has moved
Figure 5.24: MDS solution for no duration subgroup for sine wave vowels. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by third dimension frequency offglide. (c) First dimension frequency by uninterpretable fourth dimension (which appears to be duration for three of the frequency levels).
to the first dimension from the second, accounting for more of the variability in the solution than frequency. Stimuli duration values has a correlation of .989 with the first dimension, and frequency has a correlation of .944 with the second.

![MDS solution for first dimension duration subgroup for narrowband fricatives. First dimension duration by second dimension frequency.](image)

Figure 5.25: MDS solution for first dimension duration subgroup for narrowband fricatives. First dimension *duration* by second dimension *frequency*.

The second dimension subgroup analysis yielded the MDS solution in Figure 5.26. Frequency and duration dimensions are in the same order as in the original group solution. Frequency has a correlation of .943 with the first dimension, and duration has a correlation of .966 with the second.

Figure 5.27 shows the MDS solution for the weak-to-no duration subgroup. There is no duration dimension—just the frequency first dimension, and the total frequency change second dimension that moved from the third dimension in the group
solution. Frequency has a correlation of .917 with the first dimension, while total frequency change has only a correlation of .591 with the second.

5.5.3 Synthetic vowels

Subjects with duration as their first or second dimension in their individual solutions were combined into the strong duration subgroup (subjects 1, 4, 5, 6, 8, 10, 12, 13, and 16). Subjects with duration as their third or higher dimension were classified as belonging to the weak duration subgroup (subjects 3, 7, 9, 11, 15, and 17). And subjects with no duration dimension became the no duration subgroup (subjects 2 and 14). Stress and $R^2$ values are shown in Table 5.18.

The MDS solution for the strong duration subgroup is shown in Figure 5.28, with frequency and duration for the dimensions. Comparing this with the group MDS solution for synthetic vowels in Figure 5.19 reveals that duration has moved
Figure 5.27: MDS solution for weak-to-no duration subgroup for narrowband fricatives. First dimension frequency by second dimension total frequency change.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>High Stress</th>
<th>High R²</th>
<th>Mid Stress</th>
<th>Mid R²</th>
<th>Low Stress</th>
<th>Low R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.318</td>
<td>0.553</td>
<td>0.352</td>
<td>0.535</td>
<td>0.350</td>
<td>0.494</td>
</tr>
<tr>
<td>3</td>
<td>0.234</td>
<td>0.673</td>
<td>0.227</td>
<td>0.705</td>
<td>0.292</td>
<td>0.572</td>
</tr>
<tr>
<td>4</td>
<td>0.182</td>
<td>0.763</td>
<td>0.195</td>
<td>0.754</td>
<td>(singularity)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.164</td>
<td>0.792</td>
<td>0.175</td>
<td>0.789</td>
<td>0.147</td>
<td>0.826</td>
</tr>
</tbody>
</table>

Table 5.18: Stress and R² values for subgroup MDS solutions on speech vowels
to the second dimension from the third, accounting for more of the variability. Frequency values have a correlation of .937 with the first dimension, and duration has a correlation of .875 with the second.

![MDS Solution for Strong Duration Subgroup](image)

Figure 5.28: MDS solution for strong duration subgroup for synthetic vowels. First dimension *frequency* by second dimension *duration*.

The MDS solution for the weak duration subgroup is shown in Figure 5.29, with frequency, frequency offglide, and duration for the dimensions. This is very similar to the original group solution. Frequency has a correlation of .969 with the first dimension, frequency offglide has a correlation of .857 with the second, and duration has a correlation of .940 with the third.
Figure 5.29: MDS solution for weak duration subgroup for synthetic vowels. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension duration.

Figure 5.30 for the no duration subgroup shows that there is no dimension for duration. Only frequency and total frequency change make up the solution. Frequency has a correlation of .860 with the first dimension, and total frequency change has a correlation of .916 with the second.

5.5.4 Synthetic fricatives

The first dimension duration subgroup was composed of subjects 2, 4, 6, 7, 10, 13, and 17. The second dimension duration subgroup was composed of subjects 1, 5, 8, 9, 11, 15, and 16. The weak-to-no duration subgroup was composed of subjects 3, 12, and 14. Stress and R² values are shown in Table 5.19.

The MDS solution for the strong duration subgroup is shown in Figure 5.31, with duration, and frequency for the dimensions. Comparing this with the group MDS solution for synthetic fricatives in Figure 5.20 reveals that duration has moved
Figure 5.30: MDS solution for no duration subgroup for synthetic vowels. First dimension frequency by second dimension total frequency change.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>High</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stress</td>
<td>R^2</td>
<td>Stress</td>
<td>R^2</td>
<td>Stress</td>
<td>R^2</td>
<td>Stress</td>
<td>R^2</td>
</tr>
<tr>
<td>2</td>
<td>0.310</td>
<td>0.557</td>
<td>0.326</td>
<td>0.520</td>
<td>0.356</td>
<td>0.445</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.251</td>
<td>0.638</td>
<td>0.244</td>
<td>0.639</td>
<td>0.266</td>
<td>0.583</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.201</td>
<td>0.725</td>
<td>0.192</td>
<td>0.729</td>
<td>0.212</td>
<td>0.674</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.187</td>
<td>0.747</td>
<td>0.173</td>
<td>0.763</td>
<td>0.195</td>
<td>0.701</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.19: Stress and R^2 values for subgroup MDS solutions on speech vowels
to the first dimension from the second, accounting for more of the variability in the solution than frequency. Duration values have a correlation of .966 with the first dimension, and frequency has a correlation of .885 with the second.

![MDS solution for first dimension duration subgroup](image)

Figure 5.31: MDS solution for first dimension duration subgroup for synthetic fricatives. First dimension *duration* by second dimension *frequency*.

The MDS solution for the second dimension duration subgroup, shown in Figure 5.32 is nearly identical to the group solution. Frequency has a correlation of .897 with the first dimension, and duration has a correlation of .949 with the second.

Figure 5.33 displays the MDS solution for the weak-to-no duration subgroup, which has no duration dimension. Frequency and an uninterpretable dimension make up the solution, which implies that a one dimensional solution would be adequate. Frequency has a correlation of .952 with the first dimension.
Figure 5.32: MDS solution for second dimension duration subgroup for synthetic fricatives. First dimension frequency by second dimension duration.

Figure 5.33: MDS solution for weak-to-no duration subgroup for synthetic fricatives. First dimension frequency by uninterpretable second dimension.
5.6 Possible explanations for perceptual differences

This section will examine what factors can possibly explain these perceptual subgroups in frequency and duration cue usage. There might be differences in frequency thresholds, or there may be a perception-production link. There may be subdialectal divergence, or listeners’ perceptual processes may have been affected by musical, foreign language, or synthetic speech experience. GLM analyses and linear regressions were run on the arcsine transformed percent correct data, using these factors as predictors.

5.6.1 Audition

Subjects’ audiograms are shown in Figure 5.34. Most listeners have audiograms that are typical for young, healthy ears (flat profile with thresholds at 20 dB or better). Subject 14 is most noticeably different from the others with notches in both ears at 750–2000 Hz. However, the stimuli used in this experiment contrast for frequencies below 500 Hz (for vowels) and above 3000 Hz (for fricatives), so subject 14 may not stand out from the others for these stimuli—all subjects heard the test tones of 250 Hz (near the center of vowel F1 values) and 4000 Hz (near the center of fricative F5 values) at 20 dB or better.

Linear regressions were run on discrimination scores of tokens that differed only by frequency, using hearing levels as the predictor variables. For each pair that compared adjacent frequency levels, the percent correct value underwent the arcsine transform. Hearing levels for the left and right ears at 125, 250, and 500 Hz were used to predict non-speech and speech vowel frequency discrimination performance, and levels at 3, 4, and 6 kHz were used for the non-speech and speech fricative frequency
Figure 5.34: Audiograms for subjects 1–17.
discrimination performance. None of the results were significant overall ($F(6, 10) = 1.448, p = .288$, $F(6, 10) = 1.574, p = .251$, and $F(6, 10) = 2.098, p = .144$ for the lowest, middle, and highest adjacent frequency pairs, respectively, of non-speech vowels; $F(6, 10) = 1.330, p = .329$, $F(6, 10) = 1.151, p = .402$, and $F(6, 10) = .825, p = .576$ for the lowest, middle, and highest adjacent frequency pairs, respectively, of non-speech fricatives; $F(6, 10) = .642, p = .696$, $F(6, 10) = 1.051, p = .449$, and $F(6, 10) = .401, p = .862$ for the lowest, middle, and highest adjacent frequency pairs, respectively, of speech vowels; and $F(6, 10) = 1.196, p = .382$, $F(6, 10) = 1.043, p = .453$, and $F(6, 10) = 3.003, p = .060$ for the lowest, middle, and highest adjacent frequency pairs, respectively, of speech fricatives).

This indicates that differences in performance cannot be attributed to differences in hearing thresholds.

5.6.2 Commutation

In the commutation test, subjects were considered to demonstrate a contrast if they correctly identified each word in a pair 8 times or higher out of 10 (the lower bound of the 95% confidence interval for 8 is 5.02, which is just above 50% chance), or each word in a triple 7 times or higher out of 10 (the lower bound for 7 is 3.66, which is just above 33.3% chance). Although the results of this test can indicate that a subject does not produce enough of a distinction between two words (or cannot perceive the slight distinction produced), it cannot with certainty show whether one word’s pronunciation or perception has merged in the middle with another, shifted towards another, or vice versa, because of the design of the answer sheet, and the fact that the word frequencies were not balanced. The answer sheet always had the
word choices in the same order (e.g., *cot* on the left and *caught* on the right), so it is possible that one subject who could not make a distinction between the choices might tend to circle the right response, while another subject might tend to circle the left. Individual results are shown in Table 5.20, as well as in Appendix B along with the vocoid trajectories.

<table>
<thead>
<tr>
<th></th>
<th>heed</th>
<th>hid</th>
<th>head</th>
<th>pin</th>
<th>pen</th>
<th>feel</th>
<th>fill</th>
<th>which</th>
<th>witch</th>
<th>cot</th>
<th>caught</th>
<th>tour</th>
<th>tore</th>
<th>doll</th>
<th>dull</th>
<th>pool</th>
<th>pull</th>
<th>pole</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>3</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>7</td>
<td>0</td>
<td>10</td>
<td>9</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>3</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.20: Subject scores for the commutation test. Values that are not better than chance are circled. One or more circled responses per pair or triple indicates that the subject does not produce or perceive that contrast.

All subjects correctly identified their /i/-/ɪ/-/ɛ/ (*heed-hid-head*) words. Only three subjects (1, 9, and 17) failed to correctly distinguish [ɪ] (*pin*) from [ɛ] (*pen*). Only subject 7 failed to identify /il/-/ɪl/ (*feel-fill*) words correctly. Although she
was born and raised in Columbus, her parents were born in Greece. Greek has no
tense/lax contrast, which might explain why she scored low for this comparison, but
she still distinguished *heed* and *hid* correctly. None of the subjects showed a reliable
*w/-w* (*which-witch*) contrast. Only six subjects (5, 9, 12, 13, and 17) showed an
*a/-a* (*cot-caught*) contrast. Fewer than half the subjects maintained an *u/-u* (*tour-tore*) distinction, but all three subjects who merged [ɪ]-[ɛ] managed to keep this
contrast. The five subjects with an *l/-l* (*doll-dull*) contrast were 5, 6, 13, 15, and 16. None of the subjects had a significant three-way contrast for *ul/-ul/-ol* (*pool-pull-pole*) except for 13.

The results of the commutation test indicate that even in this relatively ho-
mogeneous group of subjects (Ohio State University students who grew up in the
northern portion of Columbus, Ohio) there is still dialectal variation. Therefore, in
evaluating individual differences in speech perception, these differences will need to
be taken into account.

An attempt was made to determine whether or not the dialects of subjects’
parents could explain the individual results for [ɪ]-[ɛ] and *a/-a*. Using the dialect
maps of Labov (in press), whether parents distinguished these contrasts or not was
estimated by their childhood places of residence. Only the mothers’ regions signifi-
cantly correlated with subjects’ distinction of [ɪ]-[ɛ] (*τ*-b*(N = 17) = .717, *p* < .05),
although fathers’ and mothers’ regions also significantly correlated with each other
(*τ*-b*(N = 17) = .717, *p* < .05).

