Sensorimotor Adaptation in Speech Production

by

John Francis Houde

B.S., Electrical Engineering
California Institute of Technology, 1985
M.S., Computer Science
Duke University, 1990

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1997

Author ....John F. Houde..........
Department of Brain and Cognitive Sciences
February 5, 1997

Certified by ....Michael I. Jordan..........
Michael I. Jordan
Professor
Thesis Supervisor

Accepted by ....Gerald E. Schneider..........
Gerald E. Schneider
Professor
Chairman, Department Graduate Committee

Sensorimotor Adaptation in Speech Production

by

John Francis Houde

Submitted to the Department of Brain and Cognitive Sciences
on February 5, 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis investigates *sensorimotor adaptation* (SA) in speech production: how speakers alter their speech production to compensate for distortions of their normal auditory feedback. Two studies were conducted that exhibit the existence and properties of speech SA and demonstrate its potential for examining phonetic structure in speech production.

In both studies, auditory feedback was distorted by an apparatus that shifts speech formant frequencies with minimal (16ms) processing delay. Via a microphone, subjects whispered into this apparatus, and it produced a formant-shifted version of their whispered speech that was fed back to them via earphones. Subjects whispered to minimize bone conduction of their actual speech feedback.

Study 1 exhibited the basic speech SA phenomenon. It found that subjects adjusted their productions of [ɛ] in CVC utterances to compensate for altered feedback. They subsequently retained these adjustments when whispering while masking noise blocked their feedback. These retained adjustments are called adaptation.

Study 2 revealed other properties of speech SA. It found that compensating production changes were apparently retained for more than a month. It also found that, although subjects differed greatly in how much they compensated, none reported noticing that their feedback was altered. This suggests that subjects' vowel perceptions may have adapted.

Study 2 also investigated how adaptation of [ɛ] in one CVC word context affected other words' productions. Subjects' production of [ɛ] in other words was affected, showing that these words share a common representation of the production of [ɛ]. Subjects' productions of other vowel were also affected, showing that the vowels' production representations are not independent, possibly because they share common features. These investigations showed how speech SA can be used to examine phonetic structure in speech production.

Thesis Supervisor: Michael I. Jordan
Title: Professor

To mother and father.

Acknowledgments

First and foremost, I would like to thank Professor Michael Jordan, my advisor. I feel privileged to have been part of his lab, and to have had his guidance in my studies. I'd also like to thank Prof. Jordan for a number of reasons specifically related to this thesis work. When I first proposed doing this work, it was unknown whether speech would exhibit sensorimotor adaptation (SA). Prof. Jordan had the initial faith that the development of the experimental apparatus would lead to significant results. In addition, he has been a constant source of knowledge and ideas concerning how speech SA relates to key issues in speech and motor control. Finally, I would like to thank Prof. Jordan for his support and guidance in my preparation of the thesis draft.

I would also like to thank my other thesis committee members. To conduct this research, I needed guidance in the fields of speech motor control and physiology, speech acoustics, and sensorimotor adaptation. MIT just happens to have the foremost authorities in these fields: Joseph Perkell, Kenneth Stevens, and Richard Held, respectively. Thus, in this sense alone, these are ideal committee members. However, I also feel they are ideal because it had been such a pleasure interacting with them. Dr. Perkell, Prof. Stevens, and Prof. Held have always had an open door and a friendly manner that encouraged me to seek them out whenever I had questions. In particular, Dr. Perkell's interest in my work and his encouragement of it has been extremely important to me.

There are many other people I would also like to thank for having helped make this thesis possible. Dr. Robert Houde – my father – provided key guidance on the spectrum analysis and speech synthesis approaches I used. Yaoda Xu was my first lab assistant and was a significant help in getting the initial speech SA experiments running. Greta Buck did excellent technical editing of my thesis drafts.

I also owe special thanks to Zoubin Ghahramani, Philip Sabes, and Daniel Wolpert, who have been great friends and colleagues in motor psychophysics in Jordan Lab. In particular, Daniel and Zoubin's investigations of reaching sensorimotor adaptation were what initially inspired me to consider the value of conducting similar adaptation

experiments in speech.

Finally, I'd like to thank Janice Ellertsen for her ability to guide me through the MIT administration and keep my academic ship afloat through numerous changes in funding sources. This allowed me to concentrate on the research that has led to this thesis.

Contents

1 Introduction **25**

1.1 Sensorimotor Adaptation in Reaching 25

1.2 Sensorimotor Adaptation in Speech 27

1.3 Thesis Overview 30

2 Design of the Speech SA Experiment **33**

2.1 Defining the Task and Feedback Transformation 33

2.1.1 The Similarity Between Vowel Production and Reaching . . . 33

2.1.2 Perceivability and Compensatability Requirements 36

2.2 Design of the SA experiment 40

2.2.1 Design of a Reaching SA Experiment 40

2.2.2 Prompting the Subject to Perform a Specific Task 43

2.2.2.1 Task Prompting in Reaching SA Experiments 43

2.2.2.2 Task Prompting in Speech SA Experiments 43

2.2.3 Exposing the Subject to the Feedback Transform 44

2.2.4 Intercepting the Subject's Feedback 45

2.2.5 Recording Task Performance 46

2.3 Summary of the Speech SA Experiment Design 47

3 Apparatus and Methods **49**

3.1 Overview 49

3.2 Transforming the Acoustic Feedback 52

3.3 Constructing the Feedback Transformation 56

3.3.1 Defining the Feedback Transformation 57

3.3.1.1 Defining [i]–[ɑ] Path Position 57

3.3.1.2 Path Projection and Deviation 62

3.3.1.3 The Complete Transform Definition 64

3.3.2 General Construction Procedure 68

3.3.3 Transformations Used 69

3.4 Utterance Data Acquisition 71

3.4.1 Initial Setup 71

3.4.2 Data Acquisition 73

3.5 Utterance Data Analysis 75

3.5.1 Formant Timecourse Analysis: Avgrams 76

3.5.1.1 Avgram Creation 76

3.5.1.2 Time Alignment 78

3.5.1.3 Other Avgram Limitations 79

3.5.2 Vowel Formant Analysis: Vowel Plots 79

3.5.2.1 Creating a Vowel Plot 80

3.5.2.2 Combined Vowel Plots 81

3.5.3 Path Projection and Path Deviation Analysis 84

3.5.3.1 Vowel (Path Projection, Deviation) Data 84

3.5.3.2 ANOVAs 85

3.5.3.3 Mean Path Projection and Path Deviation Change . 85

3.5.3.4 Mean Compensation 86

3.6 Chapter Summary 87

4 Study 1: Existence of Speech Sensorimotor Adaptation 91

4.1 Introduction 91

4.2 Methods 93

4.3 Subjects 95

4.4 Results and Discussion 96

4.4.1 Individual Subject Results 97

4.4.1.1 Subject MB Results 98
4.4.1.2 Subject BM Results 102
4.4.1.3 Subject MF Results 106
4.4.1.4 Subject JK Results 110
4.4.1.5 Subject JG Results 113
4.4.1.6 Subject BK, MK, JD, And ST Results 117
4.4.2 Across-Subject Analysis 123
4.4.2.1 Across-Subject Plots 123
4.4.2.2 Statistical Tests 128
4.5 Summary and Conclusions 131
4.5.1 Do Subjects Adapt? 132
4.5.2 How Much do Subjects Adapt? 133
4.5.3 Methodological Issues 133

5 Study 2: Timecourse and Generalization of Speech Sensorimotor Adaptation 135
5.1 Introduction 135
5.1.1 Confirming the Existence of Speech SA 136
5.1.2 Examining the Timecourse of Speech SA 136
5.1.3 Generalization of Speech SA 137
5.1.3.1 Context Generalization 137
5.1.3.2 Target Generalization 138
5.2 Methods 139
5.2.1 Prompted Words 139
5.2.1.1 Training Words 139
5.2.1.2 Testing Words 140
5.2.2 Timecourse of the Experiment 142
5.2.2.1 Timecourse of Each Epoch 142
5.2.2.2 Epoch Sequence 143
5.2.3 Post-Experiment Interview 145

5.2.4 Subjects 146

5.3 Results and Discussion 147

5.3.1 Compensation and Adaptation Results 148

5.3.1.1 Mean Compensation, Adaptation, and Path Deviation Change 148

5.3.1.2 Path Projection Analysis 156

5.3.1.3 Subjects' Awareness of the Altered Feedback 166

5.3.1.4 Discussion 167

5.3.2 Timecourse Results 172

5.3.2.1 Timecourse Plots 172

5.3.2.2 Train and Test Phase Analysis 174

5.3.2.3 Ramp Phase Analysis 175

5.3.2.4 Discussion 179

5.3.3 Generalization Results 181

5.3.3.1 The Testing Word Set 181

5.3.3.2 Assessing Generalization 182

5.3.3.3 Overall Generalization 191

5.3.3.4 Context Generalization 193

5.3.3.5 Target Generalization 196

5.3.3.6 Discussion 198

5.3.4 Summary 199

5.4 Individual Subject Results 202

5.4.1 Subject CW 202

5.4.1.1 Compensation and Adaptation Results 202

5.4.1.2 Timecourse Results 207

5.4.1.3 Generalization Results 211

5.4.2 Subject RS 222

5.4.3 Subject OB 229

5.4.4 Subject SR 236

5.4.5 Subject RO 244

5.4.6 Subject TY 251

5.4.7 Subject VS 258

5.4.8 Subject AH 261

6 Discussion **265**

6.1 Existence of Speech SA 265

6.1.1 Perception of the Altered Feedback 266

6.1.2 Compensation 267

6.1.3 Adaptation 268

6.1.4 Summary 269

6.2 Using Speech SA to Address Phonetic Structure Questions 270

6.2.1 Context Generalization 271

6.2.1.1 Word Frequency Differences 271

6.2.1.2 Syllables Versus Phonemes 271

6.2.2 Target Generalization 272

6.3 Conclusions 272

Appendices **275**

A Formant Estimation, Analysis, and Subject Pretesting **275**

A.1 Formant Estimation 276

A.1.1 What are Formants? 276

A.1.2 Formant Estimation Problems 277

A.1.3 The Chosen Estimation Procedure 281

A.1.4 Discussion 286

A.1.4.1 An Articulatory Explanation 286

A.1.4.2 Implications for Vowel Perception Theories 291

A.2 Formant Data Analysis 292

A.2.1 Word Production Data Records 293

A.2.2 Calculating Mean Formants of Formant Data 293

A.2.3 Analysis of the Method 297

A.3 Subject Pretesting 299

A.3.1 Measuring the F1 Range 300

A.3.1.1 Vowel Data Acquisition 301

A.3.1.2 Vowel Data Analysis 301

A.3.2 Measuring Path Vowel Formants 304

A.3.3 Assessing Feedback Transformation Fidelity 306

B Signal Processing **309**

B.1 Acquisition 310

B.1.1 Sampling Rate 310

B.1.2 Frame Size 310

B.2 Spectral analysis 312

B.2.1 Creating the Magnitude Spectrum 313

B.2.2 Processing the Magnitude Spectrum 313

B.3 Formant Estimation 317

B.4 Formant Alteration 319

B.5 Synthesis 319

B.5.1 Creating the Vocal Tract Impulse Response 319

B.5.2 Generating a Glottal Source Function 320

B.5.3 Simulating the Response of the Vocal Tract: Generating Whispered Speech 320

C Resolution and Stability of Path Projections **325**

C.1 Cubic-Spline [i]–[ɑ] Path Definitions 325

C.2 Resolution and Stability Issues 327

C.3 Impact on the Feedback Transformations 327

C.4 Impact on Data Analysis 328

D Equipment **331**

D.1 Detailed Apparatus Description 331

D.2 Equipment Specifications 332

List of Figures

1-1 Phases of a hypothetical reaching SA experiment. 26

1-2 Spectrogram of the word "guess". 31

1-3 An (x,y) plot of vowel (F1,F2) frequencies. 32

2-1 The results of the Peterson and Barney study, showing the average F1, F2, and F3 frequencies for a number of vowels, averaged over 33 male speakers. 34

2-2 The vowels of Figure 2-1, where the x and y position of each vowel is determined by its F1 and F2 values, respectively. 35

2-3 Examples of a non-compensatable feedback transformation and the proposed perceivable, compensatable transformation. 37

2-4 The apparatus used by Held et al. for investigating the effect of visual shifts on reaching. 42

3-1 Overview of the experimental apparatus. 49

3-2 Overview of the signal processing that implements the acoustic transformation. 51

3-3 An illustration of the formant estimation procedure. 54

3-4 Definition of a subject's [i]–[ɑ] path. 58

3-5 Calculation of the path position of path point P between [ɛ] and [æ]. 60

3-6 Redescription of Figure 2-3 using the transform definition and the terms defined in this section. 61

3-7 Showing (in two- and three-dimensions) how points in formant space were represented in terms of path projection and deviation. 63

3-8 Action of the +2.0 feedback transformation and compensation for its perceived shift of F. 65

3-9 Action of the -2.0 feedback transformation and compensation for its perceived shift of F. 67

3-10 Effects of the 2p and 2m transformations on vowels produced by subject MB. 70

3-11 Avgrams of subject SR's production of the word "bep" in the baseline and test phases of Study 2. 78

3-12 A vowel plot. 82

3-13 Vowel plots of subject SR's production changes for [ɛ] in "bep" during study 2 (to be discussed later). 83

4-1 Subject MB avgram and vowel plots. 101

4-2 Subject BM avgram and vowel plots. 105

4-3 Subject MF avgram and vowel plots. 109

4-4 Subject JK avgram and vowel plots. 112

4-5 Subject JG avgram and vowel plots. 116

4-6 Subject BK avgram and vowel plots. 119

4-7 Subject MK avgram and vowel plots. 120

4-8 Subject JD avgram and vowel plots. 121

4-9 Subject ST avgram and vowel plots. 122

4-10 Compensation and adaptation responses for each subject in Study 1. 127

5-1 Mean compensation and path deviation change for each subject in Study 2. 152

5-2 Path projections of each subject's vowel productions in Study 2. . . . 164

5-3 Study 2 path projections, averaged across subjects. 165

5-4 Mean compensation and adaptation timecourses for the subjects showing the four largest mean adaptations in Study 2. 173

5-5 Analysis of the adaptation onsets seen in Figure 5-4. 178

5-6 Subject OB testing word avgrams. 188

5-7 Subject OB testing word vowel plots. 189
5-8 Subject OB testing word path projection changes. 190
5-9 Mean generalization for all Study 2 subjects used in the overall generalization test. 193
5-10 Mean generalization for the context generalization words in Study 2. 195
5-11 Mean generalization for the analyzable target generalization words in Study 2. 197
5-12 Subject CW overall compensation and adaptation responses. 217
5-13 Subject CW compensation and adaptation timecourses. 218
5-14 Subject CW testing word avgrams. 219
5-15 Subject CW testing word vowel plots. 220
5-16 Subject CW testing word path projection changes. 221
5-17 Subject RS overall compensation and adaptation responses. 224
5-18 Subject RS compensation and adaptation timecourses. 225
5-19 Subject RS testing word avgrams. 226
5-20 Subject RS testing word vowel plots. 227
5-21 Subject RS testing word path projection changes. 228
5-22 Subject OB overall compensation and adaptation responses. 231
5-23 Subject OB compensation and adaptation timecourses. 232
5-24 Subject OB testing word avgrams. 233
5-25 Subject OB testing word vowel plots. 234
5-26 Subject OB testing word path projection changes. 235
5-27 Subject SR overall compensation and adaptation responses. 239
5-28 Subject SR compensation and adaptation timecourses. 240
5-29 Subject SR testing word avgrams. 241
5-30 Subject SR testing word vowel plots. 242
5-31 Subject SR testing word path projection changes. 243
5-32 Subject RO overall compensation and adaptation responses. 246
5-33 Subject RO compensation and adaptation timecourses. 247
5-34 Subject RO testing word avgrams. 248

5-35 Subject RO testing word vowel plots. 249

5-36 Subject RO testing word path projection changes. 250

5-37 Subject TY overall compensation and adaptation responses. 253

5-38 Subject TY compensation and adaptation timecourses. 254

5-39 Subject TY testing word avgrams. 255

5-40 Subject TY testing word vowel plots. 256

5-41 Subject TY testing word path projection changes. 257

5-42 Subject VS overall compensation and adaptation responses. 259

5-43 Subject VS compensation and adaptation timecourses. 260

5-44 Subject AH overall compensation and adaptation responses. 262

5-45 Subject AH compensation and adaptation timecourses. 263

A-1 Comparison of voiced and whispered vowels. 280

A-2 An illustration of the formant estimation procedure. 283

A-3 Comparison of original and resynthesized whispered vowels. 285

A-4 The time waveform of [ɛ] from which spectra of the whispering styles were calculated. 290

A-5 Spectra of the closed and open-glottis portions of the whispered [ɛ] waveform shown in Figure A-4. 290

A-6 Illustrating how word data records were created. 295

A-7 Steps in the calculation of mean formants of a set of formant data. 296

A-8 Showing the separate histograms for each formant's data from subject BK. 298

A-9 The complete formant histogram made by adding together the individual formant histograms of Figure A-8. 299

A-10 Composite histograms used to measure the F1 range of two subjects. 303

A-11 Composite histograms used to measure path vowel formants of two subjects (the same shown in Figure A-10). 305

B-1 Showing how the incoming speech data was processed into magnitude spectra. 315

B-2 Key steps in processing the magnitude spectrum of frame N from the previous figure. 316

B-3 Showing how formants were estimated from the spectra in Figure B-2. 318

B-4 The impulse response functions created from the formant estimates of Figure B-3. 322

B-5 Showing how the next frame of output synthesized speech was constructed from the convolution of the vocal tract impulse response and one frame of glottal source function. 323

C-1 Comparison of line-segment and cubic-spline [i]–[ɑ] paths made from the same path reference points. 326

C-2 Resolution of path projections. 329

C-3 Stability of path projections. 330

D-1 Diagram of the equipment used in the experimental apparatus. 333

List of Tables

4.1 Study 1 path projection and deviation ANOVA tests. 128

4.2 Study 1 mean compensation and path deviation changes, averaged across subjects. 130

5.1 Study 2 path projection and deviation ANOVA tests. 154

5.2 Study 2 mean compensation and path deviation changes, averaged across subjects. 155

5.3 Dates on which subjects were run in the experiments of Study 2. . . . 161

5.4 Study 2 Ramp phase compensation and adaptation timecourse regression results. 176

5.5 Showing, for each subject in Study 2, which words' vowel formants could not be properly estimated. 192

5.6 ANOVA tests of path projection change for the context generalization words in Study 2. 194

5.7 ANOVA tests of path projection change for the target generalization words in Study 2. 196

A.1 Path vowel formant frequencies for subjects JI and GL. 306

Chapter 1

Introduction

This thesis investigates *sensorimotor adaptation* in speech production. This investigation consists of a series of studies that (1) exhibit existence of the basic phenomenon, (2) investigate some of its properties, and (3) demonstrate its potential for examining phonetic structure in speech production.

1.1 Sensorimotor Adaptation in Reaching

Sensorimotor adaptation (SA) is the modification of motor task performance resulting from exposure to altered sensory feedback. To illustrate the SA phenomena, a hypothetical experiment exhibiting SA in reaching is described. SA in reaching has been shown in numerous experiments (see [Welch, 1978] and [Welch, 1986] for reviews). This hypothetical experiment is chosen for it's close analogy to the speech SA experiments used in this thesis.

Consider an apparatus that allows a subject's view of his hand to be blocked, or to be viewed through a prism that shifts the image of his hand. Using this apparatus, we can exhibit reaching SA with an experiment consisting of three phases: a baseline phase, a training phase, and a testing phase. Figure 1-1 illustrates these phases.

In the baseline phase, the subject's visual feedback is blocked and he is directed to reach "straight ahead". Where he reaches is marked. (Note: the subject cannot see this mark.)

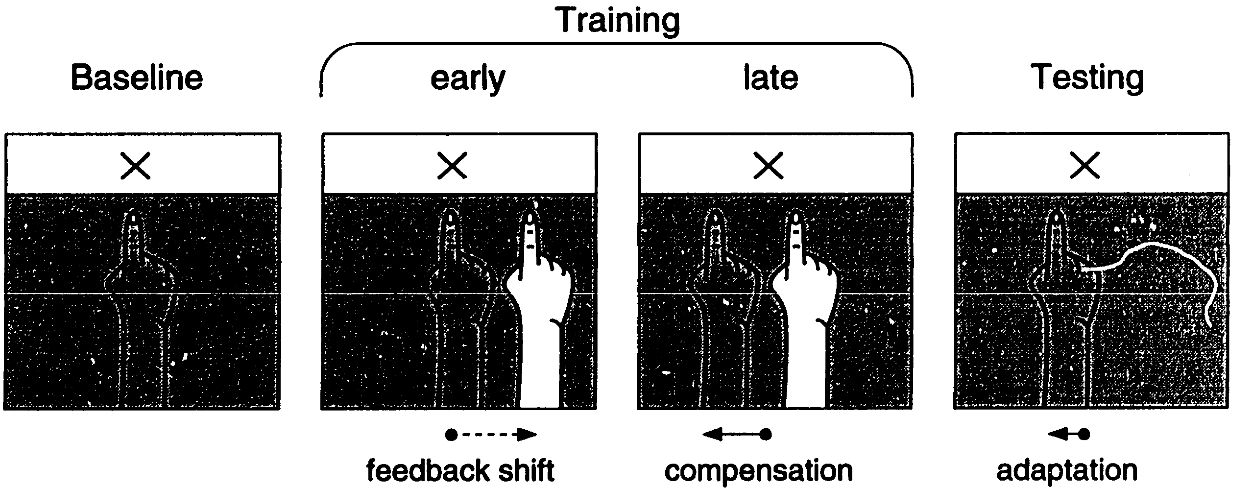


Figure 1-1: Phases of a hypothetical reaching SA experiment. The leftmost panel shows the baseline phase. In this phase, a subject makes a straight-ahead reach without visual feedback (as suggested by the gray hand picture in the gray background). An "x" shown in the top of the panel marks the position he reaches. For reference, this mark is shown in each panel. The middle two panels show the training phase. In the training phase, the subject sees a shifted image of his actual hand position. In both training phase panels, actual hand position is shown as a gray hand picture in the gray background, while the shifted hand position image is shown as the white hand picture. The arrow below the early training phase panel shows the magnitude and direction of the hand image shift. The arrow below the late training phase panel shows the subject's shift of hand position that partially compensates for the hand image shift. The rightmost panel shows the testing phase. This phase is a repeat of the baseline phase. The arrow below this panel shows the amount of compensation gained in the training phase that the subject retains in subsequent reaches without visual feedback – i.e., his *adaptation.*

In the training phase, the subject is directed to continue to make straight-ahead reaches, but now he is permitted to see his hand through a prism that introduces a visual feedback shift. Early in the training phase, when the subject reaches straight ahead, his hand image appears shifted to one side. Depending on the size of this shift, he may or may not be aware of this image shift.

Regardless of his awareness, as he continues to make reaches, he gradually adjusts them so that, by late in the training phase, his "straight ahead" reaches are shifted from their original position. This shift is in the direction opposite to that of the feedback shift, and brings the subject's hand image closer to his baseline straight-ahead position. This shift will be called the subject's *compensation* for the feedback alteration.

In the subsequent testing phase, the subject's visual feedback is blocked and he is once again asked to reach straight-ahead, Now, under the same conditions he experienced in the baseline phase, the subject's straight-ahead reaches are shifted. Thus, even with feedback blocked, the subject has retained some of the compensation he developed in the training phase. This retained compensation will be called *adaptation.*

This adaptation of a motor task performance resulting from altered sensory feedback exposure is called *sensorimotor adaptation.*

1.2 Sensorimotor Adaptation in Speech

This thesis investigated SA in a different motor task – speech production. It was hypothesized that speech production, like reaching, would exhibit adaptation in response to altered sensory feedback – in this case, altered auditory feedback.

Within this potentially broad domain of SA in speech, this thesis focused specifically on SA in vowel production. This is because there exists a convenient representation of vowel sounds which lends itself to sensory feedback alteration – the representation of vowels in terms of their formant frequencies.

If one looks at a spectrogram of an utterance containing a vowel, one sees isolated bands of spectral energy. The bands are called formants, and their frequencies generally correspond to the principal resonances of the vocal tract. (An example spectrogram can be seen in Figure 1-2.) During the vowel portion of this utterance, the formants hold steady-state values whose frequencies are characteristic of the vowel. The frequencies of F3 and higher formants show little variation across vowels; the frequencies of the first two formants (F1 and F2) have the largest role in determining vowel identity. This is shown in Figure 1-3.

Thus by altering formants of a speaker's speech feedback, it should be possible to alter the speaker's auditory perception of what vowel he was producing.

The principal hypothesis examined in this thesis is that the speaker will compensate for such perceivable alterations of his normal auditory feedback. Furthermore, it is hypothesized that some of this compensation will not be achieved solely by an immediate correction response, but will instead involve long-term adjustment of parameters controlling speech. This long-term adjustment will be revealed by compensating production changes persisting in the absence of auditory feedback – i.e., adaptation as defined above.

The motivation for investigating speech SA extends beyond confirming that SA exists in other motor tasks besides reaching. Speech SA would provide a tool for examining issues of phonetic structure in speech production.

We know that producing speech sounds is part of the larger task of communicating words. There are also theories and other lines of evidence suggesting intermediate-sized representations in the production of words (e.g., syllables, phonemes – see [Levelt, 1989] and [Meyer, 1991] for reviews). There are thus a number of testable hypotheses concerning word production that we can design SA experiments to examine. To give a concrete example, consider the following hypothetical situation: we might observe adaptation of the production of a specific vowel in a specific word context (for example [ε] in "get"). We can then design experiments to investigate the following questions:

- Do we find this vowel's production is adapted in other words? If so, it suggests that words do not independently specify their complete productions as indivisible units. Rather, their productions would appear to be constructed from intermediate units of speech production (e.g., phonemes).

- Do we also find that other vowels' productions are adapted? If so, it suggests that, not only do intermediate units of speech production exist, but that their representations are not independent, perhaps because they share some common features.

These types of questions cannot be investigated in studies of reaching SA, since, unlike the production of speech sounds, it's not as clear what the larger tasks are that specify reaches.

In this thesis, the general experimental design used to investigate SA in the production of vowels is the same as that described above for reaching SA. It consists of the following three phases:

1. **A Baseline Phase,** in which a subject's produced vowel formants are measured with and without auditory feedback.

2. **A Training Phase,** in which a formant-shifting feedback transformation is introduced

3. **A Testing Phase,** in which the feedback transformation is maintained and the baseline phase formant measurements are repeated.

The experimental results of principal interest are changes in formant frequencies (testing phase - baseline phase) of a subject's vowels. Two versions of this change were looked at:

1. Compensation: formant change in vowels produced while the subject hears altered feedback of his speech.

2. Adaptation: formant change retained when these vowels are produced while the subject is prevented from hearing his speech.

1.3 Thesis Overview

In the rest of this thesis, the apparatus, methods, and experiments used to exhibit speech SA are discussed:

- Chapter 2 develops a more detailed description of the speech SA experimental design.

- Chapter 3 provides an overview of the apparatus used to shift the formants of the subject's speech feedback. It also describes the utterance data collection and analysis methods.

- Chapter 4 describes Study 1 – the preliminary study of speech SA in vowel production.

- Chapter 5 describes Study 2 – a more complete investigation of the properties of speech SA in vowel production.

- Finally, Chapter 6 summarizes and discusses the major findings of this thesis.

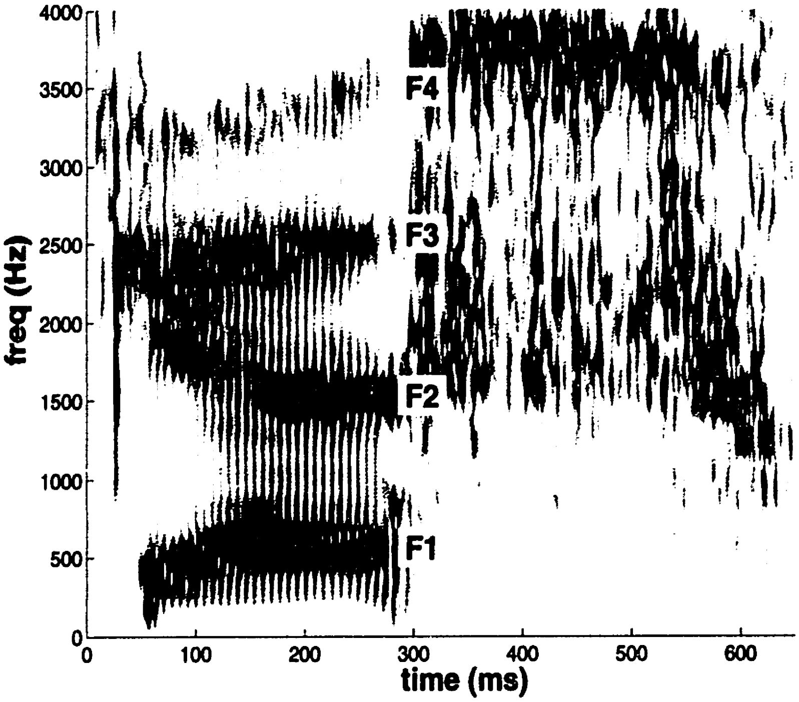


Figure 1-2: Spectrogram of the word "guess". In this plot, each vertical slice represents the utterance's magnitude spectrum at succeeding moments in time. (For any pixel in the slice, the pixel's vertical position represents a frequency, while the pixel's blackness represents spectral energy at that frequency.) Up to about 300ms (where the production of [s] begins) the spectrogram exhibits a regular pattern consisting of four dark bands. These bands are consistent peaks in the utterance's spectrum over time. These bands are called formants and are labeled F1, F2, F3, and F4, from lowest to highest frequency. The formants move as the articulators move (as in the [g]-[ɛ]transition between 0 and 150ms), but maintain steady-state values as the articulators hold position to produce the vowel (between 150 and 300 ms). (Vertical striations during the vowel are from excitatory pulses of air passing through the vibrating vocal folds.)

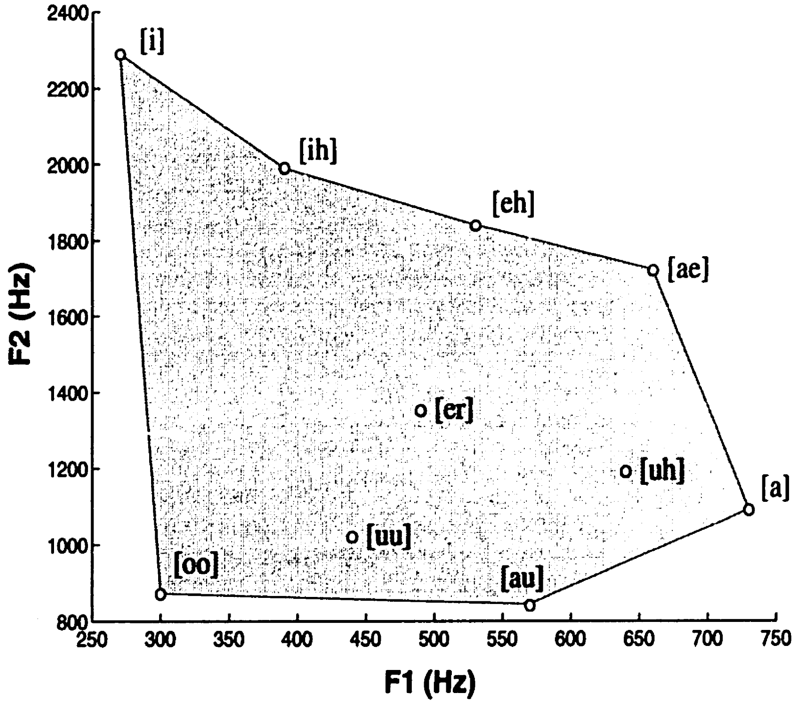


Figure 1-3: An (x,y) plot of vowel (F1,F2) frequencies. (The vowel formant data are averages from a study of 33 male speakers done by [Peterson and Barney, 1952].) Each vowel's x and y position is determined by its F1 and F2 frequencies, respectively. Each vowel occupies a unique position in the plot, and the distribution of vowel positions forms a roughly triangular region called the *vowel triangle* (shown as the shaded region).

Chapter 2

Design of the Speech SA Experiment

In this chapter, the general speech SA experiment design is described in more detail.

2.1 Defining the Task and Feedback Transformation

The objective of an SA experiment is to assess how performance of a task is affected by a feedback transformation that alters sensory feedback. Thus, before designing an SA experiment, the task and feedback transformation must be defined.

2.1.1 The Similarity Between Vowel Production and Reaching

The choice of task and transformation in the speech SA experiment was facilitated by describing vowel production in terms that reveal its similarity with reaching. This description is based on the following two observations:

1. If vowel productions are sustained long enough, vowel formant frequencies attain steady-state values. Figure 1-2 shows an example of this for the vowel [ε].

2. F1 and F2 determine most vowel identities; higher formants show little relative variation when compared across vowels. Figure 2-1 illustrates this for F1, F2, and F3.

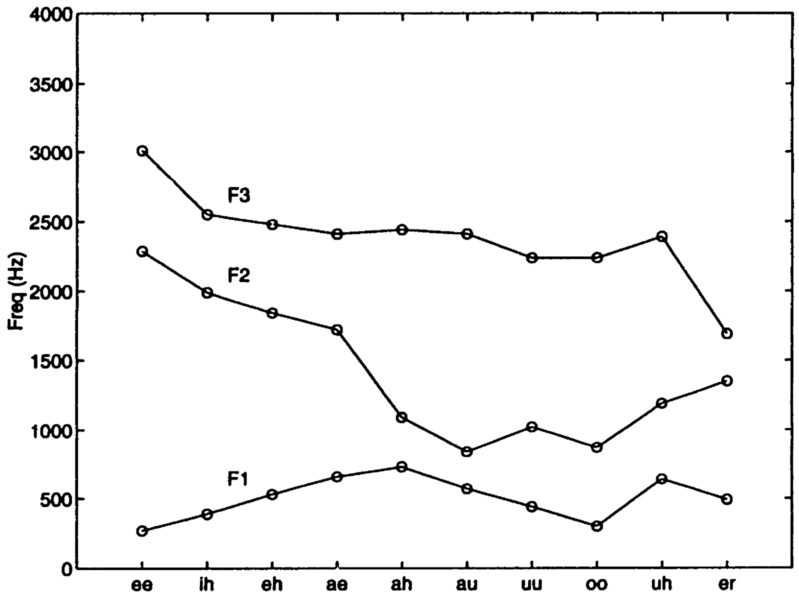


Figure 2-1: The results of the Peterson and Barney study, showing the average F1, F2, and F3 frequencies for a number of vowels, averaged over 33 male speakers. The data show that F1 and F2 vary greatly, but for all but two of the vowels, F3 does not change much [Peterson and Barney, 1952]. Due to font limitations, the following two-letter abbreviations were used, with their IPA symbol in parentheses: "ee" ([i]), "ih" ([ɩ]), "eh" ([ɛ]), "ae" ([æ]), "ah" ([ɑ]), "aw" ([ɔ]), "oo" ([ɷ]), "uu" ([u]), "uh" ([ʌ]), "er" ([ɚ]).

Thus, if vowels are plotted in a 2D plane with *x* and *y* axes determined by F1 and F2, they occupy unique positions in the plane, as shown in Figure 2-2. This plane will be referred to as either formant space, (F1,F2) space, or vowel space. Representing vowels in this manner gives a geometric interpretation to vowel production and perception: production of a desired vowel sound can be viewed as achieving a desired position in (F1,F2) space, while perception of a vowel sound can be viewed as perception of position in this space.

To enhance the analogy between vowel production and reaching, the vowel production task was restricted to the production of vowels in monosyllabic, (stop consonant)-vowel-(stop consonant) (CVC) utterances (which will be referred to as "words").

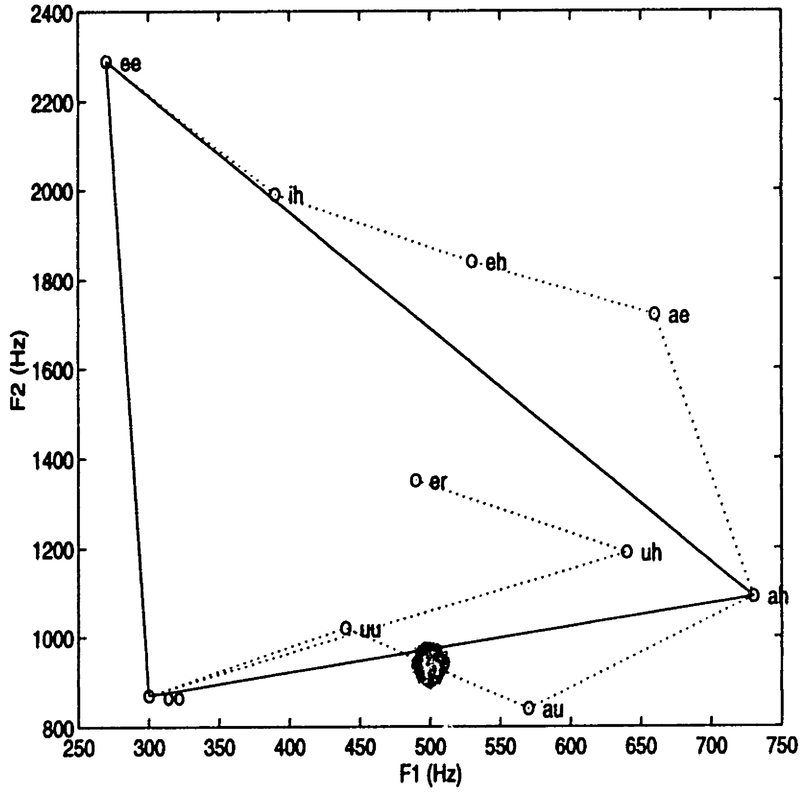


Figure 2-2: The vowels of Figure 2-1, where the x and y position of each vowel is determined by its F1 and F2 values, respectively. A dotted line connecting the points corresponding to the vowels links them in the same order in which they appear in Figure 2-1. These points are considered to form a roughly triangular region in (F1,F2) space, as indicated by the solid line, which is called the vowel triangle. (The two-letter vowel name abbreviations are defined in the caption of Figure 2-1.

Thus, the articulators move from the initial consonant's target configuration, to the vowel's, and back to the final consonant's, much as the hand moves from some initial position, to a target, and back again in a reaching movement.

2.1.2 Perceivability and Compensatability Requirements

In order to permit the modification of task performance by feedback, the feedback transformation must satisfy the following two requirements:

1. *Perceivability*: The feedback transformation must alter the subject's percept. For example, in reaching experiments, shifting a subject's perception of (x,y) position can alter his perception of having reached to the correct position.

2. *Compensatability*: It must be physically possible for the subject to compensate for the perceived error in task performance created by the feedback transformation. For example, in reaching experiments, a subject can compensate for a shift of perceived (x,y) position in one direction by shifting his reaching in the opposite direction.

Shifting the visually perceived (x,y) position produces SA in a reaching task. The parallel in vowel production might be a shift in auditorily perceived (F1,F2) position. What restrictions must be placed on vowel production and the (F1,F2) shift to insure that the shift can be perceived and compensated for? This is determined by the physical limitations on vowel production.

Studies have shown that F1 and F2 reflect the (high ↔ low) and (front ↔ back) position of the tongue body in vowel production [Borden et al., 1994]. Since there are limits to these articulatory dimensions, there are also limits to the range of (F1,F2) combinations a speaker can produce. The limited range restricts vowel productions to a region in formant space called the *vowel triangle* (as shown in Figure 2-2).

These limitations constrain which feedback transformations can be perceived and compensated for. Figure 2-3 illustrates these limits. The figure is a re-plotting of the 33-speaker data of Figure 2-2, but here we assume it represents formants of the vowels produced by a single hypothetical subject.

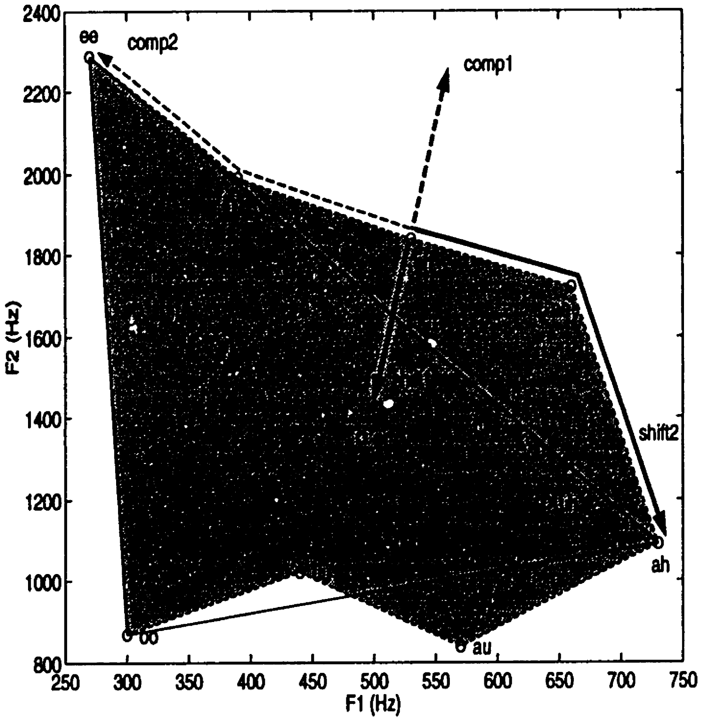


Figure 2-3: Examples of a non-compensatable feedback transformation (**shift1**) and the proposed perceivable, compensatable transformation (**shift2**). **comp1** and **comp2** are the compensation positions for these transformations. The gray region indicates the range of producible vowel sounds. See text for further explanation.

In the figure, the solid arrow labeled **shift1** shows the effect of a hypothetical feedback transformation on the vowel [ɛ]. The arrow base represents the formants of the actual vowel produced by the subject, in this case [ɛ]. The arrow tip represents the perceived formant values of [ɛ] after the feedback transformation. The effect of such a feedback transformation on [ɛ] is likely to be perceived, since it shifts the [ɛ] formants to a point within the vowel triangle. Could a speaker compensate for this shift? If this transformation is assumed to act on all other points in the (F1,F2) plane in the same way that it is shown acting on [ɛ], then the answer is probably no. Consider the the vowel sound whose formants, after being shifted by this transformation, would sound like [ɛ]. The point corresponding to this vowel sound is called the compensation position: it is labeled as **comp1** and is shown as the tip of a dashed arrow whose base is at [ɛ]. Since this point is outside the vowel triangle, it is probably not producible by the speaker. This would prevent the subject from compensating for this transformation of [ɛ].

Suppose the direction of the feedback transformation's formant shift was rotated 180 degrees, so that now **comp1** represents its action on [ɛ]. In this case, the compensation position would be at the point labeled **shift1**. Since this point is within the vowel triangle, it is likely the subject could produce this sound. Less clear is whether the effect of such a transformation is still perceivable. Recall that **comp1** now represents the perceived formants of [ɛ] after the feedback transformation. Since the point is outside the vowel triangle, it is in a region of (F1,F2) space where the subject does not hear himself producing vowel sounds. The subject therefore may not be sensitive to vowel sound differences in this region. If this were true, the formants of **comp1** might not be perceived as being significantly different from those of [ɛ].

To resolve these problems of perceivability and compensatability, task and transform were restricted in the following way:

- The task was limited to the production of CVC words in which the vowel was [i], [ɩ], [ɛ], [æ], or [ɑ].

- The feedback transformation was restricted to shifts of position along the path in formant space which connects these vowels.

To understand why this choice of task and transform satisfies the perceivability and compensatability requirements, consider the path connecting the vowels [i], [ɩ], [ɛ], [æ], and [ɑ]. Figures 2-2 and 2-3 show that it forms an edge of the vowel triangle. This path will be called the *[i]--[a] path*, and the vowels along it will be called *path vowels.*

The distribution of vowels along this path suggests that speakers can produce vowel sounds anywhere along it. This hypothesis is supported by the fact that the formant trajectory of the diphthong [ai] is roughly along the [i]–[ɑ] path. This suggests that shifts of perceived path position in one direction could be compensated for by shifts in produced vowel formants in the opposite direction along this path.

Note also that position along this path distinguishes five different vowels. This suggests that speakers would be highly sensitive to changes in perceived path position. Thus, the effects of a feedback transformation which shifts perceived path position would be expected to be quite salient to a speaker.

The arrow labeled **shift2** in Figure 2-3 shows a hypothetical example of such a path-shifting feedback transformation. The arrow shows the supposed action of this transform on the perceived formants of [ɛ]: they are shifted along the path towards [ɑ], as indicated by the tip of the arrow labeled **shift2**. If this transformation is assumed to act on all path positions with the same amount of path shift, then there would be a compensation position for the action of **shift2** on [ɛ]: i.e., there would be some path point between [ɛ] and [i] whose perceived formant values, after the feedback transformation, would sound like [ɛ]. This point is labeled **comp2** in the figure. Thus, by shifting the formants of his production of [ɛ] towards [i], a subject could compensate for the effect of the transformation and restore his original percept of a correctly produced [ɛ].

2.2 Design of the SA experiment

There are two questions that any SA experiment must address:

1. Does the subject *compensate*? Does he adjust his task performance to compensate for the perceptual shift produced by the feedback transformation.
2. Does the subject *adapt*? Does he retain his adjusted performance, even when denied sensory feedback?

An appropriate methodology for answering these questions is to record task performance in an experimental procedure consisting of:

- A **baseline phase**, in which the subject is prompted to perform specific tasks with sensory feedback blocked.
- A **training phase**, in which the subject is exposed to the feedback transformation.
- A **testing phase**, identical to the baseline phase.

The logic of this design is: by comparing task performance in the baseline and testing phases, task performance change due to exposure to the feedback transformation can be assessed. To illustrate this point more concretely, let us consider a reaching experiment.

2.2.1 Design of a Reaching SA Experiment

Held et al. [Held and Gottlieb, 1958] studied adaptation to shifts of perceived hand position (earlier work had already established that subjects could compensate for such shifts). This experiment involved three phases:

1. A baseline phase in which reaching performance without visual feedback was recorded.
2. A training phase in which the subject was exposed to a transformation of the visual feedback.

3. A testing phase, identical to the baseline phase, in which reaching performance without visual feedback was again recorded.

The experimental methodology involved blocking visual feedback and prompting for reaching movements. This was achieved by a mirror positioned between the subject's eyes and his hand. This mirror blocked the subject's view of his hand, but allowed him to see the reflected image of crosshairs in the same virtual plane as his hand. He could then be prompted to reach to specific crosshair intersections (target points) without being allowed to see his hand (see Figure 2-4).

The subject was subsequently exposed to a feedback transformation that was designed to minimize awareness of targeting errors. In the training phase, the mirror was removed. The subject could no longer see the crosshair target points, but could now see his hand. The action of removing the mirror also positioned a prism in front of the subject's eyes. The prism shifted the perceived location of his hand by 11cm. The subject then viewed his hand motion through the prism for some amount of time.

The datum of interest in the Held et al. experiment was the subject's hand position. Subjects were prompted to reach to a target point and the (x,y) coordinates of the resulting hand position were recorded. For each target point, mean hand position of reaching movements in the testing phase was subtracted from mean position of reaching in the baseline phase. The difference was then compared with the shift of the feedback transformation to assess whether the change in reaching position compensated for the effect of the transformation.

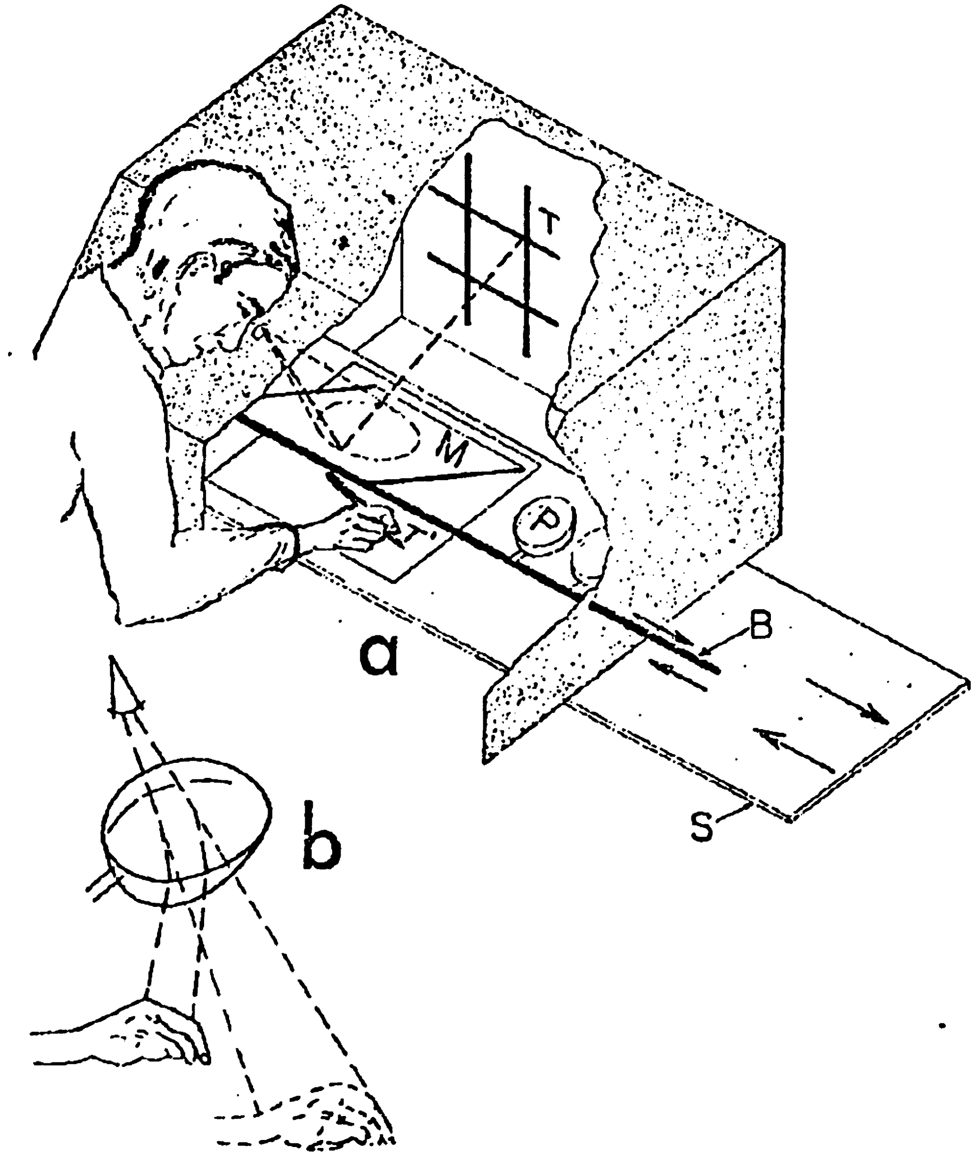


Figure 2-4: The apparatus used by Held et al. for investigating the effect of visual shifts on reach.ng. (a) Mirror M's reflections of targets T created virtual images T' of the targets in the plane of the subject's hand (and also prevented the subject from seeing his hand). M was connected via bar B to prism P; B could be pushed (along with surface S on which the subject had marked targets) so that the subject could see his hand through prism P, but not the targets nor his marking of them. (b) illustrates the visual shifting action of prism P. (Figure from [Held and Gottlieb, 1958].)

This description of the experiment design of Held et al. highlights four key design issues:

1. Prompting the subject to perform a specific task.

2. Exposing the subject to the feedback transform.

3. Intercepting the subject's feedback.

4. Recording task performance.

Each of these issues will now be discussed in more detail in the context of speech production.

2.2.2 Prompting the Subject to Perform a Specific Task

Assessing adaptation in an SA experiment requires comparing how a subject performs a task before and after being exposed to a feedback transformation. An SA experiment thus needs some method of prompting a subject to perform the same task at different times in the experiment. One such method involves showing the subject the desired sensory outcome for the task.

2.2.2.1 Task Prompting in Reaching SA Experiments

There is a difficulty associated with providing subjects with desired sensory outcomes. In particular, if a subject can see both his hand and the target, he can see directly any errors he makes in reaching to the target. In their reaching SA experiment, Held et al. were able to avoid this problem by using a training phase in which the subject sees only his hand and no target points.

2.2.2.2 Task Prompting in Speech SA Experiments

In a speech SA experiment, subjects could be prompted to produce a specific speech sound by letting them hear the speech sound they should produce. Unfortunately, this would be analogous to the presentation of visual targets in reaching SA experiments, and would thus create the same methodological problem described above. However, with speech the problem can be avoided via the use of printed words.

For a given language accent group, there are rules that relate the sounds and spellings of utterances, even if they are not meaningful words in the language [Ladefoged, 1982]. Since these rules are known by speakers of that accent group, spellings can be devised that will prompt the pronunciation of the desired speech sounds. For example, in the eastern North American accent group of the English language, the spellings:

{"peep", "pip", "pep", "pap", "pop"}

prompt CVC utterances with the vowels:

{[i], [ι], [ε], [æ], [ɑ]}

By prompting for vowel targets in this fashion (as opposed to acoustically), there is no explicit prompting for specific positions in formant space, and thus no external reference available for a subject to judge the correctness of his produced vowel formants.

For this reason, the CVC utterances were prompted visually using spelled words that elicit the desired speech sounds. This imposed the modest limitation that all subjects must come from the same language accent group. Because of the large numbers of speakers from the eastern North American accent group at MIT (where the studies were carried out), subjects were restricted to be from this accent group.

2.2.3 Exposing the Subject to the Feedback Transform

In an SA experiment, the purpose of altering the feedback delivered to the subject is to alter what he perceives to be the sensory outcomes of his motor actions. To do this, the altered feedback must be provided in a *veridical* way to the subject. In other words, the subject must believe that this feedback arises causally and immediately from his motor actions.

One key factor in determining veridicality is the amount of delay introduced by the feedback alteration. It has been found, for example, that subjects will not exhibit reaching SA if there is a delay in visual feedback of more than 300ms

[Held and Durlach, 1991]. In speech, minimal feedback delay is also important because studies have shown that feedback delays of as little as 30ms begin to disrupt speech production [Lee, 1950, Yates, 1963].

Another factor is fidelity of the altered feedback. Fidelity refers to how artificial the subject perceives the feedback to be. For reaching SA, low fidelity feedback appears to suffice; in particular, experiments have shown that a dot on a video screen showing only the subject's hand position is sufficient to induce adaptation [Welch, 1972]. Since speech SA has not been investigated previously, it was not known how impoverished the feedback could be and still induce adaptation. Thus, the fidelity criterion adopted in this study was that the subject should not perceive any significant difference between the output of the feedback transformation apparatus and normal acoustic feedback of his whispering.

2.2.4 Intercepting the Subject's Feedback

In an SA experiment, intercepting the sensory feedback the subject would normally receive is necessary for two reasons: (1) in the training phase, the subject should be exposed to the altered sensory feedback, not the normal, unaltered feedback; (2) in the baseline and testing phases, the subject should receive no sensory feedback.

In reaching SA experiments, blocking visual feedback is relatively easy. In a speech SA experiment, blocking auditory feedback is difficult. Two reasons for this are:

1. The source of light energy for exhibiting arm movements is external to the subject and under the control of the experimenter. The source of acoustic energy for exhibiting speech movements is internal (the glottis) and not directly controllable by the experimenter.

2. The bones and tissues of the body conduct acoustic energy fairly well; they do not conduct light energy well. For these (and other) reasons, acoustic speech feedback reaches the cochlea internally via the skull and soft tissues of the head as well as externally via the air, while visual feedback reaches the eyes only via the easily blockable external pathway.

As discovered by Von Békésy, the ratio of the acoustic energy transmitted to the cochlea internally via the head (commonly called "bone conduction") to that transmitted by air (called the "side tone") is about 5dB [Békésy, 1949]. Thus the internal pathway is nearly as efficient as the external one. Since there is no feasible way to interrupt the bone conduction signal,[1] there appeared to be only one option: mask out the bone conduction signal with adequate noise.

There are two problems with this option. First, the level of noise necessary to do this for voiced speech is quite high. Second, even assuming sufficiently loud masking noise to block a subject's hearing, it is not easy to introduce a feedback substitute in the presence of a mask; the substitute feedback must be provided at a higher amplitude level than the masking noise.

The solution chosen was to restrict the subject to using *whispered* speech. The overall amplitude of whispered speech is much less than that of voiced speech [Schwartz, 1970]. Indeed, we found in pilot experiments that very weak masking noise could block out most of a subject's hearing of his own whispered speech. This allowed a mildly amplified version of the substitute acoustic feedback to be mixed with the noise and fed to the subject's ears, thus avoiding high sound amplitude levels. Somewhat stronger, but still mild, masking noise could also be used by itself to block all of a subject's hearing of his own whispered speech.

2.2.5 Recording Task Performance

In an SA experiment, some aspect of the performance is recorded to assess whether the subject has compensated for the effects of the feedback transformation.

In the case of the Held et al. experiment, the feedback transformation shifted perceived hand position. The recording of task performance was limited to recording endpoint hand position (i.e., the (x,y) point the subject's hand arrived at in a reach to a target). For each experimental phase, mean endpoint hand position of all reaches to a target could then be calculated and compared. Mean hand position

[1] Air conducted signals can be attenuated by earphones.

in the testing phase could be subtracted from mean hand position in the baseline phase. Any observed difference could then be compared with the shift of the feedback transformation to assess whether the hand position change compensated for the transformation.

In the vowel SA experiment, the same compensation assessment procedure was adopted, with vowel formants being recorded instead of hand positions.

2.3 Summary of the Speech SA Experiment Design

In this chapter, the key elements of an experimental design for investigating SA in vowel production were identified, leading to the following general outline:
(1): The task of subjects is to whisper CVC utterances, where the C's are stop consonants and V is a vowel from the set of [i]–[ɑ] path vowels: {[i],[ɩ],[ɛ],[æ],[ɑ]}. These utterances are prompted orthographically with appropriately spelled words.
(2): The feedback transformation shifts perceived position along the [i]–[ɑ] path in formant space. The apparatus that accomplishes this transformation has to introduce less than 30ms of feedback delay, and the fidelity of its output has to be such that the subject notices no significant difference between it and his normal acoustic feedback.
(3): The purpose of the experiment is to determine if F1 and F2 of a subject's vowel production can be affected by experience with the feedback transformation. The design of the experiment therefore consists of the following sequence:

- **A baseline phase**, in which the subject is prompted visually to whisper CVC utterances while the formant values of his utterances are recorded. For some of the utterances, he hears (unaltered) feedback of his whispering. For others, he is prevented from hearing his whispering by masking noise.

- **A training phase**, in which the feedback transformation is introduced. In this phase, the subject is prompted visually to whisper a limited number of utterances while the feedback transformation alters the perceived formants of his whispering.

- **A testing phase**, in which the subject is again prompted visually to whisper CVC utterances while the formant values of his utterances are recorded. For some of the utterances, he hears altered feedback of his whispering. For others, he is prevented from hearing his whispering by masking noise.

(4): The subject's responses during the experiment are grouped into two types:

1. **Compensation responses**: Utterances the subject produces while he hears feedback (unaltered or altered) of his whispering.

2. **Adaptation responses**: Utterances the subject produces while masking noise prevents him from hearing his whispering.