Since there are different groupings based on the different contrasts in the
commutation test, and the only directly relevant contrast (*heed-hid*) had everyone
performing at the same level, results from the production task were investigated.
5.6.3 Production

As a part of the commutation test, subjects read the test words used in the identification and discrimination experiments *heed, hid, bath, and bass*. This provides the opportunity to confirm that the acoustic dimensions used in the perceptual stimuli based on the talker (subject 1) are used by the listeners. Acoustic measurements of subjects’ productions of *heed-hid* and *bath-bass* tokens are shown in Figures 5.35–5.38, and averages by gender are shown in Tables 5.21 and 5.22. Only the first and last repetitions of each of these tokens were measured. The middle 40% of the vowels’ F1 values were measured instead of “steady-state” values because the formant dynamics in [i] would vary within speakers from token to token. One utterance might have the typical F1 “hump” with a falling transition to the following coronal, while another would have an F1 “trough” with a following rise due to ingliding and then the falling transition to the coronal. The center 40% average would approximately cover the “steady-state” portion of the first type and the rise portion of the second type.

The F1 difference for the vowels appears to be very robust in Figure 5.35, and the duration difference shown in Figure 5.36 seems to be large as well (though subjects 17 and 9 have only a 15–20 ms difference).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Acoustic cue</th>
<th>Frequency (Hz)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[i]</td>
<td>[i]</td>
<td>[i]</td>
</tr>
<tr>
<td>Men</td>
<td>282</td>
<td>428</td>
<td>247</td>
</tr>
<tr>
<td>Women</td>
<td>324</td>
<td>483</td>
<td>274</td>
</tr>
</tbody>
</table>

Table 5.21: Average F1 and duration values of the vowels in *heed* and *hid* for men and women
Figure 5.35: Subjects’ average F1 values of the vowels in *heed* and *hid*.

Figure 5.36: Subjects’ average duration of the vowels in *heed* and *hid*. 
Measuring the fricative poles (using the same procedure as for the talker in Section 4.3) was difficult, since there were only two tokens per fricative per subject for comparison. It appears in Figure 5.37 that the listeners make the same type of F5 distinction as the talker, except perhaps for subject 5. Fricative duration in Figure 5.38 yielded the least amount of congruence between talker and listeners—subjects 7, 6, 4, and 13 had only a 5–30 ms average difference between two tokens of /θ/ and two tokens of /s/.

![Subjects' average F5 values of the fricatives in bath and bass.](image)

Figure 5.37: Subjects’ average F5 values of the fricatives in bath and bass.

Although the talker the stimuli were based on had similar vowel acoustic measurements for the commutation test as he had for his original recordings, his fricative poles are now closer together, and his durations have increased. Another problem found for the fricative poles is that the average values for the men are higher than
Subject

<table>
<thead>
<tr>
<th>Duration (ms)</th>
<th>123456789</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10111213141517</td>
</tr>
</tbody>
</table>

Figure 5.38: Subjects’ average duration of the fricatives in *bath* and *bass*.

<table>
<thead>
<tr>
<th>Acoustic cue</th>
<th>Gender</th>
<th>Frequency (Hz)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>3659 4305</td>
<td>199 293</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>3562 4291</td>
<td>208 263</td>
</tr>
</tbody>
</table>

Table 5.22: Average F5 and duration values of the fricatives in *bath* and *bass* for men and women
the women, but theoretically women should have higher values on average because their front mouth cavities are smaller than men’s on average.

The only significant effect of gender on perception \((F(1, 15) = 5.717, p < .05)\) was that men performed significantly better on discriminating speech vowel duration (78.7%) than women did (67.4%). However, there is no difference in the mean vowel durations produced by the subjects \((F(1, 15) = .100, p = .756)\).

Linear regressions were run on arcsine transformed discrimination scores of tokens that differed by frequency or duration, using differences in F1 and duration for the vowels and F5 and duration for the fricatives as the predictor variables. The prediction is that distinctions subjects make in production should be mirrored in their perceptual contrast ability. However, none of the results were significant: non-speech vowel frequency \((F(1, 15) = 2.775, p = .116)\), non-speech vowel duration \((F(1, 15) = 1.573, p = .229)\), non-speech fricative frequency \((F(1, 15) = .446, p = .514)\), non-speech fricative duration \((F(1, 15) = .253, p = .622)\), speech vowel frequency \((F(1, 15) = 1.534, p = .234)\), speech vowel duration \((F(1, 15) = .266, p = .613)\), speech fricative frequency \((F(1, 15) = .892, p = .360)\), speech fricative duration \((F(1, 15) = 1.876, p = .191)\). Although, since there were only two tokens measured for each of the four words (heed-hid, bath-bass) it may be that these examples were not representative of subjects’ average productions.

Is there a relationship between perception and production by subgroup? Would subjects in a subgroup with a strong duration dimension make a larger durational contrast in their productions than subjects in a subgroup with a weak duration dimension? A GLM on subjects’ durational and ERB frequency differences
for *heed-hid* and *bath-bass* was run with how the subjects were subgrouped by duration dimensions as the predictors. Duration dimension subgrouping for non-speech vowels had no effect on production of vowel F1 ERB frequency \( (F(2, 8) = .431, p = .664) \) or duration \( (F(2, 8) = .385, p = .692) \). Duration dimension subgrouping for non-speech fricatives had no effect on production of fricative F5 ERB frequency \( (F(2, 8) = .989, p = .413) \) or duration \( (F(2, 8) = 1.824, p = .223) \). Duration dimension subgrouping for speech vowels had no effect on production of vowel F1 ERB frequency \( (F(2, 8) = .270, p = .770) \) or duration \( (F(2, 8) = .850, p = .463) \). And duration dimension subgrouping for speech fricatives had no effect on production of fricative F5 ERB frequency \( (F(2, 8) = .468, p = .642) \) but there was a significant effect for duration \( (F(2, 8) = 10.816, p < .05) \). However, examining the pairwise comparisons with Sidak adjustment shows that subjects with duration as their first dimension made a frequency durational difference that was significantly smaller than those with duration as their second dimension.

These results indicate that similarity in use of cues perceptually by subjects does not correspond with their productions.

**5.6.4 Subdialects**

Since the different parts of the greater Columbus area have different settlement histories, an analysis was attempted to determine if geographical differences between subjects could be related to different arcsine transformed percent correct scores in the discrimination task. Because there were few subjects per community, the map was quartered, with the four subjects in Westerville and northern Columbus comprising the northeast, the five subjects in Clintonville and Gahanna comprising the southeast,
the three subjects in Upper Arlington and Hilliard comprising the southwest, and
the five subjects in Worthington and Dublin comprising the northwest. The only
significant effect was for speech fricative duration \( (F(1, 13) = 16.006, p < .05) \), with
westerners having a score of 89.4%, and easterners 77.9%. There was no east-west
difference in the duration difference between the productions of \([\theta]\) and \([s]\) \( (F(1, 15) =
.256, p = .620) \).

Therefore, there is little indication that geographical distinctions in the Colum-
bus area are the cause of individual differences in perception. Socio-economic differ-
ences were not controlled, other than all subjects being students at a public university.

5.6.5 Music, language, and speech synthesis

There were no musician/non-musician effects (through counting from 0 to 4
the number of musical instruments played, though how proficient the listeners are is
unknown) on the arcsine transformed percent correct discriminations using a multi-
variate GLM \( (F(1, 9) = .347, p = .792, F(1, 9) = .093, p = .962, F(1, 9) = .842,
p = .504 \) for frequency, duration, or both cues, respectively, for non-speech vowels;
\( F(1, 9) = .296, p = .828, F(1, 9) = .165, p = .917, F(1, 9) = .193, p = .898 \) for
frequency, duration, or both cues, respectively, for non-speech fricatives; \( F(1, 9) =
10.315, p < .05,^{10} F(1, 9) = 1.092, p = .401, F(1, 9) = .103, p = .956 \) for frequency,
duration, or both cues, respectively, for speech vowels; \( F(1, 9) = .249, p = .860,
F(1, 9) = 3.242, p = .074, F(1, 9) = 1.549, p = .268 \) for frequency, duration, or
both cues, respectively, for non-speech fricatives). Nor were there effects based on
type of musical training (e.g., vocal, stringed, wind, or percussion instruments) on

---

\(^{10}\)The pairwise comparisons with Sidak adjustment were not meaningful (listeners who could play
two instruments had better discriminations than those who could play only one or three).
discrimination. Multivariate GLMs using the number of foreign languages known, and amount of experience with synthetic speech had no significant effects as well.

None of the probable causes have been able to fully explain the individual differences in perception for frequency and duration cues of speech and non-speech vowels and fricatives. However, some of the findings might be explained if the usage of frequency and duration by another dialect is compared with that of Columbus.

5.7 Language change

Since it appears that there are no obvious factors that can explain the individual differences in perception for speakers of this dialect, how do they compare to other dialects? Little research has been done on dialectal differences in [θ] and [s], but much vowel research has been performed. Could differences in production between dialects explain these findings in the perception of [i] and [i]? Why do the subjects not perform as well at discriminating vowels by duration as they do by F1 frequency?

To determine whether this Columbus dialect produces the /i/-/i/ contrast similarly to what has been reported in previous research, men’s and women’s average values were compared to those of nearby dialects in southern Michigan (Peterson & Lehiste, 1960; Lehiste & Peterson, 1961; Hillenbrand et al., 1995). More than thirty years had passed since Peterson and Lehiste’s studies in Ann Arbor, but the results found by Hillenbrand et al. in Kalamazoo show very little change has occurred in the production of F1 and duration of these vowels. The average F1 frequencies of [i] and [i] in Lehiste and Peterson (1961) were 320 and 410 Hz, respectively, a 90 Hz difference. Adult male speakers in Hillenbrand et al. (1995) produced average F1

11Some of Hillenbrand et al.’s subjects came from surrounding states.
values at 342 and 427 Hz, respectively, an 85 Hz difference. The average durations of [i] and [i] in Peterson and Lehiste (1960) were 240 and 180 ms, respectively, a 60 ms difference. Hillenbrand et al.'s adult male speakers produced average durations of 243 and 192 ms, respectively, a 51 ms difference.

The Columbus men’s average durations were 247 and 185 ms for [i] and [i], respectively, with a 62 ms difference, are not that different from the results in Peterson and Lehiste (1960) and Hillenbrand et al. (1995). The women’s average duration difference between the vowels of 67 ms is nearly the same as the 69 ms difference between the 306 ms [i] and 237 ms [i] produced by the women in Hillenbrand et al. (1995). This is illustrated in Figure 5.39.

![Figure 5.39: Michigan and Ohio values for duration of [i] and [i].](image-url)
For frequency, the Columbus men’s average [i] F1 is 428 Hz, as shown in Table 5.21, and it is a good match with the Michigan data. However, the F1 of [i] is 282 Hz, 146 Hz lower than [i], which is about a 70% greater Hz difference than for Michigan. Comparing Columbus women’s vowel F1 with the Michigan women’s F1 in Hillenbrand et al. (1995) also shows this difference. Michigan women have an average F1 of 483 Hz for [i], which is 46 Hz higher than the value for [i]. Columbus women’s average F1 for [i] is 483 Hz, 159 Hz higher than for [i]. This difference in F1 frequency is over 245% greater than it is for Michigan women. This is shown in Figure 5.40.

Figure 5.40: Michigan and Ohio values for F1 frequency of [i] and [i].
In summation, it appears that while the [i] and [ɪ] of Columbus speakers have similar durations as those of speakers from a neighboring dialect, the F1 frequency values have much more separation for the Columbus speakers. If there is a strong perception-production link, then the prediction for the vowel perception data is that frequency will be much more important than duration for discrimination, which is what was found.
CHAPTER 6

DISCUSSION

6.1 Detection versus interpretation

The differences in subjects’ performance cannot be attributed to differences in hearing thresholds, as shown by the failure of their hearing levels to account for their frequency discrimination performance. For example, as shown in Figure 5.34, subject 3 has the lowest overall hearing thresholds, 2 has the highest for the frequencies involved, and 14 is highest overall. However, as shown in Figure 5.1, for speech vowels subject 3 had the lowest frequency discrimination scores of all subjects, while 2 and 14 had the highest (the differences in scores were not significant, but the trend should have been in reverse order if threshold hearing levels were a factor in frequency discrimination). The subjects’ performance in relation to each other differs depending on the stimuli, since for the speech fricatives it is 3 and 14 who are two of the highest scorers. Although a positive trend between hearing ability and performance was not found, neither was the possible negative trend of sensitivity to frequencies subjects have high thresholds for (Doherty & Lutfi, 1999).