For both types of responses, vowel production changes are assessed. To do this, the steady-state vowel portions of each utterance are extracted, and mean formant values computed. For each vowel, a comparison is made between mean formant values in the testing phase and those of the baseline phase. The resulting formant difference is compared with the formant shift of the feedback transformation. This comparison is used to assess whether the vowel production change compensated for the feedback transformation.

Compensatory vowel production changes seen in a subject's compensation and adaptation responses exhibit different phenomena:

- Such changes seen in his compensation responses exhibit *compensation*: they show he attempted to restore how he heard his vowel formant frequencies before feedback was altered.

- Such changes seen in his adaptation responses exhibit retained compensation: they show he retained his compensatory vowel production changes even when denied acoustic feedback. This retention of compensation is called *adaptation*: if it occurs, speech is said to exhibit SA.

Chapter 3

Apparatus and Methods

In the previous chapter, the general outline of an experiment to investigate SA in vowel production was developed. This was the basis for the actual vowel SA experiments described in later chapters. In this chapter, the apparatus, procedures, and data analysis methods used in these experiments are discussed.

3.1 Overview

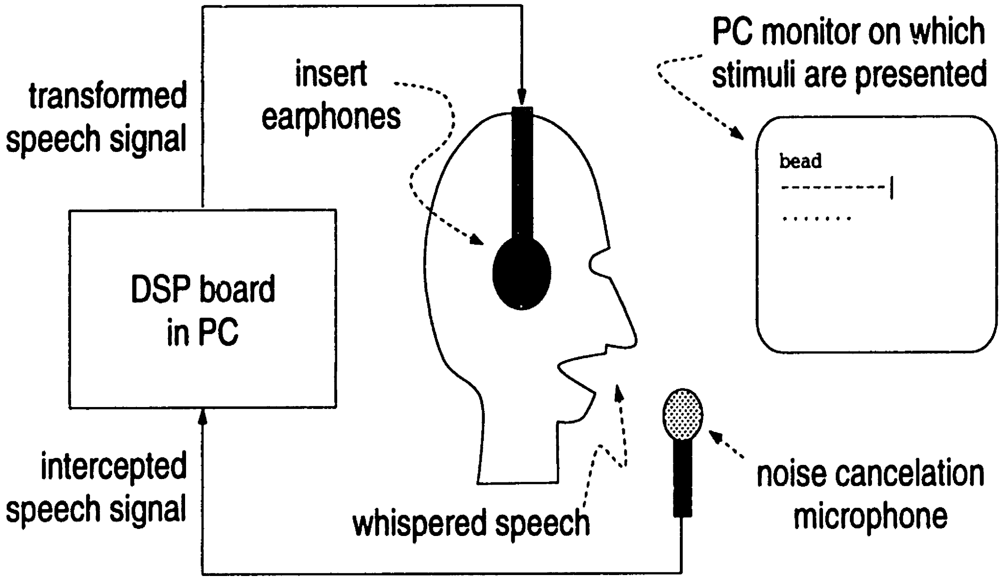

transformed speech signal
insert earphones
PC monitor on which stimuli are presented
bead
----------|
....
DSP board in PC
intercepted speech signal
whispered speech
noise cancelation microphone

Figure 3-1: Overview of the experimental apparatus.

Figure 3-1 shows an overview of the key components of the apparatus used in the experiments. The subject sits in front of a PC video monitor wearing a head-mounted microphone and earphones. Words are presented on the monitor screen for the subject to pronounce. He whispers his pronunciations of these words, and his speech is transduced by the microphone and fed as input to a digital signal processing board (called the DSP system) inside the PC. The DSP system implements the formant shifting acoustic transformation and returns the altered feedback to the subject via the insert earphones. It also records the formants of subject's utterances.

In the following sections, the process of conducting the experiments using this setup will be described. These sections will explain the four major aspects of this process:

- **Transforming the Acoustic Feedback**: The signal processing done by the DSP system to alter the formants of a subject's whisper feedback.
- **Constructing the Feedback Transformation**: The precise definition of the formant alterations used, as well as the general method for constructing them.
- **Utterance Data Acquisition**: How the experiments prompted subjects to whisper the desired utterances.
- **Utterance Data Analysis**: How data collected in an experiment was analyzed.

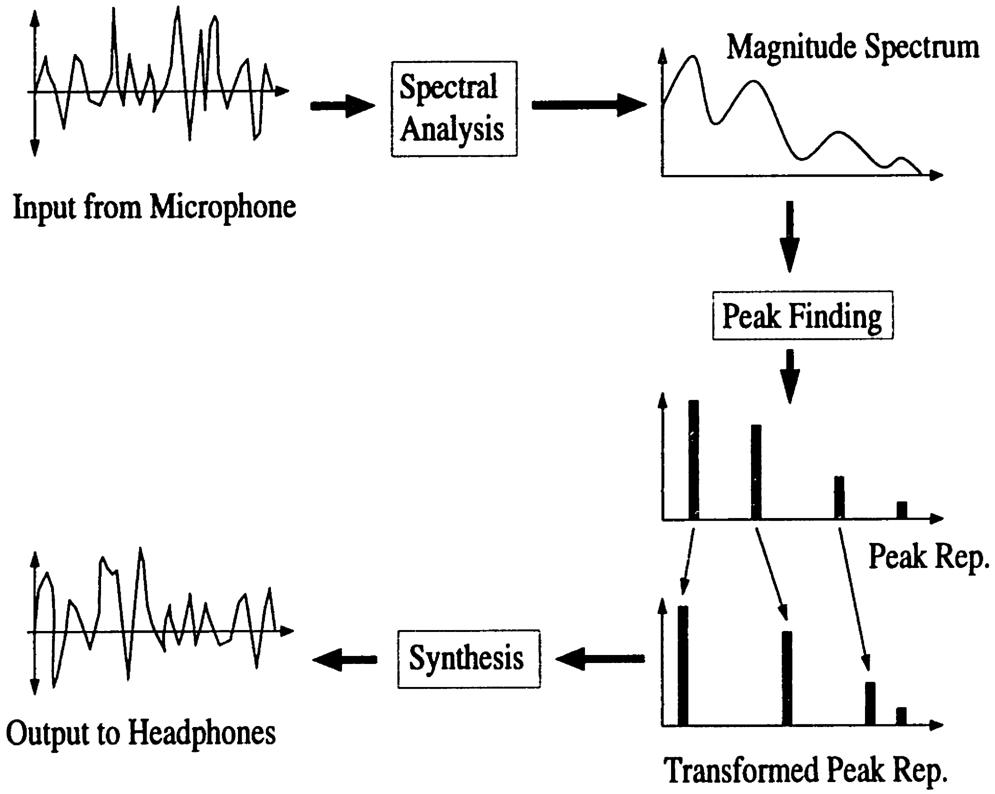


Figure 3-2: Overview of the signal processing that implements the acoustic transformation.

3.2 Transforming the Acoustic Feedback

Figure 3-2 shows an overview of the key signal processing steps running on the DSP which implements the acoustic transformation. It is an analysis-synthesis process which repeatedly:

(a) Captures from the microphone an 8ms frame of the subject's whispered speech (64 time samples at an 8KHz sampling rate).

(b) Performs a 64-channel spectral analysis of this frame, retaining only a smoothed magnitude spectrum of it.

(c) Estimates the first four formants from the magnitude spectrum.

(d) Alters the frequencies of the three lowest formants via a lookup table.

(e) Resynthesizes a new 8ms frame of whispered speech from the altered formants.

This process incurred a feedback delay of only 16ms.

A more complete description of the signal processing done in these steps can be found in Appendix B. Here we discuss further only those aspects which affect the SA experiment design.

Step (a): Acquisition Time samples of the subject's whispering were acquired from the microphone at a rate of 8KHz. 64 time samples constituted an 8ms *frame* of whispered speech data. Once a frame of data was collected, it was passed on to the spectral analysis step.

The 8KHz rate at which time samples were acquired from the microphone limited the bandwidth of the spectral processing to 4000 Hz. To improve fidelity, it was desirable to retain F4 in the synthesized feedback returned to the subject. For female and child speakers, F4 is usually above 4000 Hz, so only male speakers were used in the experiments.

Step (b): Spectral Analysis A 64-channel magnitude spectrum was computed for the current frame. This spectrum was averaged with the previous frame's spectrum and then smoothed in frequency.

Step (c): Formant Estimation From the smoothed magnitude spectrum, the frequencies and amplitudes of the first four formants were estimated.

In voiced speech, it is conventional to estimate formants from peaks in the envelope of the magnitude spectrum [Peterson and Barney, 1952]. However, complications were found in the spectrum of whispered speech which prevented this simple estimation approach. In particular, it was found that F1 was best estimated as the centroid of spectral amplitudes within a limited range of frequencies. This range was subject-specific and was called the F1 range. The higher formants – F2, F3, and F4 – were estimated as the three lowest frequency spectral peaks above the F1 range.

This estimation procedure is illustrated in Figure 3-3. Further discussion of this approach to formant estimation can be found in the appendices: Section A.1 discusses the characteristics of whispered speech that motivated the approach, while Appendix B discusses the signal processing details involved in implementing the approach.

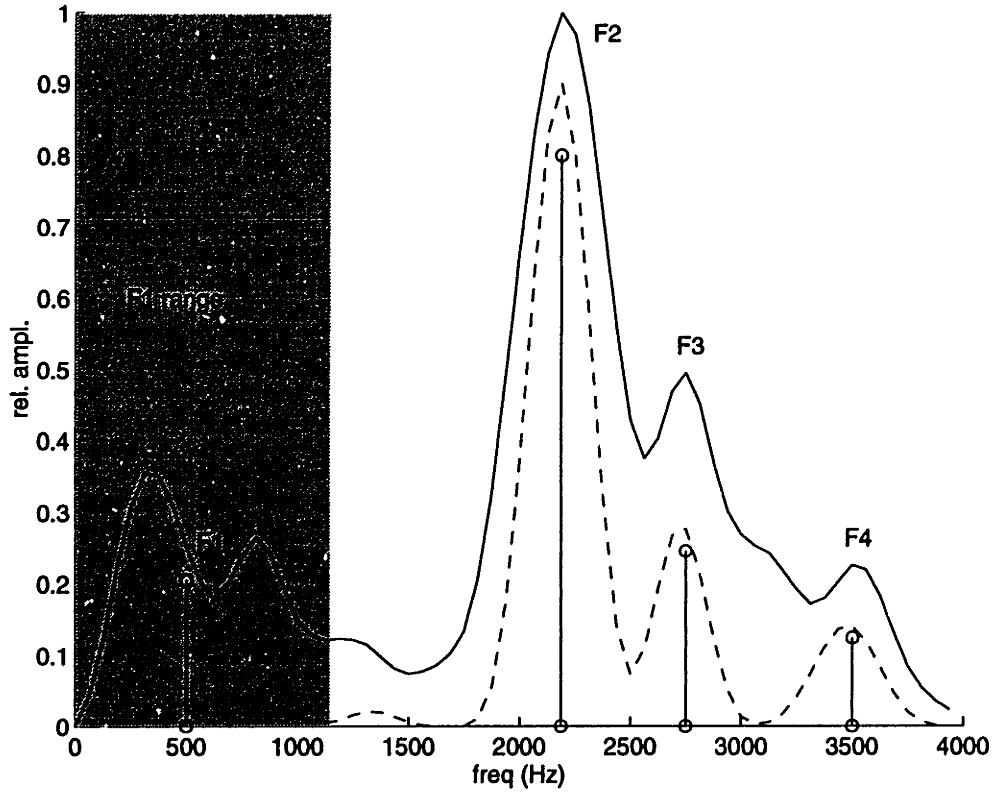


Figure 3-3: An illustration of the formant estimation procedure. The solid line shows the spectrum of one frame of the author's whispering of [i]. The circle-terminated vertical lines display the formants estimated from this spectrum. The gray region highlights the F1 range. As these lines indicate, F1 is estimated as the centroid of spectral amplitudes within the F1 range, while outside of this range, F2, F3, and F4 are estimated from the spectrum's peaks, just as they would be in a voiced spectrum. (In this range, the dashed line is a peak-enhanced version of the spectrum used to facilitate peak finding. Note also that, for display purposes, the spectra and estimated formant amplitudes have been offset vertically from each other by magnitude scaling.)

Step (d): Formant Alteration To implement the feedback transformation, a lookup table was used to shift the frequencies of F1, F2, and F3.[1]

In this approach, the desired feedback transformation was stored in a table as a finite number of transformation pairs of the form:

$$((F1,F2,F3)_{\text{original}}, (F1,F2,F3)_{\text{transformed}})$$

The process of altering the formant frequencies was thus a two step process involving:

1. Finding in the table the transformation pair whose $(F1,F2,F3)_{\text{original}}$ entry matched the frequencies of F1, F2, and F3.

2. Replacing the frequencies of F1, F2, and F3 with the values found in the $(F1,F2,F3)_{\text{transformed}}$ entry of this transformation pair.

It was possible to use a lookup table because the number of possible (F1,F2,F3) formant frequency combinations was limited. This was due to several factors:

1. The spectral analysis was discrete: only 64 frequency values between 0 and 4KHz were represented. Thus, each formant frequency could take on only one of 64 possible values.

2. Because the experiments involved CVC words restricted to the vowels [i], [ɩ], [ɛ], [æ], and [ɑ], only a limited region of (F1,F2,F3) space around these vowels needed to be considered.

By reducing the formant alteration process to a table lookup operation, it could be implemented with minimal computational overhead, which was essential given the time constraints on the signal processing. The result was that the actual computations involved in creating the table could be done off line in advance of the experiment, using a process that will be described below.

[1] In the previous chapter, the feedback transformation was specified as needing only to shift F1 and F2. However, because the DSP had sufficient processing speed and memory, the implemented feedback transfcrmation also shifted F3 to enhance the fidelity of the shifted feedback.

Step (e): Synthesis The process of synthesizing substitute acoustic feedback for the subject from the altered peak representation was based on an approach called *formant synthesis*, which is a method used in many current speech synthesis systems.[Klatt, 1980, O'Shaughnessy, 1987]

The approach is based on the idea that speech can be modeled as a *source-filter* process in which the vocal tract is seen as a time-varying linear filter, and speech is the response of this filter to the glottal source function. For any one time instance, this response is the convolution of the glottal source function with the impulse response of the vocal tract filter.

The speech synthesis process based on this therefore involved two steps:

1. Computing the filter impulse response from the formant frequencies and amplitudes.

2. Convolving this impulse response with a random impulse sequence representing the whispered glottal source function.

The synthesized speech frame was fed back to the subject via the earphones.

Thus, by controlling how the lookup table altered formant frequencies, we could control how much the synthesized feedback differed from the subject's original acoustic output.

3.3 Constructing the Feedback Transformation

As described above, feedback transformations were implemented on-line using lookup tables. For each feedback transformation used, a table was needed that specified the formant shifts of every (F1,F2,F3) combination a subject could produce. This approach off-loaded a considerable computation burden: the table could be precomputed prior to its use in the experiment.

In this section, the method used to construct these *feedback transformation tables* will be described. This is followed by a discussion of the specific feedback transformation tables actually used in the experiments.

3.3.1 Defining the Feedback Transformation

The transformation of a given (F1,F2,F3) combination was based on the general concept discussed in Section 2.1.2. There it was explained that a shift of *[i]-[ɑ] path position* should produce a perceivable feedback alteration that a subject could compensate for. (The arrow labeled **shift2** in Figure 2-3 shows an example of such a feedback alteration.) This concept was based on the idea that movement along the [i]–[ɑ] path changes phonemic value, while movement perpendicular to the path does not.

In the discussion which follows, the feedback transform definition developed from this general idea is explained.[2]

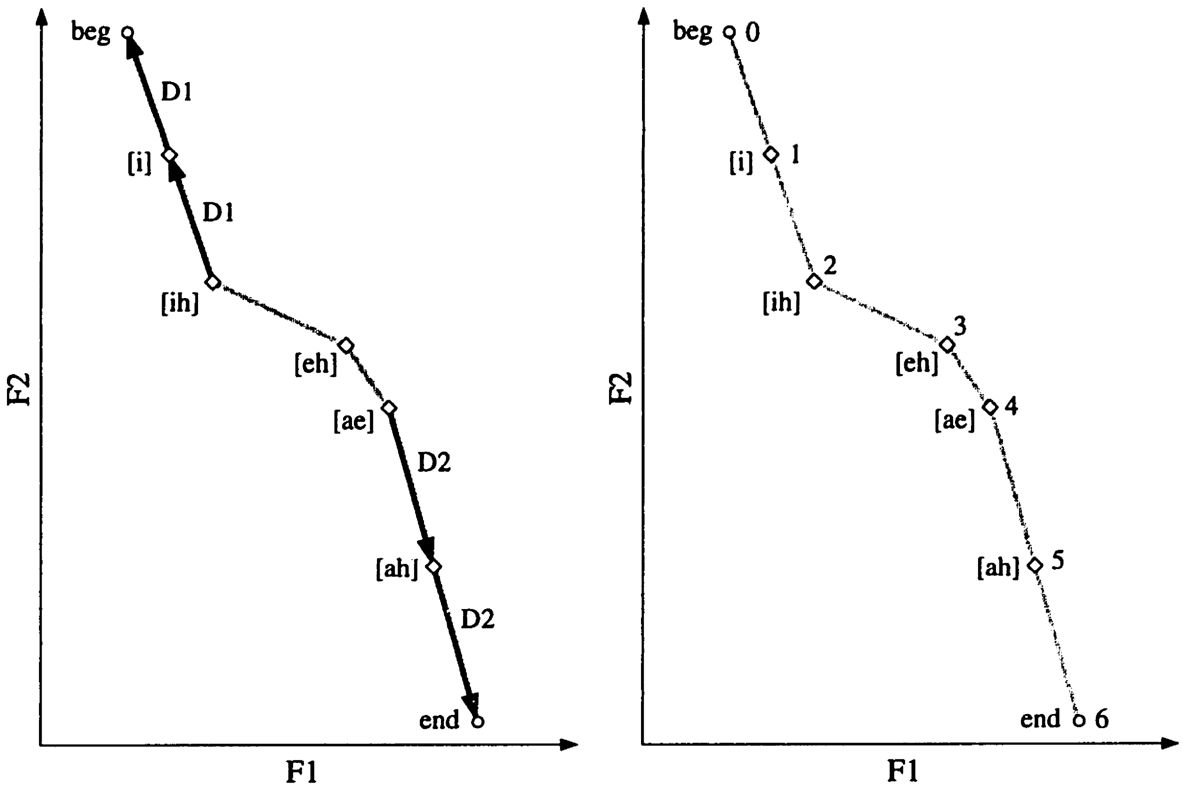
3.3.1.1 Defining [i]–[ɑ] Path Position

Defining a transformation based on the above concept required a precise definition of [i]–[ɑ] path position. This required specifying:

1. The precise definition of the [i]-[ɑ] path.

2. How position along it would be measured.

[2] In this explanation, most concepts are illustrated in (F1,F2) formant space. However, it should be kept in mind that these concepts are meant to apply equally well to (F1,F2,F3) formant space.



(a) extending the path

(b) numbering the path

Figure 3-4: Definition of a subject's [i]–[ɑ] path (see next page).

(a) shows how the path was extended beyond [i] and [ɑ]. (1) At the [i] end of the path, the difference vector D1 between [ι] and [i] was added to [i]. The resulting position was designated the path *beginning* (shown as "beg" in the plot). (2) At the [ɑ] end of the path, the difference vector D2 between [æ] and [ɑ] was added to [ɑ]The resulting position was designated the path *end* (shown as "end" in the plot).

(b) shows how the path reference points were numbered. Path reference points are shown as white symbols: path vowels as white diamonds; path extensions as white circles. The complete [i]–[ɑ] path is shown as the gray line linking the path reference points.

Defining the [i]–[ɑ] Path One problem with the concept of a feedback transformation that shifted perceived [i]–[ɑ] path position was defining how it should act on the ends of the path: [i] and [ɑ]. The chosen solution was to extend the path at both ends so that [i] and [ɑ] were no longer the path ends. Figure 3-4(a) shows how this was done.

Next, the points corresponding to path beginning, path vowels, and path end were numbered from 0 to 6 as shown in Figure 3-4(b). These numbered points are called *path reference points.* The actual curve corresponding to the [i]–[ɑ] path was created by connecting, in the numerical order, these path reference points.

Measuring Path Position In reaching SA studies, a fixed prism shifts perceived position of every (x,y) point by the same amount. There is thus a uniformity in the prism's effect on a subject's perception of (x,y) position. In the vowel SA studies, the intent was to design feedback transformations with this same uniformity of effect on a subject's perception of path position. That is, we wanted equal shifts in path position, no matter where on the path they occurred, to be equally salient to the subject.

Path position, thus, was to be measured such that equal changes in its value anywhere along the path were always equally salient to the subject. To gauge this saliency, it was assumed that shifts of path position from one path vowel to the next were all equally salient to the subject. In other words, shifting path position from [i] to [ɩ] was as salient as shifting from [ɩ] to [ɛ] or from [ɛ] to [æ] or from [æ] to [ɑ]. This assumption was based on the hypothesis that all such changes in vowel identity had equally significant linguistic consequences.[3]

Thus, position along the path was specified in relation to the path vowels. For any point on the path, its position was calculated as follows:

- If the point was a path reference point (a path vowel, or the path beginning or end point), its position was simply the number corresponding to that point.

[3] Consider, for example, how changing path vowel V in bVd changes the word from "beet" to "bit" to "bet" to "bat" to "bought".

- For any point between between two path reference points, position was expressed as a fraction of the total curve length between these path reference points. Figure 3-5 illustrates how this was done for a path point P between [ɛ] and [æ].

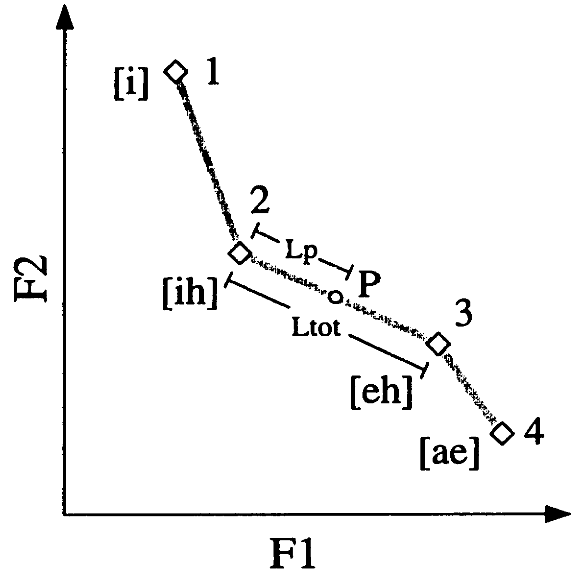


Figure 3-5: Showing how the path position of path point P between [ɛ] and [æ] was calculated as the path position of [ɛ]plus the ratio of path lengths l_{tot}/l_{part}.

By this measure, from any path vowel, a 1.0 unit path position change is a single change in vowel identity. For this reason, the units in which path position is measured are called "vowel-units".

Using this measure, each feedback transformation was defined by single number: the amount of path shift it added to each path position.[4]

This definition is consistent with the original concept described in Section 2.1.2. This is illustrated in Figure 3-6, which is a re-description of Figure 2-3 using the above transform definition.

[4] Obviously, sensible additions to this definition were needed to specify what to do when adding the shift resulted in a path position less than 0 or greater than 6 (i.e., off the path). These additional specifications were: if the resulting path position was less than 0, limit it 0; if it was greater than 6, limit it to 6.

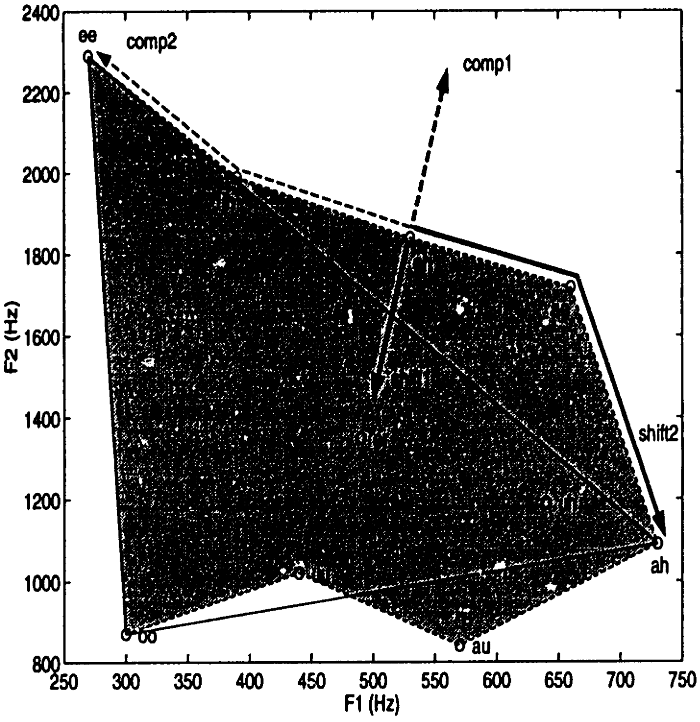


Figure 3-6: Redescription of Figure 2-3 using the transform definition and the terms defined in this section. In the figure, **shift2** represents a +2.0 feedback transformation: it adds 2.0 vowel units to the perceived position of any point on the [i]–[ɑ] path. It's action on [ε] increases [ε]'s path position 2.0 vowel units to [ɑ]. To compensate for a 2.0 vowel unit increase in perceived path position, a subject must decrease his path position by 2.0 vowel units. Thus **comp2** represents complete compensation for **shift2**'s effect on [ε]: a path position shift of -2.0 vowel units to [i]. This restores the subject's untransformed percept of [ε], since the transform's effect on [i] shifts [i] +2.0 vowel units to [ε].

3.3.1.2 Path Projection and Deviation

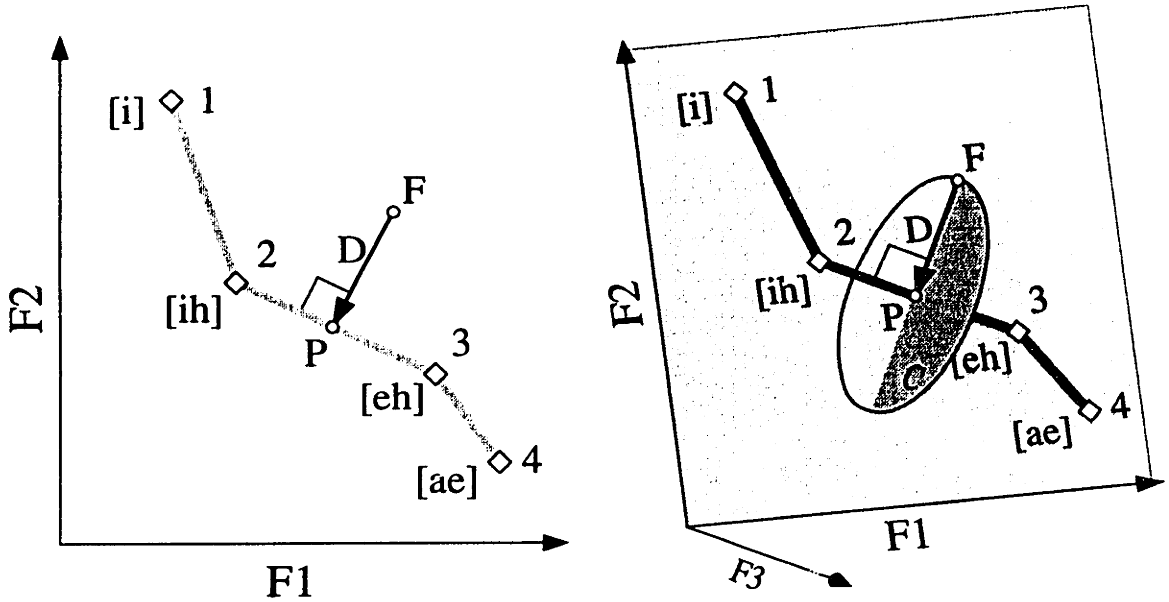
To complete the definition of the feedback transformations, their actions on points near, but not actually on, the [i]–[ɑ] path needed to be specified. This was done by representing each point F in formant space in terms of two quantities (see Figure 3-7):

1. The position P on the [i]–[ɑ] path nearest this point.
2. A vector D representing the difference between the nearest path position and this point.

Quantity 1 was called a point's *path projection*: it functioned as the path position of the point.

Quantity 2 was called a point's *path deviation*: it represented how much the point actually deviated from its path projection.

Figure 3-7 shows how F's path projection and deviation were calculated in two- and three-dimensional formant spaces.



(a) (F1,F2) space

(b) (F1,F2,F3) space

Figure 3-7: Showing (in two- and three-dimensions) how points in formant space were represented in terms of path projection and deviation.

(a) shows path projection and deviation of point F in (F1,F2) space. In this figure, a limited segment of the [i]–[ɑ] path is show as a gray line. The path point P nearest F is found by dropping perpendicular D from F to the [i]–[ɑ] path. The path position of P (see Figure 3-5) is called F's *path projection.* It is a scalar quantity. Perpendicular D is called F's *path deviation.* In (F1,F2) space, D is a scalar quantity: D's magnitude represents the perpendicular distance from F to P; D's sign represents whether F is above (+) or below (-) the path. (In the figure, F is shown above the path.)

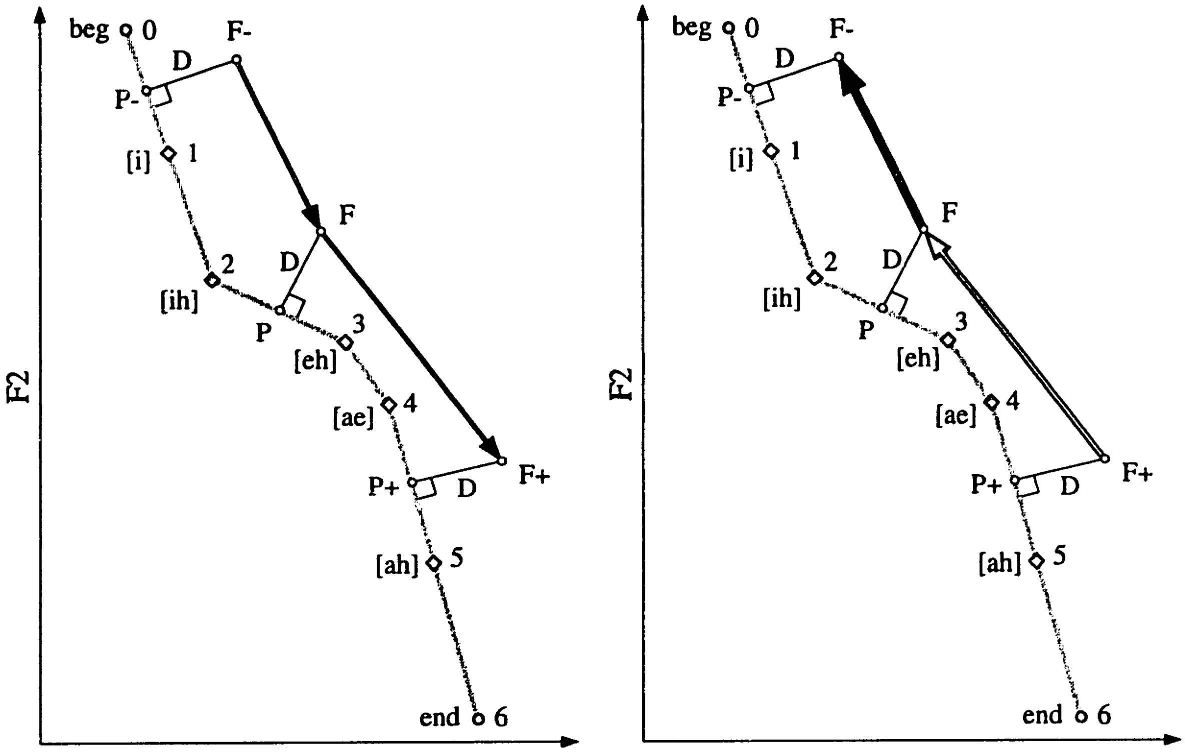
(b) shows path projection and deviation of the same point F in (F1,F2,F3) space. In this figure, the light gray (F1,F2) plane is the same plane shown in figure (a) (we assume point F and the [i]–[ɑ] path all have the same F3 values.). The F3 axis is shown rising obliquely out of the page. The path point P nearest F is still found by dropping perpendicular D from F to the [i]–[ɑ] path. F's path projection (the path position of P) is still a scalar. However, in (F1,F2,F3) space, F's path deviation D is a vector: D's magnitude still represents the perpendicular distance from F to P, but D's direction can't be specified with just a number sign. With the added F3 dimension, F could be anywhere on circle C and still have the same perpendicular distance to P. Thus, to uniquely specify point F, D's direction must be specified as a rotation angle.

3.3.1.3 The Complete Transform Definition

By representing points near the [i]–[ɑ] path in this fashion, each feedback transformation could still be defined by a single number: the amount of path shift it added to each point's path projection. Feedback transformations defined this way shifted a point's path projection but left its path deviation unaltered.

This feedback transform definition also specifies what subjects must do to compensate. As defined above, a feedback transform shifts perceived path projection by some amount and in some direction along the path. To cancel this perceived effect, a subject must shift path projection of his vowels by the same amount in the opposite direction. However, since the transform leaves path deviation unaltered, the subject's compensating production change must also leave path deviation unaltered.

Figures 3-8 and 3-9 illustrate these concepts. Each shows a magnitude 2.0 feedback transform and a subject's compensation for it: Figure 3-8 shows a +2.0 transform; Figure 3-9 shows a -2.0 transform. For reasons discussed below, these particular transforms were used repeatedly throughout the experiments. As a result, they warrant further discussion here:



(a) action of the transformation

(b) compensation

Figure 3-8: Action of the +2.0 feedback transformation and compensation for its perceived shift of F.

The +2.0 Feedback Transformation Figure 3-8(a) shows how the +2.0 feedback transformation shifts what a subject hears: if he whispers vowel sound F-, he hears sound F; if he whispers F, he hears F+.

This action of the transformation is best explained in terms of path projections and deviations. The figure shows that F-, F, and F+ have path projections P-, P, and P+, respectively. All have the same path deviation D. The transformation increases by 2.0 the path projection of all vowel sounds, but leaves their path deviations unaltered:

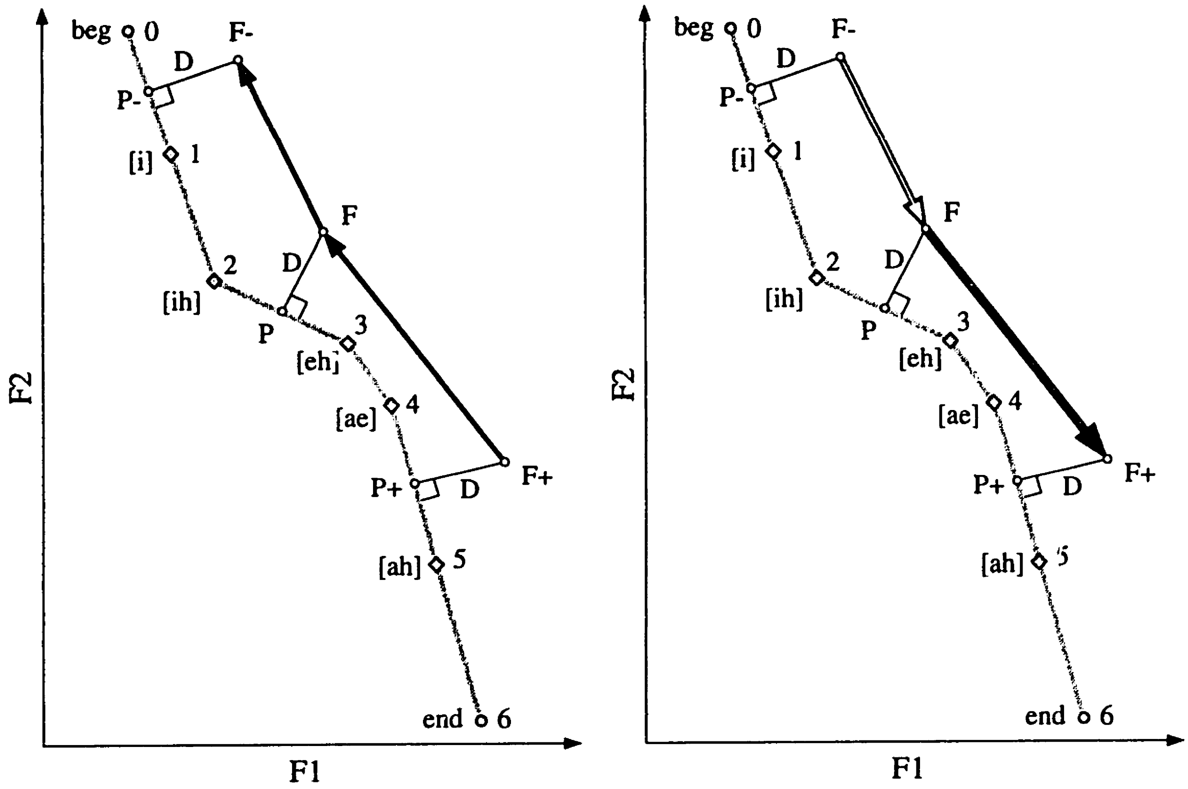
- Suppose the subject whispers F-. F- has a path projection of 0.5 (point P-) and a path deviation D. The transformation adds 2.0 to P- and leaves D unchanged. Thus, the subject hears F- shifted to a point with path projection 2.5 (point P) and the same path deviation D – this is point F.

- Suppose the subject whispers F. F has a path projection of 2.5 (point P) and a path deviation D. The transformation adds 2.0 to P and leaves D unchanged. Thus, the subject hears F shifted to a point with path projection 4.5 (point P+) and the same path deviation D – this is point F+.

From this description, it is easy to see how a subject compensates for the transformation. Figure 3-8(b) shows complete compensation for the transformation's perceived effect on F. In this figure, the large dark arrow shows the compensating production change and the hollow arrow shows how the subject hears this production change.

Normally, if a subject wishes to hear F, he whispers F. However, if he does this with feedback altered by the +2.0 transformation, he hears F+ instead. To hear F in this case, he must shift his whispering from F to F- (dark arrow): doing so shifts the sound he hears from F+ back to F (hollow arrow).

This compensation is more precisely described in terms of path projections. The transformation adds 2.0 to the path projection of the subject's vowel sounds, but does not alter their path deviations. To compensate, the subject must subtract 2.0 from his vowel sound path projections without changing their path deviations. F has a path projection of 2.5 and a path deviation D. Subtracting 2.0 from F's path projection without changing F's path deviation shifts F to F-.



(a) action of the transformation

(b) compensation

Figure 3-9: Action of the -2.0 feedback transformation and compensation for its perceived shift of F.

The -2.0 Feedback Transformation Figure 3-9(a) shows how the -2.0 feedback transformation shifts what a subject hears: if he whispers vowel sound F+, he hears sound F; if he whispers F, he hears F-. This is just the reverse of the previous situation.

The transformation subtracts 2.0 from the path projections of all vowel sounds, but leaves their path deviations unaltered:

- Suppose the subject whispers F+. F+ has a path projection of 4.5 (point P+) and a path deviation D. The transformation subtracts 2.0 from P+ and leaves D unchanged. Thus, the subject hears F+ shifted to a point with path projection 2.5 (point P) and the same path deviation D – this is point F.

- Suppose the subject whispers F. F has a path projection of 2.5 (point P) and a path deviation D. The transformation subtracts 2.0 from P and leaves D unchanged. Thus, the subject hears F shifted to a point with path projection 0.5 (point P-) and the same path deviation D – this is point F-.

Figure 3-9(b) shows a subject's complete compensation for the transformation's perceived effect on F. To hear F in this case, he shifts his whispering from F to F+ (dark arrow): this shifts the sound he hears from F- back to F (hollow arrow).

In terms of path projections, the subject compensates by adding 2.0 to his vowel sound path projections without changing their path deviations. Adding 2.0 to F's path projection without its path deviation shifts F to F+.

3.3.2 General Construction Procedure

As described in Section 3.2, a feedback transformation was stored as a table with a finite number of transformation pairs of the form:

$$((F1,F2,F3)_{\text{original}}, (F1,F2,F3)_{\text{transformed}})$$

Using the above feedback transformation definition, a feedback transformation table for a subject was constructed as follows:

1. Define the subject's [i]–[ɑ] path in (F1,F2,F3) formant space.

2. Specify a region R around the path large enough to include the subject's normal path vowel production variations.

3. Specify the desired amount of path projection shift for the feedback transformation.

4. Discretize each formant's frequency range to 64 values between 0 and 4KHz. Doing this matches the frequency resolution of the DSP spectral analysis (described in Section 3.2).

5. For every (F1,F2,F3) combination in region R:

(a) Store it as $(F1,F2,F3)_{\text{original}}$.

(b) Determine its (path projection, path deviation) representation.

(c) Shift the path projection by the amount specified for the transformation.

(d) Convert the resulting (path projection, path deviation) representation back into an (F1,F2,F3) combination.

(e) Store the result as $(F1,F2,F3)_{\text{transformed}}$.

3.3.3 Transformations Used

In an effort to elicit detectable compensating responses, subjects were exposed to large feedback transformations in the experiments.

In order to do this, the experiments restricted subjects to hearing the effect of the feedback transformation only when they were producing [ɛ]. Because of [ɛ]'s central [i]-[ɑ] path position, large feedback transformations of ±2.0 vowel units could be applied to [ɛ]without the perceived result extending beyond the range of path vowels:

- A +2.0 transform makes [ɛ] sound like [ɑ], which a subject can compensate for by producing [i] (as shown in Figure 3-6).

- A -2.0 transform makes [ɛ] sound like [i], which a subject can compensate for by producing [ɑ].

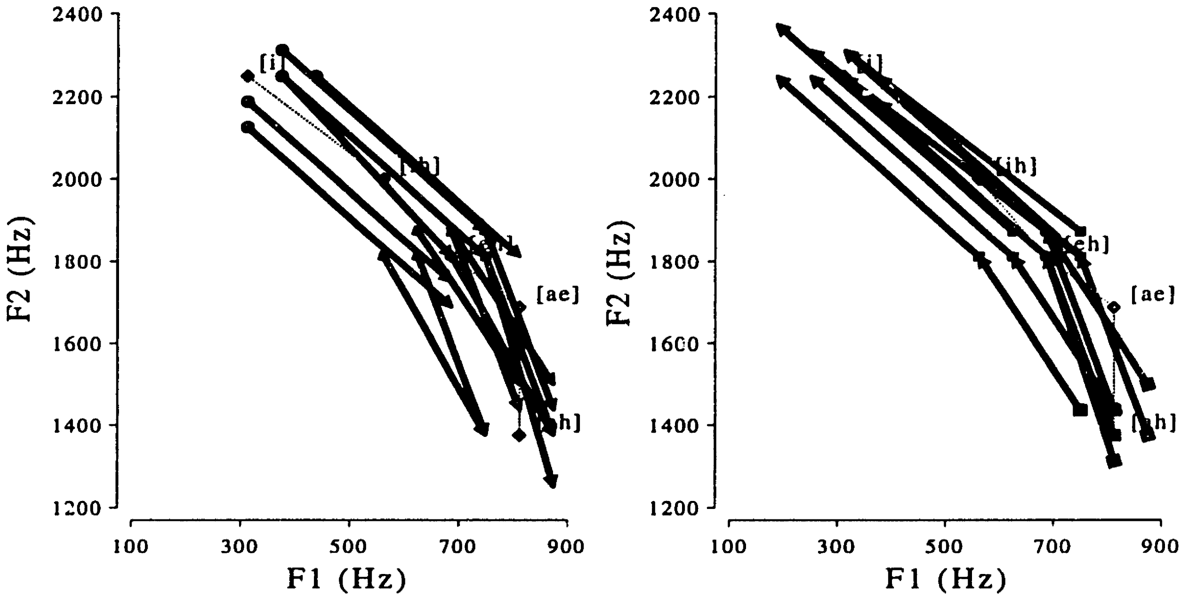
Thus, ±2.0 vowel unit transformations of [ɛ] always produced large, perceivable vowel sound differences that a subject could, in theory, completely compensate for.

Since these transformations were so commonly used, they were given names:

- The +2.0 transformation was called the **2p feedback transformation**.

- The -2.0 transformation was called the **2m feedback transformation**.

The effects of these transformation on vowel sounds and the production changes necessary to compensate for them were shown in figures 3-8 and 3-9. Figure 3-10 shows the actions of these transforms on actual vowel formant data collected from a subject.



(a) 2p transformation

(b) 2m transformation

Figure 3-10: Effects of the 2p and 2m transformations on vowels produced by subject MB. In both (a) and (b), the arrow base represents the (F1,F2) formar t values of one of the subject's vowel productions. The arrow tip represents how the transformation shifted the perceived values of these formants. (a) shows the effect of the 2p transformation on his renditions of [i]and [ε], showing that the resulting perceived shift of [i]is to [ε], and that of [ε]is to [ɑ]. (b) shows the effect of the 2m transformation on his renditions of [ɑ]and [ɑ], showing that the resulting perceived shift of [ɑ]is to [ε], and that of [ε]is to [i].

3.4 Utterance Data Acquisition

As described at the end of chapter 2, each speech SA experiment had structure consisting of a sequence of at least three phases:

1. **A baseline phase**, in which the subject was visually prompted to whisper CVC utterances while the formant values of his utterances were recorded. For some of the utterances, he heard unaltered feedback of his whispering. For others, he was prevented from hearing his whispering by masking noise.

2. **A training phase**, in which the feedback transformation was introduced. In this phase, the subject was visually prompted to whisper a limited number of utterances while the feedback transformation altered the perceived formants of his whispering.

3. **A testing phase**, in which the subject was again visually prompted to whisper CVC utterances while the formant values of his utterances were recorded. For some of the utterances, he heard altered feedback of his whispering. For others, he was prevented from hearing his whispering by masking noise.

Thus, as experienced by the subject, each experiment was simply a series of promptings to whisper words under different feedback conditions. In this section, we explain this prompting process in more detail by describing the general script of how all the experiments were conducted.

3.4.1 Initial Setup

All experiments began with the subject sitting in front of the PC's video monitor, putting on the insert earphones, and having the head-mounted noise canceling microphone put on his head. This was followed by a test of the two modes of feedback he heard throughout the experiment:

- *Mixed feedback*: a mixture of mild (40 dB SPL) masking noise (that impeded hearing of bone conducted whispering), and, at a louder level (≈50dB SPL), synthesized feedback of his whispering.

- *Noise feedback*: 60 dB SPL masking noise that completely prevented him from hearing his whispering.

The first feedback mode test was prompted on the video screen by the phrase:

```
mixed feedback test
```

Simultaneous to this, the mixed feedback mode of the DSP system was turned on, allowing the subject to hear feedback of his whispering. The formant frequencies of the synthesized feedback were, at this point, unaltered. The subject was then instructed to whisper the words “beed’, “bid”, “bed”, “bad”, and “bod”, while both he and the experimenter listened to the DSP output. This had two purposes:

1. It allowed the experimenter to determine if the subject is whispering loud enough so the DSP could properly process the whispered speech.

2. It allowed both experimenter and subject to verify that the synthesized feedback was a sufficiently good reproduction of the subject’s whispered speech.

The DSP remained in mixed feedback mode until the experimenter clicked any of the buttons on the PC’s mouse, at which point the second feedback mode test began.

The second feedback mode test was prompted on the video screen by the phrase:

```
noise feedback test
```

Simultaneous to this, the noise feedback mode of the DSP system was turned on, blocking the subject’s hearing. The subject was instructed to again whisper the words ‘beed’, “bid”, “bed”, “bad”, and “bod”. He was also instructed to pay attention to whether he could hear his own whispered speech. The experimenter then switched off the DSP output amplifier and asked the subject if he could hear his whispered speech. If he said yes, the positioning of the insert earphones and the output volume of the DSP output amplifier were checked, the DSP output amplifier switched back on, and the test was repeated. With proper earphone positioning, a masking noise amplitude of 60 dB SPL (still rated by all subjects as not uncomfortably loud) was sufficient to prevent all subjects from hearing their whispered voices. Another click by the experimenter ended this test, turned off the noise output from the DSP, and prompted the subject with displayed phrase:

```
ready to start?
```

At which point the experiment halted until the subject was ready to proceed, which he signaled by pressing any mouse button.

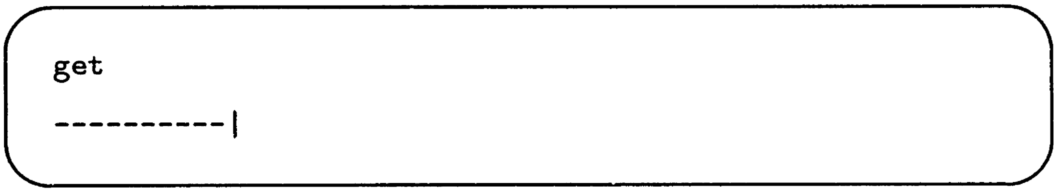
3.4.2 Data Acquisition

Past this point, the experiment was a series of utterance data acquisition *epochs*. Each epoch began with an epoch progress report: a display of how many more epochs there were in the experiment. A typical display was:

```
-------------------------- 170
```

indicating 170 epochs left to go. The experiment also halted at these points to allow the subject to take a brief break and/or drink water provided. The experiment re-commenced when the subject hit any mouse button, at which point the words to be pronounced in that epoch were prompted for.

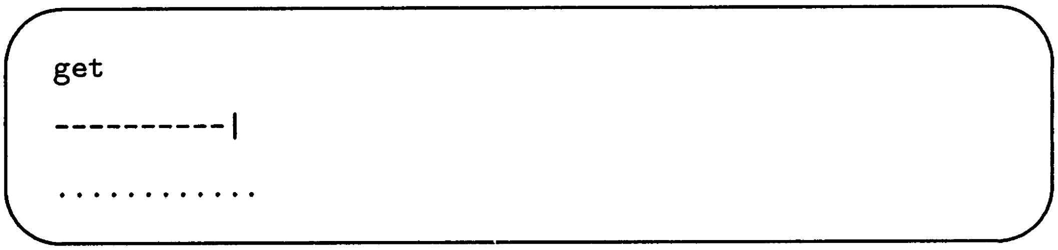
Each word prompting prompted for a desired word and a desired utterance duration. A typical prompting was:



```
get
----------|
```

where the line terminated by the horizontal bar indicated the desired duration (in this case about 500ms).

Simultaneous to this prompting, the DSP output was in either mixed or noise feedback mode. When the subject began whispering the word, an input amplitude threshold was exceeded, causing the DSP system to begin recording the utterance data, and also causing dots to be printed on the screen. These dots continued to be printed as long as the subject continued whispering. Subjects were instructed to whisper sufficiently long to make the dots end approximately where the horizontal line was. A typical display after a subject responded to the above shown prompting was:



```
get
----------|
............
```

Subjects were encouraged to exceed the horizontal mark, rather than undershoot it. The subject's completed whispering of one word triggered the prompting of the next word. After 10–20 prompted words, the end of the epoch was reached. At the epoch's end, a progress report was displayed, and the experiment would again wait for a mouse button push to continue to the next epoch. This process continued until all epochs of the experiment were completed.

3.5 Utterance Data Analysis

All the experiments were designed to investigate SA in the production of vowels in CVC words. For a specific vowel in a specific CVC word, this involved repeatedly prompting a subject to whisper that word in each experiment phase.

The data analysis averaged a word's repeated productions in an experiment phase to determine its mean production in that phase. Mean productions of the word in different phases were then compared to exhibit overall production changes. Overall production changes in the word's vowel portion were then assessed for compensation: i.e., how much they compensated for the feedback transformation effects.

In this section, this analysis process is described in more detail. In brief, it was a sequence of the following analyses:

1. Formant timecourse analysis using *avgrams.* Avgrams were spectrogram-like plots of a word's mean formant timecourses. These plots exhibited a word's mean production in an experiment phase. Comparing avgrams of the same word from different phases exhibited any overall changes in the word's production.

2. Vowel formant analysis using *vowel plots.* These plots showed, in formant space, (F1,F2) of a word's vowel portion, averaged over an experiment phase. Plots from different phases were combined to exhibit vowel production change in relation to a subject's [i]–[ɑ] path. This provided a visual assessment of how much the vowel production change compensated for the feedback transformation.

3. Vowel *(path projection, deviation)* analysis. The feedback transformations altered perceived path projections (but not deviations) of a subject's formants. Thus, resolving a vowel's production change into its path projection and deviation components allowed compensation to be quantified. The path projection component measured the amount of production change directly counteracting (or worsening) the feedback alteration. The path deviation component measured the amount of production change having no effect on the feedback alteration.

These three analysis steps are explained in more detail in the following sections.

3.5.1 Formant Timecourse Analysis: Avgrams

Formant timecourse analysis was done to exhibit how a subject's entire production of a word (or set of words) was affected by exposure to altered feedback.[5] This analysis was based on comparing *avgrams* of the word in different experiment phases. An avgram is a way of looking at a word's production, averaged over an experiment phase. It shows in spectrogram-form the word's mean formant timecourses.

3.5.1.1 Avgram Creation

Creating an avgram involved selecting an utterance data set, averaging formant frequency timecourses of the set, and plotting the results.

Selecting the Utterance Data Set The utterance data set was determined by which word and experiment phase were of interest. For example, an avgram of the word "get" in the baseline phase would use the utterance data records of all baseline-phase productions of "get".

Averaging Formant Frequency Timecourses Recall that formants were estimated for each 8ms frame of a subject's whispered speech. When a subject's utterance was recorded, the (F1,F2,F3,F4) estimates of each frame of the utterance were stored. An utterance's data record (which we will call an *utterance record*) was thus a sequence of these (F1,F2,F3,F4) estimates (which we will call *frames*[6]).

Since utterances have different durations, their data records have different *frame lengths* (numbers of frames). Thus, a special method was needed to average formant timecourses of an avgram's utterance data set.

[5] For the rest of these sections, the term "word" can also mean "set of words", with the distinction being made when important.

[6] Thus, depending on the context, the term "frame" means different things: in discussing signal processing, "frame" refers to a 64-sample chunk of speech data; in discussing data analysis, "frame" refers to the (F1,F2,F3,F4) estimate of that chunk of speech data.

The first step of this method was to find the frame length L of the minimum duration utterance record. The next step was to create separate *time bins* for frames 1 to L. Each time bin contained formant data frames from the same position in each utterance record. Thus:

- Frame 1's time bin contained frame 1 of all utterance records.
- Frame 2's time bin contained frame 2 of all utterance records.

$$\vdots$$

- Frame L's time bin contained frame L of all utterance records.

For each time bin, mean formant frequencies were then estimated from its formant data frames.[7]

Displaying the Results The time bins' mean formant estimates were then displayed in a spectrogram-like plot where:

- the horizontal axis represents time, and
- the vertical axis represents frequency.

In such a plot (see Figure 3-11), each vertical slice represents a time bin's mean formant frequency estimates. The horizontal position of the slice corresponds to the time represented by the time bin (i.e., 8ms times the the frame number of time bin). Within each slice, vertical positions of black pixels mark the time bin's mean formant frequencies.

If a formant frequency estimate doesn't change too quickly over successive time bins, its marked position over successive time slices forms a curve. This curve is called the "track" of the formant and represents the formant's timecourse.

In this way, the formant tracks in an avgram represent the mean formant timecourses of a word, averaged over an experiment phase. Avgrams of the same word in

[7] The process of estimating mean formants from a set of formant data frames is explained in detail in Section A.2.

different experiment phases are combined to show mean changes in the word's production (usually resulting from exposure to a feedback transform). An example of this is shown in Figure 3-11.

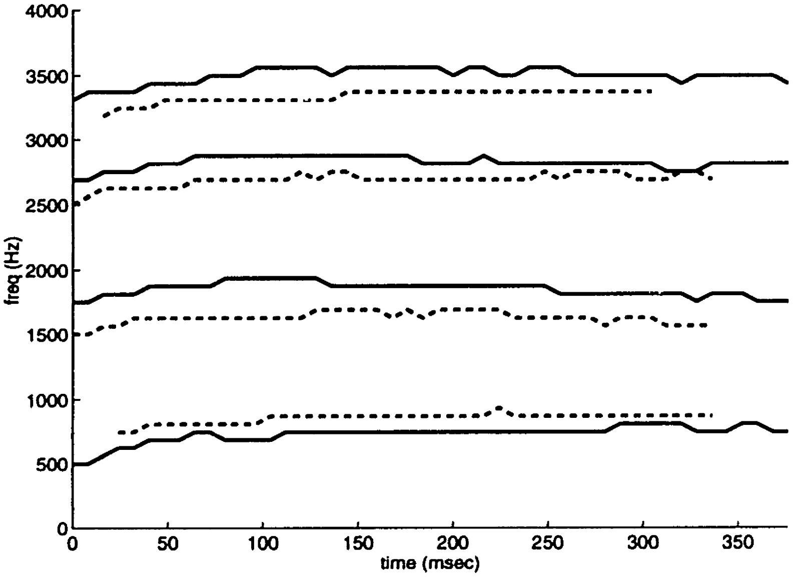


Figure 3-11: Avgrams of subject SR's production of the word "bep" in the baseline and test phases of Study 2 (to be discussed in full later). The baseline phase avgram is shown as solid lines, the testing phase avgram as dashed lines.

The plot shows all formant frequencies rise between 0 and about 80ms, but past this point maintain approximately steady-state values. This indicates transition from the initial [b] took about 80ms, which was followed by production of the steady-state vowel [ɛ] lasting about a quarter of a second.

The plot also shows how the subject has adjusted his production to compensate for a feedback transformation he was exposed to.

The strength of the feedback transformation (introduced between the baseline and testing phases) was -2.0, which pushes [ɛ] towards [i] effectively making F1 and F2 sound farther apart. The plot shows the subject compensates for this by bringing F1 and F2 closer together.

3.5.1.2 Time Alignment

An important limitation to the avgram plotting method concerns how a word's mean production is represented. Because of how they are constructed, avgrams are a good representation of a word's beginning, but a poorer representation of its end.

As described above, constructing a word's avgram involves essentially overlaying and averaging formant timecourses of each production of the word in an experiment phase. Because each production has a different duration, their data records can be aligned to either their starting points or their ending points, but not both. A time-stretching technique could have been used to normalize utterance duration and line up both ends, but it was unclear what technique would be appropriate to use.

A linear time stretch of utterances was considered, but this would stretch the formant transitions of the consonant as much as the vowel. Such time stretching misrepresents how speakers vary CVC utterance duration: normally, the consonant duration is kept constant and the duration of the steady-state vowel portion is varied [Peterson and Lehiste, 1960].

Some other variable stretching technique, such as Dynamic Time Warping [Rabiner and Levinson, 1981], could have been used, but it was decided instead to simply align the data records to their beginnings, and concentrate analysis on the word beginning. This was found to be sufficient for exhibiting formant changes due to altered feedback exposure.

3.5.1.3 Other Avgram Limitations

Another limitation of avgrams is that confidence intervals on the mean formant estimates are not shown. Adding confidence intervals around the formant tracks made avgrams overly cluttered and difficult to interpret. This limitation was offset by displaying confidence intervals in all other types of data plots.

A final avgram limitation is that the subject's response to a feedback transformation is not seen in formant space, which is where the transformation was defined. This was the motivation for the *vowel plots* described in the next section.