However, overall hearing thresholds ability might not be a good predictor of frequency discrimination. The configuration of the subjects’ auditory filters (Fletcher,
1940; Moore, 1986) would have been a better predictor, but the required testing would not have been feasible. It is possible that subjects may have different auditory filter center frequencies and bandwidths. But since there was a range of about 3 ERB for the vowel tokens from the lowest point of the low F1 to the highest point of the high F1 (shown in Figure 4.3), on average each frequency stimulus would most highly activate a different filter. The fricatives had only a 1.5 ERB range from the lowest point of the low F5 value to the highest point of the high F5 value (shown in Figure 4.4), with an average of only two or three separate filters being activated the most highly. But the fricative frequency contours would have activated different filters in different sequences (and the vowel formant contours would have done so as well), so unless a subject has cochleae that deviate greatly from average, it seems highly unlikely that these young adults with no history of hearing problems would have physiologies different enough to have a range of accuracy in discrimination ability that can span almost fifty percentage points. Even assuming that male cochleae are larger on average than female did not yield any significant differences in frequency discrimination, so any potential differences in auditory filters are probably too small to show in the results.

For duration discrimination, however, the theories on the location of this process on the perceptual cognitive pathway are more divergent. “Counting mechanisms” (e.g., Creelman, 1962) could be low level neural accumulators or higher level cognitive processes. The large range of duration discrimination abilities (larger than for frequency) suggests that the cause is differences in cognitive processes or strategies. If the range of abilities was caused by detection differences, and if duration is determined by the counting of frequency pulses, then the subjects, with no evidence of
problems in the F1 and F5 frequency ranges, must have a wide range in the accuracy of their counters. Cognitive differences is the simpler explanation: Subjects vary in their ability to compare a temporal sequence with its earlier pair stored in memory. The memory trace may not store temporal information with the same accuracy for different subjects, or their comparison strategies may be different.

Therefore, with evidence that the subjects had normal hearing thresholds for the frequencies involved, and the fact that the frequency and duration distinctions were above JND, the interpretation is that the individual differences documented in this dissertation are cognitive in nature. Whether the differences in weighting are caused by response biases, memory store size, or some combination of them cannot be determined by this study.

6.2 Identification

Comparing the subjects’ descriptions of speech and non-speech stimuli showed significant differences in the number of subjects who used duration to describe vowel tokens. Significantly fewer subjects used frequency to describe speech vowel tokens than non-speech vowel tokens, while the number of subjects who used duration descriptions did not differ between the two vowel types. Adding the fact that the majority of subjects heard words with [i] and [i] in them yields the conclusion that vowel categorization precludes the subjective contrasting of vowels by F1 frequency. However, subjects were still able to report length differences among the stimuli, which suggests that duration is not used as much as frequency in vowel categorization. Although there are few descriptives the general population can use to describe differences in vowel timbre (besides vowel identity), the question remains as to why there is a
difference between describing vowels and describing their non-speech analogues with the same timbres.

It is possible the subjects used a general auditory mode of listening while discriminating the synthetic speech tokens, using a speech mode only for the identification tasks. But as a group they performed significantly worse on frequency, and significantly better on duration contrasts for speech than for non-speech vowels.

The forced-choice results of the Ganong up-down task gives evidence that identification is not needed in the discrimination of speech stimuli. This seems to go against theories of speech categorization, but in this study the correct discrimination values are averaged over between- and within-category pairs (depending on whether the listener had more than one category). Results have not been tabulated for between versus within scores, so it may be that between-category discrimination has higher accuracy than within for each particular subject.

6.3 Cue usage for segment and speech types

Overall, non-speech discrimination scores are lower than speech scores. This implies that the ability to categorize the stimuli aids listeners in discriminating them. One problem with this conclusion is that some of the pairs might be classified identically (e.g., 41 and 42 for the vowels might both be labeled /hId/ since each one is high and non-long). This would hinder the subject in discrimination, since both tokens would be within the same class. However, examining the Ganong solutions in Appendix C reveals that in almost all cases the largest category has seven tokens or less (assuming that what the subject perceived during identification of the speech tokens is similar to what was perceived during the earlier discrimination). Since the
number of unordered pairings of the twelve stimuli is \( \binom{12}{2} = 66 \), then \( 7 \cdot 5 = 35 \) or \( 6 \cdot 6 = 36 \) yields a majority of between category comparisons. Add in the fact that there is much ambiguity between some of the categories, which increases the number of possible category distinctions (e.g., /θ/ vs. not-/θ/, /s/ vs. not-/s/, and /θ/ vs. /s/) and the likelihood that most of the discriminations are made between categories increases.

Duration scores are not significantly different for non-speech segments, but they are significantly higher for fricatives than vowels for speech. This suggests that duration is a better speech cue for fricative discrimination than vowel discrimination. Frequency or combined frequency and duration scores in non-speech are higher for vowels than fricatives, but there is no significant difference in speech between vowels and fricatives for frequency alone or frequency with duration. This suggests that the frequencies tested are used as distinguishing cues for both vowels and fricatives.

Examining speech and non-speech fricatives, it is shown that neither duration nor frequency is a stronger cue than the other, but having both leads to improved performance. This means that both frequency and duration are strong cues for distinguishing different noise tokens, whether or not they are speech. Vowels, however, are poorly distinguished by duration alone, since frequency is by far a better cue for discrimination. Duration is only effective if it is added to frequency in speech (there cannot be a significant increase in performance for non-speech because of the ceiling effect). For vowels, frequency is the strong cue, while duration is only secondary.

Returning to the combinatorial math used to determine that there are more between than within category comparisons for speech, this was for cue levels overall. It is true for pairs that differ by frequency and duration, but it is only true for
pairs that differ by one cue if the boundary corresponds (at least partially) with that cue. Investigating the Ganong solutions in Appendix C shows that the longest boundaries between vowel categories are between frequency levels, while fricatives can also contain long duration boundaries. This results in most duration comparisons for vowels to be within a category, and therefore duration differences are not fully distinctive.

### 6.4 Cue usage consistency within subjects

How consistent subjects were in using frequency and duration cues will be discussed in this section, with reference to Figures 5.1 and 5.10, and Tables 5.2–5.4 and 5.15. The subject who was most consistent overall is 10. Her percent correct scores for frequency were never significantly different from her scores for duration; her treatment of these cues in vowels was not significantly different from her treatment of them in fricatives; and her treatment of these cues in non-speech was not significantly different from her treatment of them in speech. Her only deviations from consistency are in the combination of cues. For non-speech vowels her discrimination of pairs that differ by both duration and frequency is better than duration alone but not frequency alone (but this may be a ceiling effect, judging by the Tukey HSD value), and for speech fricatives her discrimination of pairs that differ by both duration and frequency is better than frequency alone but not duration alone (although this may be a ceiling effect as well). For her MDS solutions, she always had duration as the second dimension for vowels and as the first dimension for fricatives.

One of the least consistent subject in regards to cue usage over different environments is 2. He has lower scores for duration than for frequency for non-speech
vowels, lower scores for frequency than duration for non-speech fricatives (the only subject with this result), but had no significant differences in frequency and duration scores for speech vowels or speech fricatives. Duration as a single cue is always worse than when it is combined with frequency for vowels, and frequency as a single cue is always worse than when it is combined with duration for fricatives. He is the only subject whose speech vowel frequency scores were higher than those for speech fricatives. In his MDS solutions, duration did not show up as a dimension for his speech vowels, while it was the fourth dimension for non-speech vowels. For fricatives, however, duration was always the first dimension.

Overall, most subjects were consistent at cue discrimination at least for subsets of the stimuli (e.g., for speech and non-speech vowels, or for speech vowels and fricatives). Subjects 7, 10, and 17 always individually discriminated frequency or duration with the same accuracy for all four stimuli sets. Subjects 2, 4, 5, 12, 14, and 16 individually had the same accuracy rates for frequency or duration for vowels, while 3, 11, and 13 had higher scores for frequency. Subjects 1, 3, 5, 6, 8–10, 12, and 13 individually had the same accuracy for frequency or duration for fricatives. Subject 15 always had lower scores for duration than frequency for non-speech, while 11 had the same difference but for speech. In regards to the contrast of individual cues with both together, subjects 1, 3, 5, 7, 8 11, 14–17 never increased in accuracy if duration was added to frequency. No subjects increased in accuracy if duration was added to frequency in vowels, but 2 and 4 did for fricatives. All subjects except for 2, 4, 9, and 12 did not increase in accuracy if duration was added to frequency in non-speech, while all subjects except for 2, 4, 6, 10, and 13 had the same result for speech. For
duration, all subjects increased accuracy if frequency was added for at least one stim-
uli set. Subject 16 did not increase in accuracy if frequency was added to duration
for vowels, while 1–7, 9, and 10 did not change for fricatives. Accuracy was increased
in vowels if frequency was added to duration for subjects 2, 3, 6, 9, 11-13, 15, and
17, while accuracy was increased for fricatives for 11 and 14 if frequency was added
to duration. Subjects 8, 11-15, and 17 had an increase in accuracy if frequency was
added to duration for non-speech, and 11 had that result for speech.

No subject has higher accuracy for non-speech over speech. Subjects 1, 2,
4–7, 10, 11, 13, 14 have no difference in accuracy between speech and non-speech
fricatives, while this holds true for all subjects except 6 and 12 for vowels. Subjects
1, 2, 4, 5, 7, 10, 11, 13, and 14 make no difference between speech and non-speech in
their accuracy for duration discrimination, all subjects except 3, 9, and 12 make no
difference for frequency, and all subjects except 3 have no difference between speech
and non-speech for the combination of frequency and duration

For vowels versus fricatives, subject 3 always has less accuracy on duration
for vowels than fricatives, and 2 always has more accuracy on frequency for vowels
than fricatives. All subjects except for 2, 3 and 7 have no difference between vowels
and fricatives for speech, while only 1, 7, 8, 10, 11, 15, and 17 can meet that criteria
for non-speech. All subjects but 2, 3, 6, 7, and 16 do not have duration accuracy
differences for vowels and fricatives, while 1, 8, 10, 11, 15, and 17 do not have fre-
quency differences, and all subjects except 3 have similar accuracies with frequency
and duration for vowels and fricatives.

In summary, it appears that there are many interlapping subsets of subjects
depending on whether they weight frequency greater than duration or not. However,
how a particular cue is used in speech for a particular subject does not depend on
whether it is in a vowel or a fricative for most subjects (although for the group there
is a difference between speech fricatives and vowels for duration). Most subjects do
not differ in their cue usage for speech and non-speech vowels, but a small majority
does for fricatives.

6.5 Cue usage differences among subjects

Compared to the variation within subjects, there are fewer between-subject
differences. There were no differences in performance between subjects for speech
pairs that contained both frequency and duration contrasts. This indicates that the
subjects form a single group, or dialect\textsuperscript{12} because they utilized “full” speech cues (i.e.,
F1 or F5 frequency and duration) similarly for vowels, and similarly for fricatives.
Even for the non-speech analogues, there were no differences between subjects (except
for subject 3 having a significantly lower accuracy than most subjects for the non-
speech fricatives).

But individual differences appear if subjects only have one cue for contrast.
Although there were no significant differences between subjects for frequency discrim-
ination on speech and non-speech vowels, subjects 3, 9, and 15\textsuperscript{13} performed signifi-
cantly worse than others (all subjects, three subjects, and one subject, respectively)
on speech vowel tokens that differed only by duration, and subjects 3, 6, 15, 12, 9,
and 11 performed significantly worse than others (fourteen, twelve, eight, three, one,
and one subject, respectively) on non-speech vowels. This indicates that frequency

\textsuperscript{12}or language, since the ability of members of other dialects was not tested.

\textsuperscript{13}Subjects 9 and 15 went to the same school, but since they were significantly different from each
other for frequency cues, having a similar peer group is probably not a significant factor in this case.
differences are enough to distinguish speech and non-speech vowel tokens for everyone, but a small minority of subjects are significantly less accurate than others in distinguishing vowels with only duration differences, with even more having difficulty for the non-speech versions.

For the fricatives, on the other hand, each singleton cue discrimination had subjects that scored significantly lower than others. For frequency discrimination on speech fricatives, subject 10 was the only one who scored significantly lower than others (five subjects), while for the non-speech tokens, subjects 12, 9, 6, 3, and 4 performed significantly worse than other subjects (twelve, seven, six, five, and five subjects, respectively). For duration discrimination on speech fricatives, 11 and 14 had significantly lower scores (than eight and three subjects, respectively), while for non-speech fricatives 16, 15, 12, and 7 had significantly worse discrimination than others (ten, eight, four, and two subjects, respectively). Subject 2 had higher duration discrimination scores on non-speech fricatives than seven other subjects.

Speech fricatives can be discriminated by either cue by most subjects, except for a few who do not perform as well as the others when either frequency or duration cues are missing. A larger number of subjects perform worse than others when a cue is missing for non-speech fricatives.

Subject 3 has the lowest overall accuracy—he is the lowest or one of the lowest scorers for both frequency and duration or frequency alone for non-speech fricatives, and he is the lowest scorer on duration for vowels. However, this does not mean that he cannot perform the task, because he was one of the higher frequency scorers for speech fricatives, and he was near the middle in the range of accuracies for duration in fricatives. The interpretation is that he utilizes duration information less than
other subjects in discriminating vowels (speech or non-speech), and his frequency
discrimination abilities are so low in comparison with those of other subjects that
even adding duration to frequency does not improve his standing.

Subjects 12, 15, 9, and 14 were the only other subjects who scored significantly
lower than others multiple times. Subjects 12, 15 and 9 had lower discriminability of
duration for non-speech vowels, 12, 15, and 14 had lower discriminability of duration
for non-speech fricatives, 15 and 9 had lower discriminability of duration for speech
vowels, 14 had lower discriminability of duration for speech fricatives, and 12 and 9
had lower discrimination of frequency for non-speech fricatives. This indicates that if
a subject has difficulty in duration discrimination of one type of stimuli, she is likely
to have difficulty with most of them.

To summarize, virtually all subjects had equal abilities in discriminating stim-
uli that contrasted by both frequency and duration, no matter whether they were
speech, non-speech, vowel, or fricative. However, if there is only one cue present
for contrasting a pair of stimuli, certain subjects performed significantly worse at the
task than others. For vowels, it is only duration by itself that lowered discriminability
for some listeners—frequency distinctions alone are enough for discrimination by all.
For fricatives, a small majority performs less accurately if duration is the only cue,
and another small majority performs less accurately if frequency is the only cue, with
only one subject in both sets.

6.6 Speech modality

Most theories of speech perception (as do many theories which involve humans)
are not based on individuals but are based on the ideal individual as defined by the
group average. Of course they acknowledge differences, but mostly treat the variance as error.

Since this dissertation examined the usage of cues in speech and non-speech, the results relate to the debate of whether speech perception is a separate modality from other forms of perception. One main theoretical contrast between speech and non-speech as that speech is categorical, while general audition is continuous (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The design and analysis of this experiment was in regards to the matrix of stimuli, instead of examining a possible increase in discriminability across a boundary.

Since for this study there seems to be a correlation between speech and non-speech differences, it seems more likely that speech uses the perceptual processes already available to the listener, instead of having a duplication of “errors”. Liberman et al. (1957) and others investigated categorical perception, in which speech sounds can be discriminated only as well as they can be identified as belonging to two different categories, which is in opposition to continuous discrimination of non-speech sound stimuli. But when non-speech sounds, like sine-waves, are manipulated to mimic certain speech cues, speech can be heard by some people (Remez et al., 1981), and categorized by some of them (e.g., Best et al., 1981; Johnson & Ralston, 1994; Hodgson & Miller, 1996).

If speech is special because it is categorical, then what is the implication of people having slightly different categories? In cue trading relations, boundaries can shift presumably because listeners know how cues are integrated with each other, and

\[^{14}\text{Although not all speech sounds are strictly categorical, as there is some discrimination within long (vowel) sounds, presumably because of the extra information in memory available during comparison (Pisoni, 1973).}\]

198
the loss of one is compensated by another. But what if some people do not shift because they only attend to one of the cues, or only shift slightly because one cue is weighted much more heavily than the other?

Acoustic cues are attended to for the purpose of the task at hand. In speech perception, identifying the phone that one hears in the word “?ack” as [p] does not help if the goal is to decide if the word is “back” or “pack”. In this case, should the acoustic information of [p] be classified with [b] as /b/, or with [pʰ] as /p/?¹⁵ Speech is used to convey meaning, so even if nonsense tokens are being played, what prevents the subject from trying to look up the “words” in their lexicon? If /b/- and /p/-initial words are activated by the stimuli, a discrimination task may only be finding how highly activated one neighborhood is over another. If two neighborhoods are activated, one for each of the two stimuli, then there is a high level of discrimination. If the two stimuli activate the same neighborhood, then there is low discrimination of the two. Other acoustic tasks require finer distinction, such as using frequency changes to determine from which direction a large object is heading. Classifying it as merely “left” or “right” might not give you enough information to prevent collision.

Individual (and dialect and language) differences in the classification of VOT indicate that there are no standard values for /p/ and /b/, or /pʰ/ and /p/, or /pʰ/ and /p/ and /b/. Different individuals have slightly different criteria in deciding what value(s) pushes a stimulus to one category or the other. Since different people have different language experience, their boundary locations indicate what cues weightings have served them well (and they have to change, indicated in Lasky, Syrdal-Lasky,

¹⁵I am not claiming that allophonic differences are not used in lexical decision (e.g., /b/ can devoice word initially), just that people do not have a separate category to label an allophone with for a word identification task.
and Klein’s (1975) results of infants in Spanish-speaking environments not attending to Spanish VOT differences). People within a dialect should theoretically have closer boundary locations than people across dialects, but individual differences (exposure to other dialects/languages, hearing loss, favorite speech sounds as a child) could blur the average lines for the group. One person’s boundary might approach another’s category center.

Of course, certain people can break out of the phoneme classification because they attend to a different task. People who work in the food industry can identify the components of a smell, while phoneticians can detect smaller amounts of aspiration. Whenever new people are encountered, one performs sociolinguistic analyses on their accents. In determining that a talker is from a different dialect, how can this be done without discriminating between items within a category?\(^{16}\) Noting that one person’s /t/ sounds different (it is more dental), or that another person has distinctive vowels (lax vowels with extreme ingliding) does not mean that you could not identify the /t/ or /r/. If you did not hear the phones as belonging to /t/ and /r/, you would not have understood the talker (because English does not have /t/ in opposition to /t/, and /r/ in opposition to /r/). You placed the phones on the correct side of the boundary, but you also noted that they were far from the category norm (which depends on the individual). In this instance, you are performing two tasks at once: speech recognition, and talker/dialect recognition. The second task cannot be done only between phonemes, while the first task is.

Is there evidence that speech MDS solutions have more categorization than non-speech? Comparing non-speech vowels in Figure 5.17 with the speech vowels in

\(^{16}\) Although, some of their allophones might be nowhere near yours.
Figure 5.19, there does appear to be some frequency clustering (recall that as a group, subjects have significantly lower frequency discrimination scores for speech vowels than non-speech). [i] appears to have pulled the lowest two frequency levels closer together in the speech solution than in the non-speech solution. Spacing between duration levels does not appear to be different between speech and non-speech, but the non-speech duration dimension is more compact in comparison to speech (as a group, subjects have significantly lower duration discrimination scores for non-speech vowels than speech).

There also appears to be categorization for the fricatives as well. The frequency levels for non-speech shown in Figure 5.18 are further apart than for speech shown in Figure 5.20 which have the two lowest frequencies pulled together for [θ] and the two highest pulled together for [s]. However, the group results for frequency showed that it was discriminated significantly more accurately for speech, not non-speech. But the range for tokens is wider for the frequency dimension in speech than it is for non-speech. The duration levels appear to be more evenly dispersed in speech than in non-speech, which may account for why duration discrimination was significantly better in speech, but there is no evidence of categorization. Although with only three levels, it would be difficult to determine whether there is categorization if the boundary is at the middle level.

But do these observations still hold for subgroups based on listeners’ duration cue usage? For the strong duration subgroup for vowels, the non-speech solution shown in Figure 5.22 does have less frequency and duration grouping than the speech solution shown in Figure 5.28. [i] has pulled the lower two frequency and the higher two duration levels together. The weak duration subgroup shows that non-speech
frequency in Figure 5.23 is more widely dispersed than speech frequency is in Figure 5.29, since [i] pulled the two lower frequency levels together. Duration, however, seems to be more categorical for the non-speech tokens, with the two longer levels closer together. The duration level has greater spread for speech vowels, though. For the no duration vowel subgroup, the non-speech frequency is much more widely dispersed in Figure 5.24 than the speech frequency in Figure 5.30. However, the speech vowel configuration has pushed the [i] and [i] endpoints closer together, which is not evidence for their categorization.

For fricatives, the first dimension duration subgroup’s results for non-speech in Figure 5.25 seem to have less frequency categorization than their speech results in Figure 5.31. [s] and [θ] each pull the two frequency levels adjacent to their endpoints closer to them. Duration levels appear similar, but the speech distribution may be a drop greater. Second dimension duration subgroup solutions show that the non-speech in Figure 5.26 might have a drop less frequency categorization for [s] than speech in Figure 5.32, but the lower two duration levels in non-speech seem to be closer than they are in speech. For the weak-to-no duration subgroup, the non-speech tokens have the fricative endpoints at extreme coordinates in the space shown in Figure 5.27, while they are closer together for speech in Figure 5.33.

It appears that frequency has greater categorization for speech than non-speech tokens, but only for the subgroups that have stronger duration dimensions. Duration only seems categorical in speech for the strong duration vowel subgroup. But as there are only three duration levels, if the middle level falls on the boundary, it would not be identified strongly with either endpoint.
In summary, what is shown by these data is that speech vowels and fricatives were heard as speech by nearly all the subjects, while the non-speech tokens had different interpretations by different subjects, so there is an effect of familiarity to aid in the perception of speech. The fricative quality was not as good as the vowel, so if the speech fricatives are not very speechlike, that might explain why their categorization is more similar to their non-speech analogues than the speech vowels are to theirs. The unsatisfying answer that these results seem to provide is that speech perception is more categorical than non-speech, but only for good categorizers.

6.7 Feature phonology

The two features that relate to these results are the [tense]/lax distinction and the [sibilant]/non-sibilant distinction. Since there are no duration distinctions in English that do not have a corresponding frequency difference as well (e.g., timbre for vowels, noise frequency for fricatives) it would seem that the primary cue of [tense]/lax is lower F1/higher F1 (although F2 is important as well). The secondary cue would be duration. The primary cue for [sibilant] would probably be loudness, but that was not tested. Different spectral envelopes are used to distinguish fricatives, so differences in F5 would correspond to this, and would be the primary cue (although this would also correspond to [place]). Duration would be secondary, and would have been secondary if loudness was used in the experiment.

Are the secondary cues necessary for these contrasts? The results seem to support the hypothesis that the [tense]/lax distinction of vowels has more to do with frequency than duration. No subject was able to utilize duration to a better extent than frequency. This suggests that duration is not needed (no subject increased their
discriminability score when duration was added to frequency contrasts, but they did
as a group). Historically, the /iː-/i/ length contrast became /aɪ-/i/, so phonon-
morphologically, these two phonemes would maintain the [tense]/lax distinction (e.g.,
hide-hid), while /i/-/i/ would be better described with [ATR] or peripheral/central.
But if other dialects do perceive the length distinction better in relation to frequency,
then there would still need to be a cover term.

However, for [sibilant], it appears that with these stimuli both frequency and
duration can be primary cues, for different listeners. Frequency could still be used for
[place], and duration can be collapsed into loudness (since longer tokens are louder
than shorter) which was the original configuration for [sibilant].