3.5.2 Vowel Formant Analysis: Vowel Plots

The experiments mainly focused on how feedback transformation exposure affected steady-state vowel productions. For this reason, vowel formant frequency changes were analyzed in a number of ways. These changes were quantified by path projection and deviation analysis, but they were visualized using *vowel plots.*

A vowel plot is a way of looking at production of the vowel portion of a word in an experiment phase. It was a plot of the vowel portion's mean F1 and F2 frequencies as a location in (F1,F2) formant space. This location was called the *position* of the vowel's production. Vowel plots from different experiment phases were combined to exhibit how a subject changed a vowel's production after exposure to a feedback transform.

Before discussing how vowel plots were made, their neglect of F3 is explained. Recall that the feedback transformations shifted F1, F2 *and* F3.

In whispered productions of the [i]–[ɑ] path vowels, F3 does exhibit some frequency change – as opposed to F4, which exhibits almost none (see Figure A-1). Thus, it made sense to include F3 in the formants shifted by the feedback transformation. However, it is unclear what role F3 plays in determining perceived path vowel identity. Thus, the vowel formant analyses were restricted to examining the frequencies of F1 and F2, whose role in determining vowel identity (as discussed in Chapter 2) is better understood.

3.5.2.1 Creating a Vowel Plot

The first step in creating a vowel plot was finding the the steady-state vowel portion of the selected word's mean production. This was done by determining where in the word's avgram the formants attained steady-state values. An example of this was discussed in Figure 3-11. Next, a time interval was selected that contained the steady-state vowel portion. This was called the *selected vowel interval* of the word. Mean F1 and F2 of this vowel interval was then determined and plotted in (F1,F2) formant space.

Determining Mean (F1,F2) of the Selected Vowel Interval To determine mean (F1,F2) of the word's selected vowel interval, separate formant estimates were made from this interval within each production of the word in an experiment phase. The resulting set of estimates were averaged. This was done in the following manner:

1. Mean formants were estimated for the selected vowel interval in each production of the word in the selected experiment phase. For each word production's utterance record:

   (a) The set of formant data frames corresponding to the selected vowel interval was identified.

   (b) Mean formants of these frames were estimated (again, using the the method discussed in Section A.2).

   In this way, a mean (F1,F2,F3,F4) estimate was calculated for each utterance record.

2. Statistics were calculated from the set of utterance record (F1,F2,F3,F4) estimates. These were:

   (a) Mean and Variance of F1.

   (b) Mean and Variance of F2.

   (c) Covariance of F1 and F2.

Plotting Mean (F1,F2) of the Selected Vowel Interval These statistics were then plotted in formant space in the following way (see Figure 3-12):

- A point F was plotted whose position represented mean (F1,F2).
- An ellipse E was plotted around this point that represented the standard error of the mean (F1,F2) statistic.
- For reference, the subject's [i]–[ɑ] path was included in the plot.

3.5.2.2 Combined Vowel Plots

In this way, a vowel plot represents in formant space both the mean production of a vowel and its production variations. Vowel plots from different experiment phases were combined to exhibit mean vowel production changes. Two types of combined vowel plots were used:

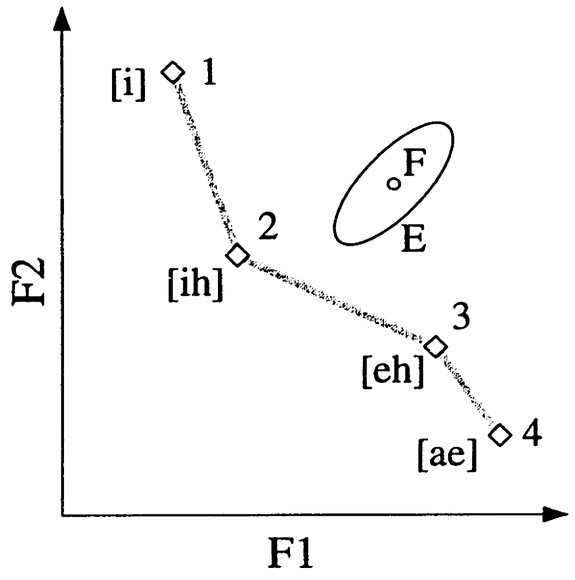
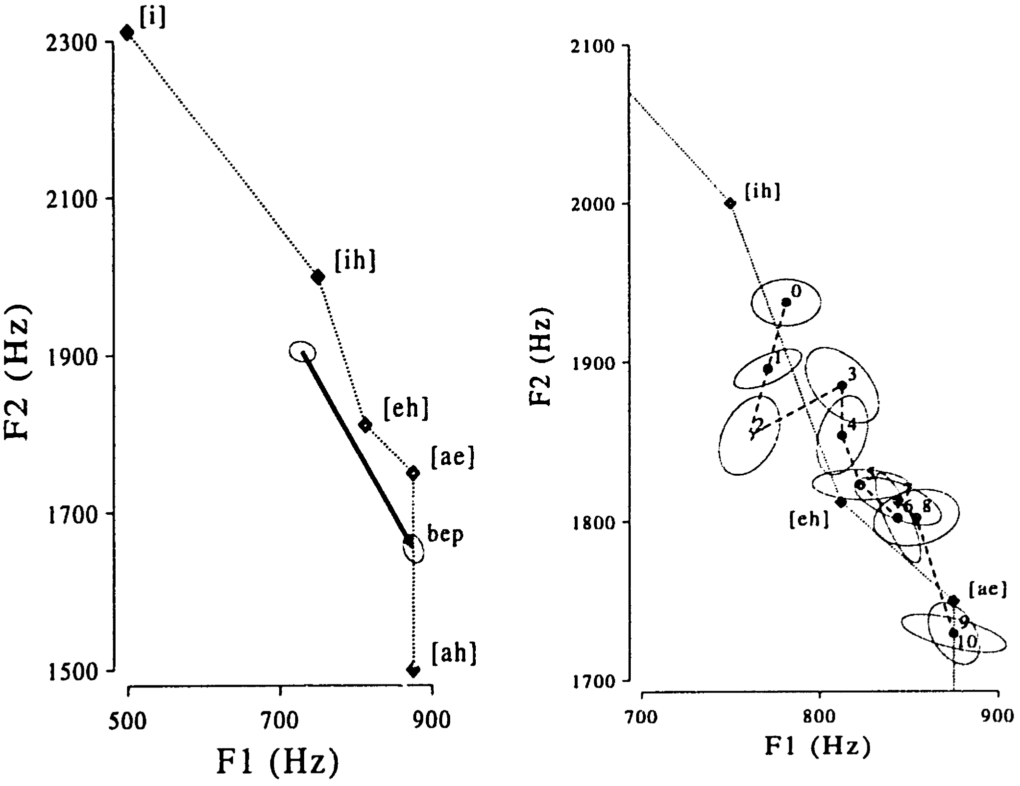


Figure 3-12: A vowel plot. Position F represents mean (F1,F2) of a selected vowel interval of a word in an experiment phase. The ellipse E around it represents the standard error of this mean (F1,F2) estimate. (The ellipse axes are the eigenvectors of the (F1,F2) covariance matrix.) For reference, a limited portion of the subject's [i]-[ɑ] is also plotted (gray line in the figure).

1. **Vowel difference plots**, which displayed the change in a vowel's mean production between two experiment phases (usually the baseline and testing phases). This mean change was shown as an arrow between the vowel's position plots in the two phases.

2. **Vowel sequence plots**, which displayed a sequence of a vowel's production changes over more that two experiment phases. This mean change sequence was shown as a line linking the vowel's position plots from successive experiment phases.

For both of these plots, a plot of the subject's [i]-[ɑ] path was also included. Examples of these plots are shown in Figure 3-13.



(a) vowel difference plot

(b) vowel sequence plot

Figure 3-13: Vowel plots of subject SR's production changes for [ɛ] in "bep" during study 2 (to be discussed later).

(a) shows a vowel difference plot of the change the subject's production of [ɛ] between the baseline and testing phases. The arrow base represents mean (F1,F2) of the baseline phase production, while the arrow tip represents mean (F1,F2) of the testing phase production. The dotted line is the subject's [i]–[ɑ] path. The plot indicates significant compensation for the 2m feedback transformation used in the experiment (compare this plot with Figure 3-9(b)).

(b) shows a vowel sequence plot of how the subject's production of [ɛ] changed over the 10 stages of the training phase. The strength of the feedback transformation was linearly changed from 0.0 to -2.0 over these 10 stages. The vowel plots of [ɛ] in each stage are labeled with stage numbers and linked with a dashed line. The progression of mean (F1,F2) positions shows the subject roughly compensated for each increase in transformation strength.

In both plots, the ellipses indicate the standard error confidence intervals around the mean measurements.

3.5.3 Path Projection and Path Deviation Analysis

The above-described plotting methods were important for visualizing a subject's response to a feedback transformation. However, they allowed only indirect, approximate estimates of how much the subject compensated for the feedback transformation. Since the feedback transformations specifically altered path projection, quantifying a subject's compensation involved analyzing path projection and deviation of a subject's vowel productions.

The first step in doing this was converting the vowel plot fo:mant data into (path projection, deviation) data.

3.5.3.1 Vowel (Path Projection, Deviation) Data

As described above, a vowel plot showed mean (F1,F2) of a selected vowel interval of a word in an experiment phase. The first step is its creation was generation of a set of utterance record (F1,F2,F3,F4) estimates. This was done by:

1. Identifying all utterance records of the word from that experiment phase.
2. Estimating (F1,F2,F3,F4) for the selected vowel interval in each utterance record.

To make the vowel plot, (F1,F2) statistics were then calculated from the set of utterance record (F1,F2,F3,F4) estimates.

To compute (path projection, deviation) data for this selected vowel interval, the set of utterance record (F1,F2,F3,F4) estimates was converted into a set of utterance record (path projection, deviation) estimates. Each (F1,F2,F3,F4) estimate was converted by computing path projection and deviation of F1 and F2.[8] This computation was similar to that shown in Figure 3-7(a). However, in this case, a smoother,

[8] The computation of path projection and deviation was restricted to (F1,F2) space for two reasons. First, for reasons explained in the last section, all vowel formant analysis considered only F1 and F2. Second, Because, (as explained in Figure 3-7(a)) in (F1,F2) space, path deviation is a scalar. Scalar representations of both path projection and deviation facilitated cross-subject comparisons of these quantities.

spline-curve version of the [i]–[ɑ] path was used.[9]

Via this process, each word production's vowel interval formant estimates were converted into vowel interval (path projection, deviation) estimates. The set of these estimates constituted the vowel interval's (path projection, deviation) data for the experiment phase.

(path projection, deviation) data for the same word's vowel in different experiment phases were then compared to measure how much a subject compensated for a feedback transformation. Several types of such comparisons were computed.

3.5.3.2 ANOVAs

ANOVAs determined statistical significance of the effect of experiment phase (baseline vs. testing) on a vowel's production. This was done by computing separate ANOVAs on the path piojection and on the path deviation data for a vowel.

3.5.3.3 Mean Path Projection and Path Deviation Change

Mean path projection for a vowel in an experiment phase was computed as mean of the vowel's (path projection, deviation) data for that experiment phase. Mean path deviation was computed similarly.

Changes in these mean estimates between two experiment phases (usually baseline and testing) quantified the amount a subject compensated for a feedback transformation. The feedback transformation definition in Section 3.3.1.3 specified that, to compensate, a subject must shift a vowel's path projection. However, to avoid introducing additional distortion, he must not alter the vowel's path deviation. Thus, by measuring a vowel's mean path projection and deviation change, how well a subject compensated could be quantified as follows:

- *Mean path projection change* was the difference in mean path projection of a vowel produced in two different experiment phases (usually testing phase minus

[9] The feedback transformation tables were constructed using a line-segment [i]–[ɑ] path definition. The advantages of instead using a spline-based [i]–[ɑ] path definition are discussed in Appendix C.

baseline phase). It measured how much a subject directly compensated for a feedback transformation's shift of perceived path projection.

- *Mean path deviation change* was the difference in mean path deviation of a vowel produced in two different experiment phases (usually testing phase minus baseline phase). It measured how much non-compensating additional distortion a subject added to a vowel's production.

3.5.3.4 Mean Compensation

To derive a number representing what fraction of a feedback transformation's perceived effect a subject compensated for, a normalized version of mean path projection change was computed. This was called *mean compensation*, and was computed as:

$$\text{mean compensation} = \frac{(\text{mean path projection change})}{-\left(\begin{array}{c}\text{path projection shift of the}\\ \text{feedback transformation}\end{array}\right)}$$

This formula is based on the fact that, for complete compensation, a subject must shift his vowel path projection by an amount equal to the feedback transformation magnitude, but opposite in direction. Its range of values represent the degree to which a subject compensates for the feedback transformation he was exposed to. If a subject's mean compensation is M:

- $M > 1.0$ means he over-compensated
- $M = 1.0$ means he completely compensated
- $0.0 < M < 1.0$ mean he partially compensated
- $M = 0.0$ means he showed no compensating change
- $M < 0.0$ means he showed anti-compensating change

Consider, for example, figures 3-8(a) and 3-9(a) above. Both figures show examples of complete compensation. In both cases, mean compensation is 1.0 because:

- In Figure 3-8(a), the path projection change is -2.0 vowel units in response to a $+2\ 0$ feedback transformation.
  This is a mean compensation of $(-2.0)/(-(+2.0)) = 1.0$.

- In Figure 3-9(a), the path projection change is $+2.0$ vowel units in response to a -2.0 feedback transformation.
  This is a mean compensation of $(+2.0)/(-(-2.0)) = 1.0$.

A final point about mean compensation is that its name depends on what response data of a subject it's computed from:

- If it's computed from his *compensation response* data – from words productions made when the subject could hear feedback of his whispering[10] – it's still called *mean compensation.*

- If it's computed from his *adaptation response* data – word productions made when the subject was prevented from hearing his whispering by masking noise – it's called *mean adaptation.*

In this way, mean adaptation measures how much of a subject's compensating production changes are retained in productions made when he can't hear himself. This conforms to the definition of adaptation specified in the introductory chapter.

3.6 Chapter Summary

In this chapter, the apparatus, procedures, and data analysis methods used in these experiments were discussed.

First, an overview of the setup used in the experiments was described, the key element of which was the DSP system that transformed the subject's acoustic feedback. The transformations were based on pre-computed tables of formant shifts of the subject's vowel sounds.

[10] The definitions of compensation and adaptation responses were given in Section 2.3.

Next, the method of constructing these feedback transformation tables was described. Here, feedback transformation were defined based on the [i]–[ɑ] path shift concept described in the previous chapter. They were defined as shifts of perceived *path projection* of a subject's vowel sounds.

Following this, the process of acquiring subjects' utterance data in the experiments was described. The description showed that, from the subject's point of view, each experiment was simply a series of visual promptings to whisper words under different feedback conditions.

Finally, the methods of analyzing a subject's utterance data from these experiments were described. These analysis methods included:

- **Formant timecourse analysis** based on *Avgrams.* Avgrams plotted, in spectrogram-like form, the mean formant timecourses of all productions of a word within an experiment phase.
- **Vowel formant analysis** based on *vowel plots.* A vowel plot was an (F1,F2) formant space plot of mean (F1,F2) of a selected vowel interval of all productions of a word within an experiment phase.
- **Vowel path projection and path deviation analysis**, based on a vowel's (path projection, deviation) data from different experiment phases. These analyses involved computing:
  - ANOVAs testing the effect of experiment phase on path projection and path deviation.
  - *Mean path projection change* between two experiment phases, to measure how much a subject's vowel production change compensated for a feedback transformation.
  - *Mean path deviation change* between two experiment phases, to measured how much non-compensating distortion a subject added to a vowel's production.

- *Mean compensation*, to measure what fraction of a feedback transformation's perceived effect a subject compensated for.
- *Mean adaptation*, to measure how much compensatory production change the subject retained when he was prevented from hearing his whispering.

The studies based on the apparatus, procedures, and data analysis methods discussed here are the subject of the next two chapters.

Chapter 4

Study 1: Existence of Speech Sensorimotor Adaptation

The purpose of Study 1 was to determine if the production of vowels exhibited the predicted sensorimotor adaptation (SA) effect. The study examined how the production of [ɛ] was affected by exposure to the -2.0 feedback transformation.

4.1 Introduction

As discussed in chapters 1 and 2, the principal questions concerning the effect of altered feedback on a subject's vowel production are:

1. Does he *compensate*? Does he adjust his vowel productions to compensate for the perceived formant shift produced by the feedback transformation?
2. Does he *adapt*? Does he retain his adjusted vowel productions, even when denied acoustic feedback?

As described in Chapter 2, an SA experiment that addresses these questions consists of the following three phases:

1. A baseline phase, in which a subject's vowel productions are measured with and without acoustic feedback. The feedback is unaltered.

2. A training phase, in which a feedback transformation is introduced that alters the subject's acoustic feedback. The subject is given experience producing vowels while hearing the altered feedback.

3. A testing phase, in which a subject's vowel productions are again measured with and without acoustic feedback . The feedback that the subject does hear is still altered.

By comparing the vowel formant frequencies produced in the baseline and testing phases, vowel production change due to exposure to the feedback transformation can be assessed. Formant changes in vowels produced while the subject hears feedback (his *compensation response*) allows us to determine compensation. Formant changes in vowels produced while the subject hears only masking noise (his *adaptation response*) measure the retention of compensation – what we have called adaptation.

Compensation and adaptation responses were analyzed using the techniques described in Section 3.5, and summarized here:

- *Avgrams* were used to examine changes in formant timecourses and to identify the steady-state vowel portions of the utterances.

- *Vowel plots* were used to examine changes in mean (F1,F2) within the steady-state vowel portions.

- *Path projection* and *path deviation* values for these steady-state vowel portions were computed. These provided scalar measures of how much subjects compensated for the feedback transformation. These measures were then compared across subjects.

Study 1 also examined subjects' capacity to compensate and adapt. Studies of reaching SA have found that mean adaptation generally reaches a limit of about 30% [Held, 1996]. To determine if vowel SA shows similar limitations, an additional training and testing phase were added to the experimental design. The principle phases of the experiment therefore were:

1. A baseline phase.
2. A first training phase.
3. A first testing phase.
4. A second training phase.
5. A second testing phase.

To summarize, study 1 sought to answer the following questions:

- Do subjects adjust their vowel productions to compensate for altered acoustic feedback?
- Do subjects retain this production adjustment when subsequently prevented from hearing feedback?
- Do subjects have a limited capacity to compensate and adapt?

4.2 Methods

Each subject was run in a single experimental session consisting of:

1. Measurement of the formant frequencies of his path vowels ([i], [ɩ], [ɛ], [æ], and [ɑ]).
2. Generation of a -2.0 formant transformation table based on these measurements.
3. Exposure, in an SA experiment, to the feedback transformation based on this table.

Steps 1 and 2 took 30 minutes. Running the SA experiment (step 3) required an additional 30 minutes. The procedures for measuring path vowel formants and generating the -2.0 formant transformation table were described in the previous chapter. Here, it is important only to recall that the -2.0 feedback transformation shifts perceived [i]–[ɑ] path position 2.0 vowel units towards [i].

The experiment assessed how exposure to this feedback transformation affected the subject's production of the vowel [ɛ]. The complete experimental session consisted of 48 epochs, each consisting of 10–20 word promptings. Each prompted word was chosen randomly from the set:

$$\mathbf{W}_{\mathbf{train}} = \left\{ \begin{array}{llll} \text{“bed”}, & \text{“bet”}, & \text{“red”}, & \text{“head”}, \\ \text{“med”}, & \text{“met”}, & \text{“ned”}, & \text{“net”}, \\ \text{“dead”}, & \text{“debt”}, & \text{“led”}, & \text{“let”} \end{array} \right\}$$

For each word prompting, the subject whispered the displayed word while the DSP was in one of two feedback modes:

- *Mixed feedback mode*, in which the subject was provided a mixture of mild masking noise and, at a louder level, synthesized feedback of his whispering.
- *Noise feedback mode*, in which the subject was provided masking noise that completely blocked his hearing of his whispering.

The 48 epochs of the experiment were divided into the following phases. These were:

1. **A 5 epoch warmup phase.** In each of these epochs, the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while hearing feedback (mixed feedback mode). This feedback was unaltered and no utterance data were recorded.
2. **A 1 epoch baseline phase.** This epoch consisted of a sequence of two periods:
   - *base1m*: in which the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while hearing feedback (mixed feedback mode). All utterance data were recorded.
   - *base1n*: in which the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while his hearing was blocked by noise (noise feedback mode). All utterance data were recorded.

   At this point, the -2.0 feedback transformation table was loaded into the DSP, so that all subsequent feedback was altered.

3. **A 20 epoch first training phase.** In these epochs, the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while hearing the altered feedback (mixed feedback mode). No utterance data were recorded.

4. **A 1 epoch first testing phase.** This epoch consisted of a sequence of two periods:

   - *test1m*: in which the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while hearing altered feedback (mixed feedback mode). All utterance data were recorded.
   - *test1n*: in which the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while his hearing was blocked by noise (noise feedback mode). All utterance data were recorded.

5. **A 20 epoch second training phase.** In these epochs, the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while hearing the altered feedback (mixed feedback mode). No utterance data were recorded.

6. **A 1 epoch second testing phase.** This epoch consisted of a sequence of two periods:

   - *test2m*: in which the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while hearing altered feedback (mixed feedback mode). All utterance data were recorded.
   - *test2n*: in which the subject whispered 10 $\mathbf{W}_{\mathbf{train}}$ words while his hearing was blocked by noise (noise feedback mode). All utterance data were recorded.

4.3 Subjects

The subjects were 14 MIT undergraduate male native speakers of North American English who were naive to the purpose of the study. Five of the subjects were subsequently excluded from the analysis of results for the following reasons:

- Subject EM was excluded because, unlike the other subjects, his path vowels formants were not measured on the same day in which the SA experiment was run.

- Subject JO was excluded because he diphthongized his production of several of the $\mathbf{W}_{\mathbf{train}}$ words.

- Subject KL was excluded because he had trouble maintaining a consistent volume of his whispering.

- Subjects WW and AC were excluded because the second testing phase of the experiment was not run on them.

4.4 Results and Discussion

Each subject's results were analyzed using the following two methods:

1. *Compensation analysis*, which analyzed production changes seen during conditions in which the subject heard feedback of his whispering (the *base1m*, *test1m*, and *test2m* periods of the experiment). These production changes are called the subject's *compensation response* to the altered feedback.

2. *Adaptation analysis*, which analyzed production changes seen during conditions in which the subject was prevented from hearing his whispering by masking noise (the *base1n*, *test1n*, and *test2n* periods of the experiment). These production changes are called the subject's *adaptation response* to the altered feedback.

Both analyses involved examining word and vowel production changes using the methods described in Section 3.5.

Word production changes were examined using an *avgram* of the subject's mean utterance formant tracks. From the avgram, a time interval was determined that contained the mean utterance's steady-state vowel portion. Data from this time interval in each utterance were used to analyze vowel production changes.

Vowel production changes were examined by first making a *vowel plot* of the subject's mean vowel formant changes. These formant changes were quantified by computing *mean compensation* and *mean path deviation change*. These two quantities were then used in across-subject assessments of vowel production change.

In the next section, we discuss plots of the individual subject results. This is followed in Section 4.4.2 by an examination of the results seen across subjects.

4.4.1 Individual Subject Results

Figures 4-1 through 4-9 show the avgram and vowel plots used in the analysis of each subject's results. Each figure shows the analysis plots for a single subject. In each figure, the left side labeled "(a) compensation" shows the compensation analysis plots, while the right side labeled "(b) adaptation" shows the adaptation analysis plots.

In the avgrams, mean utterance formant tracks from different experiment phases are shown in different line styles:

- Baseline phase formants are shown as solid lines.
- First testing phase formants are shown as dashed lines.
- Second testing phase formants are shown as dotted lines.

The gray region in each avgram indicates the time interval used to analyze vowel production changes. As suggested by the brace and downward arrow, utterance data from this time interval was processed into the vowel plot shown below the avgram.

Within the vowel plots, mean (F1,F2) values in each testing phase are compared with mean (F1,F2) values in the baseline phase:

- The first testing phase comparison is shown as the arrow labeled "test1".
- The second testing phase comparison is shown as the arrow labeled "test2".

In the following sections, the figures showing the subject results are discussed. The first five sections discuss (in order of decreasing adaptation) the five subjects showing the most adaptation. The last section discusses, as a group, the remaining four subjects who showed the least adaptation.

4.4.1.1 Subject MB Results

Figure 4-1 shows plots of subject MB's compensation and adaptation responses to exposure to the -2.0 feedback transformation. MB's large responses provide good illustrations of how the measures of path projection, path deviation, and mean compensation are derived. For this reason, his results will be described in greater detail than those of other subjects.

Compensation Analysis Consider first the compensation plots (box (a) in the figure). The avgram has three clear groups of formant tracks, showing that the amplitudes of F1, F2, and F3 were above plotting threshold for each of the experiment phases analyzed (the baseline, first testing, and second testing phases).

All three formants appear to have attained steady-state values. For F1, there is little difference between its value in the baseline phase (solid line) and in the first and second testing phases (dashed and dotted lines, respectively). In contrast to this, F2 values are consistently lower in the testing phases than in the baseline phase. On the other hand, F3 values are consistently higher in the testing phases than in the baseline phase.

The gray region of the avgram shows the time interval used to analyze vowel production changes. As suggested by the brace and downward arrow, utterance data from this time interval was processed into the vowel plot shown below the avgram.

The vowel plot shows MB's F2 response more clearly. The arrows show change in mean (F1,F2) of his productions of the vowel [ɛ]. The position of the arrows' common base indicates mean (F1,F2) of his vowel productions in the baseline phase. As indicated by the labeling ("testl" and "test2"), the positions of the arrow tips indicate mean (F1,F2) of his vowel productions in the first and second testing phases, respectively. Both arrows show changes in mean (F1,F2) that are almost completely in the negative F2 direction. The stability of these production changes are shown by the small standard error ellipses around the mean (F1,F2) positions for each experiment phase.

The vowel plot also shows MB's [i]–[ɑ] path. This allows graphical estimation of the changes in mean path projection and mean path deviation corresponding to the vowel production changes.

The plot shows that mean (F1,F2) values of his baseline vowel productions were quite close to [ɛ]. This point has minimal path deviation and projects to a position on the [i]–[ɑ] path that is close to [ɛ]. Thus, since [ɛ] has a path position of 3.0, MB's baseline vowel productions have a mean path projection of about 3.0 and a mean path deviation close to 0 Hz.

On the other hand, mean (F1,F2) of his first and second testing phase productions, respectively, are two points deviating by about 100 Hz below his [i]–[ɑ] path. Both points project to path positions near [æ]. Since [æ] has a path position of 4.0, MB's testing phase vowel productions both have mean path projections of about 4.0 and mean path deviations of about −100 Hz.

Thus, in averaging MB's testing phase results, we see he has changed mean path projection of his vowel productions by about +1.0 vowel unit and changed mean path deviation by about −100 Hz. This means he has compensated for about half of the -2.0 vowel unit path projection shift produced by the feedback transformation. We therefore estimate his mean compensation to be about 0.5.

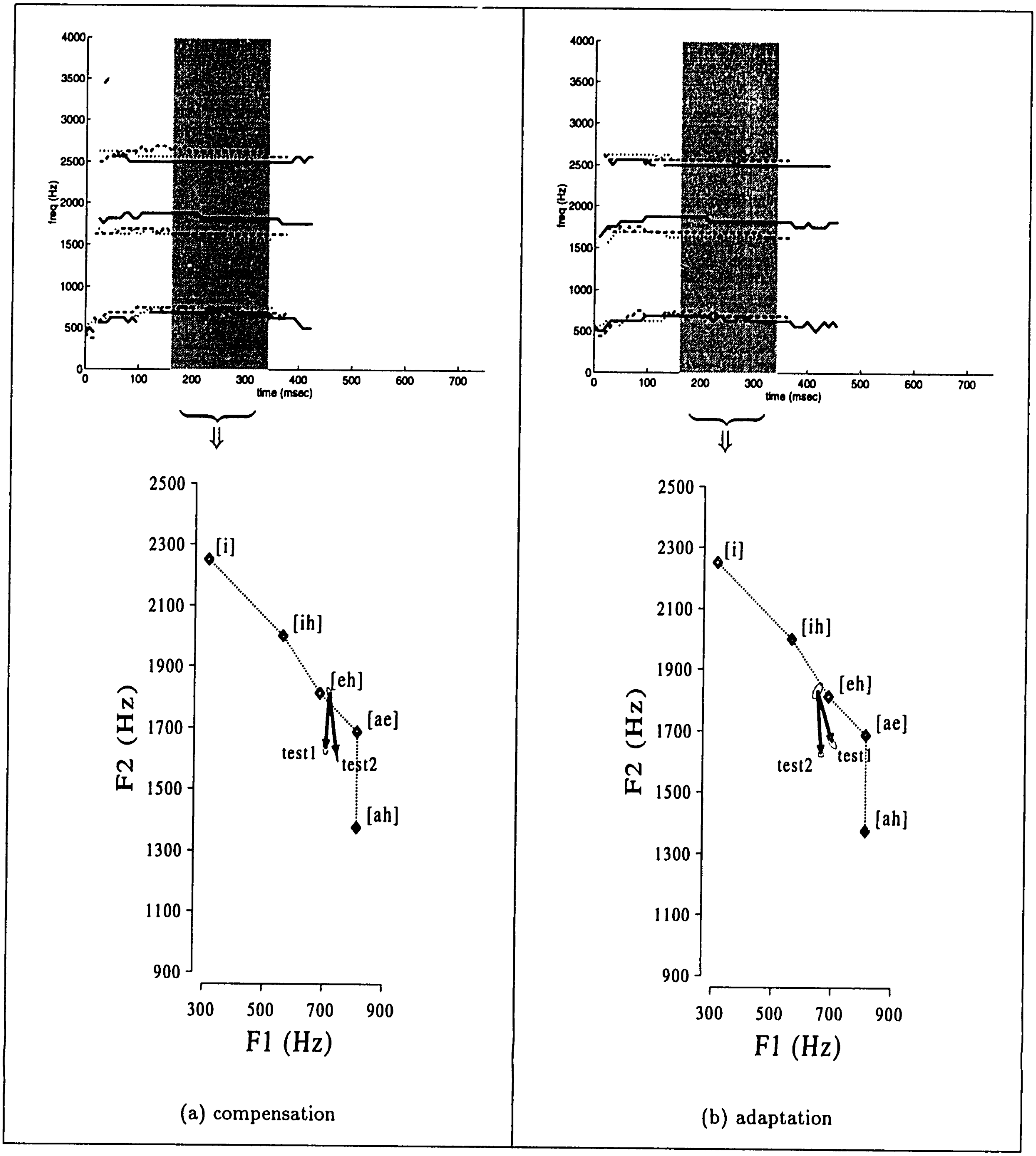
These estimates closely agree with the calculated values: the calculated mean compensation (averaged over the two testing phases) is 0.47±0.08, and the calculated mean path deviation change is −105 ± 12 Hz.

Adaptation Analysis The adaptation plots (box (b) in the figure) show essentially the same results as seen in the compensation plots. The avgram and vowel plots both show that the F2 lowering seen in MB's compensation response has carried over to his adaptation response. This is again seen in the avgram plot: the testing phase F2 formant tracks (dashed and dotted lines) are both visibly lower in frequency than the baseline F2 formant track (solid line). There is also less F3 increase apparent in this avgram.

The F2 lowering is again more easily seen in the vowel plot. The arrows labeled "test1" and "test2" show that the change in mean (F1,F2) from the baseline to the first and second testing phases is again almost completely in the negative F2 direction. Note that the "test1" and "test2" arrows have switched positions from their arrangement in the compensation vowel plot. Note also that mean baseline (F1,F2) is close to its position in the compensation vowel plot. Because of this, average path projection and deviation values for MB's adaptation response are nearly the same as those of his compensation response.

Mean compensation and path deviation change for MB's adaptation response are thus predicted to be close to the values for his compensation response. This prediction is borne out by the the calculated values for his adaptation response: calculated mean compensation is 0.45 ± 0.08, and calculated mean path deviation change is -103 ± 27 Hz.

In sum, it appears that exposure to the -2.0 feedback transformation has had substantial effect on MB's production of [ɛ]. The compensation analysis shows that he compensated for approximately half of the path shift produced by this transformation; the adaptation analysis shows that he retained this compensation even when his hearing was blocked by noise. There is also no consistent difference in (F1,F2) position between the two testing phases, suggesting that the extra exposure to the altered feedback in the second training phase had little effect on MB.



(a) compensation

(b) adaptation

Figure 4-1: Subject MB avgram and vowel plots.

4.4.1.2 Subject BM Results

Figure 4-2 shows plots of subject BM's compensation and adaptation responses to exposure to the -2.0 feedback transformation. Like subject MB, BM's responses are large. However, BM's compensation results show an F1 estimation instability that ultimately affects calculation of mean compensation and path deviation. These results illustrate how formant estimation problems can cause errors in these calculations.

Compensation Analysis Consider first the compensation plots (box (a) in the figure). In the avgram, the formant tracks for F1 and F2 are clear for all three experiment phases, but the formant track for F3 is visible only for the second testing phase (dotted line). This indicates low F3 amplitudes and suggests that all formant amplitudes may be low. Within the time interval used to analyze vowel production changes (gray region), F1, F2, and F3 all attain steady-state values. There is, however, an apparent instability in F1 estimation just prior to this interval. This instability may be due to low F1 amplitude. In the results analysis, this causes F1 to be resolved into two peaks: one at approximately 600 Hz and the other varying between 300 to 400 Hz.[1] This resolution of F1 into two peaks is most pronounced for the baseline phase data (solid lines).

F1 occasionally being erroneously resolved as two peaks in the vowel analysis interval would fool the vowel formant analysis routines: the lower F1 peak would be considered F1, while the upper F1 peak would be considered F2. F2 would appear to have an occasional large drop in frequency, and F1 would appear to have an occasional moderate drop. This would cause the mean F2 standard error to be very large and the mean F1 standard error to be moderately large.

In the vowel plot, the the unusual position and huge standard error ellipse for baseline mean (F1,F2) suggests that this F1 measurement error did occur within

[1] It's important to note the difference between formant estimation and formant data analysis. During the experiment, the DSP's formant *estimation* routine always produces a single F1 estimate. However, low formant amplitudes could cause this estimate to be unstable. After the experiment, such unstable F1 estimates could cause the formant data *analysis* routine to resolve two or more peaks in the F1 region. See Appendix A.

the vowel analysis interval. The plot shows mean (F1,F2) for the baseline phase to be the same as mean (F1,F2) for the two testing phases. However, the avgram shows this is not the case: baseline F2 is clearly higher than F2 for the two testing phases. The baseline standard error ellipse shows very large mean F2 uncertainty and moderately large mean F1 uncertainty. This is the predicted outcome if F1 is occasionally erroneously resolved into two peaks.

Thus, although it is not visible in the avgram, baseline F1 is probably briefly resolved into two peaks within the vowel analysis interval. The likelihood that this F1 measurement problem occurred within the vowel analysis interval is also supported by the visible occurrence of it in the avgram prior to the vowel analysis interval.

This F1 measurement problem also affects calculation of mean path projection and deviation. Because baseline mean (F1,F2) values have large standard errors and appear close to the testing phase values, baseline mean path projection and deviation values have large standard errors and to appear close to the testing phase values. This should make mean compensation and path deviation change small compared to their standard errors. This prediction is borne out in the calculated values: calculated mean compensation is 0.22 ± 0.14, and calculated mean path deviation change is −20 ± 44 Hz.

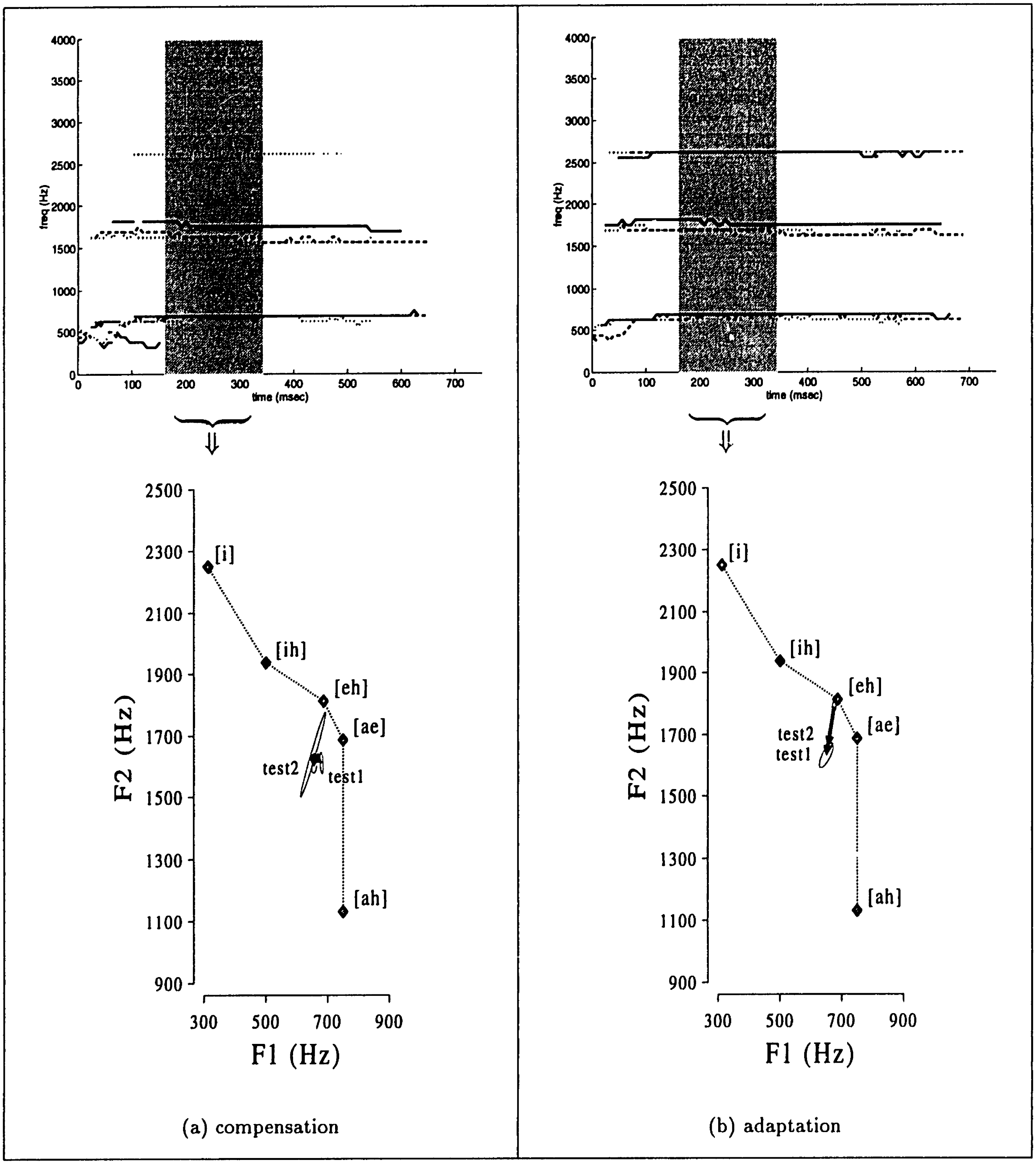
Adaptation Analysis The adaptation plots (box (b) in the figure) show results that are more robust. The avgram plot exhibits three clear groups of baseline phase formant tracks. This shows that the amplitudes of F1, F2, and F3 were always above the plotting threshold during this phase. The testing phase formant tracks show no gaps that would indicate low formant amplitudes. (However, it's possible that unseen gaps exist, because the baseline formant tracks overlay much of the testing phase formant tracks,)

Because of this more robust data, the adaptation vowel plot shows clearly what was obscured in the compensation vowel plot: a noticeable lowering of F2 in response to the feedback transformation. This result is consistent with the formant tracks visible in the avgram. It is also similar to subject MB's adaptation response.

Resolution of BM's adaptation response into its path projection and deviation components shows it also to be similar to MB's adaptation response. Baseline mean (F1,F2) position is close to [ɛ], while the testing phases' mean (F1,F2) positions are nearly equal and appear to deviate by about −100 Hz from the [i]–[ɑ] path. Both positions also appear to project to a path point near [æ]. Since all the standard error ellipses are small, the mean path projection and deviation estimates should have minimal variance.

Mean path projection change is thus about +1.0 vowel unit, while mean path deviation change is about -100 Hz. Both have minimal standard errors. The mean path projection change should translate to a mean compensation of about +0.5. These graphical estimates again closely agree with the calculated values: the calculated mean compensation is 0.42 ± 0.07 and the the calculated mean path deviation change is -69 ± 24 Hz.

In sum, it appears that, like subject MB, exposure to the -2.0 feedback transformation caused BM to partially compensate. In the compensation analysis, the avgram clearly shows this response. However, in the vowel plot and (path projection, deviation) calculations, this response is obscured by an F1 estimation instability. Also like subject MB, there was little difference in mean (F1,F2) between the two testing phases. This suggests that the extra exposure to the altered feedback in the second training phase also had little effect on BM.



(a) compensation

(b) adaptation

Figure 4-2: Subject BM avgram and vowel plots.

4.4.1.3 Subject MF Results

Figure 4-3 shows plots of subject MF's compensation and adaptation responses to exposure to the -2.0 feedback transformation. Unlike the previous subjects, MF shows evidence of production change in both F2 *and* F1. This makes his vowel production changes more aligned with his [i]–[ɑ] path, resulting in minimal path deviation change. MF's results are therefore a good illustration of efficient compensation for the altered feedback.

Compensation Analysis In the compensation plots, the avgram shows noticeable jitter in the resolution of the F1 and F2 formants.[2] Nevertheless, it does appear that the F1 and F2 formant tracks are closer together in the testing phases (dashed and dotted lines) than they are in the baseline phase (solid lines). The F3 formant tracks appear stable, and F3 is consistently lower in the second testing phase than in the other phases.

The convergence of F1 and F2 is also seen in the vowel plot. The "test1" and "test2" arrows show that mean (F1,F2) is nearly the same in both testing phases. The arrows also show that the change in mean (F1,F2) from the baseline to the testing phases is is aligned with the [ɛ]-[æ] path segment and is about half its length. This creates minimal change in path deviation and causes a path projection change of about +0.5 vowel units. Mean compensation is thus estimated to be about +0.25. The calculated values are in close agreement: mean compensation is 0.25±0.08, while mean path deviation change is 6 ± 16 Hz.

Adaptation Analysis In the adaptation plots, the avgram shows retention of most of the production changes seen in the compensation plots. The F1 and F2 formant tracks are again closer together in the testing phases than they are in the baseline phase. The F3 formant tracks again appear stable, but F3 lowering in the second test phase does not appear as consistent as it is in the compensation avgram.

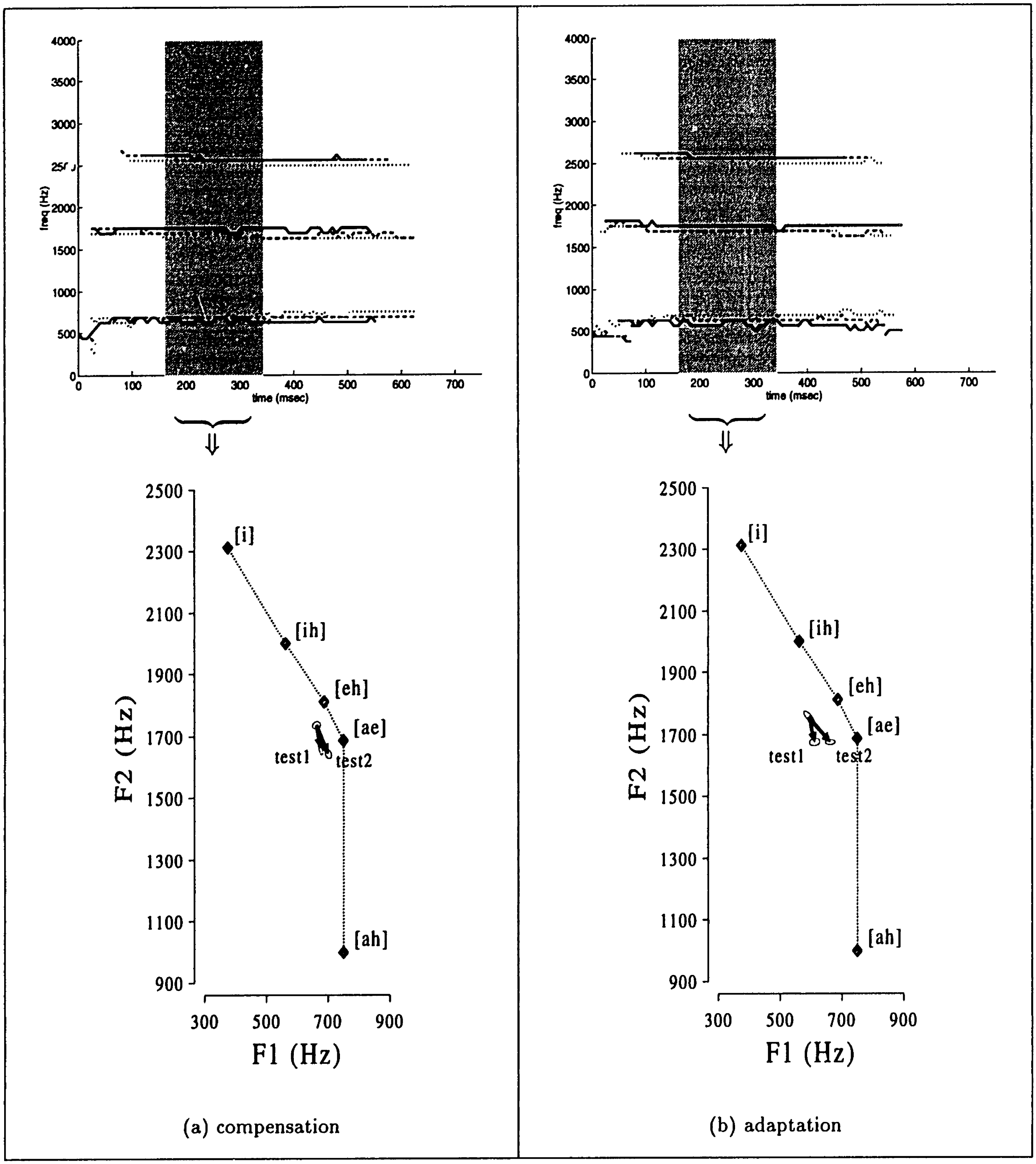
[2] The apparent discrete steps in this jitter are the result of the discretization of the frequency by the 64-channel spectral analysis done by the DSP; they do not reflect actual discrete production changes by the subject.

Further comparison of the adaptation and compensation avgrams reveals an overall production difference. For all three experiment phases, F1 appears lower in the adaptation avgram than in the compensation avgram. Since this F1 lowering is apparent in the baseline phase, the lowering may have been caused by the masking noise conditions under which the adaptation data was acquired. This suggests that hearing the masking noise caused the subject to lower his production of F1.

In the vowel plot, this general F1 lowering shifts the "test1" and "test2" arrows to the left of their compensation vowel plot positions. This is because, for all three experiment phases, mean F1 has decreased. However, the plot shows that each phase exhibits a different F1 decrease. The baseline phase shows the most F1 decrease (and thus the largest left shift), the second testing phase shows the least, and the first testing phase is between these two.

The larger left shift of the baseline mean (F1,F2) position slightly enlarges the difference between it and the mean (F1,F2) positions for both testing phases. This makes the adaptation baseline - testing phase mean path projection changes slightly larger than they were in the compensation plot. The larger baseline left shift also brings the average baseline - testing mean (F1,F2) difference into even more parallel alignment with the subject's [i]–[ɑ] path than it is in the compensation vowel plot. This makes the average mean path deviation change smaller than it was in the compensation vowel plot. However, the larger left shift of the first testing phase (as compared to the second testing phase) increases the standard error of the average mean path deviation change. These graphically estimated differences between the subject's adaptation and compensation responses are confirmed by the calculated values: for the subject's adaptation response, mean compensation is 0.35 ± 0.11 and mean path deviation change is 1 ± 19 Hz.

In sum, there are two interesting aspects of MF's response to the altered feedback. First, compared with subjects MB and BM, MF's compensation is more aligned with his [i]–[ɑ] path. This makes his compensation more efficient, in the sense that it shows minimal path deviation change. Second, MF appears to lower his production of F1 in response to hearing the masking noise. This F1 lowering is most pronounced in the baseline phase, less so in the first testing phase, and least pronounced in the second testing phase.



(a) compensation

(b) adaptation

Figure 4-3: Subject MF avgram and vowel plots.

4.4.1.4 Subject JK Results

Figure 4-4 shows plots of subject JK's compensation and adaptation responses to exposure to the -2.0 feedback transformation. JK exhibits the same F1 lowering in response to hearing masking noise that subject MF exhibited. For MF, this F1 lowering made his adaptation appear larger than his compensation. For JK, however, it appears possible that he exhibits no compensation, and that his adaptation is completely explained by the F1 lowering.

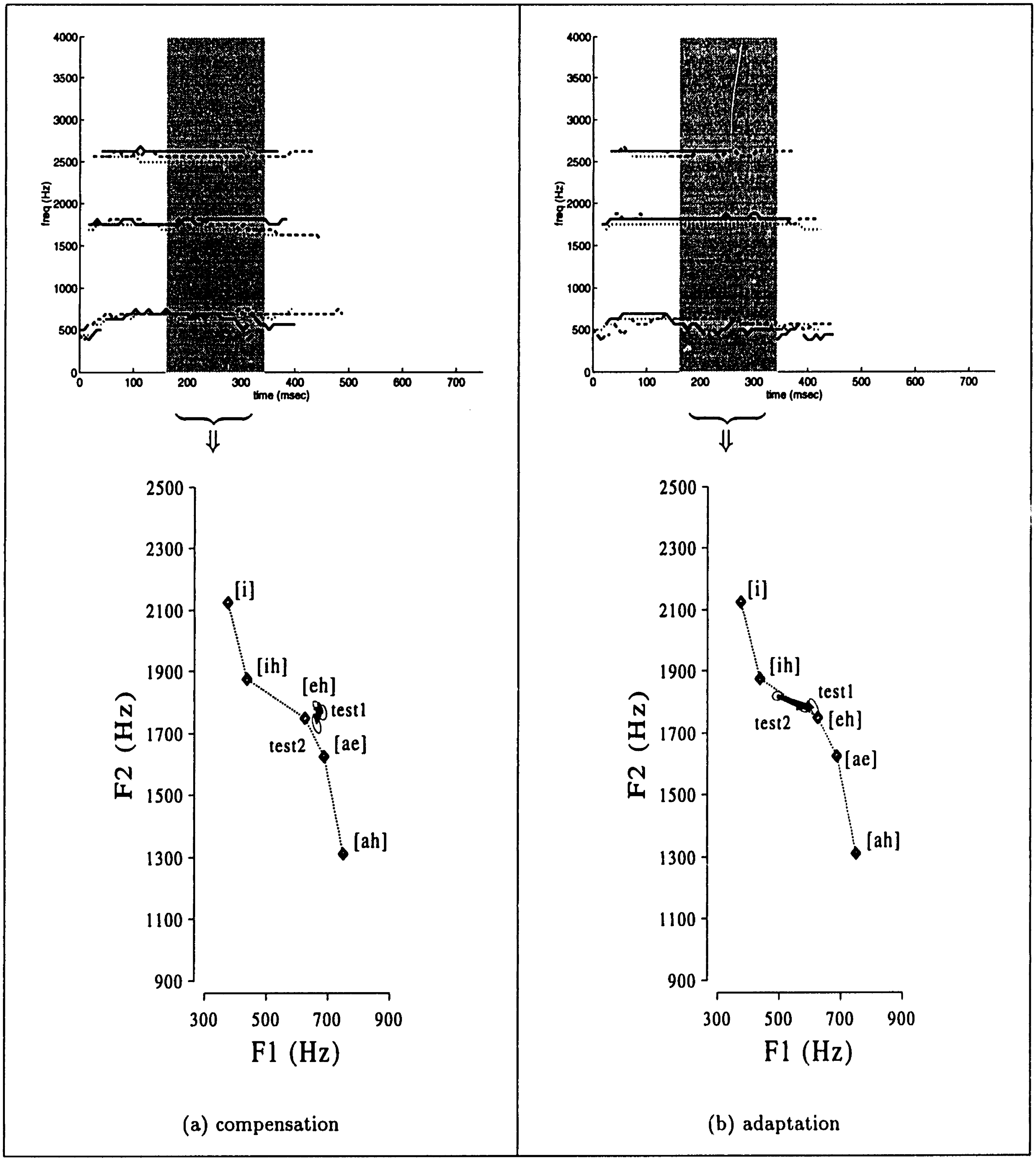
Compensation Analysis In the compensation plots, the avgram shows gaps and jitter in the formant tracks, suggestive of low-amplitude formants. F1 appears nearly the same for all experiment phases until near the end of the vowel analysis interval (gray region). At this point, F1 shows a large burst of jitter in the baseline phase (solid line). F2 appears similar for all experiment phases until about a third of the way into the vowel analysis interval. At this point, the testing phase F2 values begin descending below the baseline values. F3's formant tracks appear stable, but they show a gap in the baseline phase. Testing phase F3 values look consistently lower than the baseline values – the opposite of subject MB's results.

The vowel plot shows that, averaged over the vowel analysis interval, F1 and F2 show little change over the course of the experiment. The almost non-existent "test1" and "test2" arrows indicate that mean (F1,F2) is about the same for the baseline, first, and second testing phases. Thus, from the plot, mean compensation and path deviation change are predicted to be small compared to their standard errors. The calculated values show this is the case: mean compensation is 0.09 ± 0.08 and mean path deviation change is -18 ± 20 Hz.

Adaptation Analysis In the adaptation plots, the avgram shows more formant track gaps and jitter than those seen in the compensation avgram. Like subject MF, when JK's compensation and adaptation avgrams are compared, F1 in his adaptation avgram is seen to be lower across all experiment phases. Also like subject MF, JK's baseline F1 appears to exhibit the most lowering, although this trend is obscured by the large jitter in F1's baseline formant track. On the other hand, F2 in the adaptation avgram looks generally more stable than in the compensation avgram. Little change is seen in its value across the experiment phases. F3 apparently has such low amplitude that it falls below plotting threshold by at least midway through the vowel analysis interval.

The F1 lowering seen in the adaptation avgram has a large effect in the adaptation vowel plot. As was true with subject MF, the F1 lowering left shifts the mean (F1,F2) positions of all experiment phases, as compared to their positions in the compensation vowel plot. The testing phase positions, which are nearly equal, are shifted least and the baseline phase position is shifted most. This creates a large mean (F1,F2) change vector between the baseline and testing phases. Because of the orientation of JK's [i]–[ɑ] path, this change vector is in close alignment with the [ι]-[ε] path segment and appears to be roughly half of the segment's length. Thus, from the plot, mean compensation is predicted to be about 0.25 and path deviation change is predicted to be small compared to its standard error. The calculated values show this is the case: mean compensation is 0.27 ± 0.07 and mean path deviation change is 25 ± 20 Hz.

In sum, one possible interpretation of JK's results is that the calculated value of 0.09 ± 0.08 may represent JK's true amount of compensation, while the calculated value of 0.27 ± 0.07 may only be a result of his lowering of F1 in response to the masking noise.



(a) compensation

(b) adaptation

Figure 4-4: Subject JK avgram and vowel plots.

4.4.1.5 Subject JG Results

Figure 4-5 shows plots of subject JG's compensation and adaptation responses to exposure to the -2.0 feedback transformation. Like subject BM, JG's compensation results show an F1 estimation instability that ultimately affects calculation of mean compensation and path deviation. But whereas BM's estimation instability causes F1 to be resolved into two peaks, JG's estimation instability causes F1 to be missed. This F1 dropout causes brief formant mislabeling that makes JG's path deviation change in his compensation response to appear large and have a large standard error.

Compensation Analysis In the compensation plots, the avgram shows evidence of low formant amplitudes. At the utterance's beginning, baseline F1 is briefly resolved into two peaks. Thereafter, F1 appears stable and slightly lower in the baseline phase than in the testing phases. F2 appears stable throughout the mean utterance and higher in the baseline phase than in the testing phases. Thus, in the testing phases, F1 and F2 appear to approach each other. F3 has such low amplitude that it is rarely above the plotting threshold for any experiment phase.

The vowel plot exhibits the F1 dropout effect. Baseline mean (F1,F2) is close to [ɛ] and first testing phase mean (F1,F2) is practically on the subject's [i]–[ɑ] path, halfway between [ɛ] and [æ]. Both mean (F1,F2) positions have small standard error ellipses. The resulting mean (F1,F2) change vector (the arrow labeled "test1"), appears reasonably well aligned with the [ɛ]-[æ] path segment and is somewhat less than half of the segment's length. From this, we would estimate a corresponding mean compensation of somewhat less than 0.25, and small mean path deviation change.

However, the second testing phase mean (F1,F2) position is far from the [i]–[ɑ] path. This appears to be caused by a brief formant mislabeling (F2's peak being labeled as F1, F3's peak being labeled as F2) resulting from F1 dropout.[3] Both testing

[3] F1 dropout as the likely cause is suggested by inconsistencies between the avgram and the vowel plot. The vowel plot shows that mean F2 in the second testing phase is slightly higher than in the baseline phase. It also shows that mean F1 in the second testing phase is much higher (by about 200 Hz) than in the baseline phase. Neither of these things are evident in the avgram plot. However, if F1 amplitude momentarily fell below analysis threshold, it would be missed. F2's peak would then be the lowest spectral peak, and F3 would be the second lowest. This would cause

phase positions appear to project to the same path position, but the second testing phase position appears to have a large mean path deviation with a large standard error. When the first and second testing phase positions are averaged together, the mean compensation estimate is unchanged, but mean path deviation is now estimated to be large with large standard error. This is confirmed in the calculated values: mean compensation is 0.20 ± 0.05 while mean path deviation change is 65 ± 81 Hz.