6.8 Sound change

Over time, differences in listeners’ perceptions can result in the change of
phonological contrasts. If the weighting of perceptual cues is somehow altered (e.g.,
from the misanalysis of cue importance), then the production of these cues could
be modified to match this new usage. Or, if production is leading the change (e.g.,
from the modeling of prestigious talkers), then perception would need to adapt to the
new pronunciation. And a third alternative could be perception and production are
modified in lockstep, each forcing the other along. Whatever the leading cause of a
sound change, the effect is that perceptual cue usage will be changed as a result.

Changes in perceptual cue weights can account for many observed sound
changes. Mergers can be explained by the lack of attention to a particular cue which
creates a loss of contrast (e.g., final devoicing). Splits can be the result of attend-
ing to a cue that was previously non-contrasting (e.g., tonogenesis). And changes

204
in phonetic realization can also be explained, such as the case of Columbus /i/-/ı/ contrast.

6.8.1 Sub-phonemic sound change in Columbus

The Columbus talkers had vowel durations similar to the Michigan talkers, but had a greater difference in F1 values of [i] and [ı]. Although there is no perceptual data for frequency and duration discrimination of Michigan vowels, it seems likely that they would be less dependent on frequency in relation to duration since their frequency differences in F1 are not as robust as the F1 differences for Columbus. If this study were run on both dialects the results found might be similar to those regarding the use of duration in Swiss French as opposed to its lack of use in standard French (Miller & Grosjean, 1997).

Focusing on possible future change within the Columbus talkers, it seems that the /i/-/ı/ contrast may become less perceptually dependent on duration. It is actually less dependent on duration now than what these results show, because the short and long tokens were decreased and increased, respectively, by a standard deviation from the talker’s productions. Four subjects were significantly poorer at discriminating vowels by duration than by frequency, and the group as a whole had the same effect. These four subjects in addition with five others gained a significant increase in discriminability if frequency differences were added to the duration differences. Assuming that General American uses a tense/lax contrast for /i/ and /ı/ with frequency and duration differences, this Columbus dialect might be heading towards a contrast by F1 (and F2) frequency, without the use of a duration contrast.
Although there is a range of ability for the fricatives, with one subject being significantly better at duration contrasts, and two at frequency contrasts, Columbus listeners have no overall significant difference in the use of frequency or duration. It may be possible for Columbus talkers to shift to no duration differences, but the alternative of having no frequency differences may not be articulatorally or acoustically possible.

6.8.2 Prestigious listeners

Many theories of linguistic change through social adaptation, such as accommodation theory (e.g., Giles, 1973), or the modeling of prestigious talkers (e.g., Labov, 1963), have the talker as the focal point of change. These theories posit that listeners, noticing that the speech of one or more talkers is different from some standard form or from their own speech, attempt to mimic some of the characteristics of the talkers’ productions if they want to conform.

Alternatively, could listeners be at the center of change? Sound change caused by listeners’ misperceptions (e.g., Ohala, 1981) is one way to explain certain historical processes, and the different Columbus subgroups’ perceptual use of vowel cues could be part of this type of change. But what if, instead of prestigious talkers, there are prestigious listeners? Then, instead of people accommodating by imitating the speech of others, they accommodate by trying to produce the cues needed by the listener for comprehension. Could certain sound changes happen because of “poor” listeners, who cause those around them to change their productions?
6.9 Conclusion

In this dissertation it has been found that people use cues differently for discriminating phonological contrasts. They also are different in their consistency of cue weighting across different speech contrasts. Columbus listeners’ low accuracy in distinguishing tense/lax vowels by duration calls into question the use of duration as a cue in this contrast, at least for this dialect. Duration may need to be designated as a secondary cue, with frequency as primary. And finally, these individual differences in cue perception may be linked to the spread of language change.
APPENDIX A

QUESTIONNAIRES

Sound Differences Experiment Background Questionnaire

Name: ___________________________________ E-mail/Phone: ___________________________________

Age: ____ Sex: ____ Race: _______________ Class: 201_371___

Where have you lived (part of town, e.g. northern Worthington), and how old were you at the time?:

Place: Ages:

___________________________ _______

___________________________ _______

What are the names of the schools (pre-school and up) you have attended?:

Any speech or hearing disorders? If yes, please list: No Yes: ________________________________

010901a

Figure A.1: Initial recruitment questionnaire
Sound Differences Experiment Background Questionnaire II

Name: ___________________________________

Where were your parents/caretakers born, and where did they grow up?:

Who: Birth place: Childhood place(s):

________________ ________ ______________________ ________________________________________

________________ ________ ______________________ ________________________________________

Do you know any foreign languages? How old were you when you started learning? How well can you read/write and speak/comprehend?

(Circle: (1) not at all, (2) poorly, (3) passably, or (4) fluently)

Language: Age: Reading: Writing: Speaking: Comprehension:

________________ ________ 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

________________ ________ 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

________________ ________ 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

Do you sing or play musical instruments? If yes, please list: No Yes: ____________________________________

Are you left-handed or right-handed?: Left Right

Figure A.2: Secondary questionnaire for participants
Sound Differences Experiment Background Post-Test Questionnaire

Name: ___________________________________

Have you heard synthetic (computer) speech before?:  No  Yes

Could you list where/what types you heard (weather channel, phone system, etc.):
____________________________________________________________________________________

Have you heard sine-wave speech before?:  No  Yes

Which Linguistics classes have you taken?:

Intro. 201, 202, 601  Historical 385, 611  Socioling. 361, 365, 661.x
Phonetics 500, 600.x  Phonology 503, 603.x  Psycholing. 371, 615  Field methods 550, 650.x

Any Speech & Hearing classes?: _________________________________________________________

020219

Figure A.3: Post-test questionnaire
APPENDIX B

VOCOID TRAJECTORIES

Example vocoid durations followed by example vocoid spaces of the subjects are displayed in this appendix. These are based on values averaged over only two to six tokens per vocoid sequence, so they provide only a rough idea of what a subject’s set of vocoid durations or vocoid space might be.

Each trajectory contains only the center 80% of the vocoid or voicoid sequence in order to minimize (but not eliminate) effects of the surrounding context. Every 10% of the way through a vocoid sequence is labeled phonemically. The first point (i.e., 10% of the way through the vocoid) is indicated by a “○”, and the last point (90%) is indicated by a “△”. Vocoid sequences (e.g., /il/, /o1/, or /au/) include the following liquid or glide in the trajectory, not just the vowel (although for these subjects a postvocalic /l/ is highly vocalized).

Captions will include results from the commutation test, such as “/a/ = /ɔ/” (i.e., /a/ and /ɔ/ were significantly misidentified with each other) and “/u1/→/ɔ1/” (i.e., /u1/ was significantly misidentified as /ɔ1/, but /ɔ1/ was significantly identified correctly).

All subjects show a fronted /u/, which is not fronted before /l/, and is fronted before /ɔ/ only for subjects 2, 5, and 17. All subjects have an /e/ that glides to a
lower F1 than /i/. For some subjects, /a'/ and /a"/ might be better transcribed as [aʰ] and [æʰ], respectively.

Figure B.1: Duration of vocoids for subject 1 (talker).
Figure B.2: Duration of vocoids for subject 2.

Figure B.3: Duration of vocoids for subject 3.
Figure B.4: Duration of vocoids for subject 4.

Figure B.5: Duration of vocoids for subject 5.
Figure B.6: Duration of vocoids for subject 6.

Figure B.7: Duration of vocoids for subject 7.
Figure B.8: Duration of vocoids for subject 8.

Figure B.9: Duration of vocoids for subject 9.
Figure B.10: Duration of vocoids for subject 10.

Figure B.11: Duration of vocoids for subject 11.
Figure B.12: Duration of vocoids for subject 12.

Figure B.13: Duration of vocoids for subject 13.
Figure B.14: Duration of vocoids for subject 14.

Figure B.15: Duration of vocoids for subject 15.
Figure B.16: Duration of vocoids for subject 16.

Figure B.17: Duration of vocoids for subject 17.
Figure B.18: Center 80% of vocoids of subject 1 (talker). Results of commutation test: i→e; a=a; ol=ʌl; ol→ʌl
Figure B.19: Center 80% of vocoids of subject 2. Results of commutation test: α→γ; ɔ→ʊ; ɔl→ol.
Figure B.20: Center 80% of vocoids of subject 3. Results of commutation test: α=ɔ; ɔi→œi; œl→λl; ol→ul
Figure B.21: Center 80% of vowels of subject 4. Results of commutation test: $a \rightarrow o$; $u \rightarrow o$; $\alpha \rightarrow \alpha$.
Figure B.22: Center 80% of vowels of subject 5. Results of commutation test: ol→dl
Figure B.23: Center 80% of vocoids of subject 6. Results of commutation test: $a\sim\sigma$; $o\sim\sigma$ ($[\sigma]$ is lower than $[\sigma]$); $ol\rightarrow ul$
Figure B.24: Center 80% of vocoids of subject 7. Results of commutation test: il→il; α=ɔ; Al→Al; ul→ul→ol ([ul] is higher than [ul])
Figure B.25: Center 80% of vocoids of subject 8. Results of commutation test: a=ɔ; u=ɔ; Al→al; ul→ol→ol
Figure B.26: Center 80% of vocoids of subject 9. Results of commutation test: ï→ë; ɔl→ɔl; ɔl→ul
Figure B.27: Center 80% of vocoids of subject 10. Results of commutation test: α=ɔ; ɔi=ɔi; ɔl=ɔl; ul=ɔl=ɔl
Figure B.28: Center 80% of vocoids of subject 11. Results of commutation test: $a=\sim; \text{al} \to \text{Al}; \text{ul} \to \text{ol} \to \text{ol}$
Figure B.29: Center 80% of vocoids of subject 12. Results of commutation test: o→ (\[ul\] is lower than [ol]); a1→a1
Figure B.30: Center 80% of vocoids of subject 13. Results of commutation test: All vocoid sequences distinct.
Figure B.31: Center 80% of vocoids of subject 14. Results of commutation test: ɔ→α; ʊi=ɛ; ɑl→ɬ; ʊl→ɔl
Figure B.32: Center 80% of vocoids of subject 15. Results of commutation test: $\alpha \Rightarrow u \Rightarrow o;\ \dot{\text{o}} \Rightarrow \text{ol}$.
Figure B.33: Center 80% of vocoids of subject 16. Results of commutation test: $a \rightarrow \varepsilon; \text{ol} \rightarrow \text{ol}$
Figure B.34: Center 80% of vocoids of subject 17. Results of commutation test: ɪ=ɛ; ʌl→ol; ʊl→ʊl→ol
APPENDIX C

GANONG UP-DOWN SOLUTIONS

What follows are the results of the Ganong up-down forced-choice identification task for each of the four stimuli types. A grid will display the stimuli space, with the lower left token having the lowest values of frequency and duration, and the higher right token having the highest. Rows represent frequency, and columns represent duration. A “□” will indicate the endpoints of a continuum (i.e., tokens 41 and 13 for [i] hid and [i] heed, respectively, and tokens 11 and 43 for [θ] bath and [s] bass, respectively.

If a subject indicates hearing the same word for both the up and down portions of the continuum, then the associated phone will be placed in the grid (e.g., if bath was heard by the subject for token 11 both times it was played, then [θ] will be placed in the appropriate position). If the subject does not hear the same word both times, then the position in the grid for that token will be left blank (e.g., the subject heard bath one time but bass the second time).

Determining whether a subject depends on frequency or duration values of the phones to identify the words involves the orientation of the boundary of a phone. If the boundary is horizontal, then this indicates that a specific frequency cut-off defines whether a token is identified as a particular phone or not. If the boundary is vertical,
then duration is the defining cue. A diagonal boundary indicates that both frequency
and duration are important in determining membership.

Subjects reported that this task was more difficult than the discrimination
task. These results do not compare well with the individual MDS solutions in Ap-
pendix D. For the speech stimuli, only 65% of the vowels and 59% of the fricatives
seem to have similar frequency and duration usage between the boundaries in the
Ganong solutions and the point clusters in the MDS solutions. Since there were only
two repetitions of each token in the Ganong task, and five for each possible pair in
the discrimination task, the Ganong task may not be sensitive enough for compari-
son. Subject mistakes in answering cannot be distinguished from actual confusion in
identity with only two responses per token.