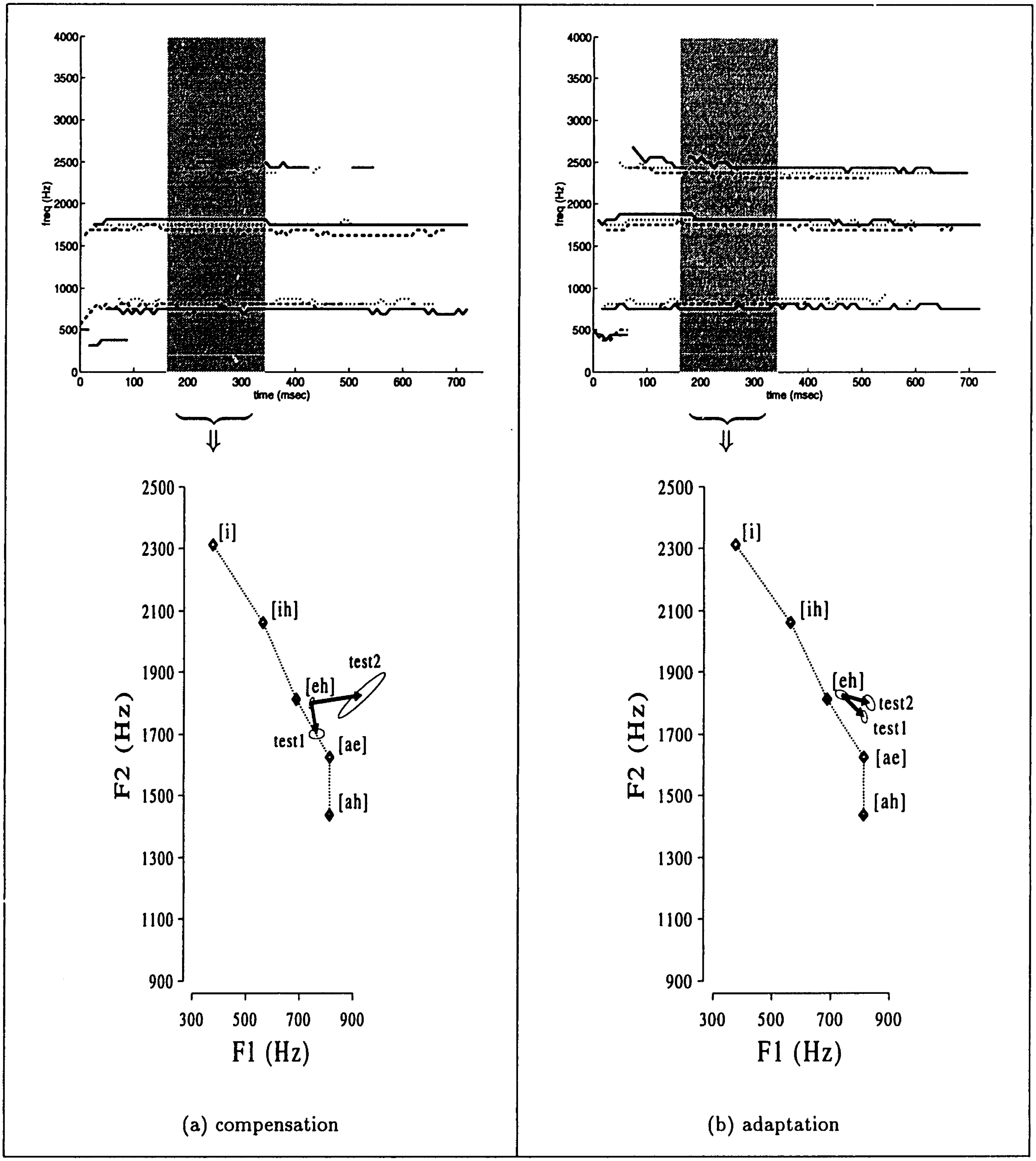
Adaptation Analysis In the adaptation plots, the avgram shows evidence of higher overall formant amplitudes in the subject's adaptation responses. F3, which was hardly visible in the compensation avgram, shows complete formant tracks for all three experiment phases. F3 in the testing phases appears lower than in the baseline phase. The F1 and F2 formant tracks also appear complete, although F1 is still resolved into two peaks at the beginning of the mean utterance. F1 and F2 also appear closer together in the testing phases than in the baseline phase.

The vowel plot shows that the stronger formant amplitudes have apparently eliminated the F1 dropout. Neither the baseline nor the first testing phase mean (F1,F2) positions are very different from their compensation vowel plot positions (although the first testing phase mean (F1,F2) position now visibly deviates from the [i]–[ɑ] path). The prominent difference between the adaptation and compensation vowel plots is that, in the adaptation vowel plot, the second testing phase mean (F1,F2) position no longer affected by F1 dropout. It now has a small standard error ellipse and is close to the first testing phase position. The net result is that mean compensation in JG's adaptation response should be similar to mean compensation in his compensation response. However, mean path deviation change of his adaptation response should be smaller and have much smaller standard error. These predictions are confirmed by the calculated values: for JG's adaptation response, mean compensation is 0.19 ± 0.06

F2 to be mislabeled as F1 and F3 to be mislabled as F2. It would then appear that both F1 and F2 experienced momentary huge frequency increases. This would significantly raise the mean and standard error of both F1 and F2 within the vowel analysis interval. In the vowel plot, this would be seen as a diagonal shift of mean (F1,F2) up and to the right. The standard error ellipse would also be large and diagonally oriented. This exactly describes the second testing phase mean (F1,F2) position.

and mean path deviation change is 44 ± 26 Hz.

In sum, it appears that JG produced moderate compensations with minimal path deviation changes, but this is obscured in the compensation analysis by F1 dropout. It is also clear that he retained much of his compensations in his adaptation response.



(a) compensation

(b) adaptation

Figure 4-5: Subject JG avgram and vowel plots.

4.4.1.6 Subject BK, MK, JD, And ST Results

Figures 4-6 through 4-9 show the avgram and vowel plots of the rest of the subjects. These subjects all showed smaller compensation and adaptation responses than subjects MB, BM, MF, JK, and JG. In many cases, these responses look non-significant – as evidenced by the comparable sizes of the mean (F1,F2) change arrows and standard error ellipses in the vowel plots. The responses which do look possibly significant are all consistent with the trends seen more clearly in the plots of subjects with larger responses.

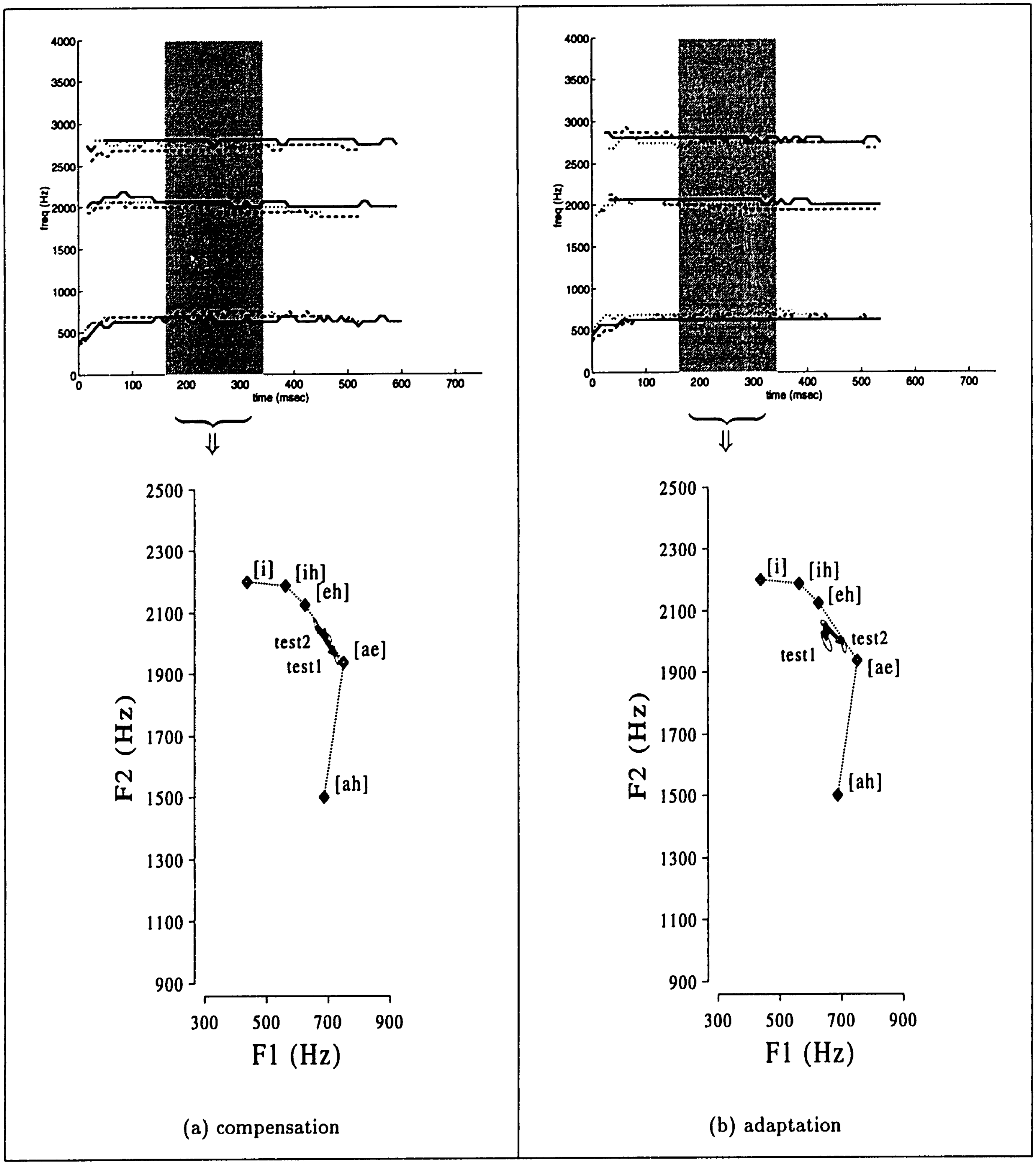
For these reasons, each of the remaining four subjects' plots will be discussed only briefly.

Subject BK Results Figure 4-6 shows plots of subject BK's response to the altered feedback. Overall, the plots show evidence of compensation and adaptation. In both the compensation and adaptation avgrams, F1 and F2 appear closer together in the testing phases than in the baseline phase. In the compensation avgram, F3 appears lower in the testing phases than in the baseline phase. However, in the adaptation avgram, F3 appears approximately the same for all three experiment phases. Both vowel plots show mean (F1,F2) changes that are roughly aligned with the subject's [i]–[ɑ] path. All changes appear to compensate for the path projection shift of the feedback transformation. Curiously, the magnitude of these changes in the two testing phases reverse between the compensation and adaptation vowel plots. In the compensation vowel plot, the first testing phase exhibits greater mean (F1,F2) change than the second testing phase. In the adaptation plot, the situation is reversed. It is possible, however, that this reversal is an artifact of the generally small mean (F1,F2) changes: note, for example, that in the adaptation vowel plot, the mean (F1,F2) change vector for the first testing phase is smaller that the first testing phase's standard error ellipse.

Subject MK Results Figure 4-7 shows plots of subject MK's response to the altered feedback. Overall, the plots show evidence of slight compensation and adaptation. MK's results exhibit three prominent features. First, there is little trace of F3 in either of the avgrams, possibly indicating low formant amplitudes. Second, in the compensation vowel plot, mean (F1,F2) ɔositions for all three experiment phases are near each other and centered around [æ] on MK's [i]–[ɑ] path. This is odd since the subject was supposedly whispering words containing the vowel [ɛ]. It suggests possible problems measuring MK's path vowel formants, or that MK has less accurate perception of the vowel sounds in his synthesized feedback. The third prominent results feature is the difference between the baseline mean (F1,F2) positions in the compensation and adaptation vowel plots. Relative to the compensation plot, baseline mean (F1,F2) position in the adaptation plot is shifted diagonally down and to the left. It also has a huge standard error ellipse. As explained in the discussion of subject BM's results, these effects are likely caused by F1 being resolved into two peaks somewhere within the vowel production analysis interval.

Subject JD Results Figure 4-8 shows plots of subject JD's response to the altered feedback. The compensation plots show slight evidence of a compensating response: in the avgram, F1 and F2 appear closer together in the testing phases, while in the vowel plot, the mean (F1,F2) change vectors are both small but oriented in the compensating direction. The adaptation plots, however, show little evidence of any retention of this compensation in JD's adaptation response.

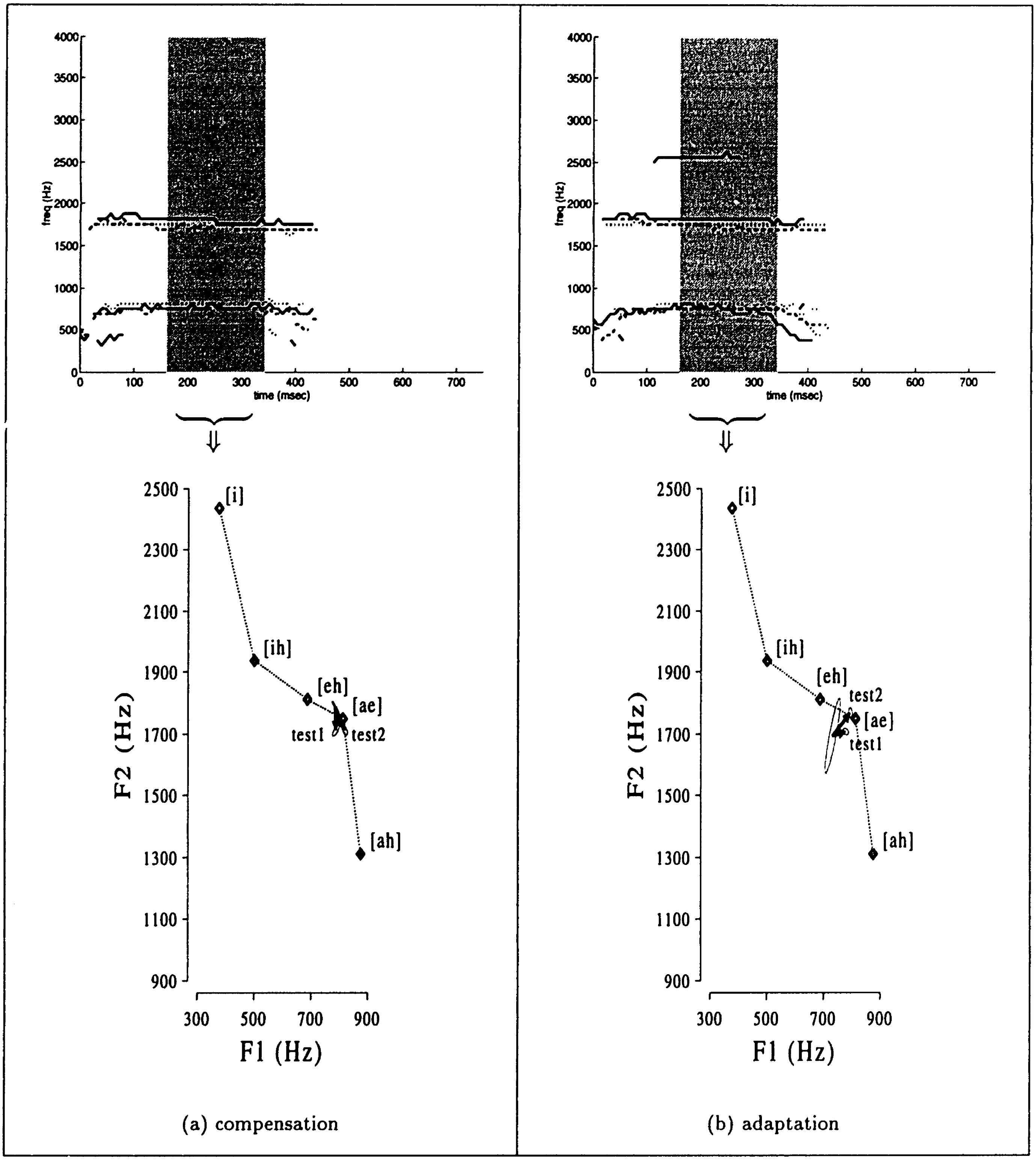
Subject ST Results Figure 4-9 shows plots of subject ST's response to the altered feedback. For this subject, neither the compensation plots nor the adaptation plots show evidence of any significant production changes.



(a) compensation

(b) adaptation

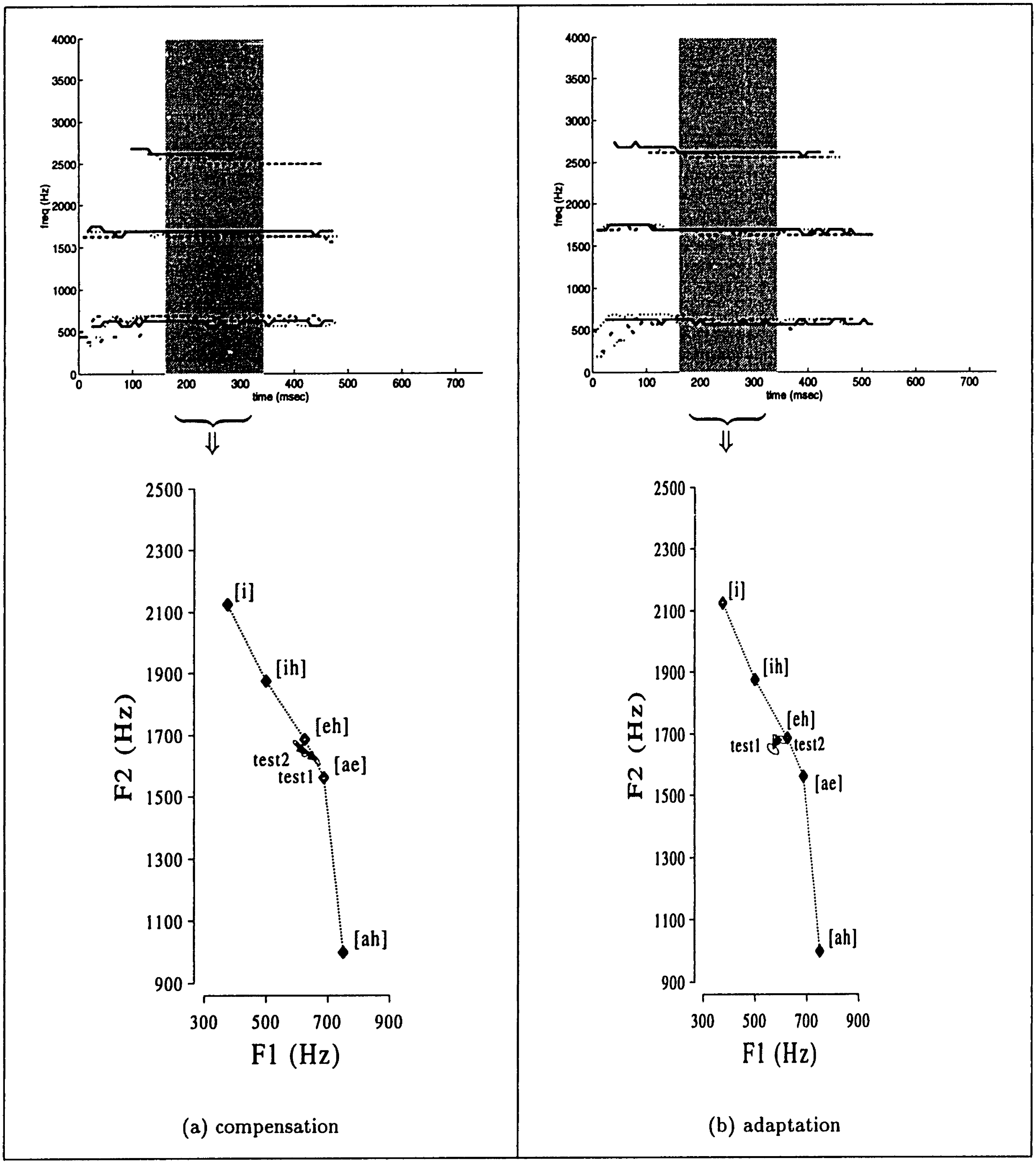
Figure 4-6: Subject BK avgram and vowel plots.



(a) compensation

(b) adaptation

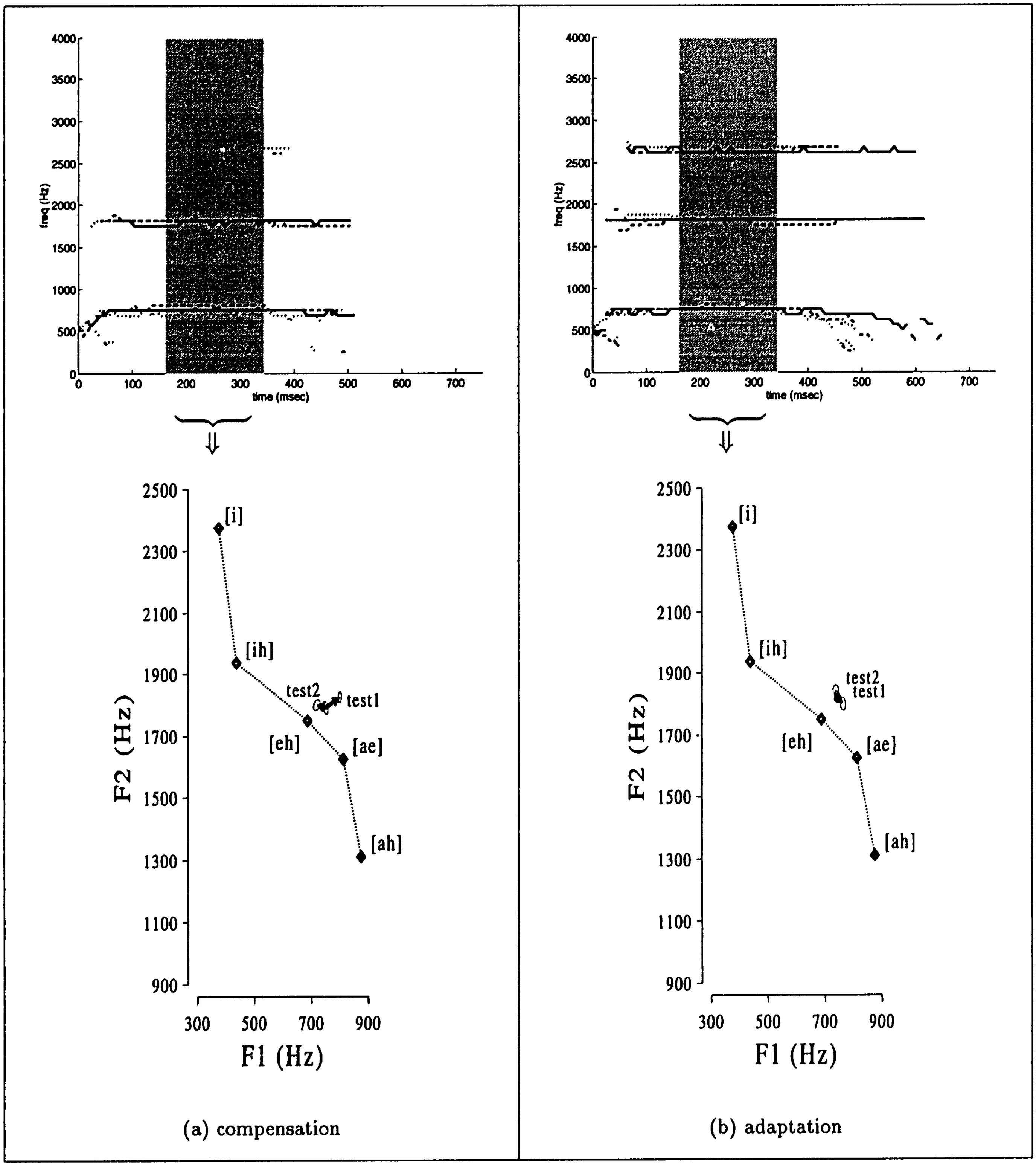
Figure 4-7: Subject MK avgram and vowel plots.



(a) compensation

(b) adaptation

Figure 4-8: Subject JD avgram and vowel plots.



(a) compensation

(b) adaptation

Figure 4-9: Subject ST avgram and vowel plots.

4.4.2 Across-Subject Analysis

In the anaiyses of individual subject results, the following consistent trends are seen:

- In the compensation analyses, it appeared that subjects compensated to varying degrees for the perceived effects of the -2.0 feedback transformation.

- In the adaptation analyses, it appeared that subjects retained most of their compensation, even when whispering with feedback blocked by masking noise. That is, subjects appeared to adapt.

- In both analyses, there did not appear to be a significant difference between the first and second testing phases in subject's responses, suggesting that the second training phase had no effect.

In addition, these trends appear to be reflected accurately by the calculated measures of path projection and deviation. This allows us to examine these trends further by collapsing results across subjects.

4.4.2.1 Across-Subject Plots

Consider first Figure 4-10, which plots, for each subject, mean compensation and path deviation change averaged over the first and second testing phases.

The figure is composed of two columns:

- The left column shows plots of subjects' compensation responses. Within this column:
  - The top plot (Figure 4-10(a)) shows mean compensation.
  - The bottom plot (Figure 4-10(b)) shows mean path deviation change.
- The left column shows plots of subjects' adaptation responses. Within this column:
  - The top plot (Figure 4-10(c)) shows mean adaptation – i.e., mean compensation seen in subjects' adaptation responses.

- The bottom plot (Figure 4-10(d)) shows mean path deviation change.

In each plot, the same ordering of subjects is used and a line is shown connecting subjects' mean values. This line is used to make more salient the pattern of results across subjects; it does not imply any dependencies among the subjects.

Mean Compensation The main feature of Figure 4-10(a) is consistent with the trend seen in the individual subject data: mean compensation of all but one subject is positive, showing that all but one subject adjusted his production of [ε] to compensate for the feedback transformation.

The plot also shows some discretization of compensation values across subjects. The plot shows subjects BM, MF, JG, BK, MK, and JD all appear to exhibit approximately the same mean compensation of 0.2. However, analysis of BM's individual plots suggests that F1 measurement errors (F1 being briefly resolved into two peaks) make BM's mean compensation appear lower than it really is; BM's true mean compensation is probably closer to MB's mean compensation. In addition, the confidence intervals of JK's mean compensation show it is not significantly different from zero. If these adjustments are made to BM's and JK's results, it appears all subjects exhibited one of approximately three mean compensations:

- Subjects MB, BM: mean comp. ≈ 0.5
- Subjects MF, JG, BK, MK, JD: mean comp. ≈ 0.2
- Subjects JK, ST: mean comp. ≈ 0.0

Mean Path Deviation Change in Compensation Responses Figure 4-10(b) shows that for all subjects but MB, mean path deviation change in their compensation responses is not significantly different from zero. However, these values should be adjusted to reflect two F1 measurement errors seen in the individual subject plots:

- In the vowel analysis of BM's results, F1 was probably briefly resolved into two peaks. This made BM's mean path deviation change look smaller and its standard error look larger. It appeared that BM's true mean path deviation change is closer to MB's and has a smaller standard error.

- In the vowel analysis of JG's results, F1 was probably briefly missed. This made JG's mean path deviation change as well as its standard error look larger. It appeared that JG's true mean path deviation change is closer to zero has a smaller standard error.

Taking into account these adjustments leaves the overall pattern largely unchanged: most subjects still show approximately zero change in path deviation in response to the feedback transformation.

This pattern is consistent with the hypothesized response of subjects to the feedback transformation. Since the transformation shifts perceived path projection, subjects were expected to compensate by shifting path projections of their productions of [ɛ]However, since the transformation does not alter perceived path deviation, the hypothesis predicts no influence on subjects' path deviations. The near-zero or inconsistent path deviation changes seen across subjects are in agreement with this prediction.

Mean Adaptation Figure 4-10(c) plots mean adaptation. Mean adaptation of all but one subject is positive, showing that all but one subject retained his compensatory productions when auditory feedback was blocked by masking noise. This is again consistent with the trend seen in the individual subject data:

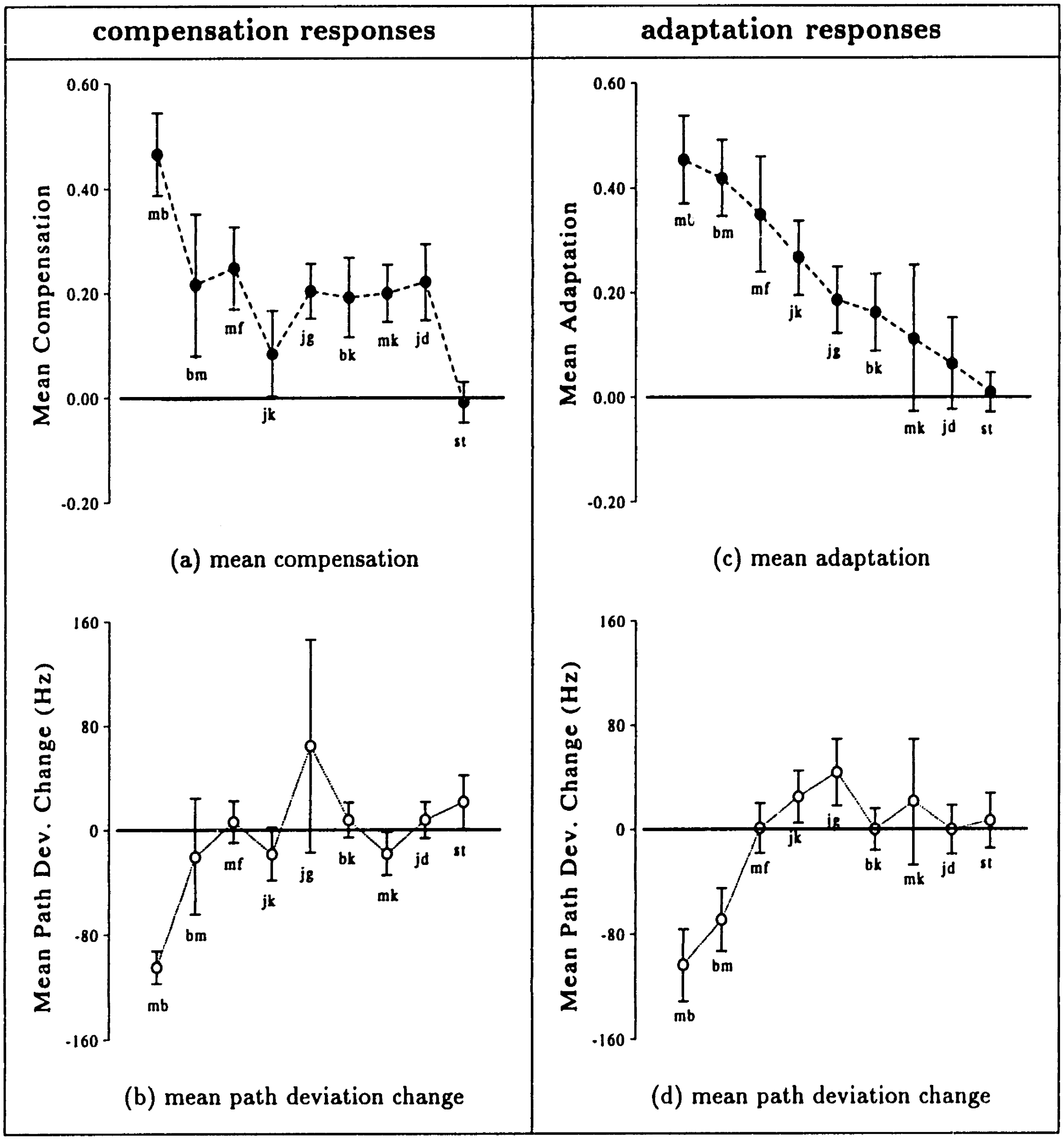
In the plot, subjects are ordered by decreasing amount of adaptation (this same ordering was used in all the plots of Figure 4-10). This ordering shows the pattern of adaptation to differ from that seen in the compensation plot: here, the distribution of subjects' mean adaptation almost uniformly covers the range from 0.0 to 0.5. The distribution of actual adaptation values is probably slightly less uniform than it appears, since JK's actual adaptation was probably zero.[4] Taking this into account,

[4] JK's calculated mean adaptation was probably not due to true adaptation. Instead, it was probably an artifact of his lowering of baseline F1 in response to the masking noise. See the discussion of his individual results for more details.

however, there is still no evidence of grouping int ı three distinct compensation values, as seen in the compensation plot.

There is no apparent theoretical basis for predicting (1) quantization of the compensation values, (2) the uniform distribution of adaptation values, or (3) the difference between the two distributions. In the next chapter, we will discuss further the distributions of compensation and adaptation values. Here, we note that the standard error bars around each mean value are big enough that random variation may partially explain the observed shapes of their distributions.

Mean Path Deviation Change in Adaptation Responses Figure 4-10(d) shows that, like the compensation responses, path deviation change is not significantly different from zero for most subjects' adaptation responses.



(a) mean compensation

(c) mean adaptation

(b) mean path deviation change

(d) mean path deviation change

Figure 4-10: Compensation and adaptation responses for each subject, averaged across the first and second testing phases. (a) and (b) are mean compensation and path deviation change for each subject's compensation response. (c) and (d) are mean compensation and path deviation change for each subject's adaptation response. In each plot, mean values are indicated by dots. Small bars around each dot indicate confidence intervals. (Note that mean compensation of a subject's adaptation response is called "mean adaptation". Note also mean compensation is a dimensionless ratio, while mean path deviation change is measured in Hz (see Section 3.5.3).

4.4.2.2 Statistical Tests

To test the statistical significance of the trends seen in the plots, two measures of across-subject path projection and deviation change were tabulated:

- F ratios and p values derived from ANOVA tests.
- Mean compensation and path deviation change, averaged across subjects.

ANOVA Tests The ANOVA tests determined if subjects' path projection and deviation values changed significantly between experiment phases. Table 4.1 summarizes the F ratios and p values of these tests.

response type	experiment phases	significance of changes in			
	compared	path proj.		path dev.	
		F(1,8)	p <	F(1,8)	p <
compensation	base vs. test1	33.338	**0.000**	0.862	**0.380**
	base vs. test2	13.947	**0.006**	0.018	**0.896**
	test1 vs. test2	0.279	**0.612**	0.508	**0.496**
adaptation	base vs. test1	23.782	**0.001**	1.731	**0.225**
	base vs. test2	13.218	**0.007**	0.000	**0.991**
	test1 vs. test2	0.205	**0.663**	1.793	**0.217**

Table 4.1: Path projection and deviation ANOVA tests.

Each row of the table reports a specific ANOVA test, done separately on path projection and path deviation of subjects' results. In each row:

- The first column indicates which subject responses were tested (compensation or adaptation).
- The second column indicates the experiment phases that the test compared ("base" means baseline phase, "test1" means first testing phase, and "test2" means second testing phase).
- The third and fourth columns (F ratio and p value) indicate the path projection test results.

- The fifth and sixth columns indicate the path deviation test results.

The table values demonstrate the statistical significance of the trends seen in the plots. Consider first path projection (columns 3 and 4). Across all subjects' compensation and adaptation responses:

- First and second testing phase path projection values differed significantly from their baseline phase values (p values of 0.000, 0.006, 0.001, and 0.007).
- First and second testing phase path projection values did not differ significantly from each other (p values of 0.612 and 0.663).

In contrast to these results, none of the tests of path deviation change (columns 5 and 6) evidenced any significant change: the best p value seen in any test is 0.217.

In sum, the ANOVA tests show:

- Significant path projection change, but insignificant path deviation change, between either testing phase and the baseline phase.
- Insignificant change in either path projection or deviation between the two testing phases.

These results indicate exposure to altered feedback has a significant effect on subjects' path projections but not on their path deviations. This reinforces the results seen in the data plots and is the expected result if subjects compensate.

The results also indicate subjects' compensations reach limits within the first testing phase. Additional altered feedback exposure (of the second training phase) does not significantly improve subjects' compensations.

Mean Compensation and Path Deviation Change Table 4.2 summarizes mean compensation and path deviation change measurements averaged across subjects.

As indicated by column 1, the first three rows of the table show mean measurements of subjects' compensation responses. In these rows:

- Row 1 shows mean measurements calculated from comparing subjects' baseline and first testing phase productions.

response type	experiment phases compared	compensation μ	compensation σ_μ	path dev. change (Hz) μ	path dev. change (Hz) σ_μ
compensation	base vs. test1	**0.21**	0.04	**-15**	16
	base vs. test2	**0.20**	0.05	**3**	23
	average:	**0.20**	0.04	**-6**	15
adaptation	base vs. test1	**0.23**	0.05	**-17**	13
	base vs. test2	**0.22**	0.06	**0**	20
	average:	**0.22**	0.05	**-8**	16

Table 4.2: Mean compensation and path deviation changes, averaged across subjects.

- Row 2 shows the same measurements calculated from comparing subjects' baseline and second testing phase productions.

- Row 3 averages the measurements made for rows 1 and 2.

Rows 4,5, and 6 show similar mean measurements of subjects' adaptation responses.

The last four columns of the table show the mean measurements: columns 3 and 4 show mean compensation (μ) and its standard error (σ_μ), while columns 5 and 6 show mean path deviation change and its standard error.

Consider first the mean compensation results (columns 3 and 4):

- The average mean compensation (row 3) is 0.20 – five times its standard error of 0.04. This suggests an overall compensating production response of subjects to the feedback transformation.

- The average mean adaptation[5] (row 6) is 0.22 – more than four times its standard error of 0.05. This suggests an overall retention of subjects' compensating production response even when acoustic feedback is blocked.

- The difference in mean compensation between the first and second testing phases (rows 1 and 2) is $0.21 - 0.20 = 0.01$, which is small compared to either testing

[5] Recall again that mean compensation of subjects' adaptation response is called "mean adaptation" and represents the amount of compensation retained when subjects' feedback was blocked.

phase's standard error (0.04 or 0.05). This suggests there was no effect of the second training phase on amount of compensation.

- The difference in mean adaptation between the first and second testing phases (rows 4 and 5) is 0.23 – 0.22 = 0.01, which again is small compared to either testing phase's standard error (0.05 or 0.06). This suggests there also was no effect of the second training phase on amount of retained compensation.

Next, consider mean path deviation change (columns 5 and 6 in the table): mean path deviation change between either testing phase and the baseline phase was never significantly bigger than its standard error.

In sum, the table's mean compensation and path deviation change results are consistent with trends seen in the previous plots and ANOVA tests. They also complement the ANOVA test results in two ways:

1. The ANOVA tests indicated significant change in subjects' path projection values between the baseline and testing conditions. The positive mean compensation values of Table 4.2 show that this change was in the direction that compensated for the feedback alteration.[6]

2. The ANOVA tests suggested that subjects' compensations reached limiting values with the first testing phase. The mean compensation results show that the average compensation limit reached by subjects was about 0.2 - roughly comparable to the average compensation seen in reaching SA [Held, 1996].

4.5 Summary and Conclusions

Recall that Study 1 investigated several questions concerning the effect of the -2.0 feedback transformation on subjects' production of the vowel [ɛ]:

1. Do subjects compensate for the perceived effects of the feedback transformation?

[6] See Section 3.5.3.4 for an explanation of mean compensation.

2. Do they retain this compensation even when denied acoustic feedback of their whispering – i.e., do they adapt?

3. If subjects do adapt, what is their maximum degree of adaptation?

4.5.1 Do Subjects Adapt?

In Study 1, auditory feedback provided to the subject was altered by the -2.0 feedback transformation. This transformation shifts perceived vowel sound path projections along a subject's [i]–[ɑ] path towards [i]. To compensate, the subject had to shift the path projections of his vowel productions towards [ɑ]. However, to avoid introducing additional distortion, he had to leave unchanged the path deviations of his vowel productions.[7]

Thus, to provide strong evidence of compensation, Study 1 should find that exposure to the feedback transformation caused the following changes in subjects' production of [ɛ]:

1. A significant compensating change in path projections.

2. A non-significant change in the path deviations.

To provide evidence of adaptation, these same changes should persist when auditory feedback is blocked.

The results show that these were in fact the findings of Study 1. The implication is that exposure to the feedback transformation caused a retained adaptation of subjects' production of [ɛ]. Not only did they produce compensatory articulations of [ɛ] when they could hear how its sound was altered, they persisted in producing these compensatory articulations when they were subsequently prevented from hearing their whispering.

[7] See Section 3.3.1.3 for a more detailed explanation of the action of the -2.0 feedback transformation and the vowel production changes that compensate for its effects.

4.5.2 How Much do Subjects Adapt?

If subjects do adapt, does their compensation and/or adaptation achieve a maximum value less than complete adaptation?

Study 1 included a second training and testing phase to investigate this. If subjects continued to increase their compensation for the feedback transformation in the second training phase, then they should exhibit significantly more compensation (and possibly more adaptation) in the second testing phase than in the first testing phase. No such significant differences were found. The implication is that subjects achieved a maximum compensation and adaptation within the first testing phase. Since this maximum was always substantially less than 1.0, it appears that subjects do not completely compensate or adapt to the feedback transformation.

This result is consistent with studies of reaching SA, which also find subjects incompletely adapting to a feedback transformation [Welch, 1986, Kornheiser, 1976, Held, 1996].

4.5.3 Methodological Issues

Certain methodological limitations of Study 1 influenced the design of later experiments. The most important of these is the lack of a control experiment. Without a control experiment, Study 1 does not definitively show that it is precisely exposure to the feedback transformation that causes the observed adaptation. It could be, for example, that merely amount of time spent in the experiment caused the observed responses.

Another potential problem with the experiment design is the way the feedback transformation was introduced. Because the -2.0 feedback transformation was introduced abruptly at the start of the training phase, it is likely that subjects were aware of the sudden feedback change. Thus, it is likely that during the experiment subjects were aware that their feedback was being altered. This complicates the picture of what cognitive processes were involved in causing subjects to adapt. A cleaner experimental design would avoid making the subject aware of the altered feedback.

In spite of these methodological issues, however, the results of Study 1 did provide evidence for the existence of speech SA. This evidence was strong enough to warrant conducting more extensive investigations of speech SA.

Chapter 5

Study 2: Timecourse and Generalization of Speech Sensorimotor Adaptation

The results of Study 1 strongly supported the existence of speech sensorimotor adaptation (speech SA), which warranted a more detailed investigation of speech SA that was carried out in Study 2.

5.1 Introduction

The investigations of Study 2 had the following objectives:

1. Confirmation of the existence of speech SA in a controlled experiment.

2. Examination of the timecourse of speech SA – examining how compensation and adaptation develop during an speech SA experiment.

3. Investigation of how speech SA generalizes – investigating how adapting a vowel's production in one word affects its production in other words and the production of other vowels.

5.1.1 Confirming the Existence of Speech SA

Study 1 provided strong evidence of the existence of SA in speech, but it had several methodological weaknesses. Study 2 sought to confirm the existence of speech SA with an experiment design that avoided these weaknesses.

One shortcoming of Study 1 was the lack of a control experiment. This left open the possibility that the observed SA effect was not due to adaptation to the feedback transformation, but rather to unrelated factors. Study 2 therefore included use of a control experiment to isolate the cause of the production changes exhibited in Study 1.

Another shortcoming of Study 1 was that the feedback transformation was introduced abruptly at the start of the training phase. It is likely that this sudden feedback change was noticed by subjects, which raises the possibility that subjects used some conscious strategy to compensate – a possibility we wished to avoid in Study 2.

In reaching SA experiments, this problem is avoided by gradual introduction of the feedback alteration [Howard, 1968]. In Study 2, this same technique was used: the feedback alteration was introduced in a series of unnoticeable increments. Subjects were also interviewed, post-experiment, to assess their awareness of the altered feedback.

5.1.2 Examining the Timecourse of Speech SA

Gradual introduction of the feedback transformation allows several questions to be examined concerning the SA effect's timecourse.

The first question concerns whether categorical perception affects subjects' compensation. Several studies of vowel perception have exhibited poorer sensitivity to vowel sound changes within vowel category regions[1] than between them [Kuhl, 1991]. A possible consequence of this poor within-category sensitivity is that subjects would

[1] In forced-choice experiments, subjects can be made to categorize vowel sounds. If these categorized sounds are plotted in formant space, they can be seen to divide the space into regions: within each region, all sounds are classified as the same vowel. Each vowel, therefore, has a category region associated with it.

only compensate when vowel sounds were shifted to different category regions. This predicts the timecourse of subjects' compensation response to an increasing feedback alteration would exhibit steps. Such steps would occur when the feedback alteration had increased to a point where it shifted vowels into the next category.

The other timecourse question concerns adaptation. Study 1 showed that compensation was, in general, greater than adaptation. This difference suggests different mechanisms may underlie compensation and adaptation. One testable characteristic of these mechanisms is whether they have different responses to the increasing feedback alteration.

5.1.3 Generalization of Speech SA

Speech SA can be used to examine phonetic structure issues in speech production. If a vowel's production has been adapted, then the process controlling its production has been altered. If a vowel's adaptation *generalizes* – i.e., its adaptation in one utterance causes production changes in different utterances, then the vowel's altered control process must also used in the production of other utterances. In this way, speech SA generalization can be used to observe organization of the processes controlling utterance productions. This allows inferences to be made about what phonetic representations could underlie the observed organization.

This ability to use speech SA to analyze phonetic structure issues in speech production was cited in Chapter 1 as a key motivation for studying speech SA. In Chapter 1, two speech SA generalization experiments were proposed, each investigating a different type of generalization. A major objective of Study 2 was to carry out these proposed experiments.

5.1.3.1 Context Generalization

The first proposed experiment was an investigation of *context generalization*: how a vowel's adaptation in one word context affects its production in other word contexts. The purpose of investigating context generalization is to examine the mechanisms underlying word production.

Suppose adaptation of a vowel in one word does not affect its production in other words. Then the process which controls the vowel's production in one word must not be used in other words. This suggests that words have independent, direct means of controlling their productions.

On the other hand, suppose adaptation of a vowel in one word does affect its production in other words. Then the process controlling the vowel's production in one word must be also used in other words. This sharing of the vowel's control process suggests that a common vowel representation is used to access the shared control process. This suggests, more generally, that words specify their productions indirectly via shared, intermediate production unit representations (e.g. phonemes).

Thus, in an investigation of context generalization, different word production mechanisms are implied by the possible experiment outcomes.

5.1.3.2 Target Generalization

The second proposed experiment was an investigation of *target generalization*: how one vowel's adaptation affects the production of other vowels. The purpose of investigating target generalization is to examine the structure of vowel representations.

Suppose adaptation of one vowel does not affect the production of other vowels. Then the process controlling the adapted vowel is not used in the production of the other vowels. This suggests vowels may have independent representations.

On the other hand, suppose adaptation of one vowel does affect the production of other vowels. Then the adapted vowel's altered control process must be used in the production of other vowels. This shows that vowel representations are not independent: that they do not specify entirely different vowel production control processes. This dependence of vowel representations would suggest the representations share some set of common features.

Thus, in an investigation of target generalization, different vowel representations are implied by the possible experiment outcomes.

5.2 Methods

The purposes of Study 2 were served by a single experiment. This experiment consisted of 422 epochs, where each epoch consisted of 10 word promptings. For the average subject, this meant the experiment had a duration of about 2 hours.

Each word prompting included prompting for a target duration of 300ms. In whispering the target word, the subject attempted match this target duration. While whispering, the subject heard one of the following in his earphones:

- unaltered feedback (0.0 feedback transform),
- altered feedback (-2.0 or +2.0 transform, depending on the subject), or
- masking noise to prevent him from hearing his whispering.[2]

5.2.1 Prompted Words

In Study 2, subjects were prompted to whisper words from a set of training words and from a set of testing words. These word sets were called $\mathbf{W}_{\mathbf{train}}$ and $\mathbf{W}_{\mathbf{test}}$, respectively. They differed in the type of feedback the subject heard while whispering them:

- Training words ($\mathbf{W}_{\mathbf{train}}$) were whispered while the subject heard feedback of his whispering or while his hearing was blocked by noise. Data from these word productions were used to assess compensation and adaptation.
- Testing words ($\mathbf{W}_{\mathbf{test}}$) were whispered only while the subject's hearing was blocked by noise. Data from these word productions were used to assess generalization of adaptation.

5.2.1.1 Training Words

The set of training words ($\mathbf{W}_{\mathbf{train}}$) was a set of four CVC [ɛ] words:

[2] Both the altered and unaltered feedback also included low-level masking noise. This was done to hamper the subject's ability to hear his actual whispering. See Section 3.4 for more details.

$$\mathbf{W}_{\textbf{train}} \quad = \quad \{ \quad \text{“pep”}, \quad \text{“peb”}, \quad \text{“bep”}, \quad \text{“beb”} \quad \}$$

These words were chosen for several reasons. First, both the beginning and ending consonants are bilabials (produced with the lips meant less interfering coarticulation with the vowel (produced with the tongue body). This tended to result in longer, cleaner steady-state vowel portions of the utterances.

The second reason for choosing these words was that their whispered productions are acoustically similar, but their phonetic representations and spellings are noticeably different. This, it was hoped, would allow collection of nearly identical utterance data while providing a sufficiently varied task to hold the subject's attention.

These words are acoustically similar in whispered speech because their only differing feature is the voicing of their consonants. [p] and [b] are both articulated with the lips the same way, they differ only in their voicing features such as voice onset time and amount of prevoicing. Considering only the upper vocal tract, the articulations of the $\mathbf{W}_{\textbf{train}}$ words are almost identical, producing nearly identical utterance data. It was hoped, however, that the phonetic, orthographic, and conceptual differences between these words, (along with their random presentation) would force the subject pay attention to which word he was whispering.

5.2.1.2 Testing Words

The set of testing words ($\mathbf{W}_{\textbf{test}}$) consisted of two subsets: $\mathbf{W}_{\textbf{test}}$**-context** and $\mathbf{W}_{\textbf{test}}$**-target**.

The $\mathbf{W}_{\textbf{test}}$**-context** subset was used to assess context generalization. This consisted of four CVC words, one of which was "pep" - a member of $\mathbf{W}_{\textbf{train}}$, and the rest in which the vowel [ɛ] was retained but the consonant context was varied. This subset specifically contained:

$$\mathbf{W}_{\textbf{test}}\textbf{-context} \quad = \quad \{ \quad \text{“pep”}, \quad \text{“peg”}, \quad \text{“gep”}, \quad \text{“teg”} \quad \}$$

The $\mathbf{W}_{\mathbf{train}}$ word "pep" was included in this set so that there would be a $\mathbf{W}_{\mathbf{train}}$ word prompted for during the same part of each epoch as the other $\mathbf{W}_{\mathbf{test}}$ words. This allowed production changes of $\mathbf{W}_{\mathbf{test}}$ words to be compared with production changes of a $\mathbf{W}_{\mathbf{train}}$ word whispered under the same conditions.

The other $\mathbf{W}_{\mathbf{test}}$**-context** words were chosen to assess how context-specific the adaptation of [ε] in the training words was. The questions of interest were: does adaptation of [ε] in the bilabial CVC training words affect only the production of [ε] in other words with the same bilabial CVC context, or will it also affect [ε] in (1) words sharing only the same initial CV, (2) words sharing only the final VC, or (3) all words with the same V ([ε])? The set of words besides "pep" that were chosen to be in the $\mathbf{W}_{\mathbf{test}}$**-context** subset were designed to answer these questions:

- No $\mathbf{W}_{\mathbf{test}}$**-context** word besides "pep" shared the same complete bilabial CVC $\mathbf{W}_{\mathbf{train}}$ word context. Thus, if no $\mathbf{W}_{\mathbf{test}}$**-context** word besides "pep" is affected by $\mathbf{W}_{\mathbf{train}}$ word adaptation, then the adaptation of [ε] is specific to the complete bilabial CVC $\mathbf{W}_{\mathbf{train}}$ word context.

- "peg" is the only $\mathbf{W}_{\mathbf{test}}$**-context** word (besides "pep") that shares the same initial bilabial CV syllable of the $\mathbf{W}_{\mathbf{train}}$ words. However, its final consonant is not bilabial. If only "peg" is affected by $\mathbf{W}_{\mathbf{train}}$ word adaptation, then the adaptation of [ε] is specific to the initial CV context.

- "gep" is the only $\mathbf{W}_{\mathbf{test}}$**-context** word (besides "pep") that shares the same final bilabial VC syllable of the $\mathbf{W}_{\mathbf{train}}$ words. However, its initial consonant is not bilabial. If only "gep" is affected by $\mathbf{W}_{\mathbf{train}}$ word adaptation, then the adaptation of [ε] is specific to the final VC context.

- "teg" was chosen because neither of its consonants are the bilabial; it shares only the same vowel as the $\mathbf{W}_{\mathbf{train}}$ words. This provides an important control condition to compare with the generalization results seen in "peg" and "gep": if adaptation generalization is selective to either the CV or VC context, then $\mathbf{W}_{\mathbf{train}}$ adaptation should not affect "teg".

The $\mathbf{W_{test}}$**-target** subset was used to assess target generalization. This consisted of four CVC words which shared the same bilabial consonant context of the $\mathbf{W_{train}}$ words but varied the vowel:

$$\mathbf{W_{test}\text{-}target} = \{ \text{“peep”}, \text{“pip”}, \text{“pap”}, \text{“pop”} \}$$

Besides allowing for simultaneous testing for context and target generalization, the inclusion of the $\mathbf{W_{test}}$**-target** word set made a more varied distribution of speech sounds for the subject to produce. This variety, along with the four-word $\mathbf{W_{train}}$ set, was intended to minimize the chance of articulatory changes resulting from over-repetition of any one speech sound.

5.2.2 Timecourse of the Experiment

The experiment consisted of 422 epochs, where each epoch consisted of 10 word promptings. As mentioned above, for the average subject this meant the experiment's duration was about 2 hours.

5.2.2.1 Timecourse of Each Epoch

The timecourse of each epoch is shown in the following table:

part	words	word set	subject	data
	prompted		heard	collected
1.	4	$\mathbf{W_{train}}$	feedback	
2.	1	$\mathbf{W_{train}}$	feedback	•
3.	1	$\mathbf{W_{train}}$	noise	•
4.	4	$\mathbf{W_{test}}$	noise	•

The table shows that each epoch was divided into four parts. In Part 1, the subject whispered 4 $\mathbf{W_{train}}$ words while he heard feedback of his whispering. No utterance data were collected. In Part 2, the subject whispered 1 $\mathbf{W_{train}}$ word, again while he heard feedback. This time, the subject's utterance data were collected. In Part 3, the subject whispered 1 $\mathbf{W}_{\mathbf{train}}$ word, this time with feedback blocked by masking noise. Again, his utterance data was collected. Finally, in Part 4, the subject whispered 4 $\mathbf{W}_{\mathbf{test}}$ words, again with feedback blocked by masking noise, and again his utterance data was recorded.

This timecourse gave each epoch the following features:

- It consisted of 10 word promptings, the first 6 of which were randomly selected from $\mathbf{W}_{\mathbf{train}}$, while the last 4 were randomly selected from $\mathbf{W}_{\mathbf{test}}$.

- The subject whispered the first 5 words (all from $\mathbf{W}_{\mathbf{train}}$) while he heard feedback of his whispering. He whispered the last 5 words (1 from $\mathbf{W}_{\mathbf{train}}$, followed by 4 from $\mathbf{W}_{\mathbf{test}}$) while prevented from hearing his whispering by masking noise.

- To minimize the space used to store data, only the last of the $\mathbf{W}_{\mathbf{train}}$ words whispered while the subject heard feedback was recorded. All other utterances produced by the subject were recorded.

Thus, over any sequence of epochs, promptings to produce $\mathbf{W}_{\mathbf{train}}$ words were interleaved with promptings to produce $\mathbf{W}_{\mathbf{test}}$ words. Since the $\mathbf{W}_{\mathbf{test}}$ words had different consonants and vowels, this insured there were no long intervals where the subject repeatedly whispered the same vowel or word. Also, over any sequence of epochs, what the subject heard switched every five words between feedback of his whispering and masking noise. This insured that there were no long intervals where the subject repeatedly whispered words under the same feedback conditions.

In sum, the epoch timecourse was designed to keep the whispering task and whispering conditions as varied as possible throughout the experiment.

5.2.2.2 Epoch Sequence

The experiment's 422 epochs were divided into a sequence of five phases, as shown by the following table:

phase	time	(epochs)	feedback was
Warmup	10min	(36)	unaltered
Baseline	17min	(60)	unaltered
Ramp	20min	(66)	altered
Train	1 hour	(200)	altered
Test	17min	(60)	unaltered

The warmup phase consisted of 36 epochs, which, for the average subject, had a duration of 10 minutes. In this phase, when the subject heard feedback of his whispering, the feedback was unaltered (i.e., the 0.0 feedback transformation was used).

The purpose of the warmup phase was to provide time for the subject to acclimate to the experimental conditions. It was expected that 10 minutes would be sufficient time for this acclimation to occur and the subject's whisperings to stabilize.

The baseline phase consisted of 60 epochs, which, for the average subject, had a duration of 17 minutes. In this phase, when the subject heard feedback of his whispering, the feedback was unaltered (i.e., the 0.0 feedback transformation was used).

The purpose of the baseline phase was to collect baseline data of the subject's whisperings before his feedback was altered. Utterance data collected in this phase was compared with data collected in later phases to assess whether the subject changed his whispering in response to the altered feedback.

The ramp phase consisted of 66 epochs, which, for the average subject, had a duration of 20 minutes. This phase was further subdivided into 11 *stages*, each of which was 6 epochs long (approximately 2 minutes in duration).

Within each stage, the amount of feedback alteration was held constant. In the first stage (Stage 0), when the subject heard feedback of his whispering, it was unaltered (i.e., the 0.0 feedback transformation was used). Over the next 10 stages, however, the amount of feedback alteration was incremented between stages. In this way, the amount of feedback alteration was linearly increased to its maximum magnitude in 10 stages. Thus, in the last stage (Stage 10), when the subject heard feedback of his whispering, it was maximally altered. This maximum feedback alteration was either -2.0 or +2.0 vowel units, depending on the subject (see discussion of subjects below).

The purpose of the ramp phase was to gradually introduce the feedback alteration to which the subject would be exposed for the rest of the experiment. As discussed in Section 5.1 above, the main reason for gradual introduction was to minimize the subject's awareness of the altered feedback. A second reason was to allow analysis of how a subject's compensation and adaptation developed in response to an increasing feedback alteration.

The train phase consisted of 200 epochs, which, for the average subject, had a duration of 1 hour. In this phase, when the subject heard feedback of his whispering, the feedback was altered by either the -2.0 or +2.0 feedback transformation (depending on the subject).

The purpose of the train phase was to give the subject roughly one hour of exposure to the full-strength feedback transformation.

The test phase consisted of 60 epochs, which, for the average subject, had a duration of 17 minutes. In this phase, when the subject heard feedback of his whispering, the feedback was altered by either the -2.0 or +2.0 feedback transformation (depending on the subject).

Except for the altered feedback, the test phase was essentially a repeat of the baseline phase. The purpose of the test phase was to collect utterance data that could be compared with utterance data from the baseline phase. Formant changes seen in this comparison were used to assess whether the subject changed his whispering in response to the altered feedback.

5.2.3 Post-Experiment Interview

At the end of the experiment, the subject was interviewed briefly. In this interview, the subject was asked a variety questions about his experience in the experiment. The questions concerned different aspects of the experiment (e.g. "Were the prompted words readable on the video monitor?").

The purpose of asking a variety of questions was to disguise the importance of questions concerning the subjects feedback. These questions were:

- Did the feedback sound correct?
- Were there any problems with how it sounded?
- Was the feedback too loud?

These questions were posed to determine if the subject noticed anything unusual about the feedback or any change in his vowel productions.

5.2.4 Subjects

The experiment was run on 8 male native speakers of North American English who were either undergraduate or graduate students at MIT. All were naive to the purpose of the study, and none was a subject in study 1.

Each subject passed the pretest and screening procedure described in Section A.3. This pretest was performed on a separate day. The pretest measured formants of subjects [i]–[ɑ] path vowels, which were needed to construct the feedback transformations. The pretest's screening also insured that all subjects had strong formants for most vowels, and that the -2.0 and +2.0 transformations of their vowels sounded correct.

Two experiments were performed with each subject: a real experiment with either the -2.0 or +2.0 transformation, and a control experiment. This control experiment was identical to the real experiment except that a strength 0.0 feedback transformation (no feedback alteration) was used throughout the experiment.

In the real experiment, half the subjects were run with the +2.0 transformation, and half were run with the -2.0 transformation:

- Subjects RS, CW, TY, and VS were run with the +2.0 transformation.
- Subjects AH, OB, RO, and SR were run with the -2.0 transformation.

5.3 Results and Discussion

The results of Study 2 are discussed in two sections: in this section, the overall experiment results are discussed. If individual subject results are of interest, they can be found in Section 5.4, which compiles plots and descriptions of each subject's individual results.