<table>
<thead>
<tr>
<th></th>
<th><strong>Non-speech</strong></th>
<th></th>
<th><strong>Speech</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>i i i</td>
<td></td>
<td>i i i</td>
<td>s s s</td>
</tr>
<tr>
<td></td>
<td>0 0</td>
<td>1 1 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>i i i</td>
<td>0 0 0</td>
<td>i i i</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>

Table C.1: Subject 1’s (talker) Ganong up-down solutions. All four stimuli types seem
to have categorization by frequency, but non-speech [θ] and speech [i] do not contain
the longest token at their respective boundary frequencies. None of the non-speech
fricatives were unanimously identified as [s].
Table C.2: Subject 2’s Ganong up-down solutions. The non-speech results are a bit unclear, though it seems that the vowels are classified by frequency, while the fricatives may be classified by frequency and duration. For speech vowels, [i] is classified by frequency, while [I] is defined by both frequency and duration values close to the endpoint. Speech fricatives are classified by frequency.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowels</td>
<td>Fricatives</td>
</tr>
<tr>
<td>[I] i i i</td>
<td>s s</td>
</tr>
<tr>
<td>i i i</td>
<td>0</td>
</tr>
</tbody>
</table>

Table C.3: Subject 3’s Ganong up-down solutions. Non-speech [i] has a frequency boundary, while non-speech [i] is only unanimously acceptable at the endpoint’s frequency and duration values. There were no tokens identified as [θ] for the non-speech fricatives, while [s] has no interpretable boundary. The speech tokens need both frequency and duration for identification because [i] and [θ] can advance into another frequency level (pushing [i] and [s] back) only at the shortest duration level.
Table C.4: Subject 4’s Ganong up-down solutions. The non-speech tokens seem to have duration and frequency boundaries, while the speech tokens seem to have mostly frequency boundaries. However, [i] was perceived at the second frequency level if it was at the longest duration, but [θ] could not be perceived at the second frequency level at the longest duration.

Table C.5: Subject 5’s Ganong up-down solutions. The category boundaries for non-speech tokens are defined by frequency, but none of the fricative tokens were unanimously identified as [s]. The speech categories also seem to have mainly frequency boundaries, but [i] can be heard at a higher frequency level only at the longest duration, while [s] cannot be heard at the shortest duration.
Table C.6: Subject 6’s Ganong up-down solutions. The non-speech tokens seem to be defined by duration, but [i] is not heard unanimously at the highest frequency, and [s] is not heard unanimously at the lower two frequency levels. Speech fricatives also seem to be defined by duration, but speech vowels seem to have categories defined by frequency and duration, clustered at the endpoints.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>1</td>
<td>i</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
</tbody>
</table>

Table C.7: Subject 7’s Ganong up-down solutions. For the non-speech vowels, [i] have high F1 or short duration, while the only unanimously labeled [i] was the endpoint. The non-speech fricative results are unclear, but speech fricatives appear to be divided by duration, although [θ] does not appear at the highest frequency level. Speech vowels seem to have [i] classified by frequency and duration that spreads throughout the lower right triangle of the space, while [i] is only labeled at the highest frequency level.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
</tbody>
</table>

242
Table C.8: Subject 8’s Ganong up-down solutions. The vowel solutions are not clear. The non-speech fricatives have no unanimously identified [θ], but [s] never occurs at the shortest duration. Speech fricatives also are categorized by duration, but [θ] occurs in some of the mid durations as well as short.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>[ ]</td>
<td>s s</td>
</tr>
<tr>
<td>i i i</td>
<td>s</td>
</tr>
<tr>
<td>i i i [ ]</td>
<td>s</td>
</tr>
</tbody>
</table>

Table C.9: Subject 9’s Ganong up-down solutions. Non-speech vowels have a frequency boundary, while non-speech fricatives have a frequency boundary but allow [θ] at a higher frequency at the shortest duration. Speech vowels have a frequency boundary but with [i] at a higher frequency with longer duration. Speech fricatives can be divided with a frequency boundary, but there may be duration involved to increase the number of perceived short [θ]s and long [s]s.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>[ ] i i [ ]</td>
<td>s s s</td>
</tr>
<tr>
<td>i i i θ s s</td>
<td>s</td>
</tr>
<tr>
<td>i i i θ θ θ</td>
<td>i i i θ θ</td>
</tr>
</tbody>
</table>

243
Table C.10: Subject 10’s Ganong up-down solutions. Frequency and duration boundaries divide both non-speech and speech vowels, with [i] occurring at short durations or mid durations at higher frequencies, while [i] is clustered around its endpoint. The fricative token identifications are unclear.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>[i]</td>
<td>[θ]</td>
</tr>
<tr>
<td>[i]</td>
<td>[θ]</td>
</tr>
<tr>
<td>[i]</td>
<td>[θ]</td>
</tr>
</tbody>
</table>

Table C.11: Subject 11’s Ganong up-down solutions. Duration appears to separate the non-speech vowel categories, and frequency appears to do so for the non-speech fricatives, but [s] was never identified unanimously. For speech vowels, frequency and duration boundaries restrict [i] close to its endpoint, while [θ] occupies the frequency and duration levels around it. For its endpoint, [s] was not unanimously labeled, while [t] occupied all other tokens.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>[i]</td>
<td>[θ]</td>
</tr>
<tr>
<td>[i]</td>
<td>[θ]</td>
</tr>
<tr>
<td>[i]</td>
<td>[θ]</td>
</tr>
</tbody>
</table>
Table C.12: Subject 12’s Ganong up-down solutions. Non-speech vowels might be split by duration, while non-speech fricatives and the speech tokens appear to be categorized by frequency.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>I I I</td>
<td>s s s</td>
</tr>
<tr>
<td>I I I</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>

Table C.13: Subject 13’s Ganong up-down solutions. The non-speech tokens and the speech fricatives are uninterpretable, but the speech vowels appear to be separated by frequency.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>I I I</td>
<td>s 0 s</td>
</tr>
<tr>
<td>I I I</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>

Table C.14: Subject 14’s Ganong up-down solutions. Vowel tokens appear to be categorized by frequency, while fricative tokens use frequency but can have [0] at a higher frequency level at the shortest duration.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vowels</strong></td>
<td><strong>Fricatives</strong></td>
</tr>
<tr>
<td>I I I</td>
<td>s s s</td>
</tr>
<tr>
<td>I I I</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>
Table C.15: Subject 15’s Ganong up-down solutions. Vowel tokens are split by frequency. [s] is never unanimously identified in the non-speech fricatives, but [θ] appears to have a frequency boundary. Speech fricative responses are uninterpretable.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowels</td>
<td>Fricatives</td>
</tr>
<tr>
<td>[i] [i] [i]</td>
<td>[s] [s] [s]</td>
</tr>
</tbody>
</table>

Table C.16: Subject 16’s Ganong up-down solutions. Vowels are categorized by frequency, but non-speech [i] is perceived at a lower frequency for the shortest duration. Non-speech fricatives have frequency boundaries, but [θ] can occur at a higher frequency at the shortest duration. Speech fricatives, on the other hand, appear to be categorized by duration.

<table>
<thead>
<tr>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowels</td>
<td>Fricatives</td>
</tr>
<tr>
<td>[i] [i] [i]</td>
<td>[s] [s] [s]</td>
</tr>
</tbody>
</table>
Table C.17: Subject 17’s Ganong up-down solutions. Non-speech vowels are separated by duration, but are clustered around their endpoints. Non-speech fricatives have a frequency boundary, but [θ] can occur at a higher frequency for the shortest duration. Speech vowels are categorized by frequency, while speech fricatives are uninterpretable.
APPENDIX D

INDIVIDUAL MDS SOLUTIONS

This appendix contains the multidimensional solutions run individually for each subject for each of the four stimuli sets. The resulting dimensions provided the cue ordering data in Table 5.13, and correlations between stimulus values and dimensions provided the cue fitting data in Table 5.14. Stress and $R^2$ values are shown in Table D.1.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Non-speech</th>
<th></th>
<th>Speech</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowels</td>
<td>Fricatives</td>
<td>Vowels</td>
<td>Fricatives</td>
</tr>
<tr>
<td></td>
<td>Stress</td>
<td>$R^2$</td>
<td>Stress</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Subject 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.567</td>
<td>0.273</td>
<td>0.575</td>
<td>0.244</td>
</tr>
<tr>
<td>2</td>
<td>0.333</td>
<td>0.563</td>
<td>0.336</td>
<td>0.496</td>
</tr>
<tr>
<td>3</td>
<td>0.279</td>
<td>0.663</td>
<td>0.231</td>
<td>0.662</td>
</tr>
<tr>
<td>4</td>
<td>0.137</td>
<td>0.875</td>
<td>0.180</td>
<td>0.758</td>
</tr>
<tr>
<td>5</td>
<td>0.096</td>
<td>0.928</td>
<td>0.159</td>
<td>0.800</td>
</tr>
<tr>
<td>6</td>
<td>0.074</td>
<td>0.954</td>
<td>0.149</td>
<td>0.820</td>
</tr>
</tbody>
</table>

(Table D.1 continues)
Table D.1 (continued)

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Non-speech</th>
<th></th>
<th>Speech</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowels</td>
<td>Fricatives</td>
<td>Vowels</td>
<td>Fricatives</td>
</tr>
<tr>
<td></td>
<td>Stress R²</td>
<td>Stress R²</td>
<td>Stress R²</td>
<td>Stress R²</td>
</tr>
</tbody>
</table>

Subject 2

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.595</td>
<td>0.233</td>
<td>0.553</td>
<td>0.299</td>
<td>0.604</td>
<td>0.219</td>
<td>0.567</td>
<td>0.253</td>
</tr>
<tr>
<td>2</td>
<td>0.342</td>
<td>0.564</td>
<td>0.298</td>
<td>0.609</td>
<td>0.334</td>
<td>0.534</td>
<td>0.297</td>
<td>0.604</td>
</tr>
<tr>
<td>3</td>
<td>0.288</td>
<td>0.648</td>
<td>0.225</td>
<td>0.704</td>
<td>0.237</td>
<td>0.686</td>
<td>0.189</td>
<td>0.777</td>
</tr>
<tr>
<td>4</td>
<td>0.178</td>
<td>0.808</td>
<td>0.187</td>
<td>0.766</td>
<td>0.174</td>
<td>0.797</td>
<td>0.148</td>
<td>0.840</td>
</tr>
<tr>
<td>5</td>
<td>0.100</td>
<td>0.921</td>
<td>0.172</td>
<td>0.791</td>
<td>0.130</td>
<td>0.867</td>
<td>0.129</td>
<td>0.867</td>
</tr>
<tr>
<td>6</td>
<td>0.077</td>
<td>0.949</td>
<td>0.168</td>
<td>0.799</td>
<td>0.108</td>
<td>0.901</td>
<td>0.110</td>
<td>0.899</td>
</tr>
</tbody>
</table>

Subject 3

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.574</td>
<td>0.317</td>
<td>0.555</td>
<td>0.240</td>
<td>0.435</td>
<td>0.566</td>
<td>0.555</td>
<td>0.284</td>
</tr>
<tr>
<td>2</td>
<td>0.257</td>
<td>0.757</td>
<td>0.345</td>
<td>0.459</td>
<td>0.270</td>
<td>0.752</td>
<td>0.313</td>
<td>0.541</td>
</tr>
<tr>
<td>3</td>
<td>0.197</td>
<td>0.844</td>
<td>0.261</td>
<td>0.592</td>
<td>0.186</td>
<td>0.839</td>
<td>0.224</td>
<td>0.683</td>
</tr>
<tr>
<td>4</td>
<td>0.128</td>
<td>0.912</td>
<td>0.219</td>
<td>0.677</td>
<td>0.169</td>
<td>0.870</td>
<td>0.170</td>
<td>0.784</td>
</tr>
<tr>
<td>5</td>
<td>0.106</td>
<td>0.932</td>
<td>0.204</td>
<td>0.712</td>
<td>0.167</td>
<td>0.879</td>
<td>0.133</td>
<td>0.849</td>
</tr>
<tr>
<td>6</td>
<td>0.097</td>
<td>0.942</td>
<td>0.194</td>
<td>0.739</td>
<td>0.167</td>
<td>0.884</td>
<td>0.119</td>
<td>0.875</td>
</tr>
</tbody>
</table>