Discussion of the overall experiment results is divided into three sections, each analyzing a different aspect of the experiment data:

- **Compensation and adaptation results:** analysis of subjects' overall compensation and adaptation.
- **Timecourse results:** analysis of how subjects' compensation and adaptation developed in response to gradual introduction of altered feedback.
- **Generalization results:** analysis of how adaptation of the training words affected production of the testing words.

5.3.1 Compensation and Adaptation Res**u**lts

We begin by considering those results pertinent to the fundamental characteristics of the speech SA effect. This discussion is divided into three sections. First, measures of mean compensation, adaptation, and path deviation change are discussed. Second, the path projection measures from which compensation and adaptation are computed are examined in more detail. Third, the results concerning subjects' awareness of the altered feedback are discussed. Finally, the findings of these sections are summarized and explained in a theory of speech SA.

5.3.1.1 Mean Compensation, Adaptation, and Path Deviation Change

The mean compensation, adaptation and path deviation change results are analyzed in two ways. First, plots displaying these response measures for each subject are discussed. The significance of the trends seen in these plots are then assessed in statistical tests of the results collapsed across subjects.

Across-Subject Plots Figure 5-1 shows mean compensation and path deviation change for each subject's response to the altered feedback. The figure's layout is the same as that of Figure 4-10. The left column shows plots of subjects' *compensation responses*: vowel production changes observed when subjects could hear feedback of their whispering. The right column shows plots of subjects' *adaptation responses*: vowel production changes that subjects retained when prevented from hearing their whispering by masking noise. In each column, the top plot shows mean compensation while the bottom plot shows mean path deviation change.

Results from both the real experiments (in which feedback was altered by either the +2.0 or -2.0 transformation) and control experiments (in which only the 0.0 transformation was used) are shown for each subject. In each graph, a solid line links subjects' results from the real experiment and a dotted line links the control experiment results.

Considering first the real experiment results (solid lines linking filled circles), we see they are similar to those of Study 1[3]. For subjects' compensation responses (left side of the figure):

- Mean compensation (Figure 5-1(a)) is consistently positive, but its amount varies widely across subjects.
- Mean path deviation change (Figure 5-1(b)) is inconsistent across subjects in its direction.

For subjects' adaptation responses (right side of the figure):

- Mean adaptation (Figure 5-1(c)) appears highly correlated with mean compensation, and is positive for all but one subject.[4] For each subject, it also appears that adaptation is generally less than compensation.
- Mean path deviation change (Figure 5-1(d)) is again inconsistent across subjects in its direction.

The differences between these results and those of Study 1 are largely a matter of scale: In Study 2, mean compensation is generally greater, while mean path deviation change is generally smaller. This is likely due to the much longer training phase of study 2.

In Study 1, two conclusions were drawn from the results pattern: (1) subjects selectively alter path projections of their vowel productions to compensate for the altered feedback; (2) they retain their altered productions when whispering with feedback is blocked with masking noise. These conclusions are equivalent to stating that speech exhibits SA, as defined in Chapter 1.

Study 2 manipulated more experimental factors than Study 1, and the effects of these factors on subjects' performance provide stronger and more detailed support of the Study 1 conclusions.

[3] For the analogous plot of Study 1's results, see Figure 4-10

[4] Recall that *mean adaptation* is the term used to refer to mean compensation of a subject's adaptation response.

In Study 2, half the subjects were exposed to the -2.0 feedback transformation, and half to the +2.0 transformation (see Section 5.2.4). Mean compensation is measured relative to the direction of the feedback transformation. The measure is positive for compensating production changes, independent of the transformation's shift direction.[5] The plots show that this measure is generally positive across all subjects. This indicates subjects generally altered their production so as to oppose the shift of the transformation, regardless of the direction of that shift. It is therefore unlikely that the observed compensatory behavior is the result of some drift in subjects' vowel productions not related to altered feedback exposure.

Study 2 also included running each subject in a control experiment in which no feedback transformation was applied (i.e. the 0.0 transformation was used). The results from the control experiments provide further evidence that subjects compensate.

For mean compensation and adaptation, subjects showed a consistent difference between the real and control experiments. Figure 5-1(a) shows that, for most subjects, mean compensation is significantly larger in the real experiment than in the control experiment. Figure 5-1(c) shows the same is true for mean adaptation.

For mean path deviation change, subjects showed no consistent difference between the real and control experiments. Some subjects show no significant difference between real and control experiments (RS, RO, and AH in Figure 5-1(b); RS, OB, RO, TY, VS, and AH in Figure 5-1(d)). For others, path deviation change in the real experiment was greater that in the control experiment (CW, OB and VS in Figure 5-1(b), CW in Figure 5-1(d)). For still others, path deviation change in the control experiment was greater (SR, TY in Figure 5-1(b), SR in Figure 5-1(d)).

These control experiment comparisons show that the only aspect of subjects' responses consistently affected by exposure to altered feedback is mean compensation and adaptation. This is exactly the predicted effect, since the feedback transformations selectively alter only path projections. The results of Study 2 therefore offer strong evidence that subjects change vowel productions specifically to compensate for

[5] For the definition of mean compensation, see Section 3.5.3.4.

alterations of their feedback.

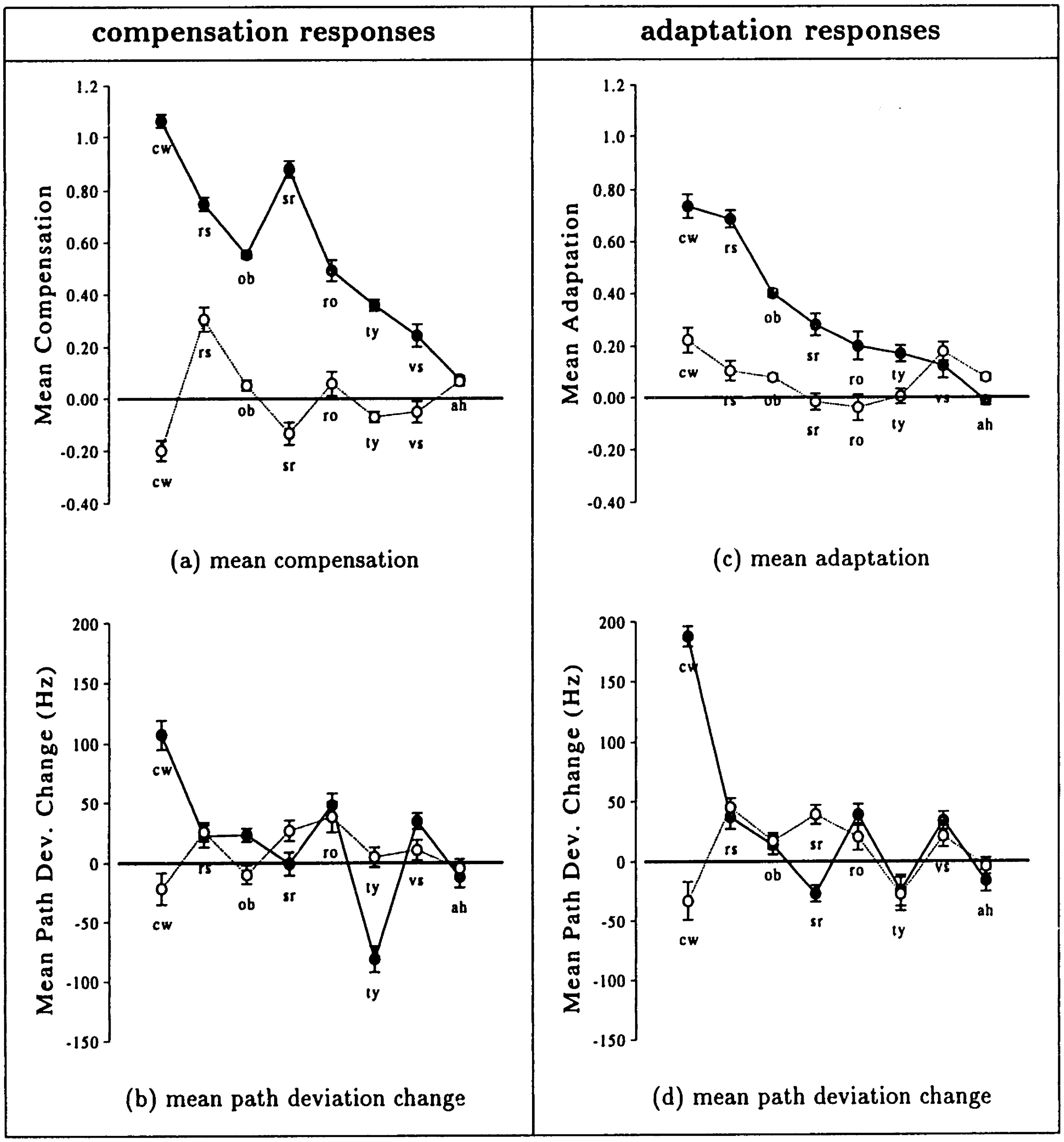


Figure 5-1: Mean compensation and path deviation change for each subject. Plots (a) and (b) show mean compensation and path deviation change for each subject's compensation response. Plots (c) and (d) show the same for each subject's adaptation response. In each plot, black dots linked by a solid line indicate real experiment data, while white dots linked by a dotted line indicate control experiment data. Small bars around each dot indicate confidence intervals. (Note that mean compensation of a subject's adaptation response is called *mean adaptation.* Note also mean compensation is a dimensionless ratio, while mean path deviation change is measured in Hz. See Section 3.5.3 for further explanation.)

Tabulated Statistics As with Study 1, statistical significance of the trends discussed above was tested with two measures of across-subject path projection and deviation change:

- F ratios and p values derived from ANOVA tests.
- Mean compensation and path deviation change, averaged across subjects.

ANOVA Tests Table 5.1 summarizes the F ratios and p values of ANOVA tests of subjects' responses. As indicated by column 1 of the table, subjects' compensation and adaptation responses were analyzed separately. The table's first three rows report tests of subjects' compensation responses:

- Row 1 tests significance of the experiment phase factor in the real experiment. This test determined if subjects' path projections (columns 4 and 5) and path deviations (columns 6 and 7) changed significantly between the baseline and testing phases. The test shows the path projection change to be highly significant ($p < 0.002$) and the path deviation change to be insignificant ($p < 0.877$).
- Row 2 tests significance of the same factor in the control experiment. In this case, neither path projections ($p < 0.951$) nor path deviations ($p < 0.681$) showed significant change.
- Row 3 tests significance of the interaction between the experiment and phase factors. This test determined if response changes in the real experiment differed significantly from those in the control experiment. The test shows there was a significant difference in subjects' path projection responses between real and control experiments ($p < 0.006$), but there was an insignificant difference in subjects' path deviation responses ($p < 0.772$).

Rows 4, 5, and 6 report results of the same tests of subjects' adaptation responses:

- Row 4 shows that, in the real experiment, there was highly significant path projection change ($p < 0.011$) and insignificant path deviation change ($p < 0.290$).

- Row 5 shows that, in the control experiment, there was marginally significant path projection change ($p < 0.047$) and insignificant path deviation change ($p < 0.411$).

- Row 6 shows that, there was a significant difference in subjects' path projection responses between real and control experiments ($p < 0.023$), but there was an insignificant difference in subjects' path deviation responses ($p < 0.215$).

response type	experiment	factor	effect on			
			path proj.		path dev.	
			F(1,7)	p <	F(1,7)	p <
compensation	real	phase	22.325	**0.002**	0.026	**0.877**
	control	phase	0.004	**0.951**	0.184	**0.681**
	both	expr-phase	15.362	**0.006**	0.091	**0.772**
adaptation	real	phase	11.590	**0.011**	1.307	**0.290**
	control	phase	5.819	**0.047**	0.764	**0.411**
	both	expr-phase	8.369	**0.023**	1.858	**0.215**

Table 5.1: Path projection and deviation ANOVA tests.

In sum, the ANOVA tests show two characteristics of subjects' path projection changes:

1. In the real experiment, these changes were significant.

2. The changes seen in the real experiment were significantly different from those seen in the control experiment.

These characteristics are seen in subjects' compensation and adaptation responses.

On the other hand, none of the ANOVA tests showed any significant path deviations changes.

Mean Compensation and Path Deviation Change Table 5.2 summarizes the values of mean compensation (columns 3 and 4) and mean path deviation change (columns 5 and 6), averaged across subjects. Rows 1 through 3 in the table show average values for subjects' compensation responses. Row 1 shows average values for the real experiment, while row 2 shows average values for the control experiment. Row 3 shows the average difference per subject in mean values between real and control experiments. Rows 4 through 6 show these same average values for subjects' adaptation responses.

response type	experiment	compensation		path dev. change (Hz)	
		μ	σ_μ	μ	σ_μ
compensation	real	**0.55**	0.12	**18**	19
	control	**0.00**	0.06	**9**	7
	diff/subj	**0.55**	0.05	**9**	13
adaptation	real	**0.32**	0.10	**31**	24
	conrol	**0.08**	0.03	**10**	10
	diff/subj	**0.25**	0.05	**21**	14

Table 5.2: Mean compensation and path deviation changes, averaged across subjects.

The values seen in the table are consistent with the ANOVA results. In subjects' compensation responses, average mean compensation in the real experiment is large compared to its standard error (0.55 ± 0.12), while average mean compensation in the control experiment is small compared to its standard error (0.32 ± 0.10). In addition, the difference between a subject's mean compensation in the real experiment and his mean compensation in the control experiment is, on average, large compared to its standard error (0.55 ± 0.05). This same pattern is seen in subjects' adaptation responses.

On the other hand, all measured mean path deviation changes are all the same order of magnitude as their standard errors.

In sum, the statistical tests confirm the significance of the key trends seen in Figure 5-1. In both compensation and adaptation responses, mean compensation was significant across subjects and significantly greater in the real experiments than in the control experiments. However, in no case was mean path deviation change significant across subjects.

From the mean compensation and path deviation analysis, we conclude therefore that subjects changed vowel productions specifically to compensate for alterations of their feedback. We also conclude that these production changes are strong enough to be partly retained when feedback is blocked by noise. The results of Study 2 thus provide strong confirmation that speech exhibits SA.

5.3.1.2 Path Projection Analysis

More aspects of this SA effect can be seen by separately examining baseline and testing phase path projection values. The results are shown in figures 5-2 and 5-3.

Figure 5-2 shows subjects' mean path projections for the baseline and testing phases of both the real experiment (solid lines) and control experiment (dotted lines).[6] The left plots show mean path projection for subjects' compensation responses: the top plot shows testing phase values; the bottom plot shows baseline phase values. The right plots show the same for subjects' adaptation responses.

In each plot, comparison of subjects' path projections is facilitated by a normalization that makes path projection increases indicate compensation for all subjects. This normalization is needed because, depending on the feedback transform he is exposed to, a subject compensates by either increasing or decreasing vowel path projections:

- Subjects AH, OB, RO, and SR were exposed to the -2.0 feedback transformation in the real experiment and are referred to as the *-2.0 subjects.* The -2.0 transformation decreases perceived path projection by 2 vowel units, and the -2.0 subjects compensated for it by increasing their vowel path projections.

- Subjects RS, CW, TY, and VS were exposed to the +2.0 feedback transformation in the real experiment and are referred to as the *+2.0 subjects.* The +2.0 transformation increases perceived path projection by 2 vowel units, and the +2.0 subjects compensated for it by decreasing their vowel path projections.

To make the -2.0 and +2.0 subjects' responses comparable, the +2.0 subjects' path projections were calculated using an [i]–[ɑ] path with reversed numbering (i.e., with

[6] Recall that path projection is measured in inter-vowel intervals along the subject's [i]–[ɑ] path. Normally, on this scale, 1.0 corresponds to [i], 2.0 to [ɩ], 3.0 to [ɛ], 4.0 to [æ], and 5.0 to [ɑ]. Path projection and the [i]–[ɑ] path are discussed in detail in Section 3.3.

5.0 corresponding to [i], 4.0 to [ɩ], 3.0 to [ɛ], 2.0 to [æ], and 1.0 to [ɑ]). Reversing the numbering converts the +2.0 subjects' compensation responses from path projection decreases to path projection increases. This makes all subjects' results comparable in the plots: for all subjects, an increase in path projection indicates compensation.

These plots display several striking features. First consider Figure 5-2(b), which shows baseline mean path projections of subjects' compensation responses. For all subjects but TY, VS, and AH (the poorest adaptors), mean path projections are higher in the control experiment than in the real experiment. Of the remaining subjects, all but SR show this same difference in their baseline adaptation responses (Figure 5-2(c)). ANOVA tests show that this baseline difference is significant across all subjects for both responses ($p < 0.014$ for compensation responses; $p < 0.043$ for adaptation responses).

The reverse of this situation is true for the testing phase: in this case, path projection seen in the real experiment is generally higher than that seen in the control experiment. The observed difference is significant for compensation responses ($p < 0.020$) and marginally insignificant for adaptation responses ($p < 0.080$). However, if subjects VS and AH (the poorest adaptors) are excluded, the difference is also significant for adaptation responses.

These differences are summarized in Figure 5-3, which shows the path projections of Figure 5-2 averaged across subjects. As with Figure 5-2, increasing path projection values represent path projection changes in the compensating direction.

The figure shows subjects' compensation and adaptation in the real experiment. The solid line in Figure 5-3(a) shows subjects' compensation responses in the real experiment. These responses have an average baseline path projection close to 3.0 (the path position of [ɛ]) However, by the test phase, this average has shifted about 1.0 vowel unit in the compensating direction. Table 5.1 showed this shift is highly significant ($p < 0.002$). The solid line in Figure 5-3(b) shows subjects' adaptation responses in the real experiment. These responses have an average baseline path projection that is slightly higher than that of the baseline compensation responses. By the test phase, this average has shifted in the compensating direction, though not by as much as the shift seen in subjects' compensation responses. As reported in Table 5.1 this shift is also significant ($p < 0.011$).

Figure 5-3 also shows that, although no appreciable production changes occur in the control experiment, baseline responses are shifted in the compensating direction. The dotted line in Figure 5-3(a) shows subjects' compensation responses in the control experiment. These responses have an average baseline path projection close to 3.4. Relative to the same responses in the real experiment, this represents a significant shift in the compensating direction ($p < 0.014$ in the ANOVA tests discussed above). The figure also shows an insignificant shift in average path projection from the baseline to the test phase ($p < 0.951$ in Table 5.1). The dotted line in Figure 5-3(b) shows subjects' adaptation responses in the control experiment. These responses have an average baseline path projection that is slightly lower than that of the baseline compensation responses. By the test phase, however, average path projection has increased to equal the compensation response value. Table 5.1 showed this increase was marginally significant.

Several notable features of subjects' control experiment responses warrant further discussion.

Baseline Path Projection Lowering The first feature is the slight lowering of baseline average path projection in subjects' adaptation responses, as compared to their compensation responses. In the test phase, subjects' adaptation responses have recovered from exhibiting the lowered path projection: test phase average path projection equals the compensation response value. This recovery makes subjects appear to exhibit marginally significant adaptation in the control experiment.

One explanation for these results involves priming: in the control experiment subjects were primed to respond the same way they did in the real experiment. For each subject, the control experiment was performed many days after the real experiment. In the real experiments, subjects shifted path projections in the compensating direction, both when they heard feedback and when they heard masking noise. Subjects may have retained a tendency to make these path projection shifts that was primed by being in the same experimental setup for the control experiment. Such priming effects of context have been well documented in studies of implicit memory (for a review, see [Schacter, 1995]).

This explanation would predict path projection shifts in both compensation and adaptation responses. However, no such shifts were seen in subjects' compensation responses. Thus, we must modify the explanation by supposing that the tendency to shift path projections was inhibited when subjects could hear feedback of their whispering.

Another explanation for these results comes from the analysis of Study 1's results. There, a similar results pattern was seen in for subjects MF and JK and was attributed to *F1 lowering.*[7] Both subjects showed apparently greater adaptation than compensation (subject JK showed no compensation). This resulted from the lowered baseline path projection of their adaptation responses, as compared to their compensation responses. For both subjects, it was clear that this baseline path projection lowering resulted from F1 lowering: baseline F1 had a lower frequency in their adaptation responses than in their compensation responses. Hearing the masking noise apparently caused the subjects to initially lower their normal productions of F1. Over the course of the experiment, subjects reduced how much they lowered F1, which could be explained by their habituating to the presence of masking noise.

Of the two suggested explanations, the F1 lowering hypothesis appears more likely because it explains more of the results with fewer added assumptions. It not only explains the adaptation response path projection increase, but also explains the lowered baseline path projection, as well as the recovery in the test phase to the compensation response path projection value.

Lack of a Compensation Response Subjects do not show a compensation response in the control experiment. Average path projection is approximately 3.4 in the baseline phase and remains unchanged by the end of the experiment. Table 5.1

[7] For detail on F1 lowering beyond this paragraph's summary description, see sections 4.4.1.3 and 4.4.1.4.

confirms the lack of significant change in this case ($p < 0.951$).

On one hand, the lack of compensation response seems predictable because feedback was not altered in the control experiment. Thus, there were no feedback alteration effects to compensate for.

On the other hand, the lack of compensation response seems surprising. In the subject pretest and in the real experiment, subjects produced [ε] with a path projection of 3.0 vowel units. However, in the control experiment, baseline path projections differed, on average, by 0.4 vowel units from 3.0. Below, we will discuss possible explanations for this difference. Here, we consider why subjects apparently feel no compulsion to reset their [ε] path projections to 3.0 during the control experiment.

One explanation for the results is that subjects are insensitive to path projection differences of only 0.4 vowel units. However, as will be seen in below in Section 5.3.2, at least half the subjects showed compensations for feedback alterations of only 0.2 vowel units.

Another explanation for the results is that subjects don't retain long-term memories of their whispered vowel sounds. Without such memories, they would not have an absolute reference from which to judge correctness of the sounds of their vowel productions. In this explanation, subjects' initial articulations of [ε] would set their reference memory of what [ε] should sound like. This memory would be used for the rest of the experiment to judge sound correctness of subsequent articulations of [ε]. If, as during the real experiment, a feedback transform made the perceived sound of [ε] differ from the reference memory, compensating productions would be induced. However, during the control experiment, the sounds of articulations of [ε] were not altered. In this case, these [ε] sounds would not differ from the reference memory, and no production alterations would be induced.

The Baseline Shift The most striking feature of subjects' control experiment responses is the shift of baseline path projection in the compensating direction. For subjects' compensation responses, this shift is, on average, about 0.4 vowel units. For their adaptation responses, the shift is slightly less (as discussed above).

The shift suggests a partial retention of the productions changes in [ε] that were induced in the real experiment. As noted above, this is theoretically possible because, for each subject, the control experiment was performed after the real experiment. However, it is useful to be more specific about the time interval between real and control experiments for each subject. Table 5.3 tabulates this information: the last column of the table shows the number of days between the real experiment and the control experiment for each subject. The table shows the following features:

- For all subjects but one, the time interval between real and control experiment was greater than *30 days.*

- For the one subject whose interval was less (subject RS), the interval was only two days. Yet figures 5-2(b) and 5-2(d) show his control baseline shift was about the same as that of other subjects.

subject	when expr. run (month/day, time)		time diff.
	real exp.	control exp.	(days)
CW	4/06, 1:12 PM	5/27, 9:05 AM	51
RS	5/28, 11:41 AM	5/30, 9:01 AM	2
OB	4/03, 3:16 PM	5/21, 2:50 PM	48
SR	4/16, 2:30 PM	5/16, 1:31 PM	30
RO	4/10, 1:06 PM	5/17, 12:58 PM	37
TY	4/15, 12:43 PM	5/20, 12:55 PM	35
VS	4/13, 1:16 PM	5/16, 9:33 AM	33
AH	4/04, 3:36 PM	5/14, 12:59 PM	40

Table 5.3: Dates on which subjects were run in the experiments of Study 2. All subjects were run in 1996.

It therefore appears that most subjects' compensating production changes, induced in the real experiment, were retained over a period of more than a month.

That production changes could be retained long-term in absence of feedback requires only the assumption of stability of the speech production system. Such stability is evident in the speech of post-lingually deafened speakers: such speakers retain intelligible speech for decades after deafening [Cowie and Douglas-Cowie, 1983, Lane and Webster, 1991].

However, during the month between the real and control experiment, the subjects presumably were not denied feedback of their speech. Why didn't this month of hearing unaltered feedback of their vowels reset the subjects' productions of [ε]?

One explanation is that the speech SA experimental conditions are sufficiently novel that subjects develop representations of their vowel productions that are specific to the experiment. In this account, during the real experiment, these experiment-specific vowel representations are initially set to produce correct-sounding vowels with no feedback alteration. Later, these representations are altered to compensate for the effects of the feedback alteration that is introduced. After the experiment, these representations are sufficiently independent of their normal vowel representations that they are not completely affected by speech feedback under normal conditions. Thus, the experiment-specific vowel representations are able to retain much of their induced alterations over the month between the real and control experiments.

Another explanation is a generalization of the previous one: subjects have representations governing the control of their whispered vowels that are somewhat independent of their voiced vowel representations. To some extent this must be true: control of the glottis for whispered speech is necessarily different from glottal control for voiced speech [Titze, 1994, O'Shaughnessy, 1987].

But suppose control of other vowel tract articulators was also represented separately for voiced and whispered vowels.[8] If this were the case, whispered vowel representations would not be completely affected by voiced vowel feedback. In this account, during the real experiment, whispered vowel representations are altered to compensate for the effects of the altered feedback. After the experiment, these whispered vowel representations remain altered because: (1) they are not completely affected by feedback of voiced vowel productions and (2) whispering is an infrequent mode of speech, not likely to be used much by subjects during the month between the real and control experiments.

From the data available, it is not possible to distinguish between the above two

[8] There may be some justification for this given the differing spectral envelopes of voiced and whispered vowels (see Section A.1.2 for more detailed discussion of these spectral differences).

explanations of the retention of production changes. However, both explanations underscore the importance of studying the generality of the vowel production changes observed in the speech SA experiments of this thesis. Experiments must be designed to examine the extent to which induced vowel production changes in the experiment affect vowel productions (whispered and voiced) under normal conditions.

Thus, because it was performed after the real experiment, the control experiment may not have been as clean a control as one would want. But the results it exhibited suggest the design of future experiments to look specifically at stability and generality of the SA effect.

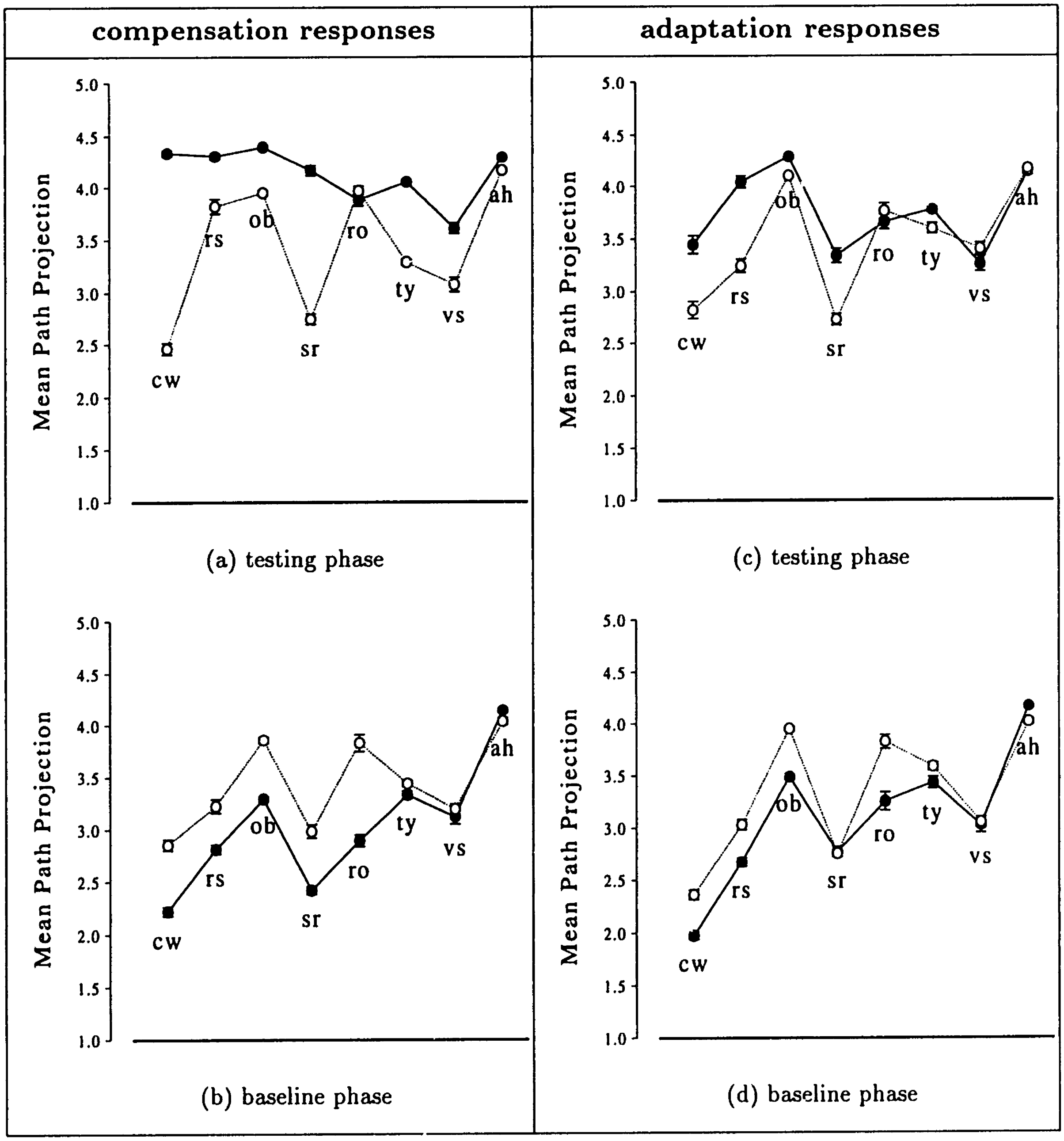
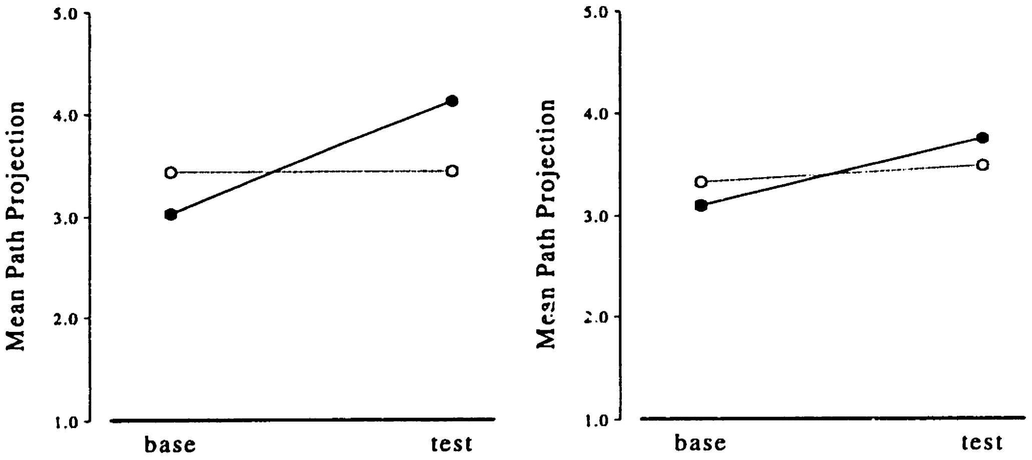


Figure 5-2: Path projections of each subject's vowel productions in the baseline and testing phases. Plots (a) and (b) show path projections for each subject's compensation response. Plots (c) and (d) show the same for each subject's adaptation response. In each plot, black dots linked by a solid line indicate real experiment data, while white dots linked by a dotted line indicate control experiment data. Small bars around each dot indicate confidence intervals.



(a) compensation responses

(b) adaptation responses

Figure 5-3: Path projections of Figure 5-2 averaged across subjects. Plot (a) shows average path projections for subjects' compensation responses. Plot (b) shows the same for subjects' adaptation responses. In each plot, the filled and open dots above the "base" label are the average path projections seen in the baseline phase of the real and control experiments, respectively. The dots above the "test" label are the average path projections seen in the test phase. The solid and dotted lines connecting the dots highlight the average path projection changes seen in the real and control experiments, respectively. (Note: confidence intervals for each average are shown but are so small that the bars representing them are obscured by the dots.)

5.3.1.3 Subjects' Awareness of the Altered Feedback

As discussed in Section 5.1.1, a key goal of Study 2 was to minimize the chance of subjects using conscious strategies to compensate for the altered feedback. For this reason, its design included (1) gradual introduction of feedback transforms and (2) post-experiment interviews to assess subjects' awareness of the altered feedback.

The results of these interviews were that no subject reported being aware of either the altering of his feedback nor his own compensatory responses to it. This suggests that the compensations and adaptations produced by subjects were not the result of conscious strategies.

However, the amount of compensation seen varied widely across subjects, with some showing little or no compensation. This raises interesting questions about the poor compensators – the subject's exhibiting little compensation: (1) why did they not compensate more, and (2) why did they not report noticing the altered feedback?

One possibility is that these subjects were somehow unable to compensate, and that the post-experiment interview was an unreliable assessment of whether subjects were consciously aware of the altered feedback. The experiment is roughly two hours long and the feedback alteration was ramped up to full strength within the first hour. During the post-experiment interview, subjects may therefore have forgotten their initial percept of altered feedback, which would have occurred minimally one hour earlier. However, it must be noted that all subjects expressed significant curiosity about the purpose of the experiment. This suggests that they were disposed to remember any unusual aspects of the experiment as clues to its purpose. Altered auditory feedback is arguably an unusual experimental aspect subjects would be likely to remember.

It is also possible that there existed some problem in feedback transform fidelity for the poorly compensating subjects.[9] However, it should be noted that (1) during subject pretest, subjects were specifically screened out whose transformed vowels had

[9] As discussed in Appendix C, there were resolution and stability problems inherent to the method of generating the feedback transformations. The magnitude of these effects depends on the geometry of a subject's path vowels in formant space. Perhaps the poorly compensating subjects' path vowels were so arranged as to exacerbate these effects.

poor fidelity, and (2) during the experiment, fidelity of the altered feedback was subjectively monitored by the experimenter. No significant transform anomalies were reported.

A second set of explanations concern speech perception. The first of these is that the poor compensators may be insensitive to the whispered vowel sound differences created by the altered feedback. Note however, that it's unlikely the subjects had a general insensitivity to these vowel sound differences: the experiment's feedback alterations were large enough to change the phonetic identity of [ε] (e.g., "pep" changed to "peep"). Insensitivity to such differences in voiced speech can sometimes be seen in non-native speakers of English, but seem unlikely in the subjects of Study 2, who were native speakers of North American English.

The second perceptual explanation for the poor compensators is that the altered feedback induced adaptation of their speech perception, not their speech production. Adaptation of speech perception has been shown in other types of experiments – specifically the selective adaptation experiments of Cooper [Cooper, 1979]. Cooper's experiments investigated shift of subjects' VOT category boundary in the perception of voiced/voiceless consonants. He found this shift could be induced by repetitive listening to one or the other of two consonants differing only by VOT. The conditions of these experiments were thus very different from the speech SA experiments of this thesis, but the existence of one type of adaptation in speech perception suggests the possibility of other types of perceptual adaptation.

5.3.1.4 Discussion

In Study 1, we concluded only that speech appeared to exhibit SA: that subjects compensate for feedback alterations, and that part of their compensation is accomplished by production changes sufficiently persistent to be observed in speech produced while their speech feedback is blocked by noise.

From the compensation and adaptation results of Study 2, we were able to confirm these basic conclusions. However, these results also revealed many more characteristics of speech SA. To summarize these characteristics, we formulate the following theory concerning subjects' response to altered auditory feedback in a speech SA experiment:

1. Perception of the altered feedback is partially offset by perceptual adaptation. The capacity to adapt perception is limited and subject-specific.
2. The perceived feedback alteration is compensated for.
3. Compensation is partly achieved by a temporary correction mechanism (active only while exposed to the altered feedback), and partly achieved by long-term adjustment of speech control.

The rationale for each hypothesis in this theory is considered in the discussion that follows.

1. Subjects' perception adapts. The perceptual adaptation hypothesis can account for two seemingly contradictory findings of Study 2: (1) the amount of compensation varied widely across subjects, yet (2) no subject reported noticing any alteration of their feedback.

The hypothesis accounts for these findings by postulating that each subject has a different capacity to adapt his perception of the altered feedback. This perceptual adaptation reduces his perception of the true amount of feedback alteration. He then produces compensations only for the perceived amount of feedback alteration. Subjects who produced large compensations are assumed to have small capacities to adapt perception, while subjects who produced small compensations are assumed to have large capacities to adapt perception.

Perceptual adaptation was not directly investigated in Study 2, so currently there is only indirect evidence of its existence. However, the perceptual adaptation hypothesis makes a testable prediction: altered feedback exposure should change a subject's perception of vowel sounds – not just those produced by the subject himself. This prediction will be tested in future experiments.

2. Subjects compensate for perceived feedback alterations. The compensation hypothesis is that subjects make production adjustments specifically to compensate for perceived feedback alterations. This hypothesis is strongly supported by three lines of evidence in Study 2 concerning subjects' compensation responses.[10]

First, across all subjects' compensation responses in the real experiment, significant changes were observed in path projections but not path deviations. This suggests subjects were compensating because the feedback transformations altered only perceived path projections, not path deviations.

Second, all significant path projection shifts were in the direction that compensated for the feedback alteration. This result is most significant because the compensating direction was not the same for all subjects. Half the subjects were exposed to the -2.0 feedback transform, which is compensated for by positive path projection shifts. The other subjects were exposed to the +2.0 feedback transform, which is compensated for by negative path projection shifts. This arrangement insured that any bias for path projections to shift in one direction could not be mistaken as compensation in all subjects.

Third, across all subjects' compensation responses in the control experiment, no significant changes were observed in either path projections or path deviations. The control experiment differed from the real experiment only in that no feedback alterations occurred. Thus, the lack of significant path projection shifts in the control experiment imply that the shifts observed in the real experiment were caused by the presence of altered feedback.

3. Compensation is achieved by both temporary and long-term speech control adjustments. In both Study 1 and Study 2, it appeared that altered feedback exposure caused subjects not only compensate, but to adapt – i.e., to retain compensating production changes in speech produced while speech feedback was blocked

[10] Recall that the term "compensation response" refers only to the feedback condition under which vowel productions were observed. The term means changes in a subject's vowel produced while he could hear feedback of his whispering – conditions when compensation for altered feedback *could* (but not necessarily would) occur.

by noise. Such retained compensation could be explained by supposing that altered feedback exposure induced long-term changes in subjects' control of their speech production. This explanation is supported by the path projection analysis, which showed significant retention of compensating production changes during the month between the real and control experiments.

It was also seen in both Study 1 and Study 2 that compensation was generally greater than adaptation for all subjects.[11] One possible explanation for this difference is that the presence of masking noise somehow causes subjects to whisper differently. This account, however, does not specify why compensation would be greater than adaptation.

Instead we hypothesize that some portion of each subject's compensation was accomplished by some temporary correction mechanism, active only in the presence of the altered feedback. The additional compensation provided by this mechanism explains why subjects' compensation responses (when they hear feedback) are generally bigger than their adaptation responses (when they hear only noise).

This explanation suggests that vowel production may be partly under auditory feedback control - a hypothesis first proposed for speech production in general by Grant Fairbanks in 1954 [Fairbanks, 1954]. Auditory feedback control has also been proposed as an explanation of subjects' compensating responses in pitch perturbation experiments [Kawahara, 1993]. In these experiments, subjects hear feedback of their speech in which the pitch is occasionally perturbed. These pitch feedback perturbations generally induce compensating changes in subjects' pitch production within 100-200ms of the onset of perturbation.

There are, however, several arguments that auditory feedback plays no direct role in the control of speech.

The first argument is that it isn't necessary to suppose auditory feedback control since speech is producible without auditory feedback. Speakers deafened in adult life retain intelligible speech [Cowie and Douglas-Cowie, 1983, Lane and Webster, 1991].

[11] For subjects in which this was not true, it was usually because of other anomalies in their adaptation responses, or because they exhibited insignificant compensation or adaptation.

Many other experiments (including those of Study 1 and Study 2) have shown that speech remains intelligible even when hearing is blocked by masking noise [Lombard, 1911, Lane and Tranel, 1971]. However, this argument does not rule out the possibility that, when available, auditory feedback control is used in speech production.

The other major argument against auditory feedback control is that it is too slow: the neural delays in processing auditory feedback probably make it unusable for the control of fast speech movements [Perkell, 1996]. But maintaining a pitch frequency or steady-state vowel does not necessarily require such fast speech adjustments. For these tasks, it therefore seems plausible their control could be partially based on auditory feedback.

We will return to the issue of auditory feedback and the hypothesized temporary correction mechanism in the next section's discussion.

5.3.2 Timecourse Results

As described in Section 5.2.2.2, in the real experiments (but not control experiments), a feedback transformation was introduced gradually over the 11 stages of the ramp phase. Within each stage, the feedback transformation's strength was held constant. Between each stage, it's strength was increased by a fixed amount. In this way, the feedback transformation's strength was linearly increased to it's maximum value over the course of the ramp phase. It's strength was held at this maximum value for the rest of the experiment.

The primary motivation for gradual alteration of feedback in the ramp phase was to minimize subjects' awareness of it. However, it also allowed examination of how subjects' compensation and adaptation developed in response to increasing feedback alterations. In Section 5.1.2, several questions concerning the timecourse of these responses were posed:

1. Would subjects' compensation responses be modulated by categorical perception?

2. Would their adaptation and compensation timecourses differ?

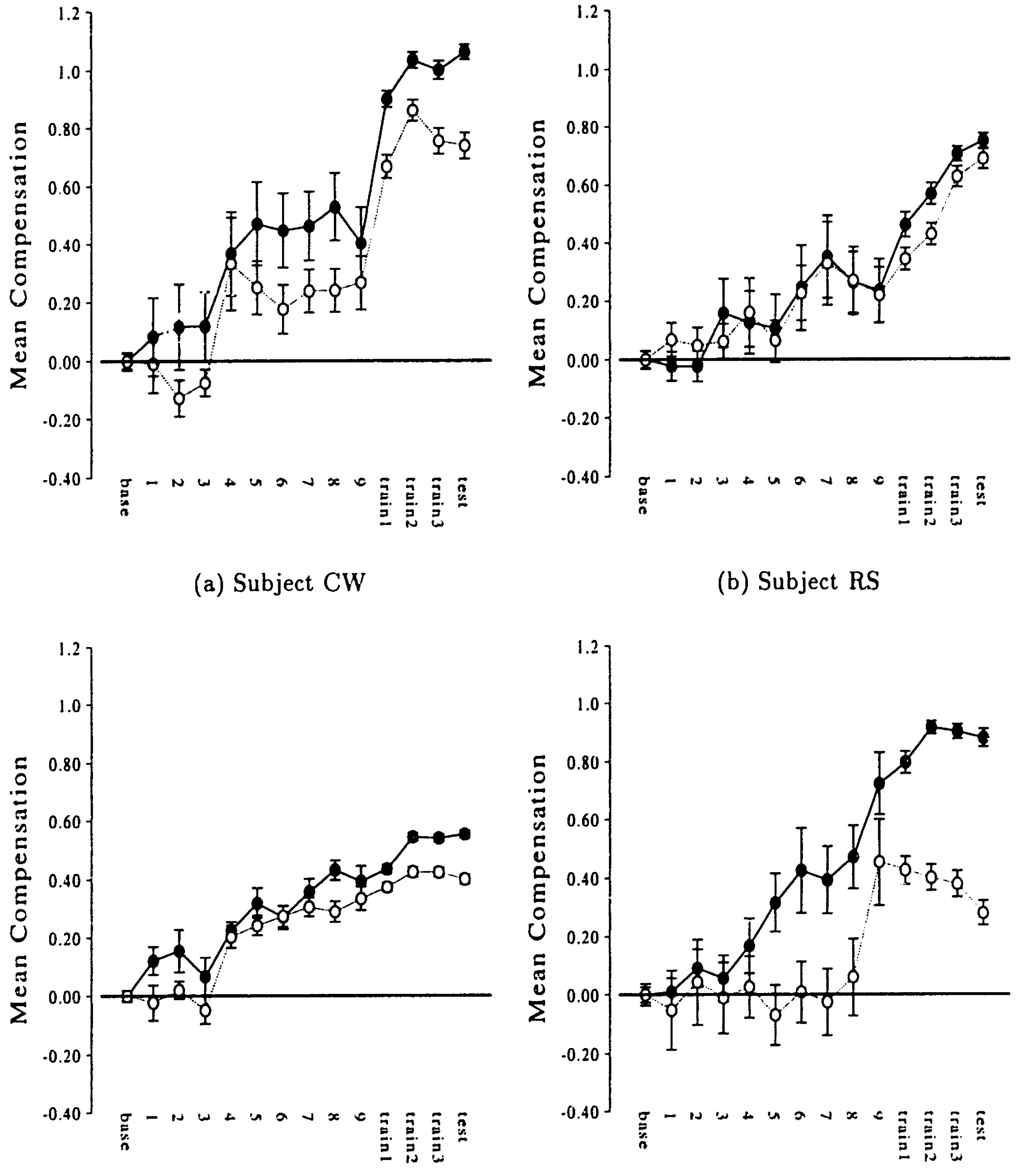
In addition, the experiment's extended (one hour) training phase allowed examination of another timecourse question: would subjects' compensating responses stabilize after extended exposure to the maximum-strength feedback alteration?

These questions are considered in the following sections.

5.3.2.1 Timecourse Plots

Only four subjects produced compensations large enough to permit analysis of their timecourses. These are shown in Figure 5-4. Each plot shows how a subject's mean compensation and adaptation developed over the experiment's timecourse.[12] This timecourse is represented on the x-axis of each plot as a succession of labeled intervals.

[12] In considering the plots of these responses, note that for each interval mean compensation and adaptation were computed relative to the baseline phase. For this reason, in every subject's plot, mean compensation and adaptation in the baseline phase (the "base" interval) is zero by definition.



(a) Subject CW

(b) Subject RS

(c) Subject OB

(d) Subject SR

Figure 5-4: Mean compensation and adaptation timecourses for the subjects showing the four largest mean adaptations. Each plot's x-axis lists the intervals in the experiment's timecourse: "base" is the baseline phase; "1"-"9" are stages 1–9 of the ramp phase; "train1", "train2", and "train3" are the 1st, 2nd, and 3rd 20-minute intervals of the train phase; "test" is the test phase. For each interval, mean compensation and adaptation are shown as black and white dots, respectively, on the y-axis. Small bars around each dot indicate confidence intervals.

It is important to note that these labeled intervals represent differing amounts of time:

- The interval labeled "base" represents the entire 17-minute baseline phase.
- The intervals labeled "1" - "9" represent ramp stages 1 - 9, which were each about 2 minutes long.
- The intervals labeled "train1", "train2", and "train3" represent successive 20-minute time intervals of the train phase.
- The interval labeled "test" represents the entire 17-minute test phase.

The differing interval durations have two consequences. First, there is less data to average in each ramp stage interval than in the other intervals. Thus, the confidence intervals of the ramp stage measurements are larger than those of the other intervals. Second, there is a timescale discontinuity at ramp stage 9: up to this point, each ramp stage interval represents another 2 minutes in the experiment; past this point, each interval represents roughly another 20 minutes in the experiment. Any apparent jumps in compensation or adaptation between ramp stage 9 and the train1 interval (as seen in subject CW's plot) are thus possibly due to the timescale discontinuity.

This timescale discontinuity is acceptable because data from the ramp phase and the rest of the experiment (the train and test phases) are analyzed separately. These separate analyses differ in purpose. The ramp phase analysis investigates how subjects' compensating responses increased in response to the increasing feedback alteration. The train and test phase analysis investigates whether subjects' responses stabilized after extended exposure to the maximum-strength feedback alteration. The train and test phase analysis is presented first because it is less extensive than the ramp phase analysis that is the main focus of timecourse analysis.

5.3.2.2 Train and Test Phase Analysis

During the train and test phase intervals ("train1", "train2", "train3", and "test"), the strength of the feedback transformation was held at its maximum value. The plots show that, within these intervals, all but one subject's responses appeared to stabilize:

- Subject CW's plot (Figure 5-4(a)) shows both his compensation and adaptation were still increasing in the first 20 minutes of the train phase, but thereafter appeared to stabilize. His mean compensation stabilized at a higher value than his mean adaptation. Subject OB and SR's plots (Figures 5-4(c) and 5-4(d)) exhibit the same basic pattern.[13]
- Subject RS's plot (Figure 5-4(b)) shows both his compensation and adaptation were still increasing throughout the train and test phases.

The results show that, for all but one subject, the test-phase compensation and adaptation measures analyzed in Section 5.3.1 are good measures of the subjects' complete potential to compensate. Not so, however, for subject RS: his results imply that, if the experiment were continued, he would continue to increase his compensation and adaptation.

5.3.2.3 Ramp Phase Analysis

Over the stages of the ramp phase, alteration of subjects' feedback was linearly increased. The basic analysis question was therefore whether subjects' compensation and adaptation linearly increased in response to it. In other words: did each feedback alteration increase induce the same amount of compensation and adaptation increase?

Qualitative examination of Figure 5-4's plots suggests this may be true for compensation but not for adaptation. Taking into account the confidence intervals, it appears plausible that each subject's compensation increased linearly during the ramp phase. However, there is much less consistency across subjects in their ramp-phase adaptation timecourses:

- Ramp stage 1 is the first experiment interval with a non-zero feedback alteration and subject RS's adaptation shows an immediate response to this. The rest of

[13] Subject SR's adaptation appears to show a slight dip in the testing phase, but, within the limits of the confidence intervals, this dip does not appear significant.

his ramp phase adaptation timecourse looks quite similar to his compensation timecourse.

- On the other hand, subjects CW, OB, and RS all exhibit delayed adaptation responses. Their responses remain at zero for several ramp stages and then suddenly increase. These sudden increases (onsets) occur between stages 3 and 4 for subjects CW and OB, but between stages 8 and 9 for subject SR. However, for each subject, the onset brings his adaptation to a value roughly equal to his compensation (taking into account the confidence intervals).

These differences between compensation and adaptation are reflected in the linear regression analysis tabulated in Table 5.4. The table shows that the linear fit of each subject's compensation is highly significant, whereas the significance results for adaptation are much more variable.

subject	compensation			adaptation		
	R	$F(1,7)$	$p<$	R	$F(1,7)$	$p<$
CW	0.850	18.257	**0.004**	0.725	7.760	**0.027**
RS	0.852	18.649	**0.004**	0.814	13.741	**0.008**
OB	0.913	35.294	**0.001**	0.909	33.119	**0.001**
SR	0.961	85.004	**0.000**	0.586	3.651	**0.098**

Table 5.4: Ramp phase compensation and adaptation timecourse regression results. Each row of the table shows regression analysis of a subject's ramp-phase compensation and adaptation. Each regression analysis is summarized by a correlation R, and a test of the significance of this correlation: $F(1,7)$ and p.

We conclude from these results that no effects of categorical perception are apparent in the subjects' compensation timecourses. As discussed in Section 5.1.2, the testable effect would be a delay in a subject's compensation response to the increasing feedback alteration. A subject would delay compensating his production of a vowel until the feedback alteration grows large enough to change the perceived phonetic identity of that vowel. No such delay is evident in the compensation results. The linearity of subject's compensation timecourses imply that, on average, subjects compensate for each feedback alteration increase, and that each increase is compensated for by the same amount.

On the other hand, many subjects do exhibit delayed adaptation responses. However, it appears unlikely that these delayed responses are caused by categorical perception. If these delayed responses resulted from failure to perceive the feedback alteration until it reached a certain magnitude, then the subjects should have delayed all responses. But no delay was evidenced in subjects' compensation responses.

One consistent feature of the adaptation onsets has already been noted: the onset brings adaptation to a value roughly equal to compensation. However, careful examination of the plots in Figure 5-4 reveals another consistent feature: where the onset occurs, mean compensation has value approximately equal to the subject's final compensation - adaptation difference. Figure 5-5 quantifies this relationship. Each point in the plot represents a different subject's results. The y-value of each point is mean compensation at the time of the subject's adaptation onset - i.e., where adaptation first exhibits a noticeable increase. The point corresponding to subject RS thus has a y-value of zero, since his adaptation begins to increase right from the beginning of the ramp phase, when both adaptation and compensation are zero. The x-value of each point is the difference between mean compensation and adaptation exhibited by the subject in the final, (test) phase of the experiment.

As indicated in the figure caption, these points fit significantly well to a straight line, and this line is quite close to the the line $y = x$. Thus, for the four analyzed subjects, amount of compensation at the onset of adaptation is nearly equal to the amount by which compensation exceeds adaptation at the experiment's end.

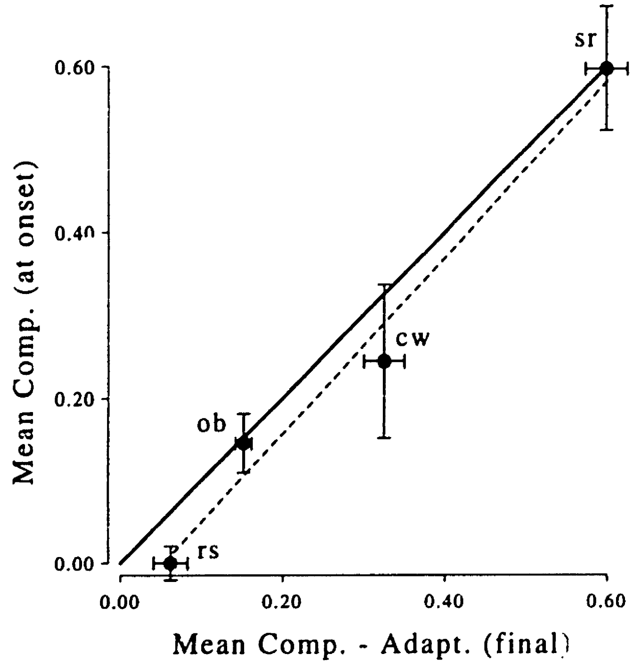


Figure 5-5: Analysis of the adaptation onsets seen in Figure 5-4. Each subject is represented by a labeled point. The point's x-value shows how much the subject's mean compensation exceeded his mean adaptation in the final (test) phase of the experiment. The point's y-value shows the subject's mean compensation at the onset of his adaptation increase. The regression line for these points (dashed line) has a y-intercept of -0.056 and a slope of 1.065, which is quite close to the line $y = x$ (solid line). The significance of the fit of the points to the regression line is: $r = 0.989, r^2 = 0.980, F(1,2) = 95.339, p = 0.010$.

5.3.2.4 Discussion

A key hypothesis of the speech SA theory proposed in Section 5.3.1.4 concerned the mechanism of compensation. It hypothesized that compensation is achieved partly by a temporary correction mechanism (active only in the presence of auditory feedback) and partly by long-term speech control adjustments. In this account, adaptation is explained as evidence of these long-term speech control adjustments.

The major results of the timecourse analysis could be explained by a slight addition to this hypothesis: supposing that subjects have a preference to compensate using the temporary correction mechanism. Only when this mechanism's capacity is reached do subjects make long-term speech control adjustments.

Thus, based on the timecourse results, we expand the speech SA theory of Section 5.3.1.4 to the following:

1. Perception of the altered feedback is partially offset by perceptual adaptation. The capacity to adapt perception is limited and subject-specific.

2. The perceived feedback alteration is compensated for.

3. Compensation is preferentially achieved by a temporary correction mechanism (active only while exposed to the altered feedback). The capacity of the temporary correction mechanism to compensate is limited and subject-specific.

4. When required compensation exceeds the capacity of the temporary correction mechanism, long-term speech control adjustments must be made – i.e., adaptation occurs.

This theory explains the subjects' timecourse results in the following way:

- Subject RS has no capacity for temporary correction. For him, all compensation for the feedback distortion must be accomplished by adjustment of long-term speech control – i.e., adaptation. This is why his adaptation response exhibits no delayed onset, and why its timecourse is so similar to his compensation timecourse: both timecourses reflect adjustment of the same mechanism.

- Subject SR, however, has a large capacity for temporary correction. For him, this capacity is not exceeded until he is induced to increase compensation beyond about 0.6. At this point (between ramp stages 8 and 9), he adapts long-term speech control enough to compensate for the feedback distortion.

- Subjects CW and OB's responses to the increasing feedback distortion are intermediate between these two extremes: Both exhibit some temporary correction capacity, which, when exceeded, triggers long-term speech control adjustment.

The above theory, however, is preliminary and will require further experiments to confirm. It also leaves unexplained some aspects of the results seen in Figure 5-4. One unexplained aspect is why adaptation rises quickly after its onset to account for all compensation. (I.e., why does the step in adaptation bring it equal to the current amount of compensation?) Another unexplained aspect is what happens to adaptation after onset. For subject OB, it appears to linearly increase after its onset, while for subjects CW and SR, it appears to stabilize.[14] These unexplained aspects will need further investigation to be understood.

[14] In subject CW's plot, both compensation and adaptation also significantly increase between ramp stage 9 and the train1 interval. As explained in Section 5.3.2.1 above, it's unclear whether these increases are rapid onsets (i.e., within one ramp stage) or more gradual increases because of the plot's timescale discontinuity at this point.

5.3.3 Generalization Results

In Section 5.2.2.1, we described how, in each epoch, the subject was prompted to produced ten different words. For the first five words, he could hear feedback of his whispering, but for the last five words, masking noise prevented him from hearing his whispering. The first six words the subject produced were randomly selected from a word set called the *training words* (abbreviated symbolically as $\mathbf{W}_{\mathbf{train}}$). The last four words were randomly selected from a set called the *testing words* (abbreviated as $\mathbf{W}_{\mathbf{test}}$). With this design, subjects produced training words under both feedback conditions: they heard feedback of their first five $\mathbf{W}_{\mathbf{train}}$ word productions but produced the sixth $\mathbf{W}_{\mathbf{train}}$ word while the masking noise blocked their hearing. On the other hand, they produced all testing words while the masking noise blocked their hearing.

Thus, when their feedback was altered, subjects only heard errors in their training word productions. Training word production changes were therefore used as direct measures of compensation and adaptation. The previous two sections (sections 5.3.1 and 5.3.2) focused on analyzing these measurements.

In contrast to this, testing word productions were not directly affected by exposure to altered feedback – they were always produced while the subject's hearing was blocked by noise. Testing word production changes therefore measured how the training word adaptations *generalized.* In this section we analyze these testing word production changes.

5.3.3.1 The Testing Word Set

As discussed in Section 5.2.1.2, the $\mathbf{W}_{\mathbf{test}}$ word set was composed of two subsets, each designed to assess a different type of generalization.

The $\mathbf{W}_{\mathbf{test}}$**-context** subset was designed to assess *context generalization*: how the adaptation of [ε] in the training words affected the production of [ε] in other words. This subset was composed of the following words:

$\mathbf{W}_{\mathbf{test}}$**-context** = { "pep", "peg", "gep", "teg" }

The $\mathbf{W}_{\mathbf{test}}$**-target** subset was designed to assess *target generalization*: how the adaptation of [ɛ] in the training words affected the production of other vowels. This subset was composed of the following words:

$\mathbf{W}_{\mathbf{test}}$**-target** = { "peep", "pip", "pap", "pop" }

A key feature of these word sets is that the word "pep" is also a training word. This was done to allow more accurate assessment of generalization. One way to assess generalization would be to compare testing word production changes with training word production changes. However, this makes word order a confounding factor in the comparison: training words were always the first six words produced in an epoch, while testing words were always the last four. Including "pep" in the testing words insured that one training word was produced under the same word-order conditions as the other testing words. We will refer to these productions of "pep" as $\mathbf{W}_{\mathbf{test}}$ "pep". Generalization was assessed by comparing production changes of the other testing words with those seen in $\mathbf{W}_{\mathbf{test}}$ "pep". This avoided the word-order confound inherent to the direct comparison of training and testing words.

5.3.3.2 Assessing Generalization

Assessing generalization requires quantifying the influence of training word adaptations on testing word productions. Because adaptations are shifts of path projection, we conservatively assume that training word adaptation primarily influences testing word path projections. We therefore represent testing word productions in terms of path projections and deviations, and assess generalization in the following ways:

1. We test generalization by assessing whether presence of altered feedback selectively affects testing word path projections.

2. We measure generalization as a comparison of testing and training word path projection changes.

To see how these assessments are made, consider figures 5-6 and 5-7, which show subject OB's generalization results.

Figure 5-6 shows avgrams of OB's testing word productions in the real experiment. In each avgram, the solid lines show formant tracks of the mean utterance in the baseline phase, while the dashed lines show formant tracks of the mean utterance in the test phase. The gray regions show the interval in each utterance that was used for vowel analysis.

Figure 5-7 shows vowel plots of subject OB's generalization results. The figure's left plot (Figure 5-7(a)) shows OB's context generalization results. Black arrows show mean vowel (F1,F2) changes (test phase - baseline phase) for his context generalization word productions in the real experiment. White arrows show the same changes in the control experiment. In a similar fashion, the figure's right plot (Figure 5-7(b)) shows OB's target generalization results. It shows mean vowel (F1,F2) changes for his target generalization word productions. (Note that, to facilitate comparisons, the vowel production change arrow for "pep" appears in both the left and right plots.)

Technical Limitations Consider first some of the technical problems with generalization assessment illustrated by OB's results.

The vowel plot and avgrams of the context generalization words illustrate the effects of coarticulation. The vowel plot shows that, in both real and control experiments, the bases of the "gep" and "teg" arrows are noticeably shifted towards [i], relative to the "peg" and "pep" arrow bases. Thus, baseline production of [ɛ] in "gep" and "teg" is different from [ɛ] in "peg" and "pep". These baseline differences confound interpretation of the context generalization results. Differences in [ɛ]'s path projection change in different words might result from differences in how [ɛ]'s adaptation generalizes to different words, or they may simply result from the baseline differences. In the avgrams, examination of baseline F2 timecourses show that the baseline differences in the vowel plot are likely due to coarticulation of the initial stop consonant with the vowel.[15]

The **W$_{\text{test}}$-target**word avgrams exhibit problems in formant estimation. In the avgram of "peep" there is no baseline F2 track, while in the avgram of "pop" there is no F2 track for either the baseline or the test phases.[16] Because of this, vowel production changes for [i] and [ɑ] could not be measured. Thus, the arrows for "peep" and "pop" are absent from the target generalization vowel plot.

Quantifying Generalization Now consider how path projection changes are used to quantify the generalization seen in subject OB's results. Figure 5-8 shows mean path projection changes (test phase - baseline) of OB's testing word vowel productions: Figure 5-8(a) shows mean path projection changes in the real experiment, while Figure 5-8(b) shows the same for the control experiment.

Figure 5-8(a) shows that, in the real experiment, all testing words exhibit significant mean path projections changes between 0.5 and 1.0. This is consistent with the production changes seen in the vowel plots (Figure 5-7). In these plots, recall that the black arrows represent mean (F1,F2) change for all testing word vowel productions in the real experiment. These arrows all have similar lengths, which shows that all testing word vowels exhibited similar formant change magnitudes. Because these arrows are all aligned with the [i]–[ɑ] path, the similar formant change magnitudes should result in similar path projection change magnitudes.

Figure 5-8(b) shows that, in the control experiment, all testing words except "pip" exhibit much smaller mean path projection changes. Again, this is consistent with the

[15] Consider the baseline F2 formant track in avgrams of "gep", "peg", and "teg". Tne timecourse of F2 in "peg" is similar to that of "pep". Since [p] is a bilabial stop, F2 is initially low but quickly transitions to a steady-state value that it holds for the rest of the utterance. This quick F2 transition is possible because the articulation of [p] does not need the tongue, so the tongue can be preset to its position for [ɛ]. In "gep", however, since [g] is a velar stop, F2 is initially high. It then takes time for the tongue to move from its position for [g] to its position for [ɛ]. This makes the F2 transition to the steady-state vowel much slower: only by the utterance's end has F2 dropped to its steady-state value in "peg". Thus, within the vowel analysis interval, average F2 in "gep" is higher than it is in "peg" or "pep". In "teg", since [t] is an alveolar stop (and thus articulated with the tongue), F2 also starts high and makes a slow (barely visible) downward transition. Thus, again, within the vowel analysis interval, average F2 in "teg" is higher than it is in "peg" or "pep". These differences in average F2 account for much of the baseline differences seen in the vowel plot.

[16] Formant estimation problems are discussed in depth in Section A.1.2.

production changes seen in the vowel plots. In these plots, the white arrows show that the vowels of all testing words except "pip" exhibit only small mean (F1,F2) changes. For "pip", its vowel production change arrow in the control experiment appears large and aligned with the [i]–[ɑ] path – essentially a slightly shifted version of its vowel production change arrow in the real experiment. Thus, mean path projection for "pip" in the control experiment is similar to its mean path projection in the real experiment.

These path projection plots illustrate the necessity of comparing production changes seen in both real and control experiments. The plots show that, for most testing words, mean path projection changes seen in the real experiment are large while those seen in the control experiment are small. For these words, real experiment path projection change appears to be a good measure of change resulting from adaptation of the training words. However, this is not the case for "pip". The plots shows that mean path projection for "pip" is equally large in both the real and control experiments. Thus, for "pip" it becomes doubtful whether its vowel production change in the real experiment is due specifically to adaptation of the training words.

Because of this, both testing and measuring generalization were done as comparisons between real and control experiment results.

Two tests of generalization were used. One test was an ANOVA of testing word vowel path projections in the real and control experiments. This test evaluated whether path projection changes differed significantly between the real and control experiments. The other test of generalization was an ANOVA of testing word vowel path deviations in the real and control experiments.

Measurement of generalization was based on calculating *mean generalization*. Mean generalization, in turn, was based on first calculating *mean relative path projection change*. For any testing word, mean relative path projection change is defined as the word's mean path projection change in the real experiment minus its mean path projection change in the control experiment. Figure 5-8 shows mean relative path projection changes for the testing word vowels of subject OB's results. This figure is essentially the difference between figures 5-8(a) and 5-8(b). The figure shows that mean relative path projection quantifies our doubts about whether "pip" has been affected by adaptation of the training words. For all other testing words, their control experiment mean path projections are small. Thus, when these are subtracted from their larger real experiment mean path projection changes, the resulting mean relative path projection changes are still substantial. However, for "pip", its similar path projection changes in both real and control experiments cancel each other, resulting in an insignificant mean relative path projection change.

Mean generalization was then calculated as a ratio of a subject's mean relative path projection changes. For every testing ($\mathbf{W}_{\mathbf{test}}$) word, this ratio compared the word's mean relative path projection change with that of $\mathbf{W}_{\mathbf{test}}$ "pep":

$$\text{mean gen.} \;=\; \frac{(\text{mean rel. path proj. change, } \mathbf{W}_{\mathbf{test}} \text{ word})}{(\text{mean rel. path proj. change, } \mathbf{W}_{\mathbf{test}} \text{ ``pep''})}$$

The key property of mean generalization computed in this fashion is that it measures the influence of training word adaptation. Since "pep" is also a training word, it is assumed that its mean relative path projection change is a direct result of training word adaptation. Because of this, the above ratio is assumed to gauge how much training word adaptation influences testing word production changes.

This method of computing mean generalization has the following three properties:

1. It is not influenced by word order effects. In each epoch, the training words come before the testing words. Thus, if generalization was measured by direct comparison with the training words, word order could influence the measurement. By instead measuring generalization with respect to $\mathbf{W}_{\mathbf{test}}$ "pep", this word-order influence is avoided. (This is also discussed in Section 5.3.3.1 above.)

2. It normalizes the directions of subjects' path projection changes. For any subject, as long as path projection changes for his testing words and $\mathbf{W}_{\mathbf{test}}$ "pep" are all in the same direction, mean generalization will always be positive. This facilitates comparison of -2.0 and +2.0 subjects (which generally exhibit opposite path projection changes).

3. It normalizes subjects' adaptations: for all subjects, mean generalization of $\mathbf{W}_{\mathbf{test}}$ "pep" is 1.0. This allows mean generalizations of different subjects to be averaged, since their adaptation differences have been normalized away.

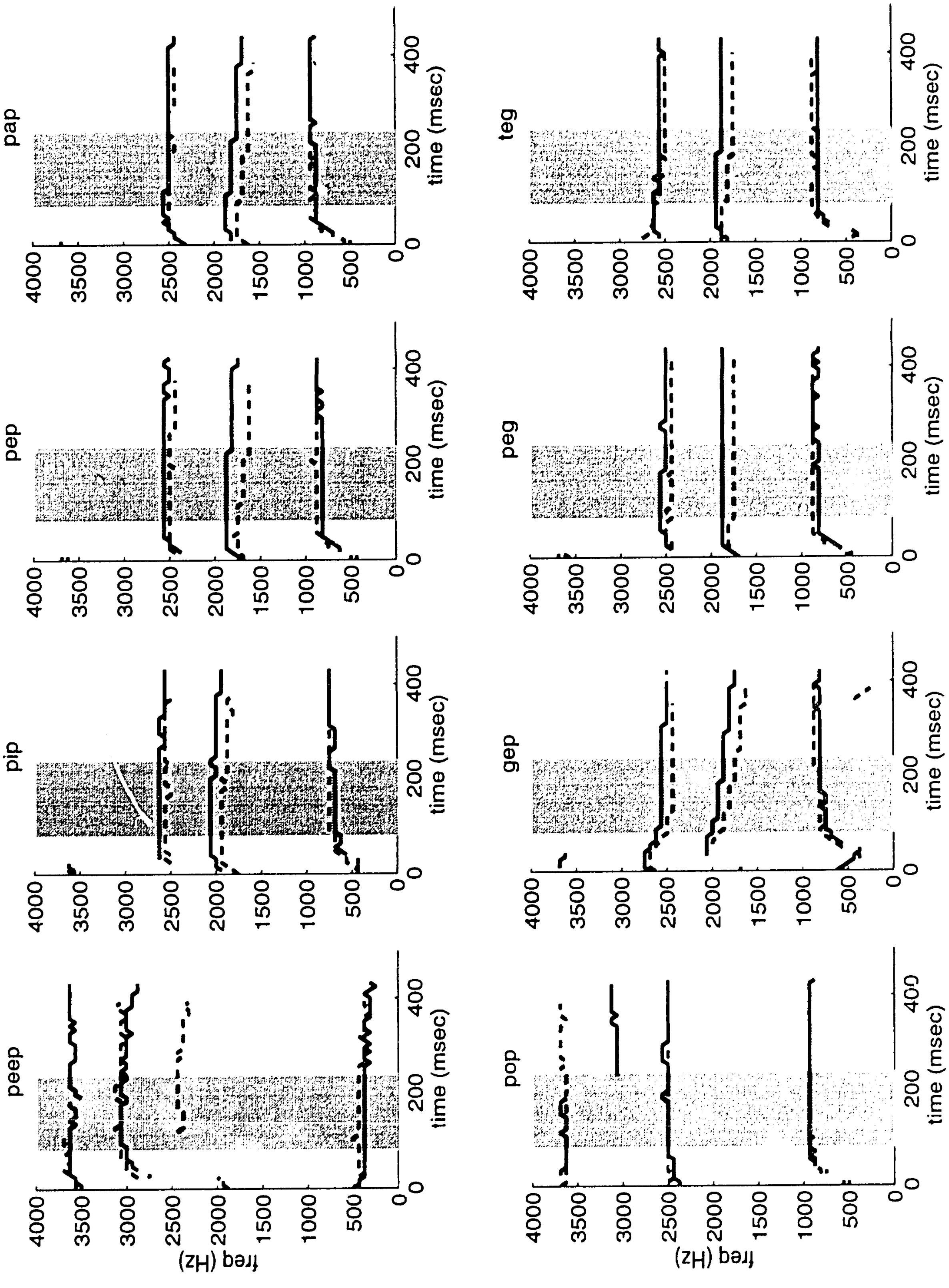
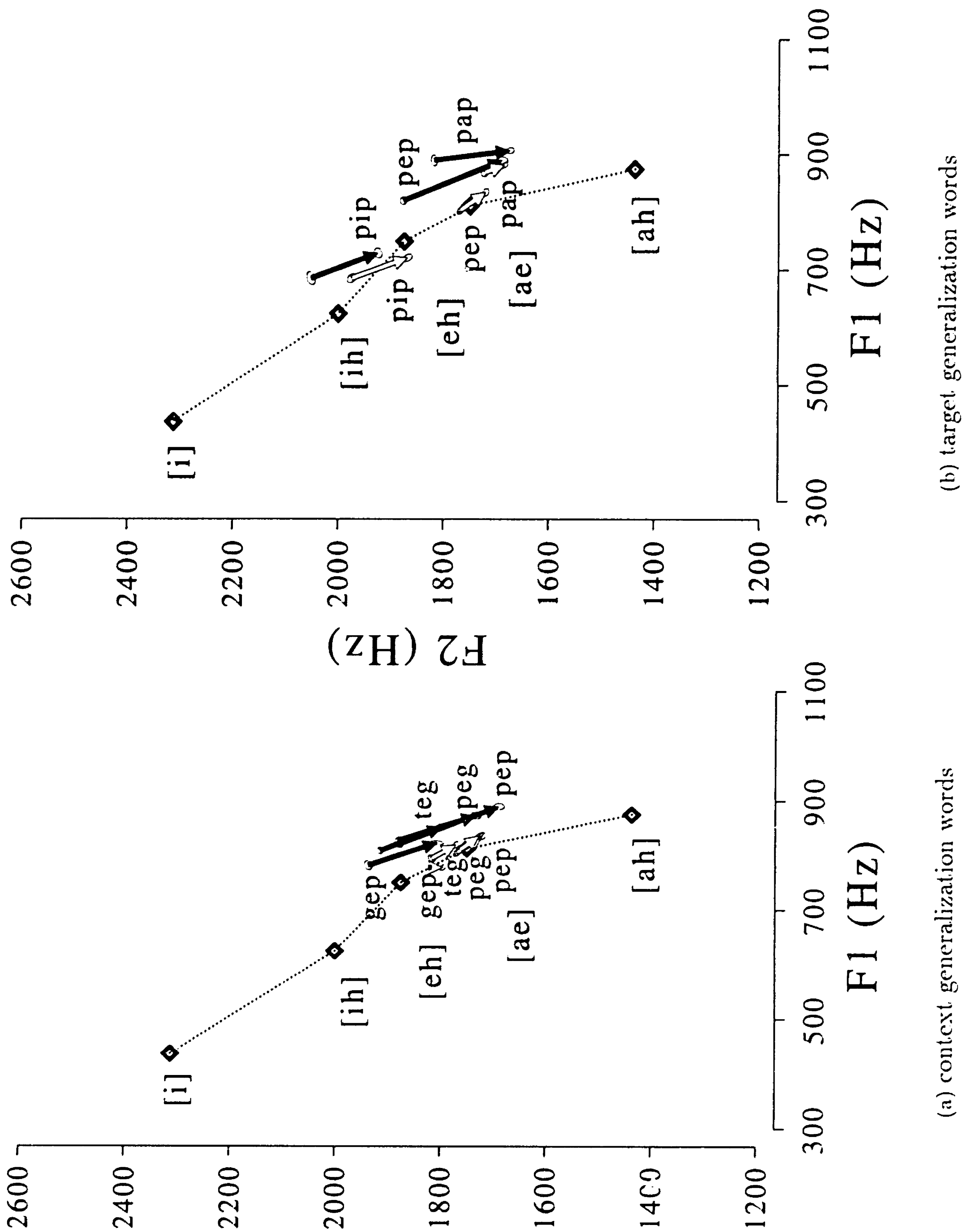


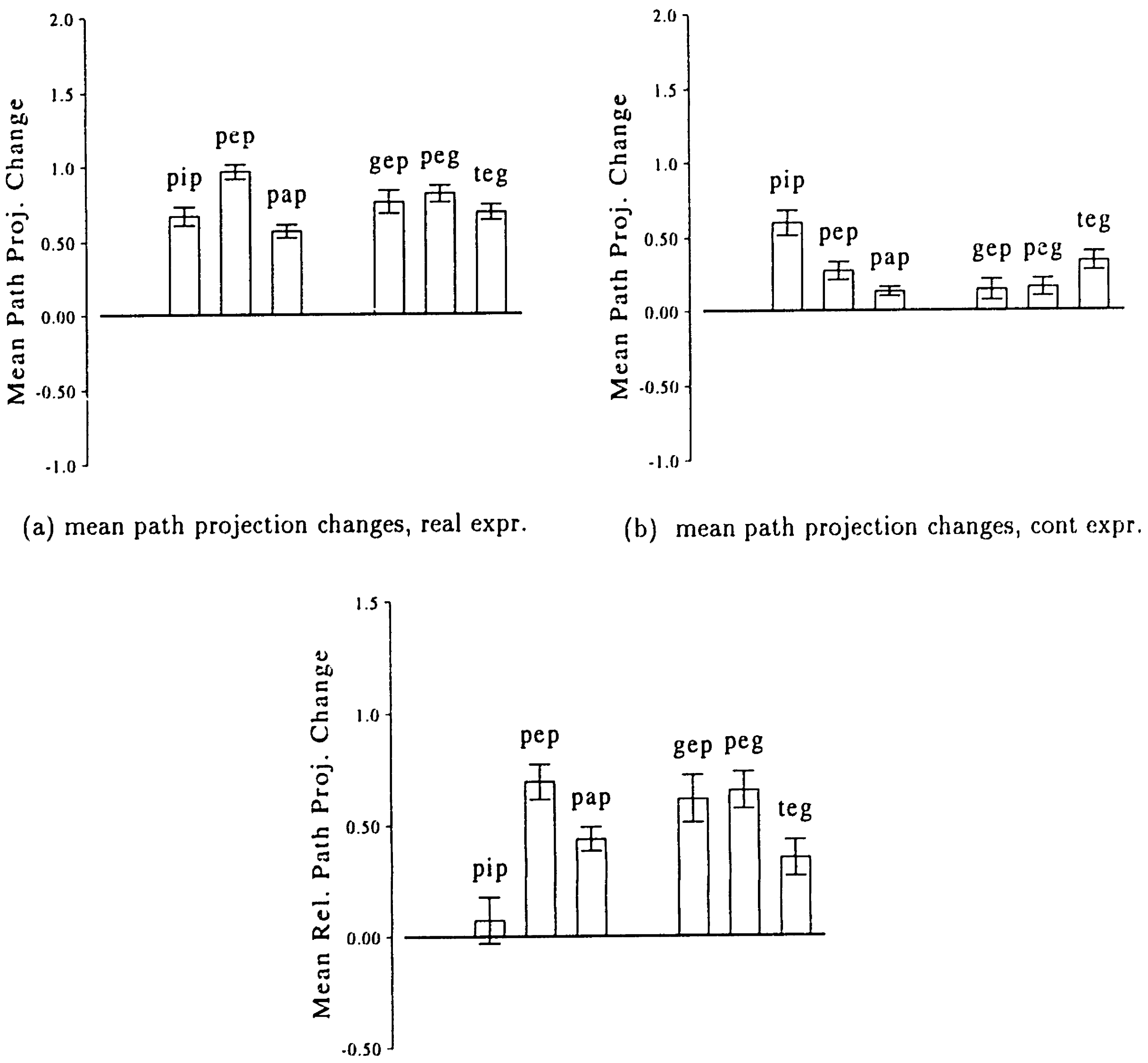
Figure 5-6: Subject OB testing word avgrams.



(b) target generalization words

(a) context generalization words

Figure 5-7: Subject OB testing word vowel plots.



(a) mean path projection changes, real expr.

(b) mean path projection changes, cont expr.

(c) mean relative path projection changes

Figure 5-8: Subject OB testing word path projection changes.

5.3.3.3 Overall Generalization

In the previous section it was hypothesized that training word adaptation would primarily affect path projections of testing word productions. This hypothesis makes the following two predictions about testing word vowel productions:

1. Their path projection changes should differ significantly between the real and control experiments but their path deviation changes should not.

2. Their path projections changes should be in the same direction as those of the training words.

Hypothesis prediction 1 was evaluated using a test of overall generalization. This test was actually two separate ANOVA tests: one which evaluated whether path projection changes differed significantly between the real and control experiments, and one which evaluated the same for path deviation changes.

Ideally, the overall generalization test would be done using all the $\mathbf{W}_{\mathbf{test}}$ word vowel productions of all subjects' results. However, a number of factors restricted which results could actually be used. One factor is the amount of adaptation subjects exhibited. Subjects VS and AH had mean training word adaptations that were very small and did not differ significantly between the real and control experiments. Consequently, these subjects were excluded. For the remaining six subjects, formant estimation problems (like those described in Section 5.3.3.2 above) prevented the analysis of the vowel productions for certain subjects' words. A summary of these words is shown in Table 5.5. In the table, an "X" shows, for each subject, which words' vowel productions were not analyzable. Crossing out the columns for "peep" and "pop" and the row for subject SR balances the table: for each remaining subject, the same testing words' results are available; for each remaining testing word, the same subjects' results are available. Thus, because the ANOVA tests require balanced tables of results, subject SR and the words "peep" and "pop" were excluded from the tests. Finally, "pep" was excluded from the tests because it is also a training word.

	context gen. words				target gen. words			
subject	pep	peg	gep	teg	peep	pip	pap	pop
CW								X
RS					X			X
OB					X			X
SR			X	X				
RO					X			X
TY					X			X

Table 5.5: Showing, for each subject, which words' vowel formants could not be properly estimated. These words are marked with an "X". (For more detailed descriptions of the problems estimating these words' formants, see the individual subject result plots of Section 5.4.)

As a result of these exclusions, the test of overall generalization was performed on the vowel production results of six subjects (CW, RS, OB, RO, and TY) and five testing words ("pip", "pap", "gep", "peg", and "teg"). When the test is run with these restrictions, its results match the prediction of the hypothesis: path projection changes in the real experiment are significantly different from those of the control experiment ($F = 8.981, p < 0.040$) but path deviation changes are not ($F = 0.362, p < 0.574$).

To evaluate hypothesis prediction 2, mean generalization was calculated for each subject used in the above generalization test. Hypothesis prediction 2 is that, for each subject, testing word path projection changes should be in the same direction as those of the training words. This is equivalent to predicting that, for each subject, mean generalization averaged across testing words is non-negative.

Figure 5-9 shows mean generalization for each subject used in the overall generalization test. For each subject, the value shown is mean generalization averaged over the testing words used in the overall generalization test ("pip", "pap", "gep", "peg", and "teg"). The figure shows that mean generalization is zero for subject RO and positive for all other subjects. Thus mean generalization is always non-negative, as predicted by the hypothesis.

Since the results agree with the hypothesis predictions, it appears that training word adaptation does selectively affect testing word productions. The success of these predictions also shows they are useful conservative criteria for assessing generalization in speech SA. By applying these criteria to different groups of testing words, we can assess whether speech SA exhibits different types of generalization.

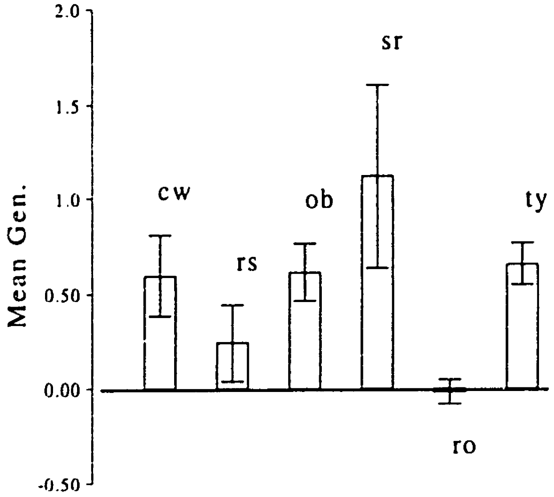


Figure 5-9: Mean generalization for all subjects used in the overall generalization test.

5.3.3.4 Context Generalization

Entries in the first four columns of Table 5.5 show subjects' analyzable productions of the context generalization words. The entries show that all productions were analyzable except subject SR's productions of "gep" and "teg". Thus, context generalization was assessed for all four context generalization words ("pep", "peg", "gep", and "teg"), but this assessment used results from only five subjects (CW, RS, OB, RO, and TY).

To test for context generalization, several ANOVA tests of path projection change were performed. First, separate tests for each context generalization word were performed. This was followed by an overall test across all context generalization words except "pep". Each test examined path projection changes observed across subjects. Table 5.6 summarizes the results of these tests. In the table, the F and p values show

how significantly different path projection changes in the real experiment were from those in the control experiment.

word	F	$p <$
pep	44.104	**0.001**
peg	8.180	**0.046**
gep	2.250	**0.208**
teg	1.390	**0.304**
overall	7.617	**0.040**

Table 5.6: ANOVA tests of path projection change for the context generalization words.

The table shows first of all, that path projection changes of $\mathbf{W}_{\mathbf{test}}$ "pep" seen in the real experiment were extremely different from those seen in the control experiment. This is consistent with the fact that "pep" was also a training word.

For the other testing words, the table shows that, as a group, their path projection changes differed significantly between the real and control experiments. This shows that training word adaptations did significantly affect vowel productions in the context generalization words.

In the individual word tests, besides "pep", only "peg" appeared to exhibit path projection changes that differed significantly between the real and control experiments. This suggests that training word adaptations may only generalize to words sharing the same initial CV. However, as discussed in Section 5.3.3.2 above, these results could also be attributed to the differing coarticulatory influences of [g], [t], and [p] on the following vowel.

To measure context generalization, mean generalization values for all context generalization words except "pep" were calculated. These values are shown in Figure 5-10. For each word, the values were averaged across all subjects used in the context generalization tests.

The figure shows that, for all three words, mean generalization is positive and exhibits roughly the same amount of variability. This agrees with the overall context generalization test results. It also allows a more specific conclusion to be drawn: that

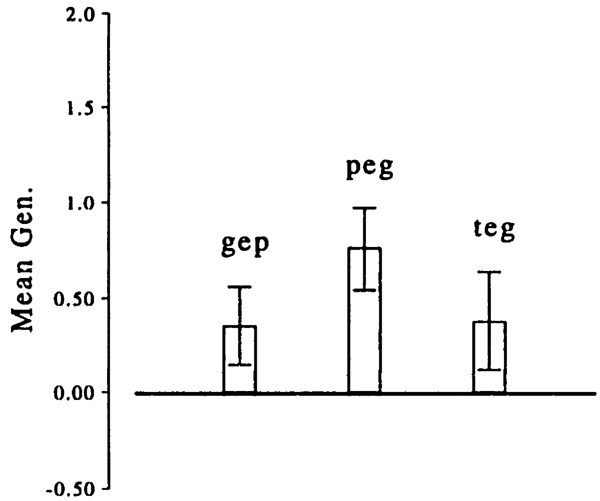


Figure 5-10: Mean generalization for the context generalization words.

adapting the production of [ɛ] in the training words causes similar production changes in [ɛ] in the testing words.

The figure also shows that mean generalization for "peg" (≈ 0.75) is noticeably larger than mean generalization for "gep" or "teg" (≈ 0.4 for both). This may suggest that adapting [ɛ]'s production in thc training words causes similar [ɛ] production changes only in words with the same initial CV. Again, however, this conclusion is confounded by the coarticulation differences discussed above.

In sum, this section's results allow only general conclusions regarding context generalization. The results suggest that adapting a vowel's production in one word context generalizes only to words sharing the same initial CV. However, this conclusion will have to be substantiated by future experiments that avoid the confounding influences of coarticulation. A more reliable conclusion to be drawn from the current results is that, overall, speech SA exhibits context generalization. In particular, when [ɛ] is adapted in training words with a common characteristic (CVC where both C's are bilabial), this adaptation generalizes to [ɛ] in testing words lacking this characteristic (CVC where at least one C is not bilabial).

5.3.3.5 Target Generalization

Entries in the last four columns of Table 5.5 show subjects' analyzable productions of the target generalization words. The entries show that, for most subjects, productions of "peep" and "pop" were not analyzable. Thus, target generalization was assessed for only two target generalization words ("pip" and "pap"), but this assessment used results from all six available subjects (CW, RS, OB, SR, RO, and TY).

The methods used to assess target generalization were similar to those used for context generalization. To test for target generalization, separate ANOVA tests for each target generalization words were performed, followed by an overall test across both target generalization words. Each test examined path projection changes observed across subjects. Table 5.7 summarizes the results of these tests.

word	F	$p <$
pip	6.734	**0.049**
pap	5.155	**0.072**
overall	14.439	**0.013**

Table 5.7: ANOVA tests of path projection change for the target generalization words.

The table shows that, as a group, the target generalization words exhibited path projection changes that differed significantly between the real and control experiments. This shows that training word adaptations did significantly affect vowel productions in the target generalization words.

In the individual word tests, however, significance levels are noticeably poorer. "pip" shows a marginally significant difference in its real and control experiment path projections changes, while "pap" shows a marginally insignificant difference. This suggests a lack of statistical power: the experiment contained enough word repetitions to confirm overall target generalization, but not enough repetitions to clearly establish generalization to particular target generalization words.

Figure 5-11 shows mean generalization values for the two target generalization words. For each word, the values were averaged across all subjects used in the target generalization tests.

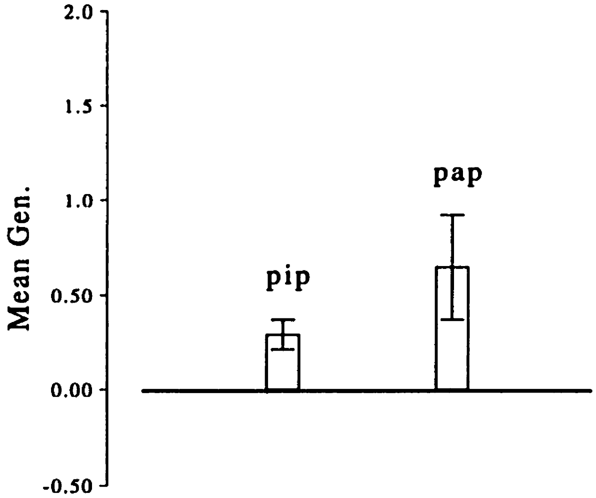


Figure 5-11: Mean generalization for the analyzable target generalization words.

The figure shows that, across both words, mean generalization is positive but exhibits different amounts of variability. This is consistent with the target generalization test results. It also allows us to specifically conclude that adapting [ɛ]'s production in the training words causes similar production changes in other vowels.

In this figure mean generalization for "pap" appears bigger than that of "pip". However, this difference does not appear significant because of the large confidence intervals for mean generalization of "pap". In fact, mean generalization of "pip" appears larger in comparison to its confidence intervals than does "pap". Thus, like the test results, the mean generalization results do not reliably indicate a difference in generalization between "pip" and "pap".

In sum, like the context generalization results, this section's results allow only general conclusions regarding target generalization. Future experiments using more test word repetitions will be needed to examine pattern of target generalization across vowels. But, from the current results, we can reliably conclude that, overall, speech SA exhibits target generalization: adaptation [ɛ] does affect the production of other vowels.

5.3.3.6 Discussion

As discussed in Section 5.1.3, the motivation for investigating speech SA generalization is its potential to reveal organization of the speech production system. In the generalization investigations of Study 2, two aspects of this organization were revealed.

The investigation of context generalization showed that adapting [ε]'s production in the training words causes similar production changes in [ε] in the testing words. Thus, part of the process controlling the [ε]'s production in the training words must be also used in the testing words. This sharing of [ε]'s control process rules out the possibility that training and testing words have independent means of controlling their productions. It suggests instead the possibility that the words use a common vowel representation to access the shared control process. This, in turn, suggests the more general conclusion that words specify their productions indirectly via shared, intermediate production unit representations (e.g. phonemes).

The investigation of target generalization showed that adapting the production of [ε] causes similar production changes in other vowels. Thus, part of the process controlling production of [ε] must also be used in the production of the other vowels. This rules out the possibility that these vowels have independent means of controlling their productions. It suggests instead a similarity of the representations used to access control of their productions. This suggests the possibility that their representations may share some common set of features.

Technical limitations prevented making more detailed conclusions concerning context and target generalization. However, many of these limitations appear to be solvable, allowing future experiments to examine in more detail the organization of the speech production system.

5.3.4 Summary

In this chapter, the results of Study 2 were presented and discussed. Study 2 was designed to confirm and characterize more fully the speech SA effect seen in Study 1. It was also designed to investigate the timecourse and generalization of speech SA. A key part of this design was the use of separate sets of words for training and testing.

- Training words were used to assess compensation and adaptation. Subjects were prompted to produce training words while hearing either (1) feedback of their whispering or (2) masking noise which blocked their hearing.

- Testing words were used to assess generalization of the adaptation of the training words. Subjects were prompted to produce testing words only while masking noise blocked their hearing.

Analysis of the diverse results of Study 2 was organized into three categories:

- Compensation and adaptation results.

- Timecourse results.

- Generalization results.

Compensation and Adaptation Results Analysis of the compensation and adaptation results confirmed the speech SA effect seen in Study 1. Comparison with control experiment results showed that subjects altered their vowel productions specifically to compensate for altered feedback. These compensations were also partially retained in productions where their feedback was blocked by masking noise (and in fact, appeared to be partially retained for the month between the real and control experiments). Thus, subjects compensated and adapted, although the amounts they exhibited varied widely, with adaptation less than (or, in one case, equal to) compensation. In spite of this range of compensations, no subject reported noticing any feedback alteration. In one sense, this was the desired result: feedback was altered gradually to reduce subjects' perception of it. In another sense, it was a surprising result given how little some subjects compensated.

Timecourse Results Analysis of the timecourse results assessed how subjects increased their compensations and adaptations in response to the increasing alteration of their feedback. This analysis showed that, on average, subjects compensated for each increase in feedback alteration. However, they did not adapt until their compensation reached a certain value. This value differed across subjects. However, for each subject, this value was approximately equal to the final amount by which the subject's compensation exceeded his adaptation. This suggested subjects may have a preference for temporarily compensating for altered feedback, but their capacity to do so is limited; once it is exceeded, they must make long-term adaptations.

To summarize the above findings, the following theory was proposed for subjects' response to altered feedback:

1. Perception of the altered feedback is partially offset by perceptual adaptation. The capacity to adapt perception is limited and subject-specific.

2. The perceived feedback alteration is compensated for.

3. Compensation is preferentially achieved by a temporary correction mechanism (active only while exposed to the altered feedback). The capacity of the temporary correction mechanism to compensate is limited and subject-specific.

4. When required compensation exceeds the capacity of the temporary correction mechanism, long-term speech control adjustments must be made – i.e., adaptation occurs.

Generalization Results Finally, analysis of the generalization results showed the potential for using speech SA to investigate phonetic issues in the organization of speech production.

Context generalization assessment showed adaptation of [ε] in the training words caused similar production changes in [ε] in the testing words. This rules out the possibility that subjects used independent processes to control training and testing word productions. It suggests instead that a common representation of [ε] is shared in the production of words containing [ε]. This is consistent with the idea that words specify their productions via intermediate representations such as phonemes.

Target generalization assessment showed adaptation of [ε] in the training words caused similar production changes in other vowels in the testing words. This rules out the possibility that subjects used independent processes to control their different vowel productions. It suggests that vowels may have similar representations, possibly because their representations share some common set of features.

5.4 Individual Subject Results

This final section is a compilation of plots and descriptions of the individual subject results in Study 2. For each subject, the following three aspects of his results are considered:[17]

1. **Compensation and adaptation results**: overall compensation and adaptation responses in the real and control experiments.

2. **Timecourse results**: compensation and adaptation response timecourses in the real experiment.

3. **Generalization results**: how adaptation of training word vowel productions affected testing word vowel productions.

Subjects are discussed in order of decreasing adaptation, beginning with subject CW. CW's strong adaptation results are good examples to use in describing the layout of the plots. CW's results also clearly show the major features of all other subjects' results. Thus, all features of CW's results are described in detail. For subsequent subjects, their results are described in terms of how their features differ from the major features seen in CW's results.

5.4.1 Subject CW

Figures 5-12 through 5-16 show the results for subject CW – the subject exhibiting the largest mean adaptation. For this subject, the real experiment was run on 4/6/96, at 1:12 PM using the +2.0 feedback transform. The control experiment was run 51 days later, on 5/27/96 at 9:05 AM.

5.4.1.1 Compensation and Adaptation Results

Figure 5-12 shows CW's overall compensation and adaptation results.

[17] These are the same aspects discussed in Section 5.3.

Plot Layouts Figure 5-12(a) shows plots of CW's overall compensation responses in the real and control experiments. The figure consists of two plots: a vowel plot and a path projection plot. Figure 5-12(a)'s caption lists calculated mean compensation for CW in the real and control experiments.

Vowel Plot The vowel plot (Figure 5-12(a), left side) shows mean formant changes (test phase - baseline) of the subject's training word vowel productions in his compensation responses. The plot has four elements:

1. The subject's [i]–[ɑ] path.

2. A black arrow labeled "real exp" representing mean formant change of vowel productions in the real experiment.

3. A gray arrow labeled "control" representing mean formant change of vowel productions in the control experiment.

4. A hollow arrow that is the feedback transformation of the "real exp" arrow, and is called the *feedback image arrow.*

For both the "real exp" and "control" arrows, ellipses around the arrow's base and tip represent standard errors of the baseline and test phase mean (F1,F2) estimates, respectively. (The standard errors are often quite small, so these ellipses are often difficult to see.)

The *feedback image arrow* shows how the subject heard his own vowel formant changes in the real experiment (the "real exp" arrow) through his altered feedback. The arrow's base shows how the feedback transformation initially distorted the subject's hearing of his baseline vowel formants (i.e., the arrow's base is the feedback transformation of the "real exp" arrow's base). The arrow's tip shows how the subject heard his vowel formants in the test phase (i.e., the arrow's tip is the feedback transformation of the "real exp" arrow's tip).

The feedback image arrow is useful for graphically exhibiting how completely a subject compensates. When compensating, the subject tries to make his test phase vowel formants sound like his baseline vowel formants did before his feedback was altered. The feedback image arrow's tip represents how the subject hears his test phase vowel formants. The "real exp" arrow's base represents how the subject heard his baseline vowel formants before his feedback was altered. Thus, when compensating, the subject tries to make the feedback image arrow's tip approach the "real exp" arrow's base.[18]

Path Projection Plot The path projection plot (Figure 5-12(a), right side) shows path projections of the formant changes shown in the vowel plot. The solid line labeled "real exp" shows vowel path projection changes in the real experiment. The line links two filled dots. The filled dot above the "base" label is mean path projection of baseline vowel productions (i.e., the path projection of the "real exp" arrow's base in the vowel plot) The filled dot above the "test" label is mean path projection of test phase vowel productions (i.e., the path projection of the "real exp" arrow's tip in the vowel plot). In an analogous fashion, the dotted line labeled "control" that links the open dots shows vowel path projection changes in the control experiment. For all dots, standard error confidence intervals are shown as bars above and below the dots. (As with the vowel plots, the stand errors are often quite small. Thus, the confidence interval bars are often so close together that they are obscured by the dots.) Note also that, as was done in Section 5.3.1.2, numbering of path positions is reversed for +2.0 subjects like CW, but not for -2.0 subjects. Doing this makes path projection increases represent compensation for all subjects.

Figure 5-12(b) shows plots of CW's overall adaptation responses in the real and control experiments. It's layout is the same as Figure 5-12(a)'s, with one exception: no feedback image arrow is shown in the vowel plot. This is because the subject never hears the vowel production changes he makes in his adaptation response.[19] Figure 5-12(b)'s caption lists calculated mean adaptation for CW in the real and control experiments.

[18] For further explanation of the feedback image arrow, see Section 3.3.1.3.

[19] Recall that, by definition, a subject's *adaptation response* refers to his word productions made while he was prevented from hearing his whispering by masking noise.

Results Analysis CW's compensation response plots (Figure 5-12(a)) show four features common to most subjects' results:

1. **Baseline formant mismatch in the real experiment**: CW's baseline [ɛ] formants in the real experiment don't match those of [ɛ] on his [i]–[ɑ] path. Figure 5-12(a)'s plots show this in two ways:

   (a) The "real exp" arrow's base is not centered on [ɛ] on the [i]–[ɑ] path. Instead, it appears closer to [æ].

   (b) Baseline real experiment mean path projection is closer to 2.0 (the value for [æ]) than 3.0 (the value for [ɛ]).

   These results indicate a discrepancy between mean baseline [ɛ] formants in the real experiment and in the subject pretest when CW's [i]–[ɑ] path vowel formants were measured. Two possible causes of this discrepancy are apparent. One is that CW changed his production of [ɛ]after the subject pretest. The other is that the discrepancy is caused by differences in vowel formant measurement procedures. In the subject pretest, vowel formants were measured as average formant values of a subject's mean production of a CVC utterance. Coarticulatory influences from the initial ([b]) and final ([d]) consonants were minimized by by prompting the subject to extend his whispering of the utterance to 500 ms. This lengthened the utterance's vowel portion. In spite of this, it is still possible that coarticulation caused mean formants of the whole utterance to be noticeably different from the true steady-state vowel formants. Vowel formant measurement in the real experiment was done using a better method: an avgram was used to identify the steady-state vowel portion of the mean training word utterance. Mean vowel formants were then estimated by averaging only over this portion of the subject's training word utterances. In future experiments, both subject pretest and the actual experiment will use this vowel formant measurement method.

2. **Compensation in the real experiment**: CW's vowel production changes in the real experiment exhibit compensation. In fact, CW's vowel production

changes are so big that they fully compensate for the path projection shift of the feedback alteration. Figure 5-12(a)'s plots show this in three ways:

(a) The "real exp" arrow is long and oriented in the compensating direction.

(b) The feedback image arrow's tip position is close to the "real exp" arrow's base position. In fact, the two positions appear to have the same path projection.

(c) In the test phase, real experiment mean path projection is approximately 2.0 vowel units greater than it is in the baseline phase.

These results are in agreement with the calculated value of mean compensation in the real experiment. As shown in the figure caption, this value is 1.06 ± 0.03.

3. **Baseline shift in the control experiment**: CW's baseline [ɛ] formants in the control experiment are shifted in the direction in which he compensated in the real experiment. Figure 5-12(a)'s plots show this in two ways:

   (a) The "control" arrow's base is shifted in direction that CW compensated in the real experiment.

   (b) Baseline control experiment mean path projection (≈ 3.0) is greater than baseline real experiment mean path projection (≈ 2.0).

   In Section 5.3.1.2, it was discussed how this shift of control experiment baseline may result from a retention of production changes that were adapted in the real experiment.

4. **Small vowel production changes in the control experiment**: In the control experiment, CW's vowel production changes are much smaller than those of the real experiment. Figure 5-12(a)'s plots show this in two ways:

   (a) The "control" arrow is much smaller than the "real exp" arrow (and, in fact, points in the opposite direction).

   (b) In the test phase, control experiment mean path projection is only slightly less than it is in the baseline phase.

These results are in agreement with the calculated negligible value of mean compensation in the control experiment. As shown in the figure caption, this value is -0.20 ± 0.04.

These same trends can be seen in CW's adaptation response plots (Figure 5-12(b)). However, there is one interesting difference: in his real experiment adaptation response, CW leaves F1 unchanged from it's baseline value. In the plots, this is indicated by the fact that the "real exp" arrow is almost completely vertical. This feature is not seen in other the subjects' results.

Thus, in comparing CW's compensation and adaptation responses, the most salient difference concerns F1: in compensating, CW adjusts both F1 and F2, but in adapting, he adjusts only F2. It appears, therefore, that CW has made long-term adjustments only to his control of F2; his adjustment of F1 is a temporary correction, present only during exposure to the altered feedback.[20]

5.4.1.2 Timecourse Results

Figure 5-13 shows the timecourse of CW's compensation and adaptation responses in the real experiment.

Plot Layouts Figure 5-13(a) shows the timecourse of of CW's compensation response. The figure consists of two plots: a vowel plot and a mean compensation plot.

Vowel Plot The vowel plot (Figure 5-13(a), left side) shows the sequence of mean formant changes in CW's training word vowel productions over the course of the real experiment. The plot shows mean vowel formants in each of a sequence of experiment intervals. Each labeled dot represents mean vowel formants during the experiment interval indicated by the label. As always, the ellipse around the dot

[20] Note also that the F1 lowering effect (see discussion of subjects MF and JK in Study 1) does not explain the results. The key feature of the F1 lowering effect is that baseline F1 is lower in a subject's adaptation response than in his compensation response. Examination of the "real exp" arrows in the vowel plots of Figure 5-12 shows this is not the case for CW.

represents the standard error of the mean formant estimates. A dashed line connects the dots of the labeled intervals in the order they occur in the experiment. A portion of CW's [i]–[ɑ] path is also shown.

As mentioned in Section 5.3.2.1, it is important to note that the labeled intervals represent differing amounts of time:

- The interval labeled "base" represents the entire 17-minute baseline phase.
- The intervals labeled "1" – "9" represent ramp stages 1 – 9, which were each about 2 minutes long.
- The intervals labeled "train1", "train2", and "train3" represent successive 20-minute time intervals of the train phase.
- The interval labeled "test" represents the entire 17-minute test phase.

Mean Compensation Plot The mean compensation plot (Figure 5-13(a), right side) shows CW's mean compensation in each of the experiment intervals shown in the vowel plot. The plot's x-axis lists the intervals in the experiment's timecourse: "base" is the baseline phase; "1"-"9" are stages 1–9 of the ramp phase; "train1", "train2", and "train3" are the 1st, 2nd, and 3rd 20-minute intervals of the train phase; "test" is the test phase. For each interval, mean compensation and adaptation are shown as black and white dots, respectively, on the y-axis.

Also as mentioned in Section 5.3.2.1, the differing interval durations have two consequences. First, there is less data to average in each ramp stage interval than in the other intervals. Thus, the ellipses and confidence intervals of the ramp stage measurements are larger than those of the other intervals. Second, there is a timescale discontinuity at ramp stage 9: up to this point, each ramp stage interval represents another 2 minutes in the experiment; past this point, each interval represents roughly another 20 minutes in the experiment. Any apparent jumps in compensation between ramp stage 9 and the train1 interval (as seen here in CW's mean compensation plot) are thus possibly due to the timescale discontinuity.

Figure 5-13(b) shows the timecourse of of CW's adaptation response. The figure consists of two plots: a vowel plot and a mean adaptation plot. The layout of the figure is exactly the same as that of Figure 5-13(a).

Results Analysis CW's compensation timecourse plots (Figure 5-13(a)) show four features common to most subjects' results:

1. **Monotonic compensation increases**: CW increases his compensation almost monotonically over the course of the experiment. Figure 5-13(a)'s plots show this in two ways:

   (a) In the vowel plot, it appears that path projections of his mean vowel formants steadily move towards the [i]-end of his [i]–[ɑ] path.

   (b) At all intervals in the mean compensation plot, mean compensation either increases or shows an insignificant decrease (e.g., at ramp stage 9).

   Monotonic compensation increases are predicted because the feedback alteration monotonically increased during the experiment. That is, the feedback alteration was increased to its maximum distortion during the ramp phase and held at this value for the rest of the experiment.

2. **Inconsistent path deviation changes**: the vowel plot shows that CW's path deviation changes were much less consistent than his path projection changes.

   Inconsistent path deviation changes are predicted because the feedback alteration affected perceived path projections, not perceived path deviations.

3. **A stable limit to compensation**: by about midway through the train phase, CW's compensation appears to have reached a stable limit. This is best seen in the mean compensation plot: it appears that mean compensation levels out to a value of 1.0 by the train2 interval.

   For CW, it can be argued that his compensation reaches a limit because he achieves complete compensation. As subsequent plots will show, other subjects reach stable compensation limits that are less than complete. This was discussed in Section 5.3.1. There, it was hypothesized that less-than-complete compensation occurs if a subject's speech perception partially adapts to the altered feedback.

4. **No clear compensation delay**: there is no clear evidence of any delay in the onset of CW's compensation response. This can be seen in the ramp phase portion of the mean compensation plot. There, the standard error confidence intervals are large enough that the jump in compensation visible between ramp stages 3 and 4 is probably not significant. In fact, a linear compensation increase is a good fit to the timecourse of mean compensation and its confidence intervals in the ramp-phase.

   The lack of any delay or discontinuous jump in mean compensation in the ramp phase was discussed in Section 5.3.2. There, it was explained that, because no clear jump in compensation was seen, there is no evidence that categorical perception affects subjects' compensations.

   Note also that, as mentioned above, the jump in mean compensation visible between ramp stage 9 and the train1 interval may not be significant: it may be an artifact of the timescale discontinuity at this point in the graph.

CW's adaptation timecourse plots (Figure 5-13(b)) show all of the same features seen in his compensation timecourse plots, except for one differing feature. The adaptation plots exhibit **a probable delay in adaptation**: there is evidence of a delay in the onset of CW's adaptation response. This can be seen in the ramp phase portion of the mean adaptation plot. There, the standard error confidence intervals are small enough that the jump in adaptation visible between ramp stages 3 and 4 is probably significant.

As subsequent plots will show, this probable delay in adaptation is seen in many other subjects' results as well. In Section 5.3.2, this delayed adaptation onset was hypothesized to be caused by subjects preferring to initially make only temporary corrections for the increasingly altered feedback.

5.4.1.3 Generalization Results

Figures 5-14 through 5-16 show plots of CW's generalization results.

Plot Layouts Figure 5-14 shows avgrams of CW's testing word productions in the real experiment. In each avgram, the solid lines show formant tracks of the mean utterance in the baseline phase, while the dashed lines show formant tracks of the mean utterance in the test phase. The gray regions show the utterance interval that was used for vowel analysis.

Figure 5-15 shows vowel plots of subject CW's generalization results. The figure's left plot (Figure 5-15(a)) shows CW's context generalization results. Black arrows show mean vowel (F1,F2) changes (test phase - baseline phase) for his context generalization word productions in the real experiment. White arrows show the same changes in the control experiment. In a similar fashion, the figure's right plot (Figure 5-15(b)) shows CW's target generalization results. It shows mean vowel (F1,F2) changes for his target generalization word productions. (Note that, to facilitate comparisons, the vowel production change arrow for "pep" appears in both the left and right plots.)

Figure 5-16 shows mean path projection changes (test phase - baseline) of CW's testing word vowel productions. Figure 5-16(a) shows mean path projection changes in the real experiment, while Figure 5-16(b) shows the same for the control experiment. Figure 5-16 shows mean relative path projection changes. This figure is essentially the difference between figures 5-16(a) and 5-16(b).

Results Analysis CW's generalization plots show two technical problems seen in most subjects' results:[21]

1. **Weak or missing formants**: analysis of vowel production changes in certain testing words can't be performed because of weak or absent formants within their vowel analysis intervals. CW's avgrams show this is the case for "pop".

[21] Note that these problems are also discussed in Section 5.3.3.2 using subject OB's generalization results.

In the avgram of "pop", baseline F2 is missing for most of the vowel analysis interval (gray region), meaning that its amplitude is below plotting threshold for most of this interval. Thus, production changes for [ɑ] in "pop" are not analyzable.

2. **Coarticulation**: analysis of vowel production changes in certain testing words is confounded by coarticulation of the initial stop consonant with the following vowel. In CW's results, both "gep" and "teg" exhibit coarticulation that affects calculation of their mean path projections. The plots show this in three ways:

   (a) In Figure 5-16(a), the mean path projection changes of "gep" and "teg" in the real experiment are noticeably lower than the mean path projection change of "peg".

   (b) The context generalization vowel plot (Figure 5-15(a)) shows that these lower mean path projection changes occur because mean (F1,F2) change vectors for "gep" and "teg" in the real experiment (black arrows) are smaller than those of "pep" and "peg". The plot shows this difference in mean (F1,F2) change vector sizes is partially due to the difference in baselines. Baseline mean (F1,F2) positions for "gep" and "teg" have higher F2 values than those of "pep" and "peg".

   (c) In Figure 5-14, the avgrams show that these differences in F2 values are probably due to coarticulation effects. For "teg", F2 in the baseline phase is initially higher than it is in "pep" and "peg". Over the course of the utterance, F2 in "teg" makes no noticeable transition down from its initially elevated position. Similarly, for "gep", F2 in the baseline phase is initially higher than it is in "pep" and "peg". F2 in "gep" then takes much longer to transition to the steady-state vowel than it does in the other words. In fact, F2 is making this downward transition throughout the vowel analysis interval. Thus, within this interval, average F2 in "gep" is higher than it is in "peg" or "pep".

These differences in F2 behavior are ascribed to coarticulation because they correlate with the different types of initial consonants. Since the initial [p] in "pep" and "peg" is a bilabial stop, F2 is initially low but quickly transitions to a steady-state value that it holds for the rest of the utterance. This quick F2 transition is possible because the articulation of [p] does not need the tongue, so the tongue can be preset to its position for [ɛ]. In "gep", however, since [g] is a velar stop, F2 is initially high. It then takes time for the tongue to move from its position for [g] to its position for [ɛ]. This makes the F2 transition to the steady-state vowel much slower. In "teg", since [t] is an alveolar stop (and thus articulated with the tongue), F2 is also initially high.

CW's generalization plots show four features of his generalization that are seen in most subjects' results:

1. **Adaptation of $\mathbf{W_{test}}$ "pep"**: CW appears to have adapted his production of [ɛ] in $\mathbf{W_{test}}$ "pep". The plots show this in four ways:

   (a) The plot of mean relative path projection change (Figure 5-16) shows, for $\mathbf{W_{test}}$ "pep", it is positive and large compared to its confidence intervals.

   (b) The mean path projection change plots (figures 5-16(a) and 5-16(b)) show this large positive mean relative path projection change results because mean path projection change for $\mathbf{W_{test}}$ "pep" in the real experiment is large but in the control experiment is small.

   (c) In Figure 5-15, the vowel plots show that these differences in mean path projection change result from differences in the size and orientation of the mean (F1,F2) change vectors for "pep". Its change vector in the real experiment (black arrow labeled "pep") is large, roughly aligned with the [i]–[ɑ] path (in fact, it is almost vertical), and is pointing in the direction that compensates for the 2.0 feedback transformation. Its change vector in the control experiment (white arrow labeled "pep") is smaller and much less aligned with the [i]–[ɑ] path.

(d) In Figure 5-14, the avgram for "pep" shows the compensating production change as an elevation of F2: F2 in the test phase (dashed line) is higher than F2 in the baseline phase (solid line).

That $\mathbf{W}_{\mathbf{test}}$ "pep" should show adaptation is predicted because it is the only testing word that is also a training word.

2. **Context generalization**: The adaptation of [ε] in $\mathbf{W}_{\mathbf{test}}$ "pep" generalizes to different word contexts – i.e., it causes similar production changes in [ε] in the other context generalization words ("gep", "peg", and "teg"). The plots exhibit this generalization as an compensatory change in the production of [ε] in the context generalization words. For each context generalization word, the generalization plots exhibit the compensatory [ε] production change in the same four ways they show adaptation of [ε] in $\mathbf{W}_{\mathbf{test}}$ "pep" (see point 1 above).

3. **Target generalization**: The adaptation of [ε] in $\mathbf{W}_{\mathbf{test}}$ "pep" generalizes to different vowel targets – i.e., it causes similar production changes in the vowels of most of the analyzable target generalization words ("peep" and "pip"). The plots exhibit this generalization as an compensatory vowel production changes in the target generalization words. For each target generalization word, the generalization plots again exhibit the compensatory vowel production change in the same four ways they show adaptation of [ε] in $\mathbf{W}_{\mathbf{test}}$ "pep".

4. **Zero generalization**: For certain testing words, the path projections of their vowel production changes in the real and control experiments are nearly equal. This makes mean relative path projection change (and thus mean generalization) zero for the this testing word.[22] CW's generalization plots show this has happened for "pap":

   (a) The plot of mean relative path projection change shows it is essentially zero for "pap".

[22] Section 5.3.3.2 describes how mean generalization is calculated from mean relative path projection changes.

(b) The mean path projection plots show this zero mean relative path projection change results from mean path projection change for "pap" in the real experiment being approximately equal to its value in the control experiment.

(c) The target generalization vowel plot shows that these equal path projections result from the orientations of the mean (F1,F2) change vectors for "pap". Its mean (F1,F2) change vector in the real experiment is large and oriented nearly vertically, while its mean (F1,F2) change vector in the control experiment is large and oriented nearly horizontally. Because of their orientation with respect to the [i]–[ɑ] path, it appears both change vectors have nearly the same path projection changes and differ only in their path deviation changes.

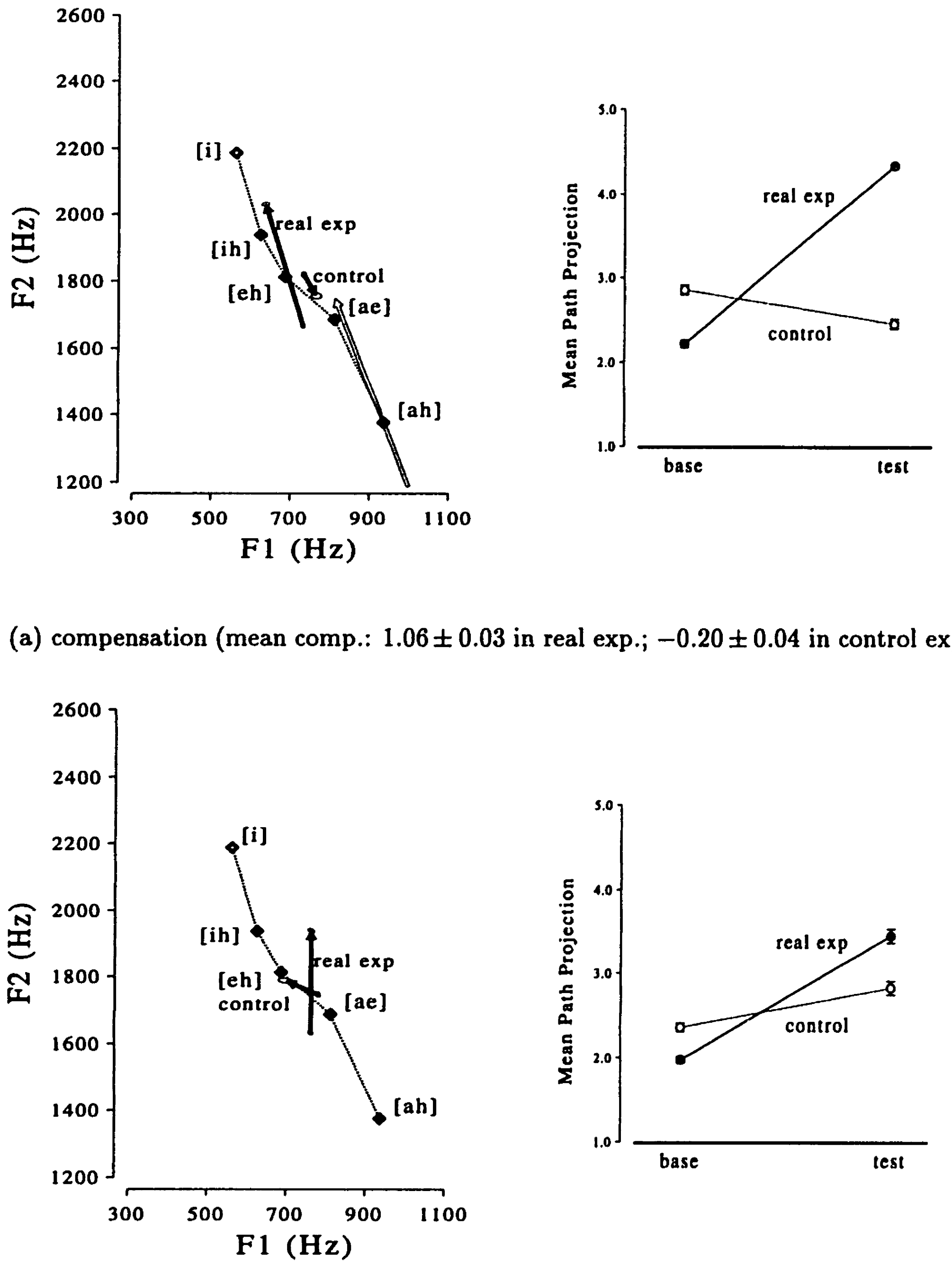
Generalization, as we have defined it, means how adaptation of training word vowel productions (as seen in $\mathbf{W}_{\mathbf{test}}$ "pep") affect testing word vowel productions. A measure of generalization should therefore gauge how correlated the vowel production changes seen in a testing word are with those seen in $\mathbf{W}_{\mathbf{test}}$ "pep".