Subject 4

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.598</td>
<td>0.239</td>
<td>0.515</td>
<td>0.358</td>
<td>0.581</td>
<td>0.256</td>
<td>0.536</td>
<td>0.310</td>
</tr>
<tr>
<td>2</td>
<td>0.323</td>
<td>0.579</td>
<td>0.274</td>
<td>0.669</td>
<td>0.306</td>
<td>0.578</td>
<td>0.284</td>
<td>0.650</td>
</tr>
<tr>
<td>3</td>
<td>0.262</td>
<td>0.675</td>
<td>0.203</td>
<td>0.769</td>
<td>0.221</td>
<td>0.727</td>
<td>0.202</td>
<td>0.772</td>
</tr>
<tr>
<td>4</td>
<td>0.148</td>
<td>0.848</td>
<td>0.167</td>
<td>0.827</td>
<td>0.164</td>
<td>0.817</td>
<td>0.147</td>
<td>0.850</td>
</tr>
<tr>
<td>5</td>
<td>0.117</td>
<td>0.891</td>
<td>0.143</td>
<td>0.874</td>
<td>0.130</td>
<td>0.870</td>
<td>0.130</td>
<td>0.872</td>
</tr>
<tr>
<td>6</td>
<td>0.093</td>
<td>0.925</td>
<td>0.139</td>
<td>0.888</td>
<td>0.120</td>
<td>0.883</td>
<td>0.113</td>
<td>0.898</td>
</tr>
</tbody>
</table>

Subject 5

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.589</td>
<td>0.231</td>
<td>0.580</td>
<td>0.198</td>
<td>0.572</td>
<td>0.221</td>
<td>0.585</td>
<td>0.189</td>
</tr>
<tr>
<td>2</td>
<td>0.342</td>
<td>0.536</td>
<td>0.341</td>
<td>0.506</td>
<td>0.326</td>
<td>0.524</td>
<td>0.348</td>
<td>0.440</td>
</tr>
<tr>
<td>3</td>
<td>0.281</td>
<td>0.625</td>
<td>0.226</td>
<td>0.687</td>
<td>0.250</td>
<td>0.631</td>
<td>0.236</td>
<td>0.633</td>
</tr>
<tr>
<td>4</td>
<td>0.169</td>
<td>0.812</td>
<td>0.188</td>
<td>0.757</td>
<td>0.195</td>
<td>0.725</td>
<td>0.186</td>
<td>0.729</td>
</tr>
<tr>
<td>5</td>
<td>0.118</td>
<td>0.890</td>
<td>0.173</td>
<td>0.790</td>
<td>0.167</td>
<td>0.777</td>
<td>0.155</td>
<td>0.792</td>
</tr>
<tr>
<td>6</td>
<td>0.095</td>
<td>0.923</td>
<td>0.168</td>
<td>0.800</td>
<td>0.153</td>
<td>0.805</td>
<td>0.133</td>
<td>0.837</td>
</tr>
</tbody>
</table>
Table D.1 (continued)

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Non-speech</th>
<th></th>
<th>Speech</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowels</td>
<td>Fricatives</td>
<td>Vowels</td>
<td>Fricatives</td>
</tr>
<tr>
<td></td>
<td>Stress</td>
<td>R²</td>
<td>Stress</td>
<td>R²</td>
</tr>
<tr>
<td>Subject 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.587</td>
<td>0.286</td>
<td>0.548</td>
<td>0.285</td>
</tr>
<tr>
<td>2</td>
<td>0.292</td>
<td>0.694</td>
<td>0.285</td>
<td>0.616</td>
</tr>
<tr>
<td>3</td>
<td>0.242</td>
<td>0.772</td>
<td>0.204</td>
<td>0.747</td>
</tr>
<tr>
<td>4</td>
<td>0.143</td>
<td>0.885</td>
<td>0.165</td>
<td>0.811</td>
</tr>
<tr>
<td>5</td>
<td>0.142</td>
<td>0.883</td>
<td>0.145</td>
<td>0.850</td>
</tr>
<tr>
<td>6</td>
<td>0.094</td>
<td>0.941</td>
<td>0.140</td>
<td>0.862</td>
</tr>
<tr>
<td>Subject 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.576</td>
<td>0.245</td>
<td>0.577</td>
<td>0.240</td>
</tr>
<tr>
<td>2</td>
<td>0.311</td>
<td>0.560</td>
<td>0.326</td>
<td>0.520</td>
</tr>
<tr>
<td>3</td>
<td>0.198</td>
<td>0.745</td>
<td>0.221</td>
<td>0.696</td>
</tr>
<tr>
<td>4</td>
<td>0.162</td>
<td>0.802</td>
<td>0.168</td>
<td>0.784</td>
</tr>
<tr>
<td>5</td>
<td>0.141</td>
<td>0.839</td>
<td>0.143</td>
<td>0.829</td>
</tr>
<tr>
<td>6</td>
<td>0.132</td>
<td>0.857</td>
<td>0.131</td>
<td>0.852</td>
</tr>
<tr>
<td>Subject 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.608</td>
<td>0.231</td>
<td>0.541</td>
<td>0.341</td>
</tr>
<tr>
<td>2</td>
<td>0.314</td>
<td>0.583</td>
<td>0.298</td>
<td>0.615</td>
</tr>
<tr>
<td>3</td>
<td>0.251</td>
<td>0.680</td>
<td>0.229</td>
<td>0.708</td>
</tr>
<tr>
<td>4</td>
<td>0.161</td>
<td>0.826</td>
<td>0.190</td>
<td>0.774</td>
</tr>
<tr>
<td>5</td>
<td>0.128</td>
<td>0.873</td>
<td>0.180</td>
<td>0.791</td>
</tr>
<tr>
<td>6</td>
<td>0.100</td>
<td>0.913</td>
<td>0.177</td>
<td>0.799</td>
</tr>
<tr>
<td>Subject 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.587</td>
<td>0.255</td>
<td>0.509</td>
<td>0.349</td>
</tr>
<tr>
<td>2</td>
<td>0.329</td>
<td>0.599</td>
<td>0.286</td>
<td>0.630</td>
</tr>
<tr>
<td>3</td>
<td>0.273</td>
<td>0.686</td>
<td>0.232</td>
<td>0.704</td>
</tr>
<tr>
<td>4</td>
<td>0.140</td>
<td>0.872</td>
<td>0.216</td>
<td>0.734</td>
</tr>
<tr>
<td>5</td>
<td>0.104</td>
<td>0.919</td>
<td>0.205</td>
<td>0.760</td>
</tr>
<tr>
<td>6</td>
<td>0.087</td>
<td>0.938</td>
<td>0.205</td>
<td>0.763</td>
</tr>
</tbody>
</table>

(Table continues)

250
<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Non-speech</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowels</td>
<td>Fricatives</td>
</tr>
</tbody>
</table>
|            | Stress | $R^2$ | Stress | $R^2$ | Stress | $R^2$ | Stress | $R^2$
| Subject 10 | 1 | 0.575 | 0.235 | 0.559 | 0.290 | 0.550 | 0.318 | 0.543 | 0.304 |
|           | 2 | 0.300 | 0.598 | 0.296 | 0.604 | 0.281 | 0.647 | 0.314 | 0.573 |
|           | 3 | 0.196 | 0.767 | 0.216 | 0.728 | 0.179 | 0.811 | 0.226 | 0.706 |
|           | 4 | 0.155 | 0.829 | 0.183 | 0.792 | 0.128 | 0.886 | 0.188 | 0.773 |
|           | 5 | 0.129 | 0.870 | 0.181 | 0.798 | 0.104 | 0.917 | 0.173 | 0.802 |
|           | 6 | 0.115 | 0.893 | 0.180 | 0.803 | 0.096 | 0.928 | 0.167 | 0.815 |
| Subject 11 | 1 | 0.556 | 0.304 | 0.575 | 0.272 | 0.580 | 0.296 | 0.552 | 0.367 |
|           | 2 | 0.324 | 0.633 | 0.341 | 0.546 | 0.307 | 0.605 | 0.285 | 0.634 |
|           | 3 | 0.269 | 0.719 | 0.219 | 0.718 | 0.203 | 0.748 | 0.190 | 0.778 |
|           | 4 | 0.140 | 0.855 | 0.176 | 0.788 | 0.166 | 0.809 | 0.154 | 0.836 |
|           | 5 | 0.109 | 0.918 | 0.157 | 0.821 | 0.125 | 0.872 | 0.138 | 0.864 |
|           | 6 | 0.092 | 0.937 | 0.150 | 0.837 | 0.105 | 0.904 | 0.132 | 0.874 |
| Subject 12 | 1 | 0.575 | 0.328 | 0.534 | 0.295 | 0.581 | 0.288 | 0.582 | 0.178 |
|           | 2 | 0.277 | 0.700 | 0.319 | 0.559 | 0.310 | 0.583 | 0.362 | 0.413 |
|           | 3 | 0.220 | 0.779 | 0.267 | 0.654 | 0.233 | 0.693 | 0.256 | 0.589 |
|           | 4 | 0.122 | 0.902 | 0.252 | 0.689 | 0.173 | 0.789 | 0.207 | 0.687 |
|           | 5 | 0.081 | 0.950 | 0.249 | 0.707 | 0.145 | 0.837 | 0.177 | 0.751 |
|           | 6 | 0.071 | 0.962 | 0.246 | 0.722 | 0.130 | 0.863 | 0.165 | 0.777 |
| Subject 13 | 1 | 0.614 | 0.211 | 0.555 | 0.289 | 0.560 | 0.237 | 0.534 | 0.306 |
|           | 2 | 0.345 | 0.539 | 0.285 | 0.636 | 0.297 | 0.603 | 0.290 | 0.619 |
|           | 3 | 0.313 | 0.594 | 0.214 | 0.732 | 0.202 | 0.760 | 0.203 | 0.750 |
|           | 4 | 0.172 | 0.821 | 0.182 | 0.782 | 0.143 | 0.848 | 0.164 | 0.813 |
|           | 5 | 0.117 | 0.901 | 0.166 | 0.812 | 0.111 | 0.899 | 0.137 | 0.861 |
|           | 6 | 0.088 | 0.940 | 0.162 | 0.822 | 0.098 | 0.919 | 0.131 | 0.874 |

(Table continues)
Table D.1 (continued)

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Non-speech Vowels</th>
<th>Non-speech Fricatives</th>
<th>Speech Vowels</th>
<th>Speech Fricatives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stress R²</td>
<td>Stress R²</td>
<td>Stress R²</td>
<td>Stress R²</td>
</tr>
<tr>
<td>1</td>
<td>0.609 0.221</td>
<td>0.487 0.402</td>
<td>0.606 0.217</td>
<td>0.560 0.352</td>
</tr>
<tr>
<td>2</td>
<td>0.353 0.550</td>
<td>0.305 0.605</td>
<td>0.349 0.496</td>
<td>0.290 0.627</td>
</tr>
<tr>
<td>3</td>
<td>0.309 0.622</td>
<td>0.223 0.720</td>
<td>0.235 0.679</td>
<td>0.213 0.727</td>
</tr>
<tr>
<td>4</td>
<td>0.190 0.794</td>
<td>0.175 0.802</td>
<td>0.188 0.756</td>
<td>0.165 0.810</td>
</tr>
<tr>
<td>5</td>
<td>0.124 0.887</td>
<td>0.159 0.833</td>
<td>0.138 0.846</td>
<td>0.139 0.851</td>
</tr>
<tr>
<td>6</td>
<td>0.106 0.910</td>
<td>0.156 0.840</td>
<td>0.114 0.884</td>
<td>0.128 0.869</td>
</tr>
<tr>
<td>Subject 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.571 0.300</td>
<td>0.535 0.388</td>
<td>0.552 0.337</td>
<td>0.553 0.275</td>
</tr>
<tr>
<td>2</td>
<td>0.298 0.664</td>
<td>0.318 0.563</td>
<td>0.324 0.585</td>
<td>0.314 0.539</td>
</tr>
<tr>
<td>3</td>
<td>0.248 0.736</td>
<td>0.231 0.693</td>
<td>0.207 0.751</td>
<td>0.206 0.724</td>
</tr>
<tr>
<td>4</td>
<td>0.149 0.863</td>
<td>0.202 0.746</td>
<td>0.156 0.833</td>
<td>0.163 0.794</td>
</tr>
<tr>
<td>5</td>
<td>0.123 0.895</td>
<td>0.188 0.776</td>
<td>0.141 0.858</td>
<td>0.144 0.828</td>
</tr>
<tr>
<td>6</td>
<td>0.116 0.905</td>
<td>0.186 0.783</td>
<td>0.131 0.875</td>
<td>0.134 0.848</td>
</tr>
<tr>
<td>Subject 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.626 0.167</td>
<td>0.466 0.490</td>
<td>0.589 0.218</td>
<td>0.579 0.244</td>
</tr>
<tr>
<td>2</td>
<td>0.396 0.403</td>
<td>0.278 0.667</td>
<td>0.332 0.515</td>
<td>0.300 0.595</td>
</tr>
<tr>
<td>3</td>
<td>0.278 0.609</td>
<td>0.191 0.801</td>
<td>0.209 0.719</td>
<td>0.221 0.707</td>
</tr>
<tr>
<td>4</td>
<td>0.203 0.751</td>
<td>0.170 0.843</td>
<td>0.168 0.789</td>
<td>0.163 0.803</td>
</tr>
<tr>
<td>5</td>
<td>0.169 0.808</td>
<td>0.162 0.864</td>
<td>0.144 0.831</td>
<td>0.129 0.860</td>
</tr>
<tr>
<td>6</td>
<td>0.135 0.867</td>
<td>0.163 0.865</td>
<td>0.129 0.861</td>
<td>0.118 0.878</td>
</tr>
<tr>
<td>Subject 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.607 0.204</td>
<td>0.567 0.247</td>
<td>0.566 0.273</td>
<td>0.590 0.206</td>
</tr>
<tr>
<td>2</td>
<td>0.324 0.533</td>
<td>0.307 0.551</td>
<td>0.344 0.508</td>
<td>0.366 0.445</td>
</tr>
<tr>
<td>3</td>
<td>0.225 0.688</td>
<td>0.227 0.676</td>
<td>0.200 0.752</td>
<td>0.237 0.660</td>
</tr>
<tr>
<td>4</td>
<td>0.172 0.783</td>
<td>0.184 0.755</td>
<td>0.148 0.838</td>
<td>0.177 0.770</td>
</tr>
<tr>
<td>5</td>
<td>0.150 0.823</td>
<td>0.160 0.806</td>
<td>0.115 0.891</td>
<td>0.146 0.824</td>
</tr>
<tr>
<td>6</td>
<td>0.137 0.847</td>
<td>0.153 0.821</td>
<td>0.103 0.908</td>
<td>0.116 0.879</td>
</tr>
</tbody>
</table>

Table D.1: Individual stress and $R^2$ values
Figure D.1: Subject 1’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by third dimension frequency offglide. (c) First dimension frequency by fourth dimension loudness.
Figure D.2: Subject 1’s narrowband fricatives MDS solution. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by third dimension duration.

Figure D.3: Subject 1’s synthetic vowels MDS solution. First dimension frequency by second dimension duration.
Figure D.4: Subject 1’s synthetic fricatives MDS solution. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension loudness.
Figure D.5: Subject 2’s sine wave vowels MDS solution. (a) First dimension frequency offglide by second dimension frequency. (b) First dimension frequency offglide by third dimension total frequency change. (c) Second dimension frequency by fourth dimension duration. (d) First dimension frequency offglide by fifth dimension loudness.
Figure D.6: Subject 2’s narrowband fricatives MDS solution. First dimension duration by second dimension frequency.

Figure D.7: Subject 2’s synthetic vowels MDS solution. First dimension frequency by second dimension total frequency change.
Figure D.8: Subject 2’s synthetic fricatives MDS solution. (a) First dimension duration by second dimension frequency. (b) First dimension duration by third dimension loudness.

Figure D.9: Subject 3’s sine wave vowels MDS solution. First dimension frequency by second dimension frequency offglide.
Figure D.10: Subject 3’s narrowband fricatives MDS solution. First dimension frequency by second dimension that might be duration.

Figure D.11: Subject 3’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by third dimension duration.
Figure D.12: Subject 3’s synthetic fricatives MDS solution. First dimension frequency by second dimension loudness.
Figure D.13: Subject 4’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension total frequency change. (c) First dimension frequency by fourth dimension duration.
Figure D.14: Subject 4’s narrowband fricatives MDS solution. First dimension *duration* by second dimension *frequency*.

Figure D.15: Subject 4’s synthetic vowels MDS solution. First dimension *frequency* by second dimension *duration*.
Figure D.16: Subject 4’s synthetic fricatives MDS solution. First dimension \textit{duration} by second dimension \textit{frequency}.
Figure D.17: Subject 5’s sine wave vowels MDS solution. (a) First dimension frequency offglide by second dimension frequency. (b) First dimension frequency offglide by third dimension total frequency change. (c) Second dimension frequency by fourth dimension duration.
Figure D.18: Subject 5’s narrowband fricatives MDS solution. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension frequency offglide.

Figure D.19: Subject 5’s synthetic vowels MDS solution. First dimension frequency by second dimension duration.
Figure D.20: Subject 5’s synthetic fricatives MDS solution. (a) First dimension *frequency* by second dimension *duration*. (b) First dimension *frequency* by third dimension *frequency offglide*. 
Figure D.21: Subject 6’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension total frequency change. (c) First dimension frequency by fourth dimension loudness.
Figure D.22: Subject 6’s narrowband fricatives MDS solution. First dimension frequency by second dimension duration.

Figure D.23: Subject 6’s synthetic vowels MDS solution. (a) First dimension duration by second dimension frequency. (b) First dimension duration by third dimension frequency offglide.
Figure D.24: Subject 6’s synthetic fricatives MDS solution. (a) First dimension *duration* by second dimension *frequency*.

Figure D.25: Subject 7’s sine wave vowels MDS solution. (a) First dimension *frequency* by second dimension *duration*. (b) First dimension *frequency* by third dimension *frequency offglide*.
Figure D.26: Subject 7’s narrowband fricatives MDS solution. (a) First dimension frequency offglide by second dimension frequency. (b) First dimension frequency offglide by third dimension total frequency change.
Figure D.27: Subject 7’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by third dimension duration. (c) First dimension frequency by fourth dimension loudness.
Figure D.28: Subject 7’s synthetic fricatives MDS solution. First dimension \textit{duration} by second dimension \textit{frequency offglide}.
Figure D.29: Subject 8’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension total frequency change. (c) First dimension frequency by fourth dimension duration.
Figure D.30: Subject 8’s narrowband fricatives MDS solution. First dimension frequency by second dimension duration.

Figure D.31: Subject 8’s synthetic vowels MDS solution. First dimension frequency by second dimension duration.
Figure D.32: Subject 8’s synthetic fricatives MDS solution. First dimension frequency by second dimension duration.
Figure D.33: Subject 9’s sine wave vowels MDS solution. (a) First dimension *frequency offglide* by second dimension frequency. (b) First dimension *frequency offglide* by third dimension *total frequency change*. (c) First dimension *frequency offglide* by fourth dimension *loudness*.  

276
Figure D.34: Subject 9’s narrowband fricatives MDS solution. First dimension duration by second dimension frequency.
Figure D.35: Subject 9’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension duration. (c) First dimension frequency by fourth dimension total frequency change.
Figure D.36: Subject 9’s synthetic fricatives MDS solution. (a) First dimension \textit{frequency} by second dimension \textit{duration}.

Figure D.37: Subject 10’s sine wave vowels MDS solution. (a) First dimension \textit{frequency} by second dimension \textit{duration}. (b) First dimension \textit{frequency} by third dimension \textit{total frequency change}.
Figure D.38: Subject 10’s narrowband fricatives MDS solution. First dimension *duration* by second dimension *frequency.*
Figure D.39: Subject 10’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension total frequency change. (c) First dimension frequency by fourth dimension loudness.
Figure D.40: Subject 10’s synthetic fricatives MDS solution. (a) First dimension duration by second dimension frequency. (b) First dimension duration by third dimension total frequency change.
Figure D.41: Subject 11’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension that might be frequency offglide. (b) First dimension frequency by third dimension total frequency change. (c) First dimension frequency by fourth dimension that might be loudness.
Figure D.42: Subject 11’s narrowband fricatives MDS solution. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension total frequency change.

Figure D.43: Subject 11’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension duration.
Figure D.44: Subject 11’s synthetic fricatives MDS solution. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension total frequency change.
Figure D.45: Subject 12’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension total frequency change. (c) First dimension frequency by fourth dimension loudness. (d) First dimension frequency by fifth dimension duration.
Figure D.46: Subject 12’s narrowband fricatives MDS solution. First dimension duration by second dimension frequency.

Figure D.47: Subject 12’s synthetic vowels MDS solution. First dimension frequency by second dimension duration.
Figure D.48: Subject 12’s synthetic fricatives MDS solution. First dimension *duration* by second dimension that might be *frequency*. 
Figure D.49: Subject 13’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension total frequency change. (c) First dimension frequency by fourth dimension loudness.
Figure D.50: Subject 13's narrowband fricatives MDS solution. First dimension *frequency* by second dimension *duration*. 
Figure D.51: Subject 13’s synthetic vowels MDS solution. (a) First dimension *duration* by second dimension *frequency*. (b) First dimension *duration* by third dimension *frequency offglide*. (c) First dimension *duration* by fourth dimension *total frequency change*. (d) First dimension *duration* by fourth dimension *loudness*. 
Figure D.52: Subject 13’s synthetic fricatives MDS solution. (a) First dimension \textit{duration} by second dimension \textit{frequency}.
Figure D.53: Subject 14’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension total frequency change. (c) First dimension frequency by fourth dimension loudness. (d) First dimension frequency by fifth dimension duration.
Figure D.54: Subject 14’s narrowband fricatives MDS solution. (a) First dimension frequency by second dimension total frequency change. (a) First dimension frequency by third dimension duration. (a) First dimension frequency by fourth dimension frequency offglide.
Figure D.55: Subject 14’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by third dimension frequency offglide.
Figure D.56: Subject 14’s synthetic fricatives MDS solution. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by third dimension duration. (c) First dimension frequency by fourth dimension loudness.
Figure D.57: Subject 15's sine wave vowels MDS solution. First dimension frequency by second dimension frequency offglide.

Figure D.58: Subject 15’s narrowband fricatives MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension total frequency change.
Figure D.59: Subject 15’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension total frequency change. (b) First dimension frequency by third dimension duration.

Figure D.60: Subject 15’s synthetic fricatives MDS solution. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension that might be total frequency change.
Figure D.61: Subject 16’s sine wave vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension total frequency change.

Figure D.62: Subject 16’s narrowband fricatives MDS solution. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension total frequency change.
Figure D.63: Subject 16’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension duration. (b) First dimension frequency by third dimension frequency offglide.

Figure D.64: Subject 16’s synthetic fricatives MDS solution. First dimension frequency by second dimension duration.
Figure D.65: Subject 17’s sine wave vowels MDS solution. (a) First dimension \textit{duration} by second dimension \textit{frequency}. (b) First dimension \textit{duration} by third dimension \textit{frequency offglide}.

Figure D.66: Subject 17’s narrowband fricatives MDS solution. First dimension \textit{frequency} by second dimension \textit{duration}.
Figure D.67: Subject 17’s synthetic vowels MDS solution. (a) First dimension frequency by second dimension frequency offglide. (b) First dimension frequency by third dimension duration.

Figure D.68: Subject 17’s synthetic fricatives MDS solution. (a) First dimension duration by second dimension frequency. (b) First dimension duration by third dimension that might be total frequency change.
BIBLIOGRAPHY


311


Makashay, M. J. (1999). *Dynamic invariance in auditory neural map formation.* (Unpublished manuscript, Ohio State University, Columbus)

Makashay, M. J. (2001). *Dialectal differences in diphthong perception.* (Unpublished manuscript, Ohio State University, Columbus)