However, in CW's results, it appears the measured zero mean generalization for "pap" may not accurately represent how correlated the vowel production changes of "pap" are with those of "pep". The target generalization vowel plot shows the mean (F1,F2) change vectors for "pap" exhibit the same orientation differences between the real and control experiments that are seen in the mean (F1,F2) change vectors for "pep". In this sense, adaptation of [ε] in "pep" has caused similar production changes in [ε] in "pap". In this case, calculation of mean generalization based on path projection changes fails to capture this production similarity.

Thus, mean generalization calculated using path projection changes may occasionally be an overly conservative measure of how related some testing and training word vowel productions are. However, zero mean generalization for a

> testing word is not always an indication of a problem with the generalization measure. As subsequent plots will show, zero mean generalization also results when, in the control experiment, a testing word's vowel production change is large while the vowel production change of "pep" is small. In such a case, mean generalization based on path projections does correctly indicate a lack of correlation between the vowel production changes seen in a testing word and those seen in "pep".

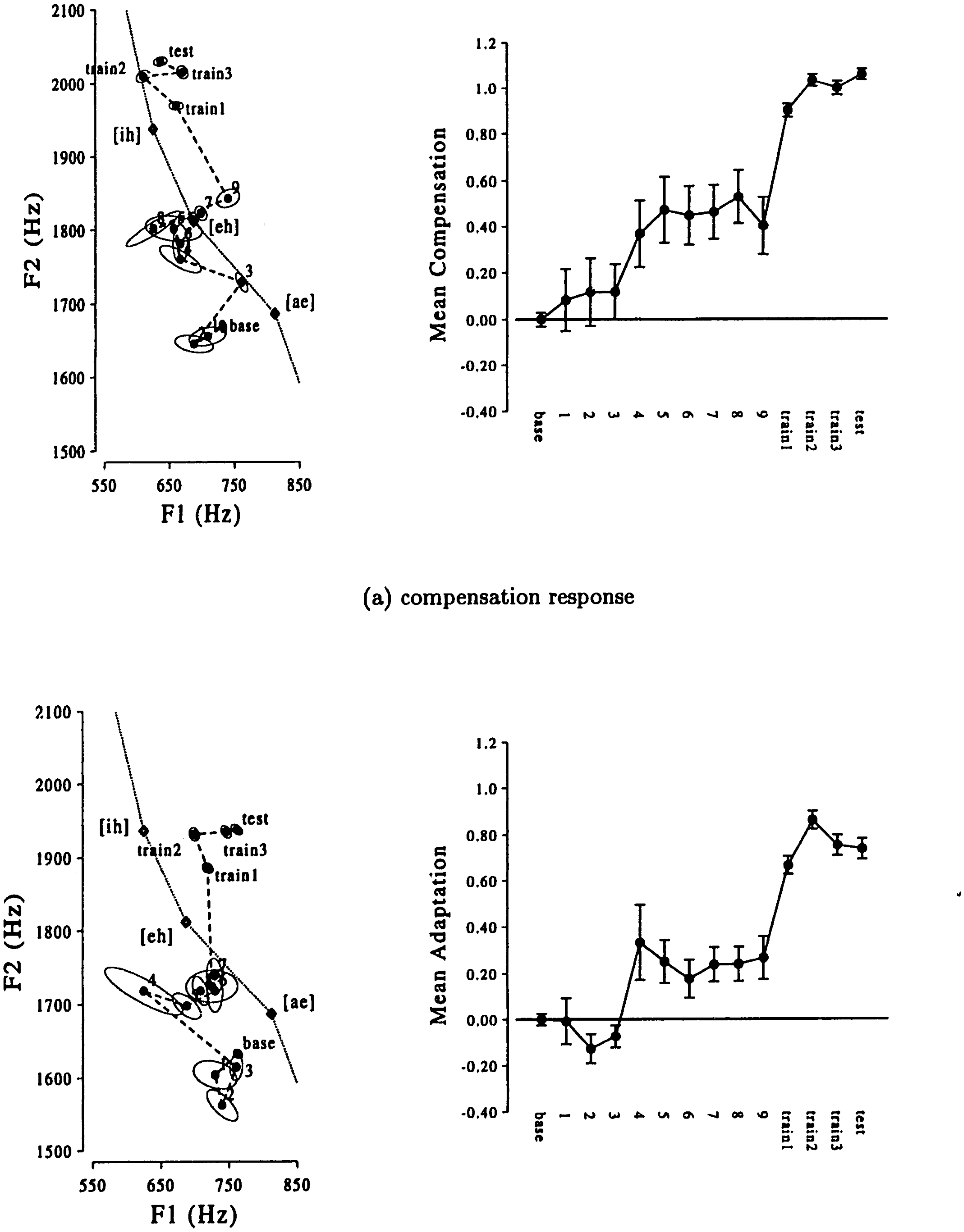
A feature of CW's generalization results that is not seen in other subjects' results is the pattern of testing word mean (F1,F2) change vectors in the control experiment. The context generalization plot shows mean (F1,F2) changes in the real experiment are principally in the increasing F2 direction, while mean (F1,F2) changes in the control experiment are principally in the decreasing F1 direction. The target generalization plot shows a more revealing pattern: control experiment mean (F1,F2) change arrows show not only an F1 decrease, but also show what appears to be a convergence towards a some common point. Such convergence is typical of a subject that ceases paying attention to correct word pronunciations: he tends towards producing the same sound for all vowels. Perhaps this occurred because, unlike the real experiment, the control experiment was run fairly early in the morning (9:05 AM) for this subject.



(a) compensation (mean comp.: 1.06 ± 0.03 in real exp.; −0.20 ± 0.04 in control exp.)

(b) adaptation (mean adapt.: 0.73 ± 0.05 in real exp.; 0.22 ± 0.05 in control exp.)

Figure 5-12: Subject CW overall compensation and adaptation responses.



(a) compensation response

(b) adaptation response

Figure 5-13: Subject CW compensation and adaptation timecourses.

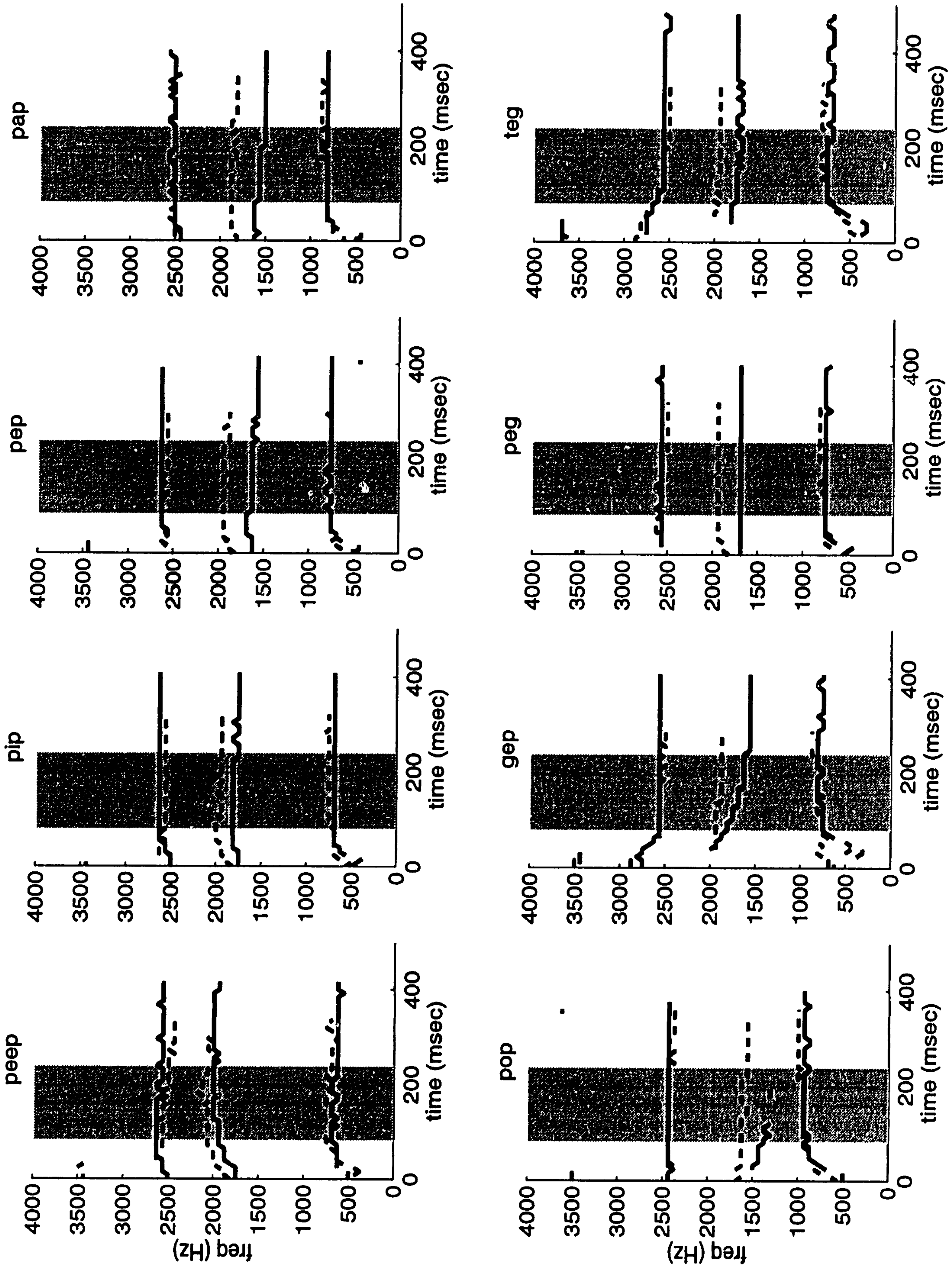
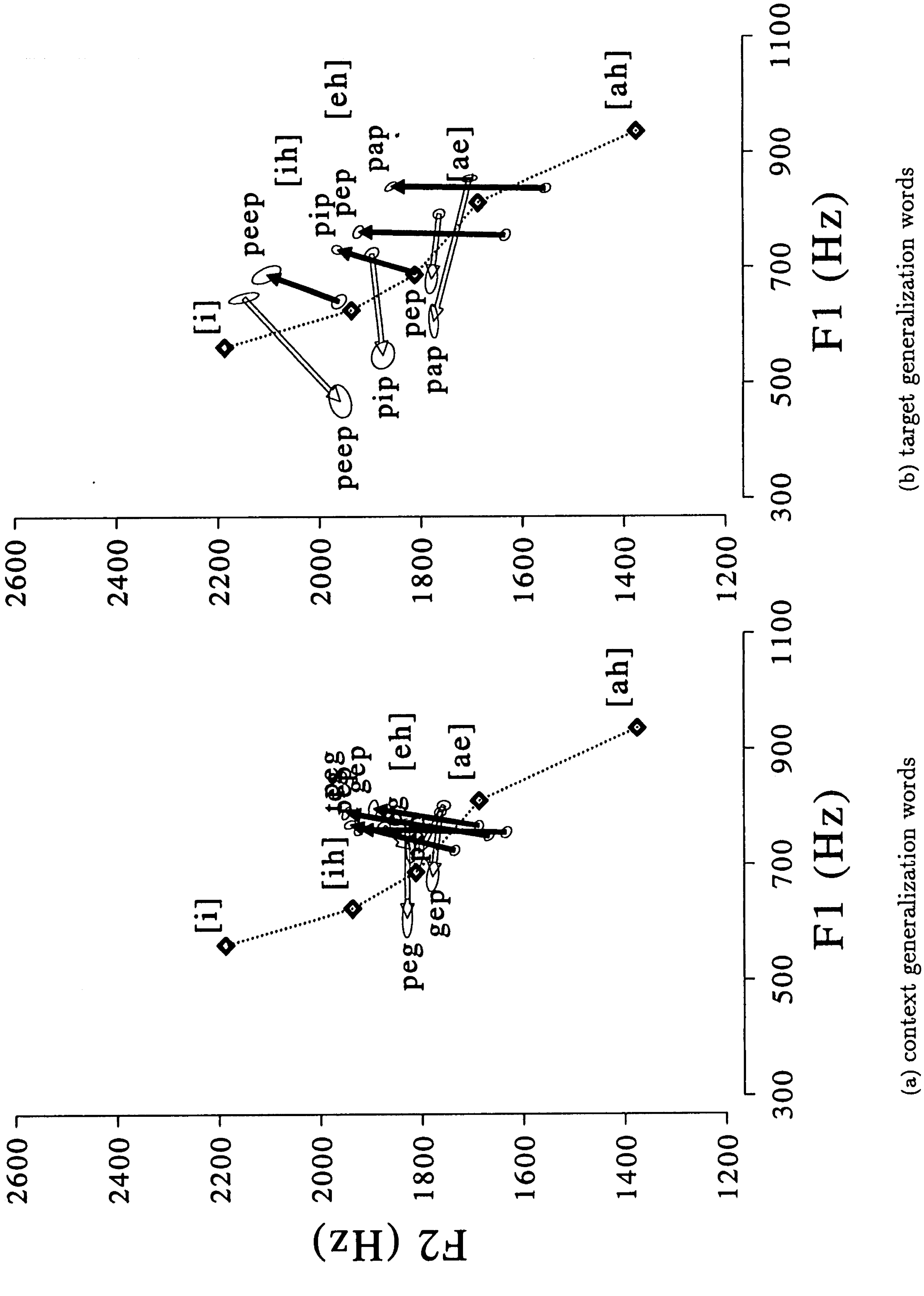


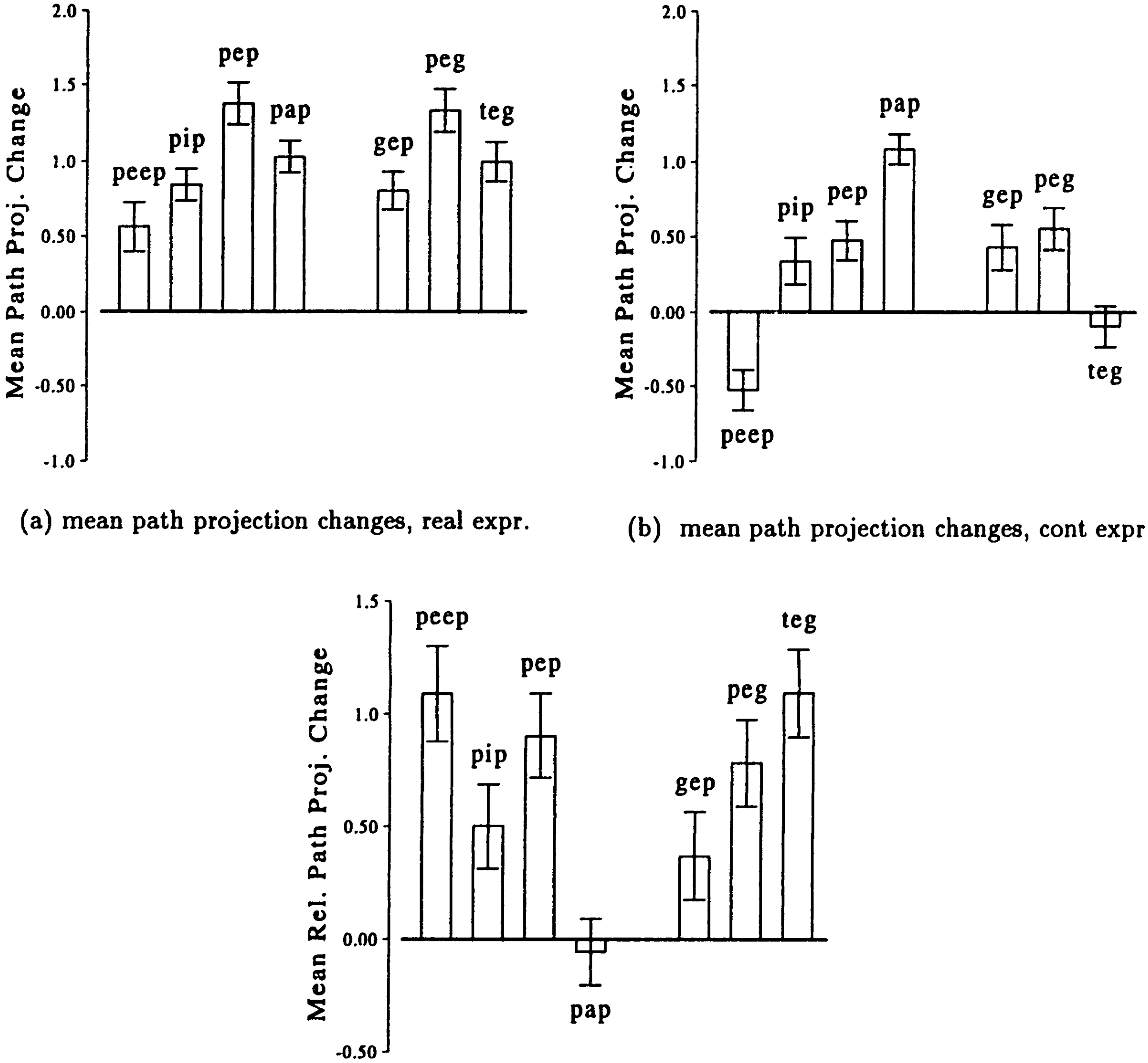
Figure 5-14: Subject CW testing word avgrams.



(a) context generalization words

(b) target generalization words

Figure 5-15: Subject CW testing word vowel plots.



(a) mean path projection changes, real expr.

(b) mean path projection changes, cont expr.

(c) mean relative path projection changes

Figure 5-16: Subject CW testing word path projection changes.

5.4.2 Subject RS

Figures 5-17 through 5-21 show the results for subject RS. For this subject, the real experiment was run on 5/28/96, at 11:41 AM using the +2.0 feedback transform. The control experiment was run 2 days later, on 5/30/96 at 9:01 AM.

Compensation and Adaptation Results Figure 5-17 shows that RS's overall compensation and adaptation results exhibit the same major features seen in CW's results. However, RS's adaptation response is not restricted to the F2 dimension.

Timecourse Results Figure 5-18 shows that RS's compensation and adaptation timecourse results exhibit all the major features seen in CW's results except for the following:

1. There is no evidence of a probable delay in adaptation. In fact, RS's adaptation timecourse is roughly similar to his compensation timecourse.

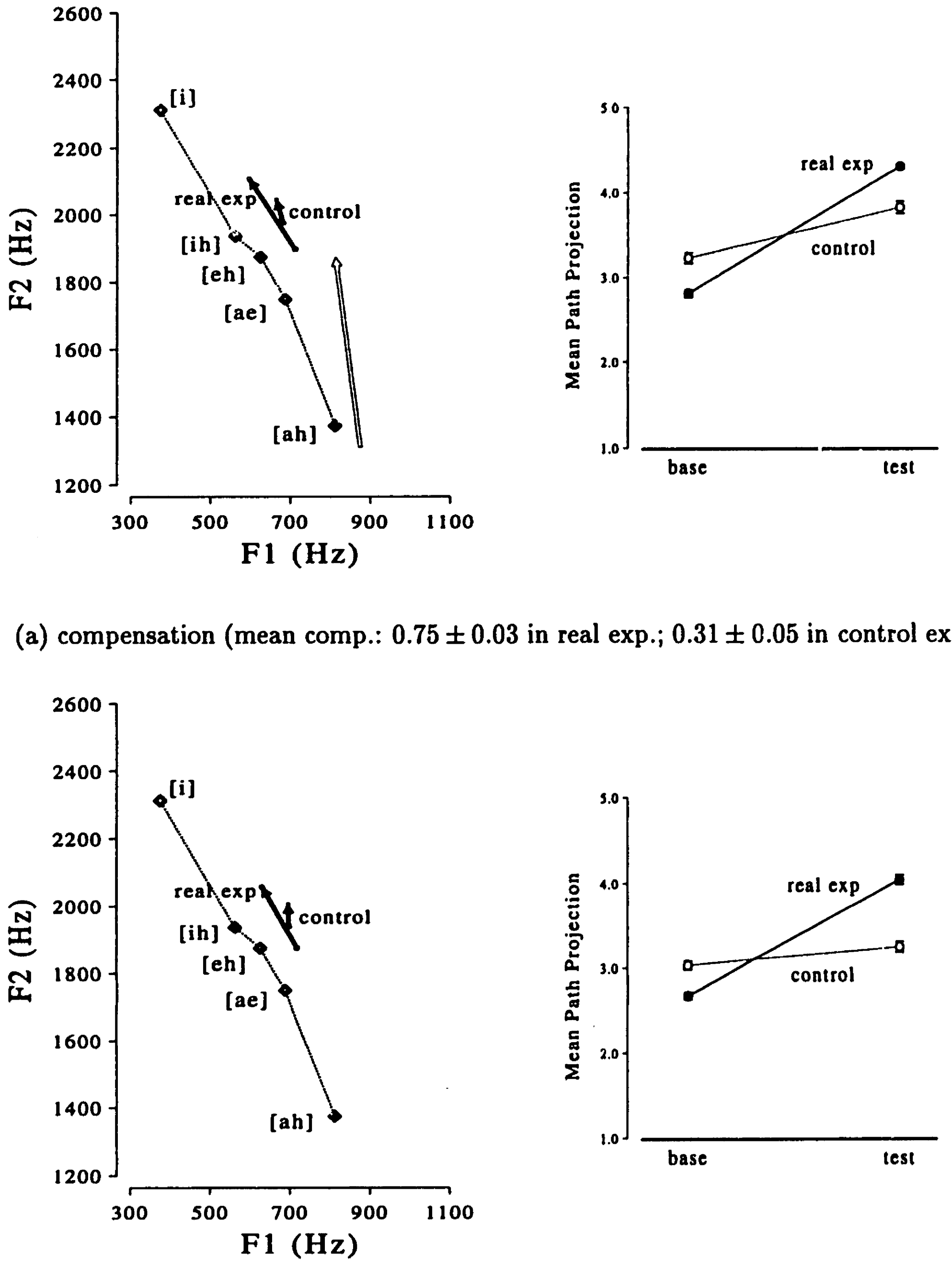
2. Neither his compensation nor his adaptation appear to have reached a stable limit by the end of the experiment. In the train and test phase intervals, RS's mean compensation and mean adaptation are still increasing, suggesting that if the experiment had continued, RS would have achieved even greater compensation.

Generalization Results Figures 5-19 through 5-21 show plots of RS's generalization results.

The first noticeable feature of RS's generalization results is seen in the avgrams (Figure 5-19): RS's utterance durations are short - about half as long as subject CW's. In most cases the shortness of RS's utterances significantly reduced how much of his utterances fell within the vowel analysis intervals (gray regions).

Other technical problems exhibited in RS's generalization results are similar to those seen in CW's results. Baseline phase F2 is missing in the vowel regions of the avgrams of both "peep" and "pop", preventing vowel analysis for these words. Coarticulation can be seen in the avgrams of "teg" and, especially, "gep". In "gep", F2 does not complete its transition to the steady-state vowel before the utterance ends.

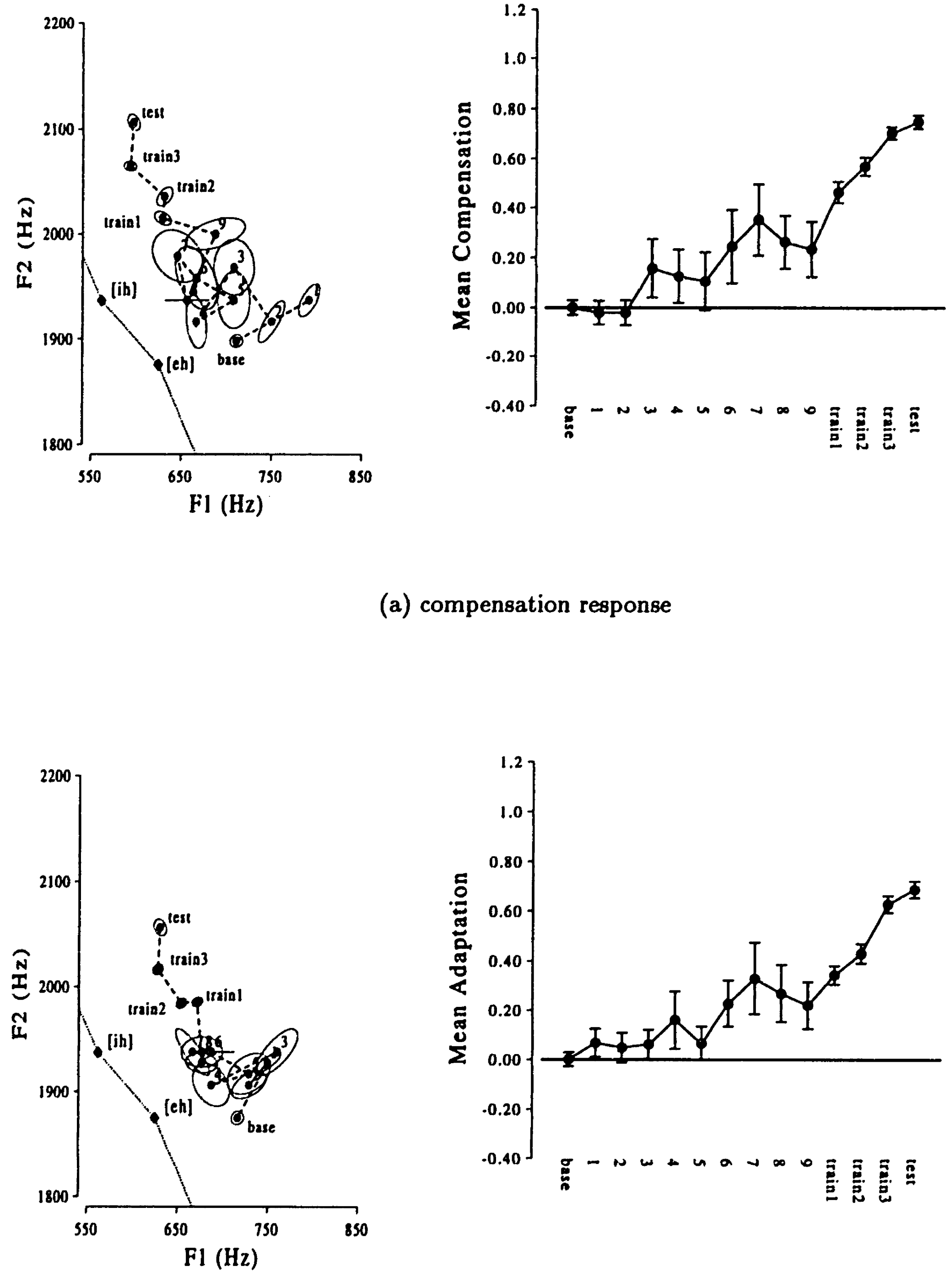
RS's generalization results also exhibit all of the major generalization features seen in CW's results. $\mathbf{W}_{\mathbf{test}}$ "pep" exhibits adaptation; "peg" exhibits context generalization: "pip" and "pap" exhibit target generalization; "gep" and "teg" exhibit zero generalization.



(a) compensation (mean comp.: 0.75 ± 0.03 in real exp.; 0.31 ± 0.05 in control exp.)

(b) adaptation (mean adapt.: 0.69 ± 0.03 in real exp.; 0.10 ± 0.04 in control exp.)

Figure 5-17: Subject RS overall compensation and adaptation responses.



(a) compensation response

(b) adaptation response

Figure 5-18: Subject RS compensation and adaptation timecourses.

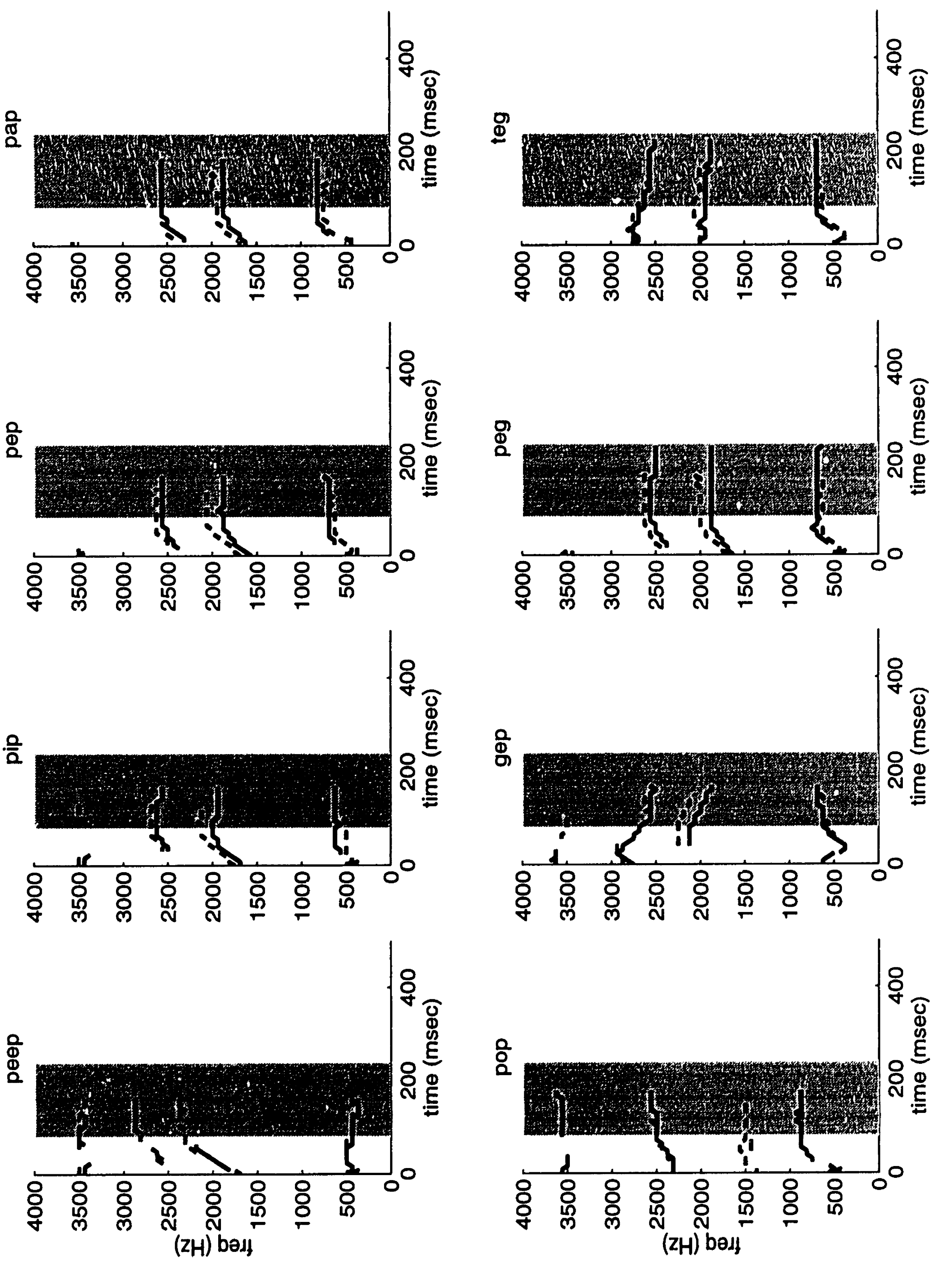
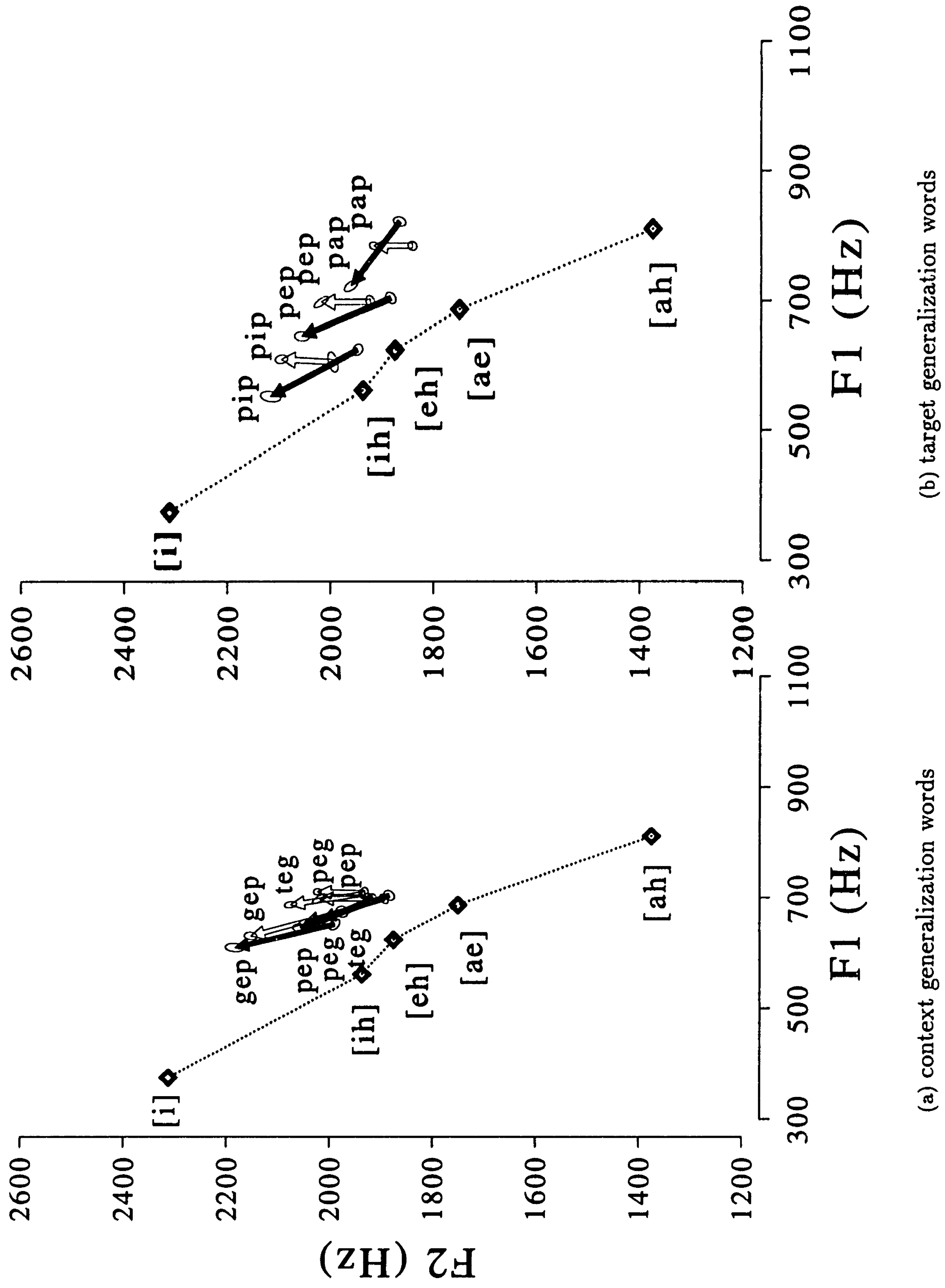
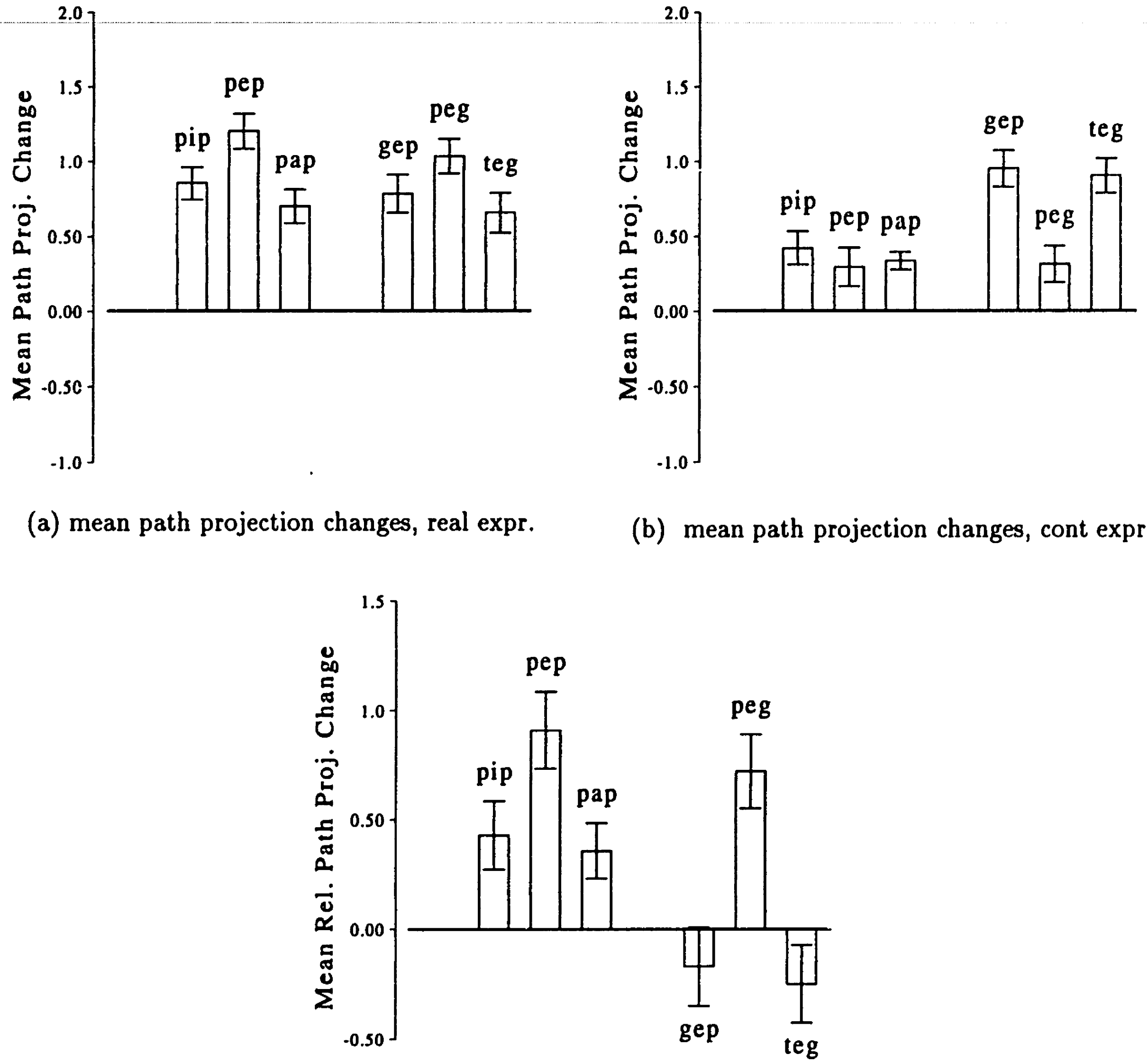


Figure 5-19: Subject RS testing word avgrams.


(a) context generalization words

(b) target generalization words

Figure 5-20: Subject RS testing word vowel plots.



(a) mean path projection changes, real expr.

(b) mean path projection changes, cont expr.

(c) mean relative path projection changes

Figure 5-21: Subject RS testing word path projection changes.

5.4.3 Subject OB

Figures 5-22 through 5-26 show the results for subject OB. For this subject, the real experiment was run on 4/3/96, at 3:16 PM using the -2.0 feedback transform. The control experiment was run 48 days later, on 5/21/96 at 2:50 PM. Subject OB showed the largest adaptation of all subject run with the -2.0 feedback transform.

Compensation and Adaptation Results Figure 5-22 shows that OB's overall compensation and adaptation results exhibit the same features described in CW's results, with the difference that OB compensates by shifting path projections in the opposite direction on the [i]–[ɑ] path.

A striking aspect of this OB's results is that he shows virtually no formant change in the control experiment in either his compensation (Figure 5-22(a)) or adaptation (Figure 5-22(b)) responses. In both figures, the "control" arrow is barely visible in the vowel plots.

Timecourse Results Figure 5-23 shows that OB's compensation and adaptation timecourse results exhibit all the same features described in CW's results. Notably, OB's mean compensation plot (Figure 5-23(a)) shows confidence intervals in the ramp stage are small enough to observe some detail of the ramp stage timecourse. Except for the dip at ramp stage 3, it appears that OB roughly linearly increased mean compensation during the ramp stage. OB's mean adaptation plot (Figure 5-23(b)) shows that, like CW, OB exhibits a probable delay in adaptation. The plot shows mean adaptation is approximately zero up to ramp stage 3, at which point it swiftly increases to equal mean compensation in ramp stage 4. Also like CW, OB's compensation and adaptation appear to reach stable limits by about midway through the train phase.

Generalization Results Plots of OB's clear generalization results were presented and discussed in Section 5.3.3.2 in order to illustrate how generalization is assessed. These plots are repeated here as Figures 5-24 through 5-26.

The technical problems exhibited in OB's results are similar to those seen in CW's results. Baseline phase F2 is missing in "peep", while test phase F2 is missing in "pop", preventing vowel analysis for these words. Coarticulation can be seen in the avgrams of "teg" and, especially, "gep".

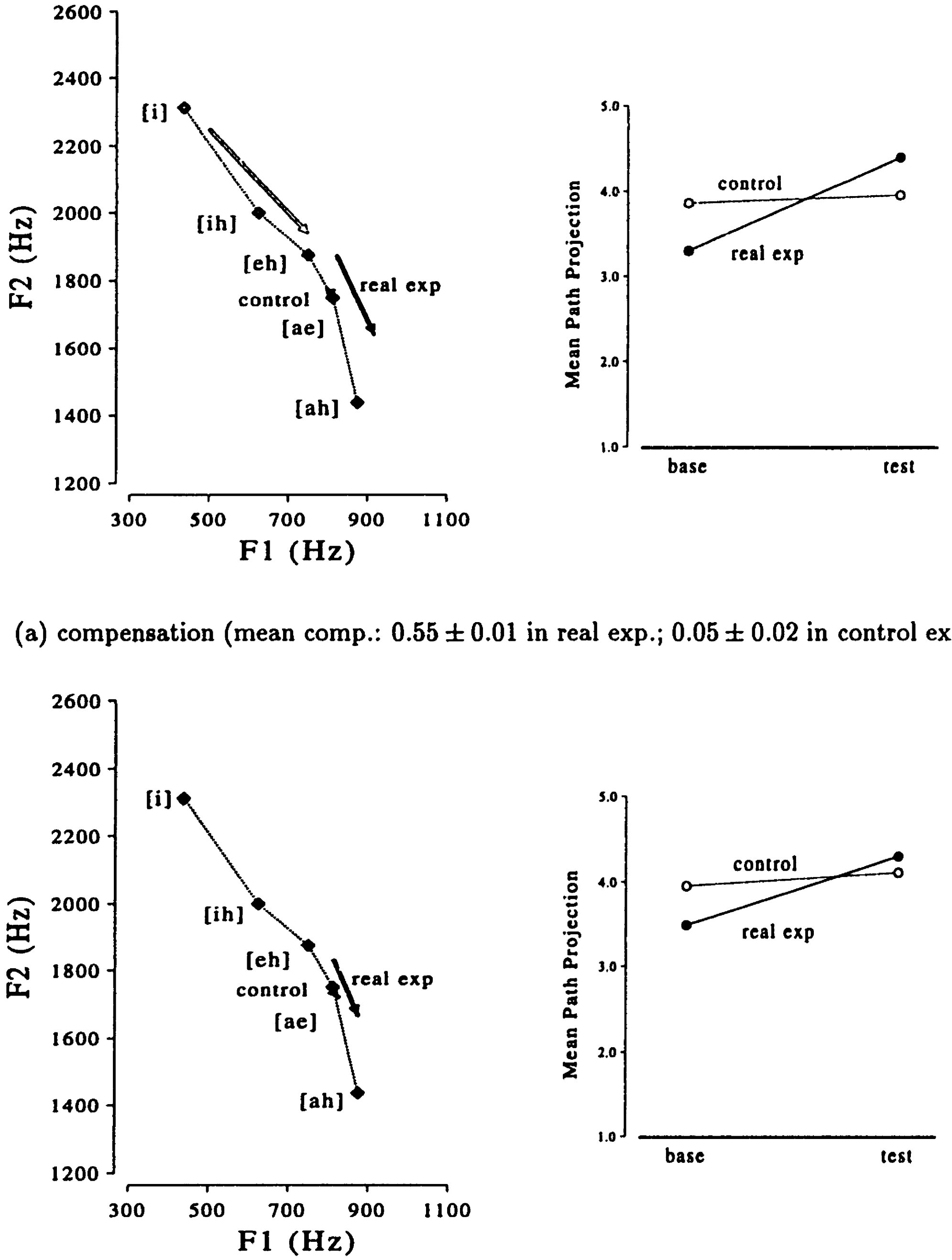
OB's generalization results also exhibit all of the major generalization features seen in CW's results. $\mathbf{W_{test}}$ "pep" exhibits adaptation (note that, in this case, adaptation is indicated by F1 increasing and F2 decreasing). All context generalization words exhibit context generalization. "pap" exhibits target generalization. "pip" exhibits zero generalization.

The zero generalization of "pip" is interesting because of how it occurs: "pip" is the only analyzable testing word whose vowel production changes in the control experiment appear substantial. The plots show this in three ways:

1. The plot of mean relative path projection change (Figure 5-26(c)) shows it to be near zero only for "pip".

2. The mean path projection change plots (figures 5-26(a) and 5-26(b)) show the zero mean relative path projection change for "pip" occurs because "pip" exhibits substantial mean path projection change in the control experiment – the highest of all analyzable testing words.

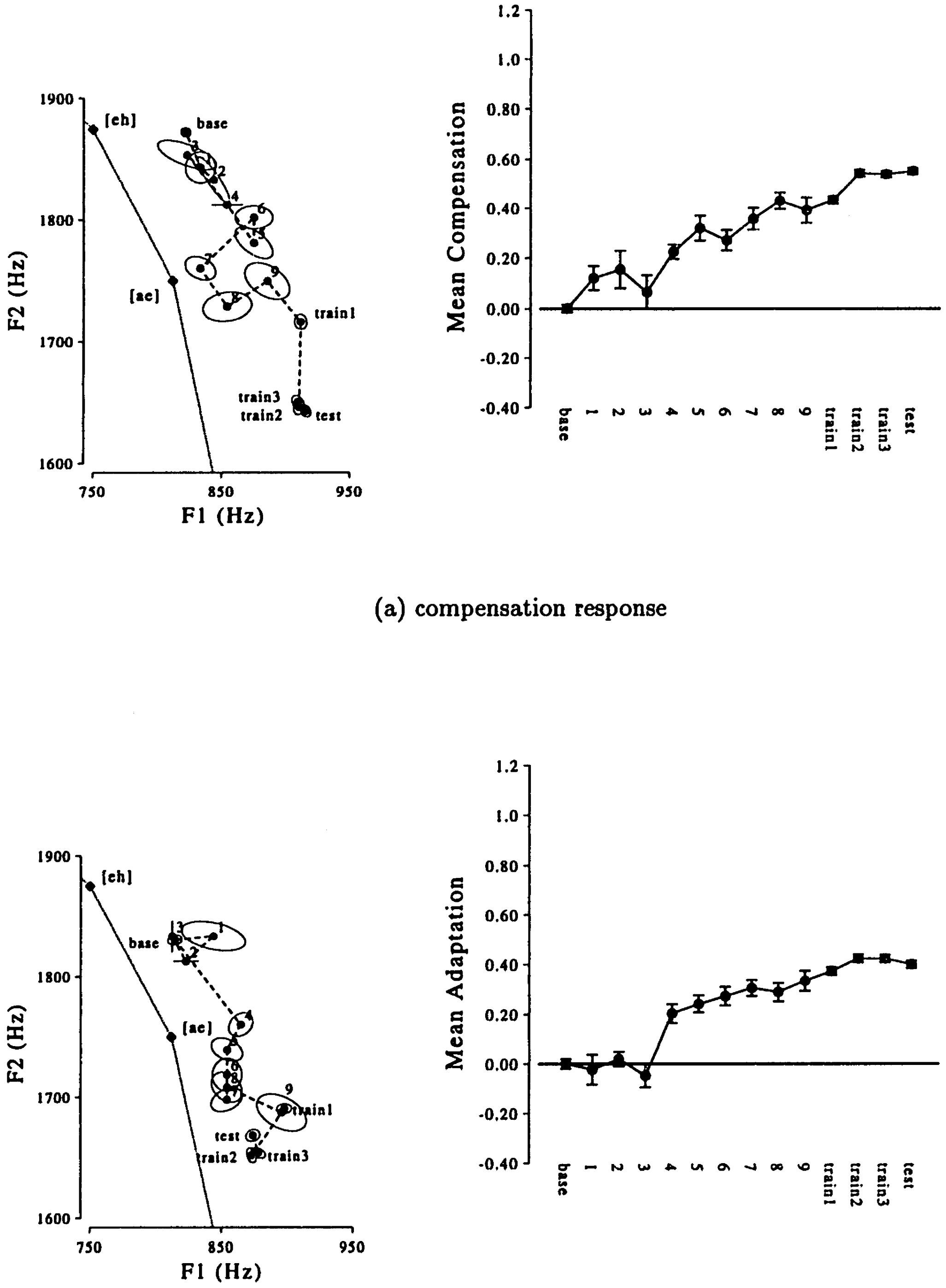
3. The vowel plots (Figure 5-25) shows the situation most clearly. They show that mean (F1,F2) change vectors in the control experiment (white arrows) are very small for all testing words except "pip". For "pip", its mean (F1,F2) change vector in the control experiment is very similar to its mean (F1,F2) change vector in the real experiment.



(a) compensation (mean comp.: 0.55 ± 0.01 in real exp.; 0.05 ± 0.02 in control exp.)

(b) adaptation (mean adapt.: 0.40 ± 0.02 in real exp.; 0.08 ± 0.01 in control exp.)

Figure 5-22: Subject OB overall compensation and adaptation responses.



(a) compensation response

(b) adaptation response

Figure 5-23: Subject OB compensation and adaptation timecourses.

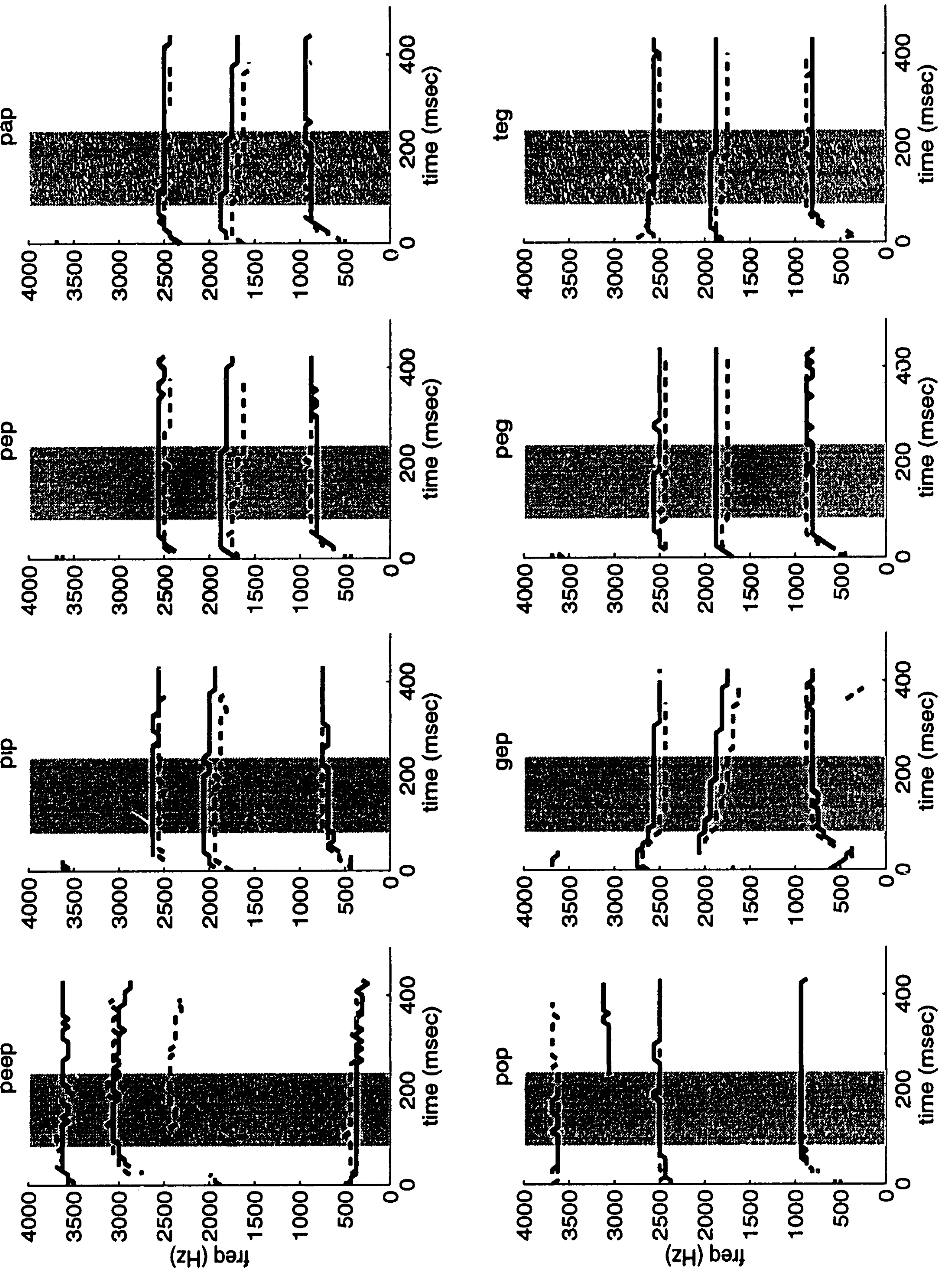
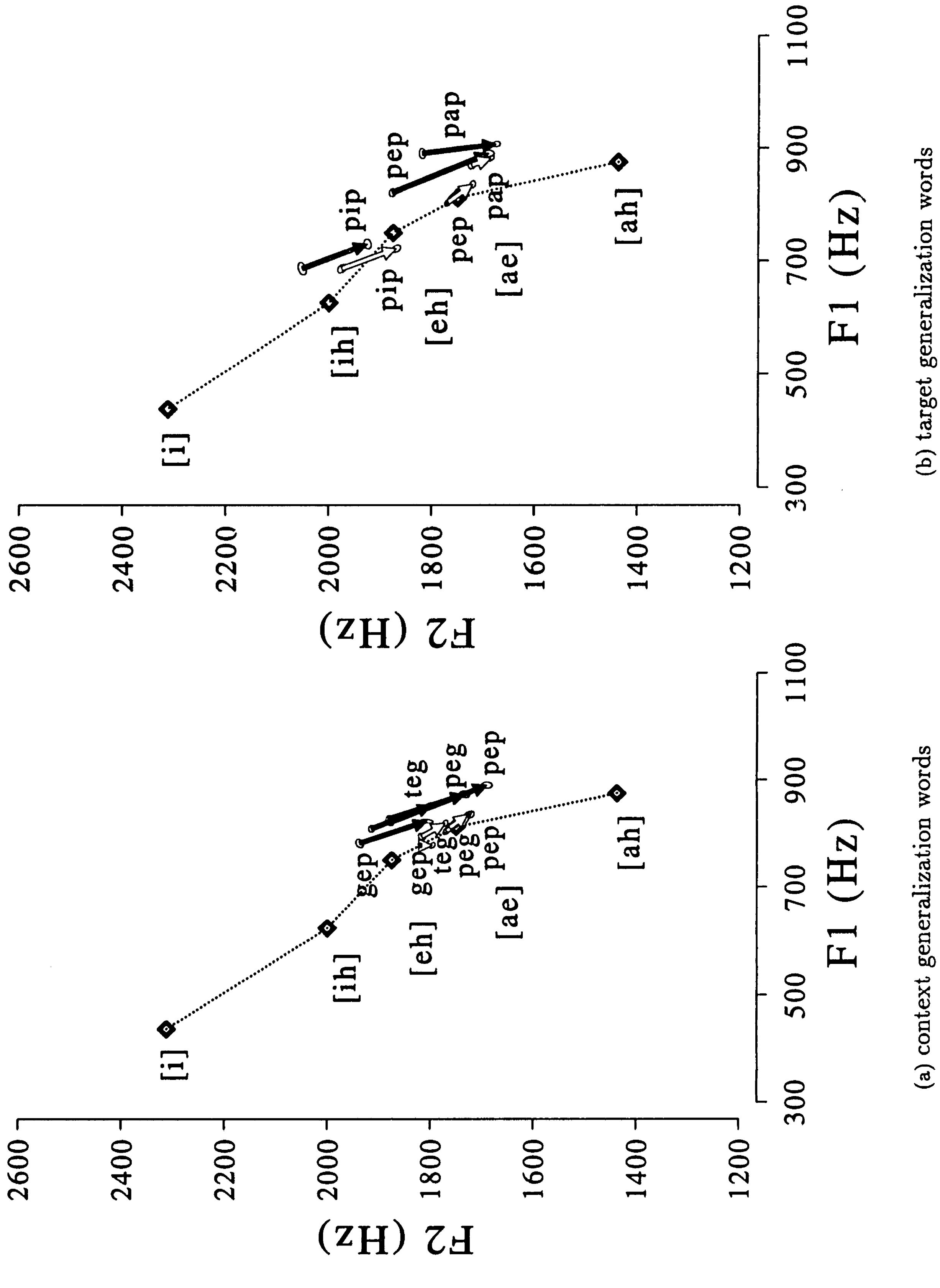


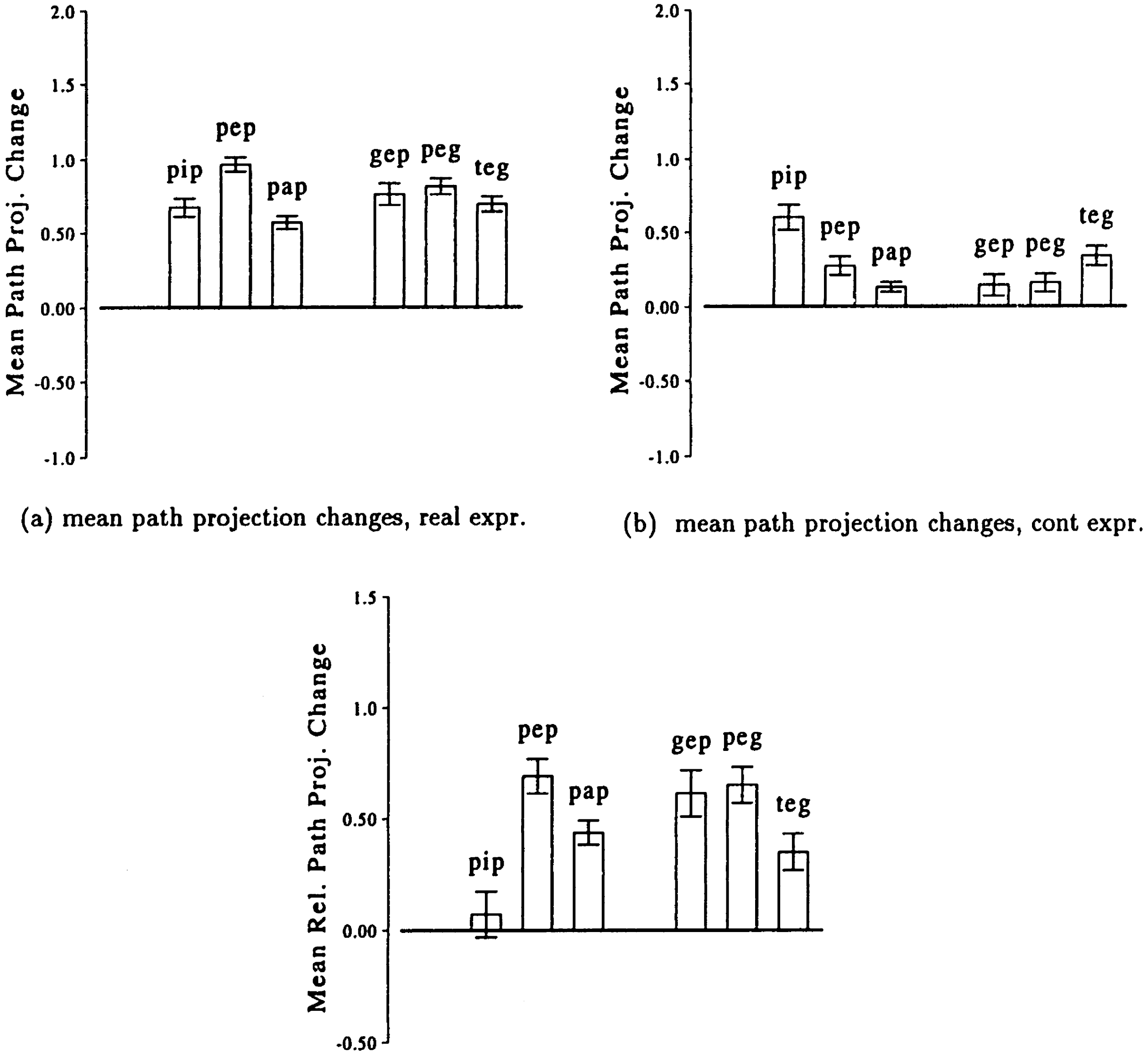
Figure 5-24: Subject OB testing word avgrams.



(a) context generalization words

(b) target generalization words

Figure 5-25: Subject OB testing word vowel plots.



(a) mean path projection changes, real expr.

(b) mean path projection changes, cont expr.

(c) mean relative path projection changes

Figure 5-26: Subject OB testing word path projection changes.

5.4.4 Subject SR

Figures 5-27 through 5-31 show the results for subject SR. For this subject, the real experiment was run on 4/16/96, at 2:30 PM using the -2.0 feedback transform. The control experiment was run 30 days later, on 5/16/96 at 1:31 PM.

Compensation and Adaptation Results Figure 5-27 shows that SR's overall compensation and adaptation results exhibit all the major features seen in CW's results. However, an interesting difference is that SR's adaptation is substantially less than his compensation. The figure's plots show this in two ways:

1. The mean (F1,F2) change vector of SR's adaptation (the "real exp" arrow in Figure 5-27(b)) is less than half the length of the mean (F1,F2) change vector of his compensation (the "real exp" arrow in Figure 5-27(a)).

2. The path projection change of SR's adaptation (the solid line labeled "real exp" in Figure 5-27(b)) is also less than half the path projection change of his compensation (the solid line labeled "real exp" in Figure 5-27(a)). Interestingly, this reduced adaptation results partly from baseline path projection of his real experiment adaptation response being significantly higher than baseline path projection of his real experiment compensation response.

It therefore appears that in his adaptation response in the real experiment SR reacted to the presence of masking noise by raising his vowel path projections. The adaptation response vowel plot shows this path projection raising corresponds to a lowering of F2 and a raising of F1. This response to the masking noise differs significantly from the F1-lowering response of subjects MF and JK in Study 1.

Timecourse Results Figure 5-28 shows that SR's compensation and adaptation timecourse results exhibit all the same features described in CW's results. However, SR's results exhibit more extreme contrasts between his compensation and adaptation timecourses.

The mean compensation plot shows no evidence of any delay in the onset of SR's compensation. Instead, his mean compensation appears to increase linearly in the ramp phase and achieves a stable limit (close to 1.0) midway through the train phase. On the other hand, the mean adaptation plot shows a very large delay in the onset of SR's adaptation. The plot shows mean adaptation is approximately zero up to ramp stage 8, at which point it swiftly increases to equal mean compensation in ramp stage 9. After ramp stage 9, SR's mean adaptation immediately levels out to a roughly stable limit for the rest of the experiment. This limit is less than 0.5.

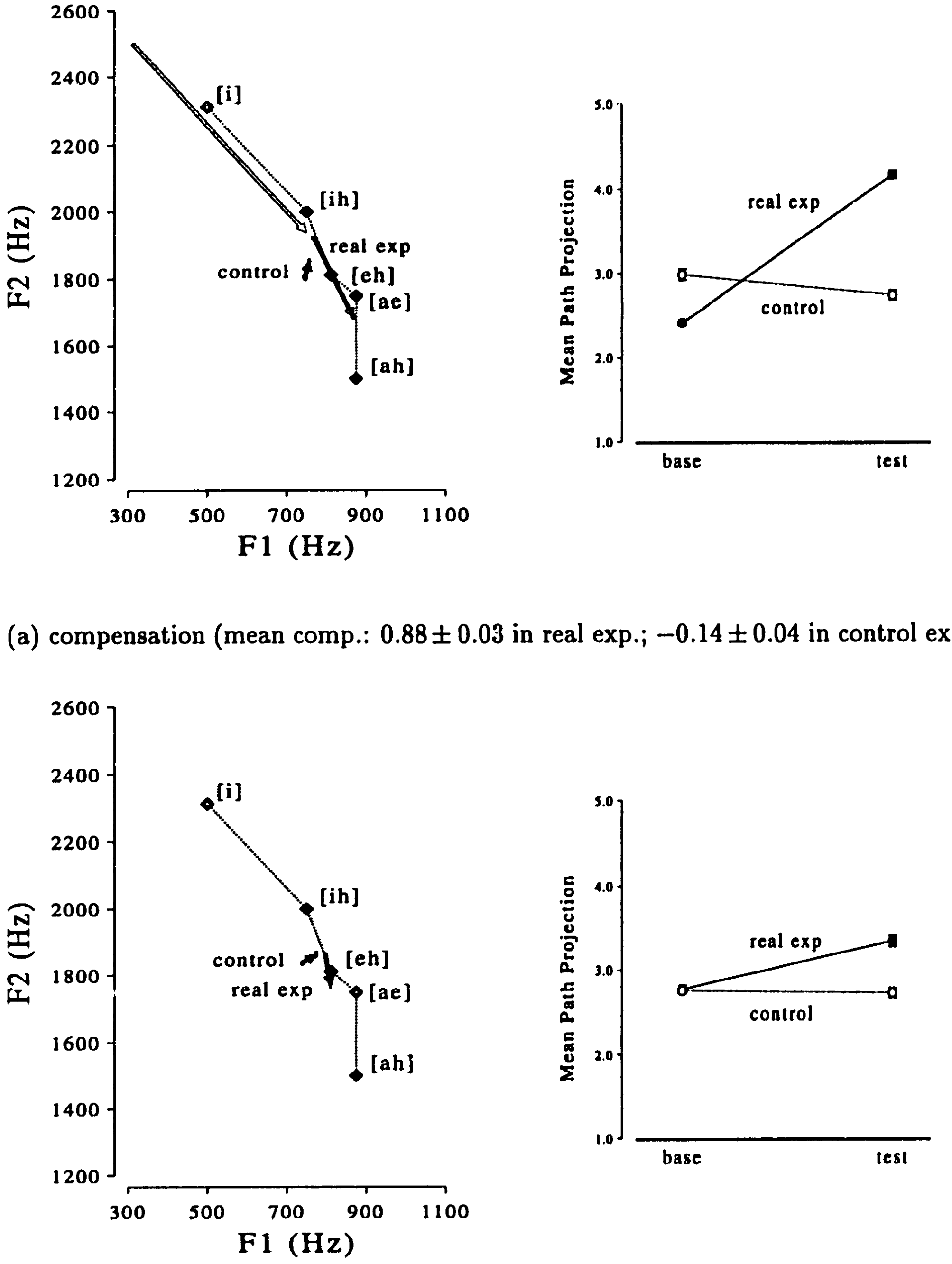
As discussed in Section 5.3.2, this extreme contrast between SR's compensation and adaptation timecourses is evidence that SR has a large capacity to compensate using only temporary production changes: only after the feedback alteration reaches near-maximum distortion is SR compelled to make long term adaptations.

Generalization Results Figures 5-29 through 5-31 show plots of SR's generalization results.

SR's generalization results do not exhibit exactly the same technical problems seen in CW's results. The avgrams of "gep" and "teg" do show the usual effects of coarticulation on F2, However, no test words were excluded because of missing formants: the avgrams of both "peep" and "pop" show analyzable formants in both baseline and test phases. A technical problem seen only in SR's results was the extreme lowering of baseline F1 in "gep" and "teg". The avgrams show this lowering is about 250 Hz in "teg" and 500 hz in "gep". Both words' avgrams also show that, after lowering, baseline F1 rises up to its normal level at the end of the utterance. This anomalous behavior suggests F1 measurement instabilities, and thus "gep" and "teg" were considered to have unanalyzable formants.

SR's generalization results do exhibit all of the major generalization features seen in CW's results. $\mathbf{W}_{\mathbf{test}}$ "pep" exhibits adaptation. "peg" exhibits context generalization. "pap" and "pop" exhibit target generalization. "peep" and "pip" exhibit zero generalization.

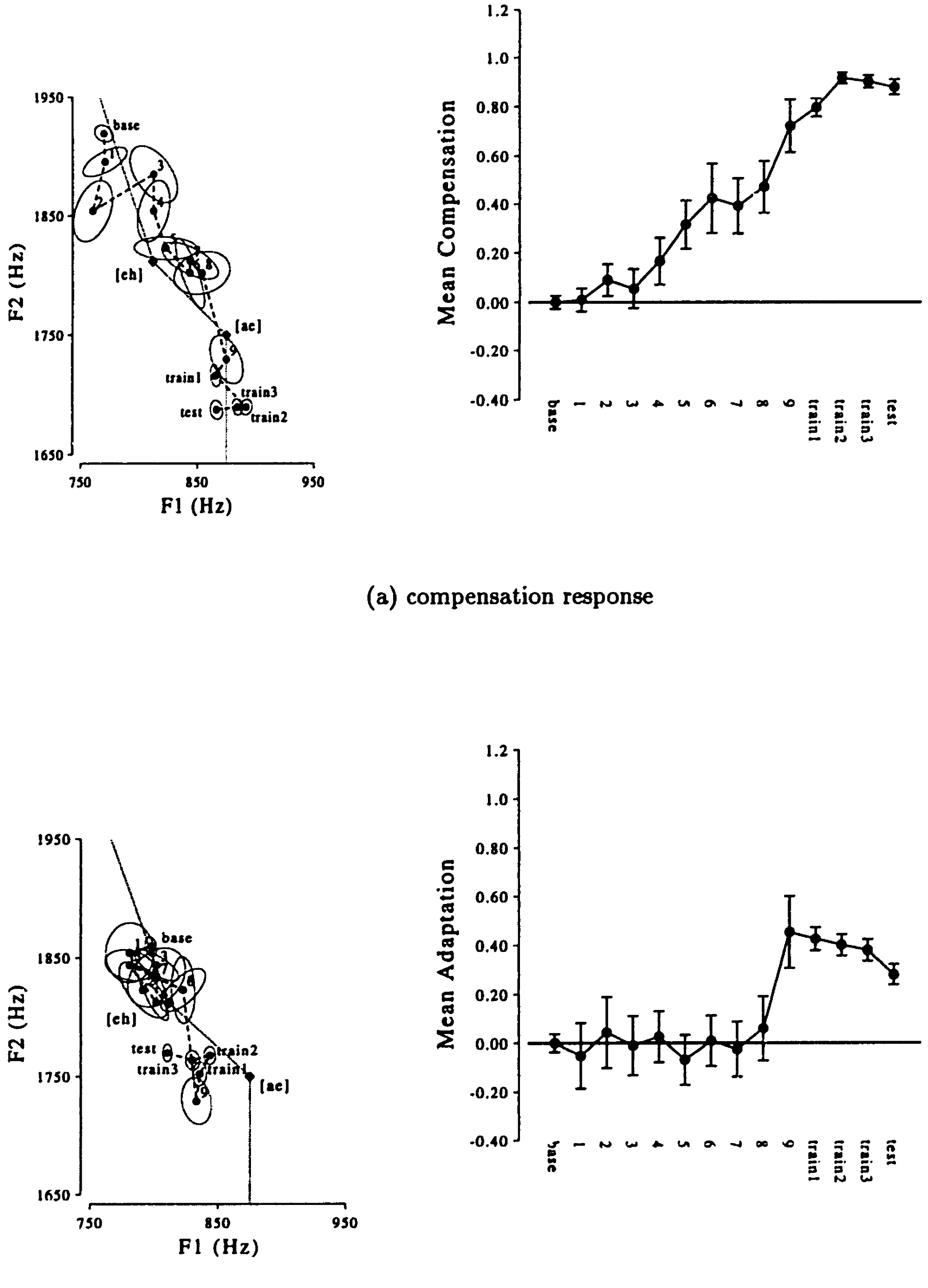
However, a caveat to SR's generalization results can be seen in the target generalization vowel plot (Figure 5-30(b)). The plot shows that, in the real experiment, mean (F1,F2) change vectors for the target generalization words show some variation in orientation, but, except for "peep", each target generalization word's vector terminates at the path vowel position corresponding to the word's vowel. This suggests that the mean (F1,F2) change vectors may represent a response to the masking noise, not adaptation and its generalization. The baseline positions of the change vectors could be a perturbation from normal (F1,F2) caused by SR hearing the masking noise. This perturbation effect disappears over the duration of the experiment, so that, by the test phase, SR's vowel productions are back to their normal (F1,F2) positions.



(a) compensation (mean comp.: 0.88 ± 0.03 in real exp.; −0.14 ± 0.04 in control exp.)

(b) adaptation (mean adapt.: 0.28 ± 0.04 in real exp.; −0.02 ± 0.03 in control exp.)

Figure 5-27: Subject SR overall compensation and adaptation responses.



(a) compensation response

(b) adaptation response

Figure 5-28: Subject SR compensation and adaptation timecourses.

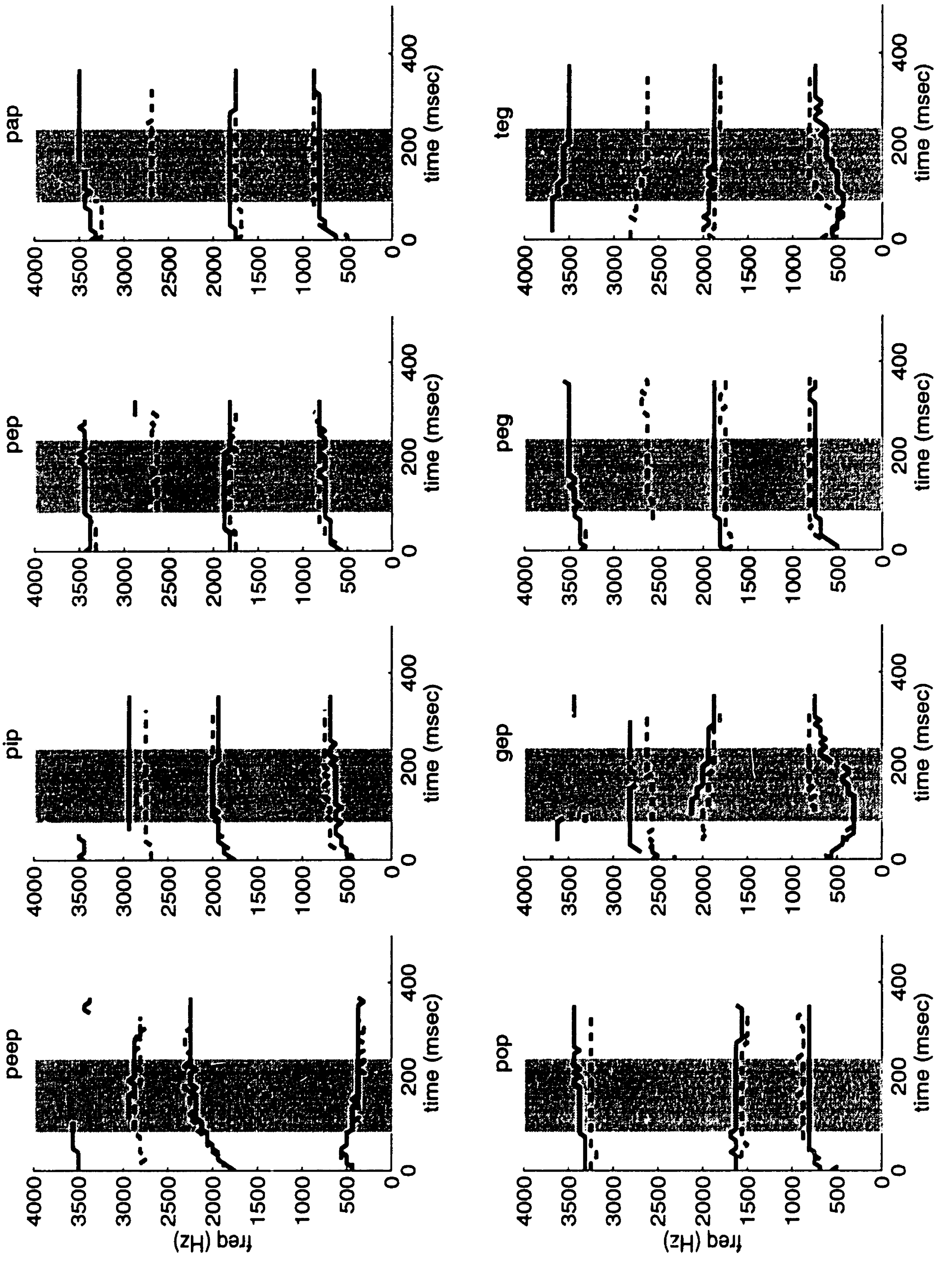
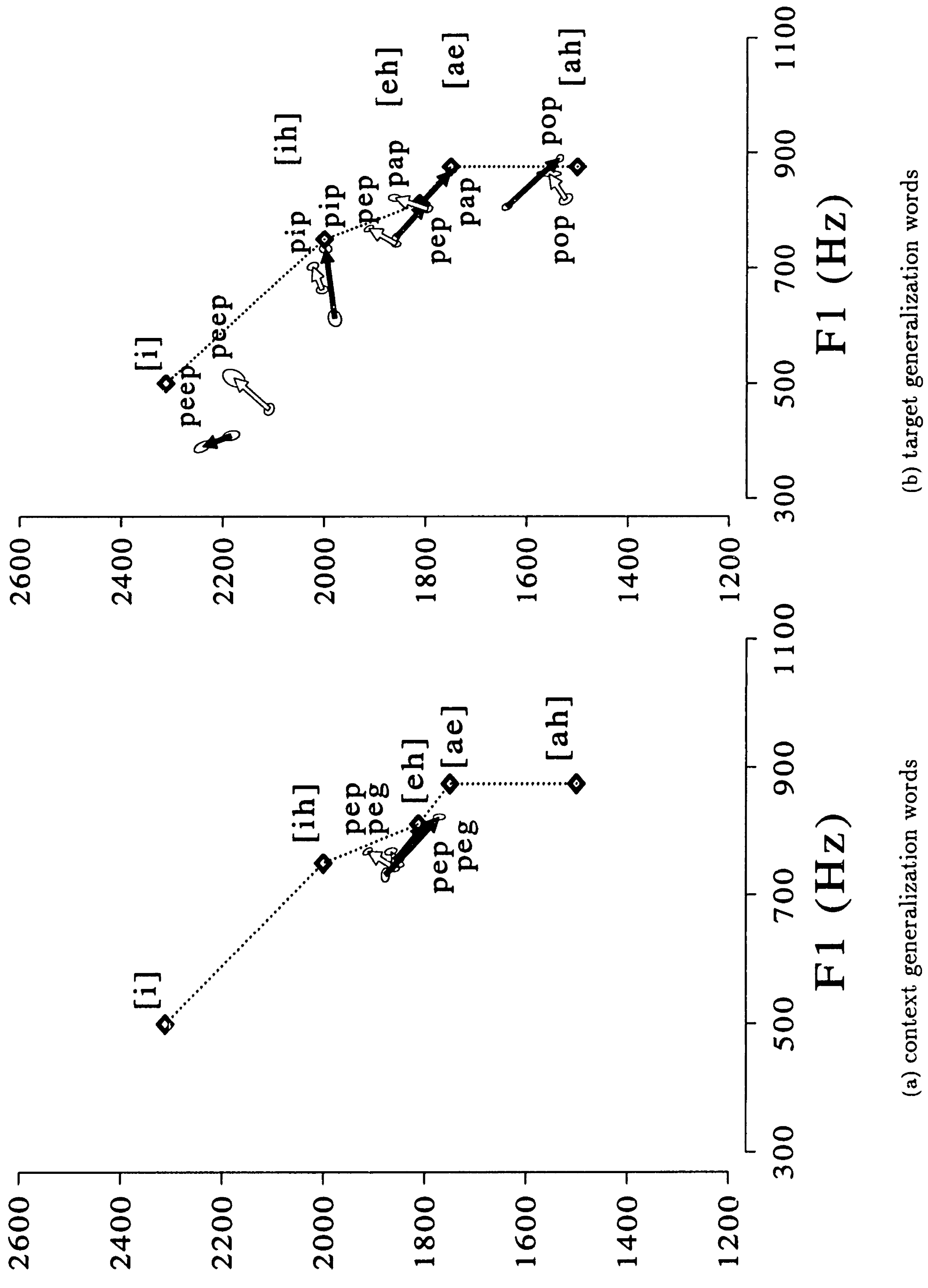


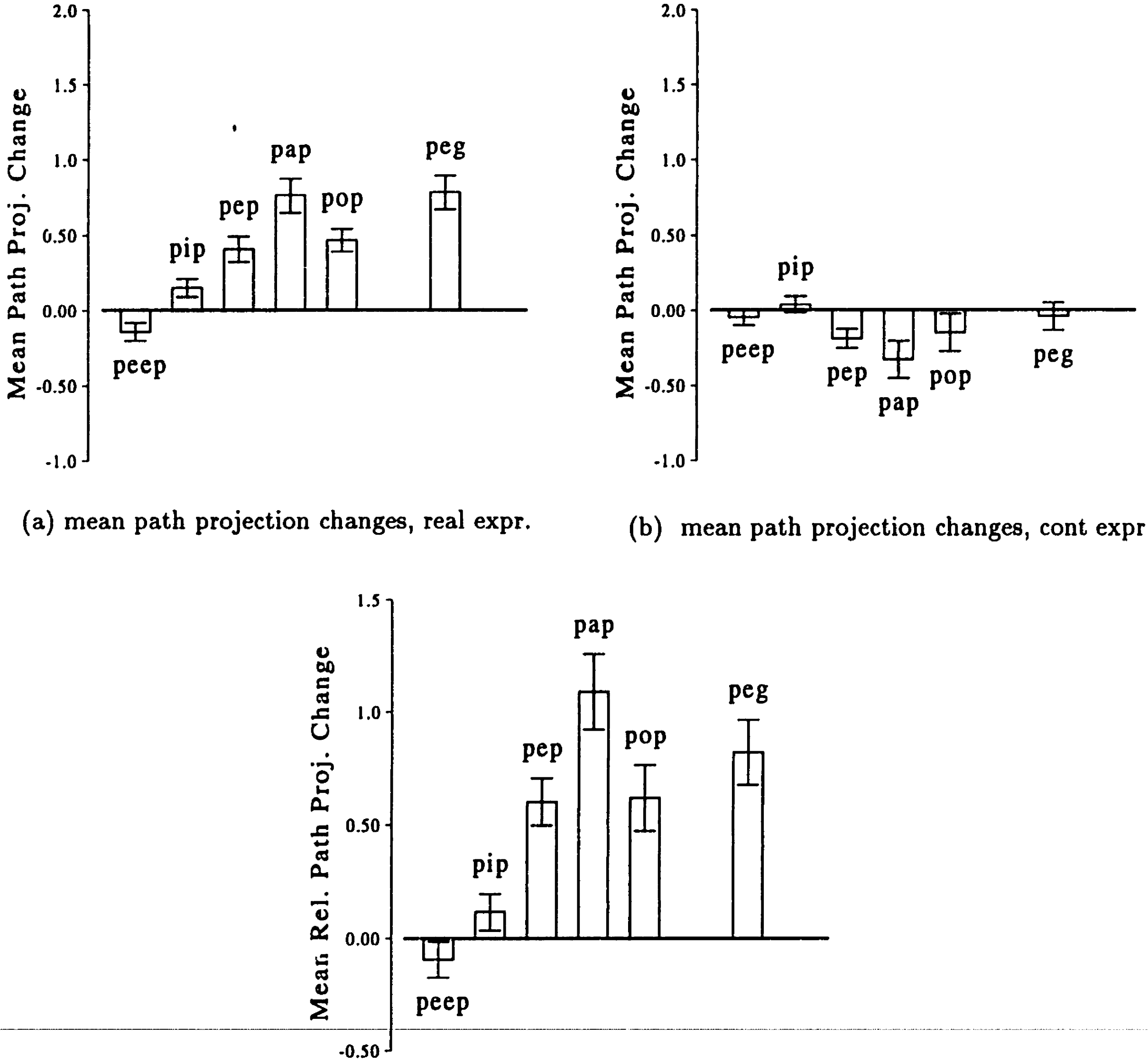
Figure 5-29: Subject SR testing word avgrams.



(a) context generalization words

(b) target generalization words

Figure 5-30: Subject SR testing word vowel plots.



(a) mean path projection changes, real expr.

(b) mean path projection changes, cont expr.

(c) mean relative path projection changes

Figure 5-31: Subject SR testing word path projection changes.

5.4.5 Subject RO

Figures 5-32 through 5-36 show the results for subject RO. For this subject, the real experiment was run on 4/10/96, at 1:06 PM using the -2.0 feedback transform. The control experiment was run 37 days later, on 5/17/96 at 12:58 PM.

Compensation and Adaptation Results Figure 5-32 shows that RO's overall compensation and adaptation results exhibit all the major features seen in CW's results. However, the vowel plots show that, in terms of mean (F1,F2) changes, RO's compensation and adaptation changes were small. The path projection plots show that these small mean (F1,F2) changes were amplified into large path projection changes. It is the proximity of [ε] and [æ] on RS's [i]–[ɑ] path that causes this amplification. Path projection is measured in terms of [i]–[ɑ] path position. Since [i]–[ɑ] path position is measured in terms of normalized inter-vowel units, the small distance between [ε] and [æ] in the (F1,F2) plot is still measured as a complete unit of path position change.[23]

Another interesting feature of RO's overall compensation and adaptation results is the baseline shift in the control experiments. For other subjects, baseline vowel path projections in the control experiment are partially shifted from their values in the real experiment. This shift is always in the direction that the subject compensated in the real experiment. RO's baseline shift in the control experiment is not partial but complete: in both his compensation and adaptation responses, baseline vowel path projections in the control experiment are nearly equal to his test phase vowel path projections in the real experiment.

Note that RO's responses in the control experiment also differ in other ways from his real experiment responses. For both compensation and adaptation responses, his control experiment mean (F1,F2) change vectors are much smaller and in different positions from his real experiment mean (F1,F2) change vectors.

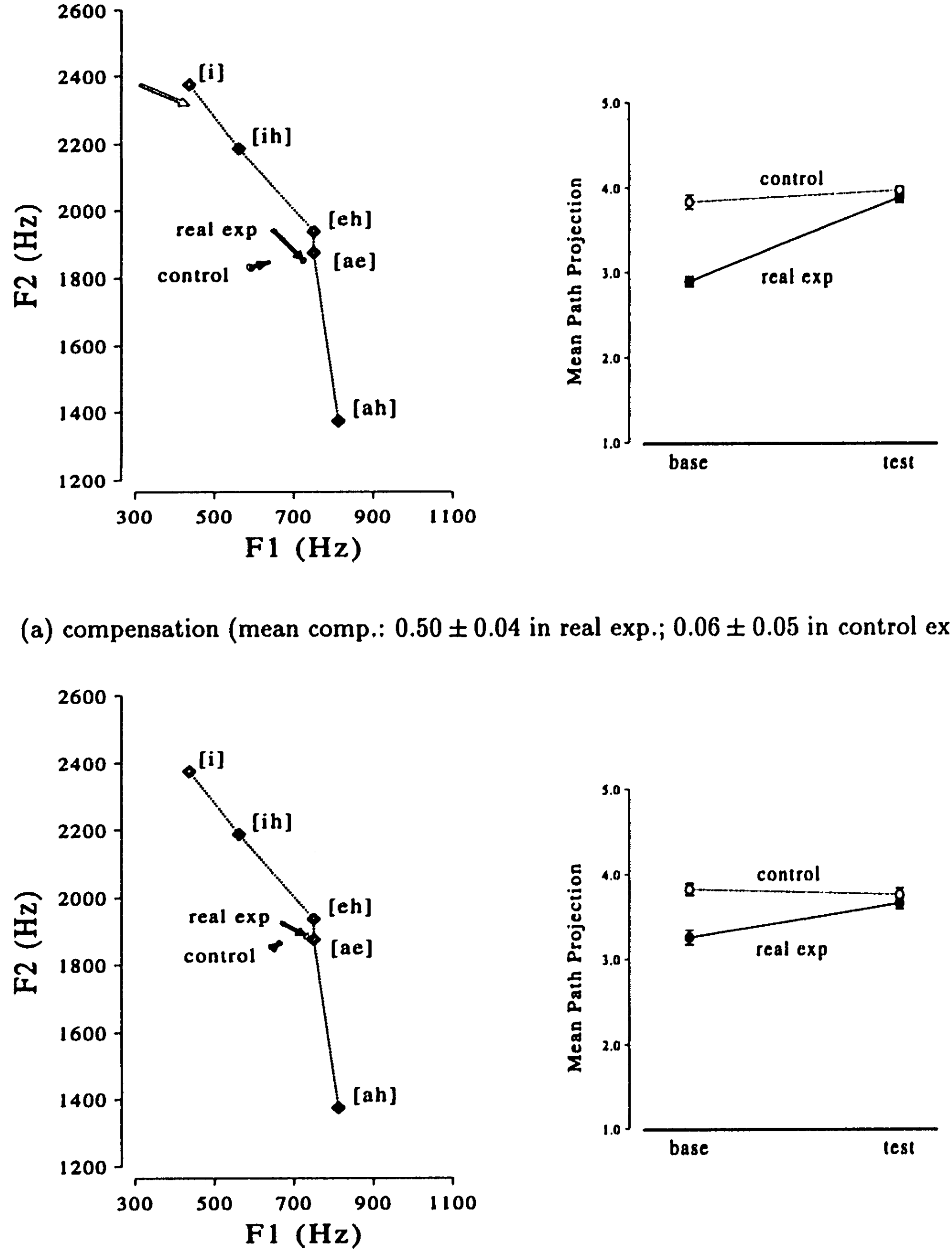
[23] For a complete discussion of path projection and path position, see Section 3.3.1.

Timecourse Results Figure 5-33 shows RO's compensation and adaptation timecourse results. RO's timecourse results are exhibit so much variability that few features can be seen in them. The mean compensation plot shows RO apparently reaches a stable limit to his compensation within the train phase of the experiment. Note, however, that even this stability is not apparent in his mean adaptation plot: his mean adaptation shows an unusual dip in its value (to approximately zero) in the train3 experiment interval.

Generalization Results Figures 5-34 through 5-36 show plots of RO's generalization results.

RO's generalization results exhibit the same types of technical problems seen in CW's results. The avgram of "peep" shows F2 is missing in both the baseline and test phases. Although not easily seen in the avgram, it turns out there are formant estimation problems for "pop" as well. The avgrams of "gep" and "teg" show the usual effects of coarticulation seen in other subjects' results.

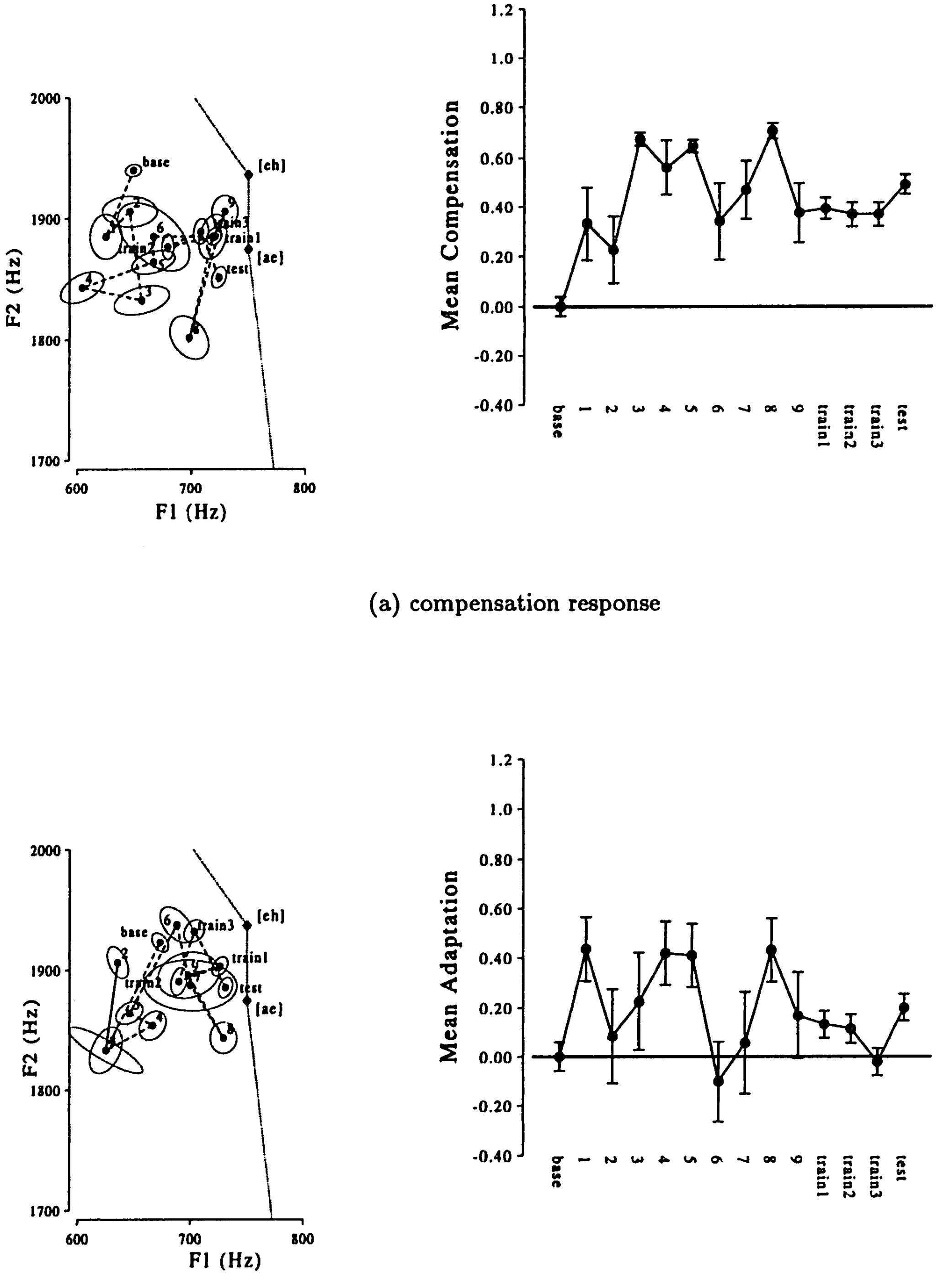
However, RO's generalization results do not exhibit the major generalization features seen in CW's results. In RS's results, none of the testing words (except possibly "pep") exhibit any noticeable production changes.



(a) compensation (mean comp.: 0.50 ± 0.04 in real exp.; 0.06 ± 0.05 in control exp.)

(b) adaptation (mean adapt.: 0.20 ± 0.05 in real exp.; −0.04 ± 0.05 in control exp.)

Figure 5-32: Subject RO overall compensation and adaptation responses.



(a) compensation response

(b) adaptation response

Figure 5-33: Subject RO compensation and adaptation timecourses.

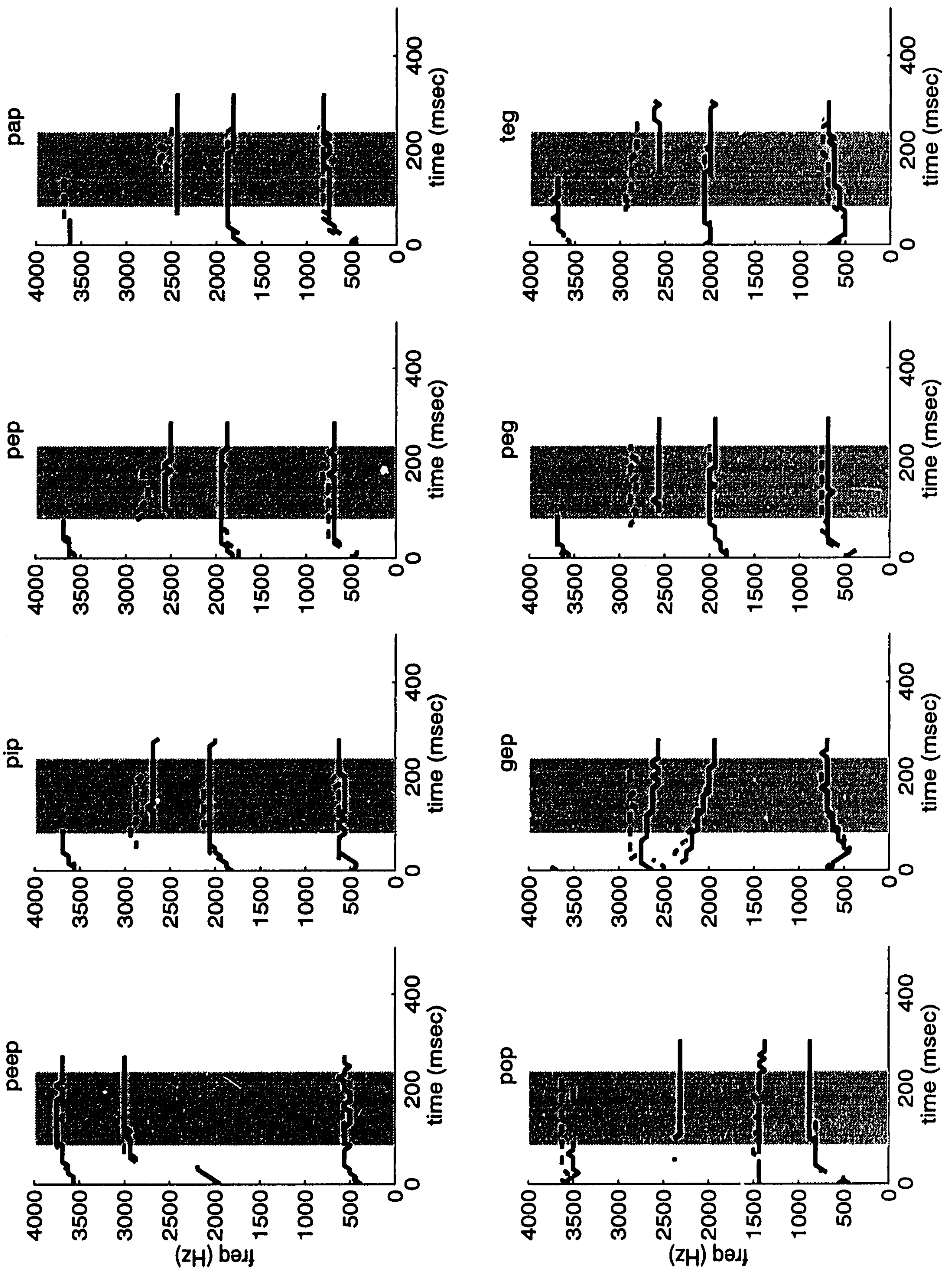
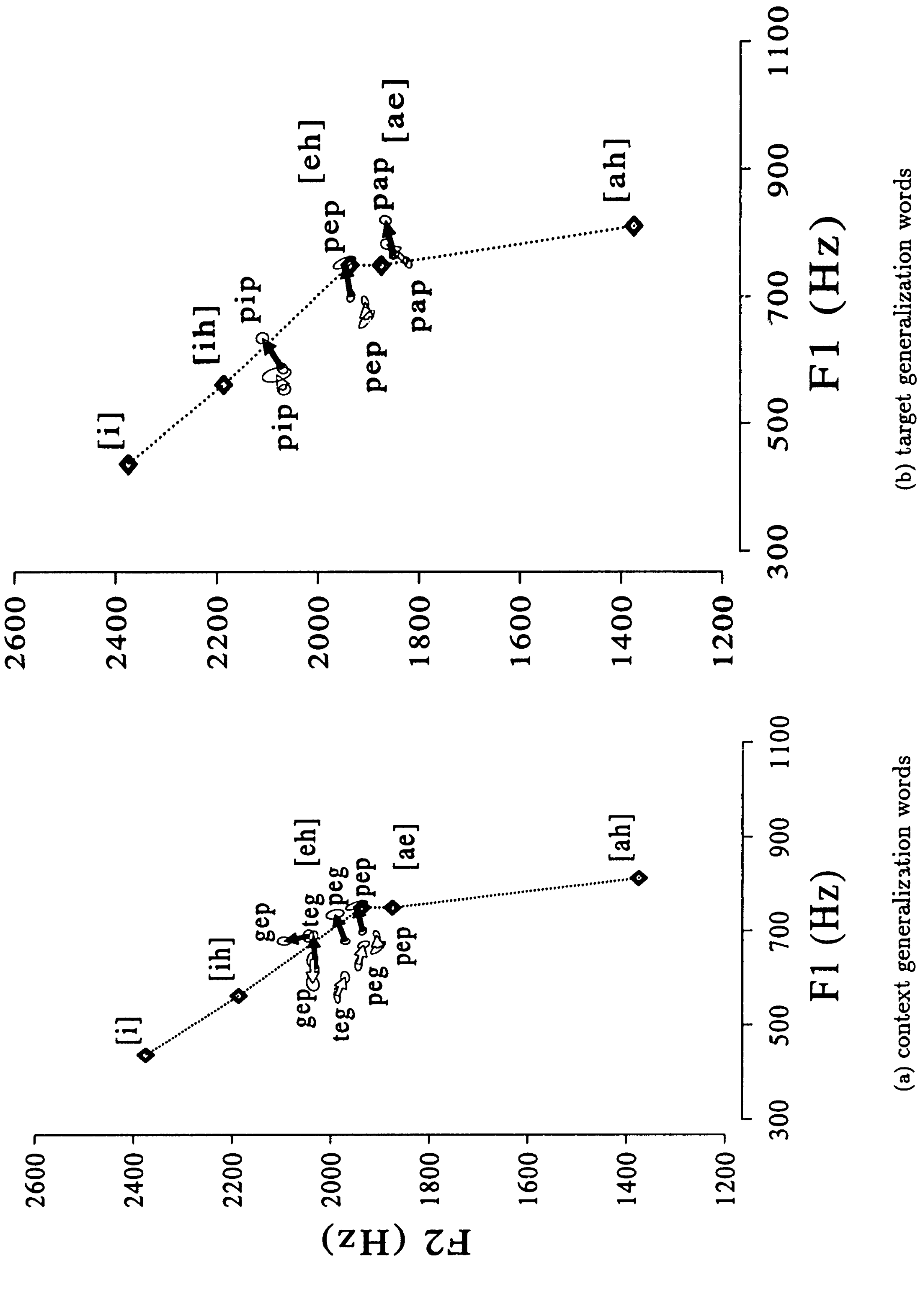


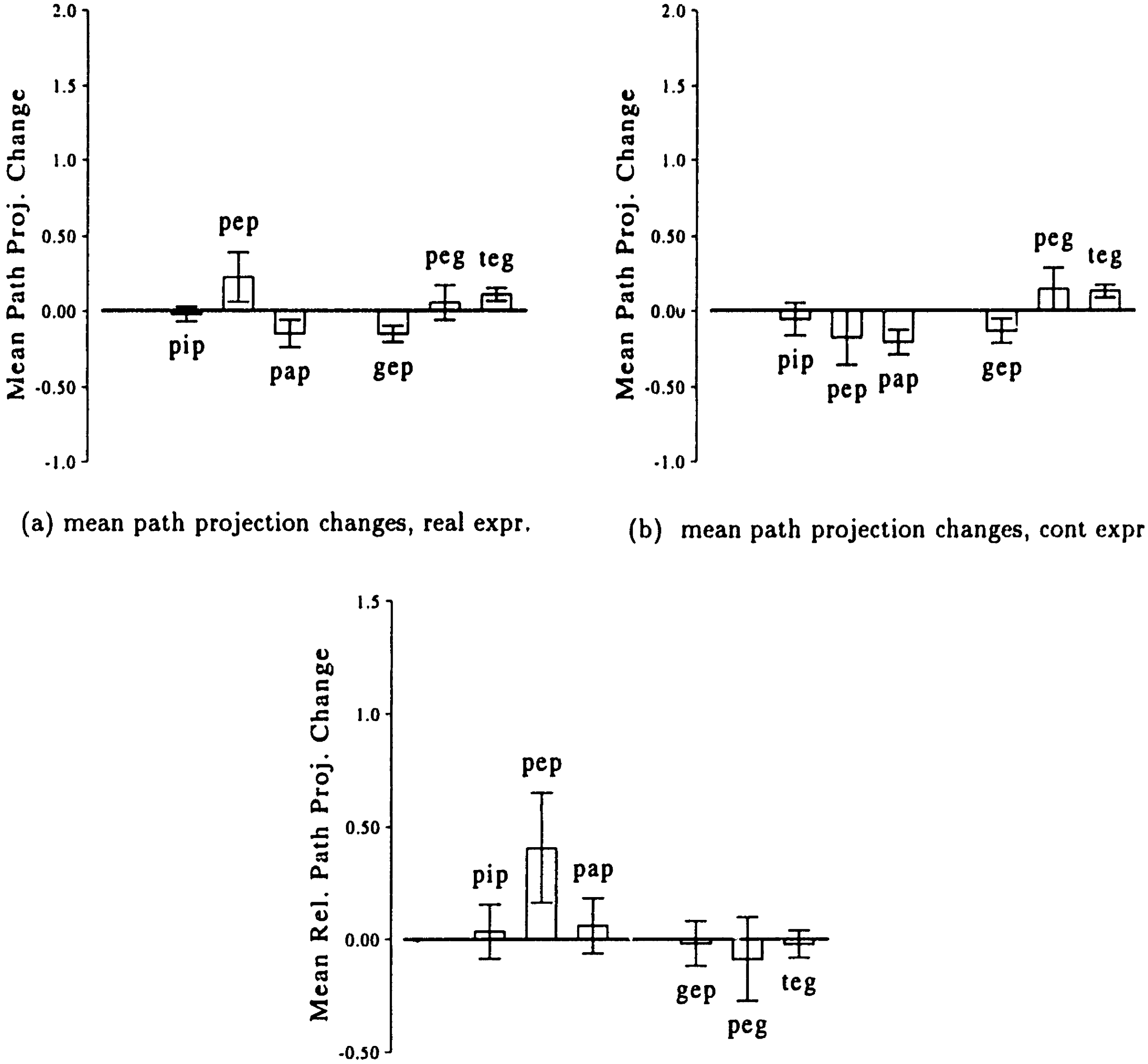
Figure 5-34: Subject RO testing word avgrams.



(a) context generalization words

(b) target generalization words

Figure 5-35: Subject RO testing word vowel plots.



(a) mean path projection changes, real expr.

(b) mean path projection changes, cont expr.

(c) mean relative path projection changes

Figure 5-36: Subject RO testing word path projection changes.

5.4.6 Subject TY

Figures 5-37 through 5-41 show the results for subject TY. For this subject, the real experiment was run on 4/15/96, at 12:43 PM using the +2.0 feedback transform. The control experiment was run 35 days later, on 5/20/96 at 12:55 PM.

Compensation and Adaptation Results Figure 5-37 shows that TY's overall compensation and adaptation results exhibit all the major features seen in CW's results.

It is interesting to compare TY's overall compensation and adaptation responses in the real experiment with RO's responses. Mean (F1,F2) changes for RO's compensation and adaptation responses occur next to a place on his [i]–[ɑ] path where the path vowels ([ε] and [æ]) are very close together. Thus, RO's small mean (F1,F2) changes are amplified into large mean path projection changes. Mean (F1,F2) changes for TY's compensation and adaptation responses occur next to a place on his [i]–[ɑ] path where the path vowels ([ι] and [ε]) are far apart. Thus, TY's mean (F1,F2) changes, which are larger than RO's, result in mean path projection changes that are smaller than RO's.

Timecourse Results Figure 5-38 shows that TY's compensation and adaptation timecourse results exhibit all of the major features seen in CW's results except one: TY's mean adaptation plot shows no evidence of a delay in his adaptation. However, the confidence intervals in the ramp phase of his mean adaptation plot are large enough that it is difficult to make any detailed conclusions about his adaptation timecourse.

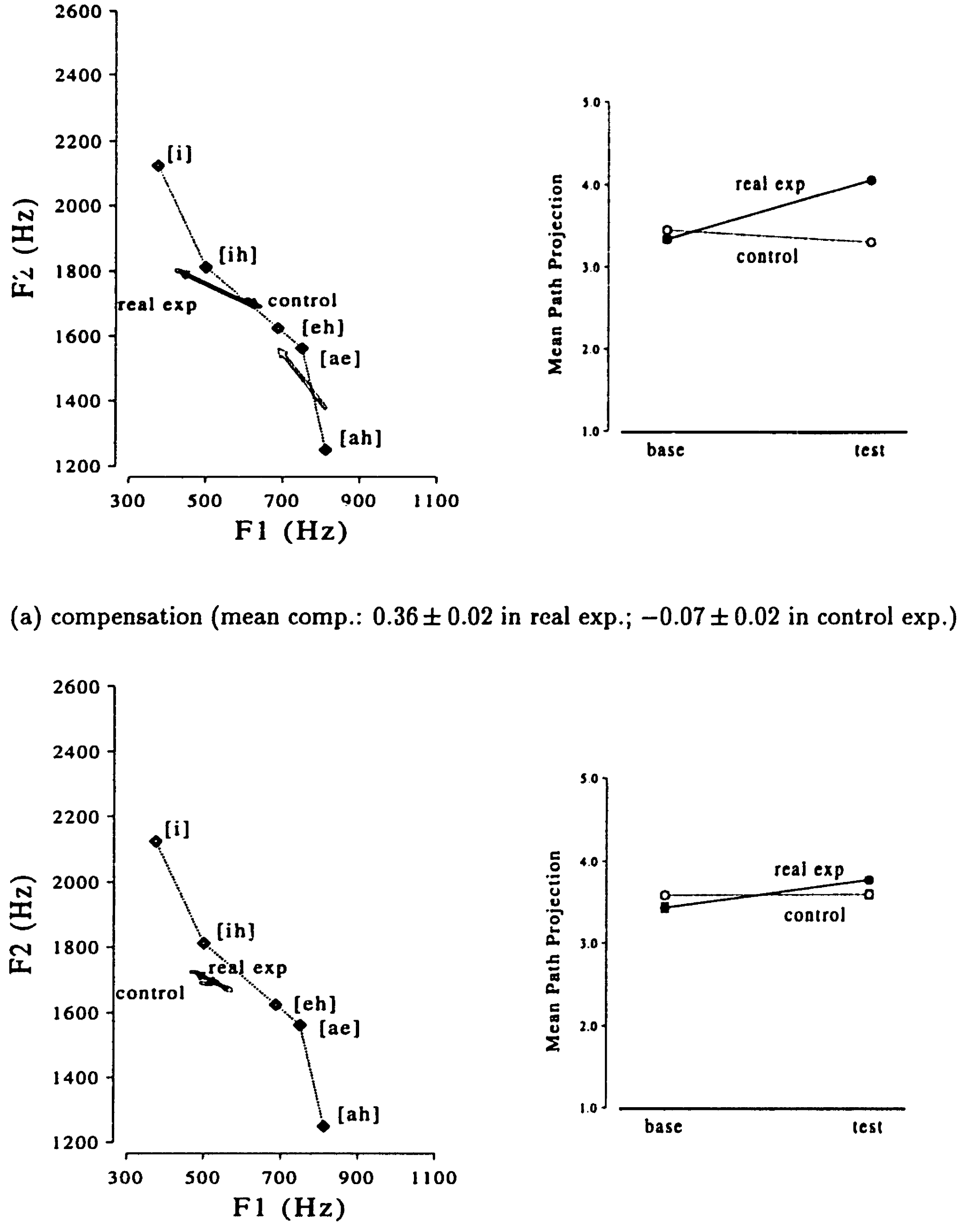
Generalization Results Figures 5-39 through 5-41 show plots of TY's generalization results.

TY's generalization results exhibit the same types of technical problems seen in CW's results. The avgram of "peep" shows F2 is missing in the test phase. The avgram of "pop" shows F2 is missing in the baseline phase. The avgrams of "gep" and "teg" show the usual effects of coarticulation seen in other subjects' results.

Note also that the avgram for "gep" shows an anomalous rise within the vowel analysis interval, followed by a lowering again outside the interval. This may be another case in which F1 is not being stably estimated, as was the case for subject RO's "gep" and "teg" avgrams.

TY's generalization results exhibit all the major generalization features seen in CW's results except one: none of TY's analyzable testing words exhibit zero generalization.

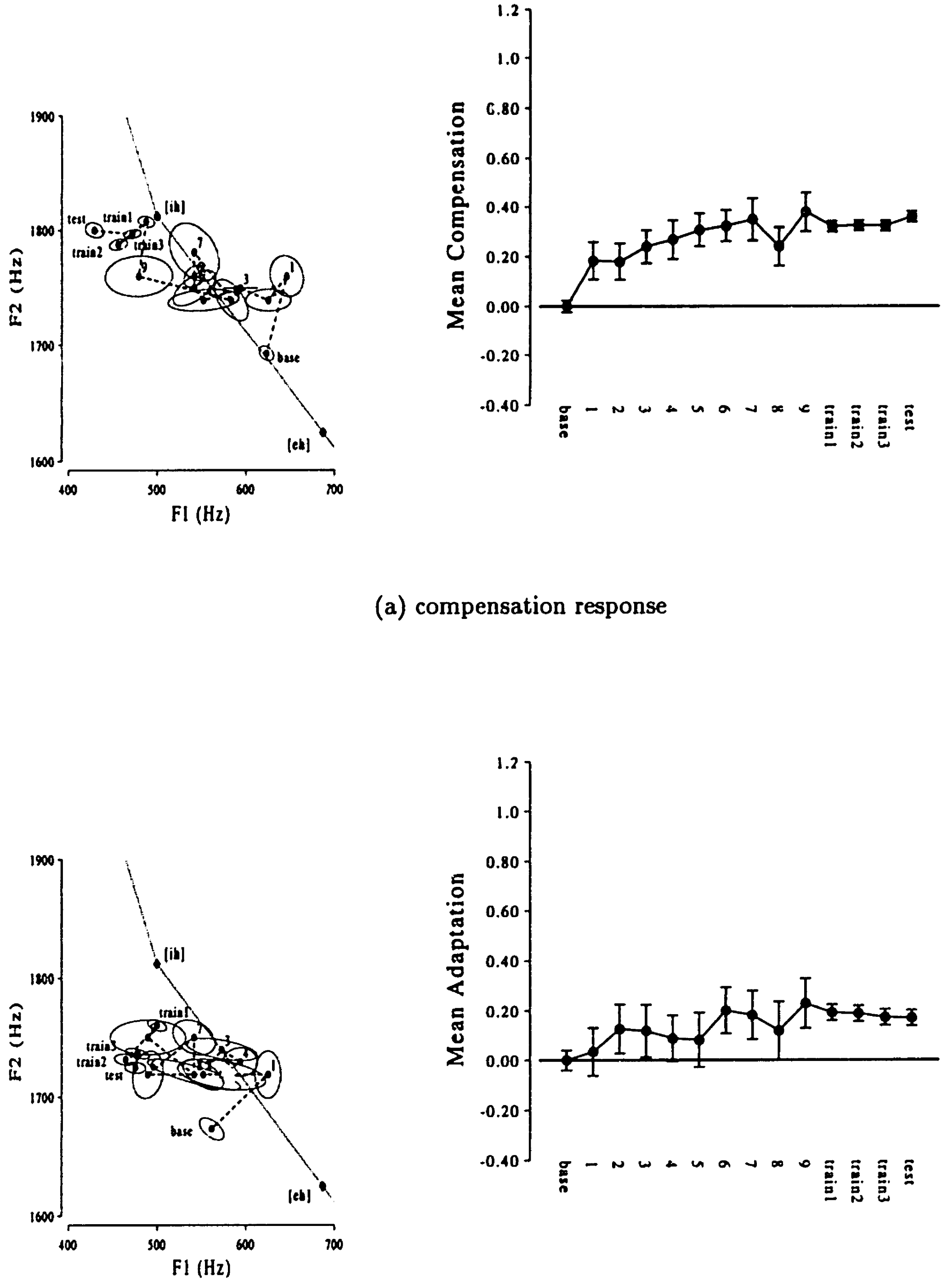
An interesting overall aspect of TY's generalization results (seen best in Figure 5-40) is that his testing word production changes in the real experiment are confined principally to the F1 dimension. This stands in marked contrast with CW's generalization results, where testing word production changes in the real experiment can be seen to be confined principally to the F2 dimension.



(a) compensation (mean comp.: 0.36 ± 0.02 in real exp.; −0.07 ± 0.02 in control exp.)

(b) adaptation (mean adapt.: 0.17 ± 0.03 in real exp.; 0.00 ± 0.03 in control exp.)

Figure 5-37: Subject TY overall compensation and adaptation responses.



(a) compensation response

(b) adaptation response

Figure 5-38: Subject TY compensation and adaptation timecourses.

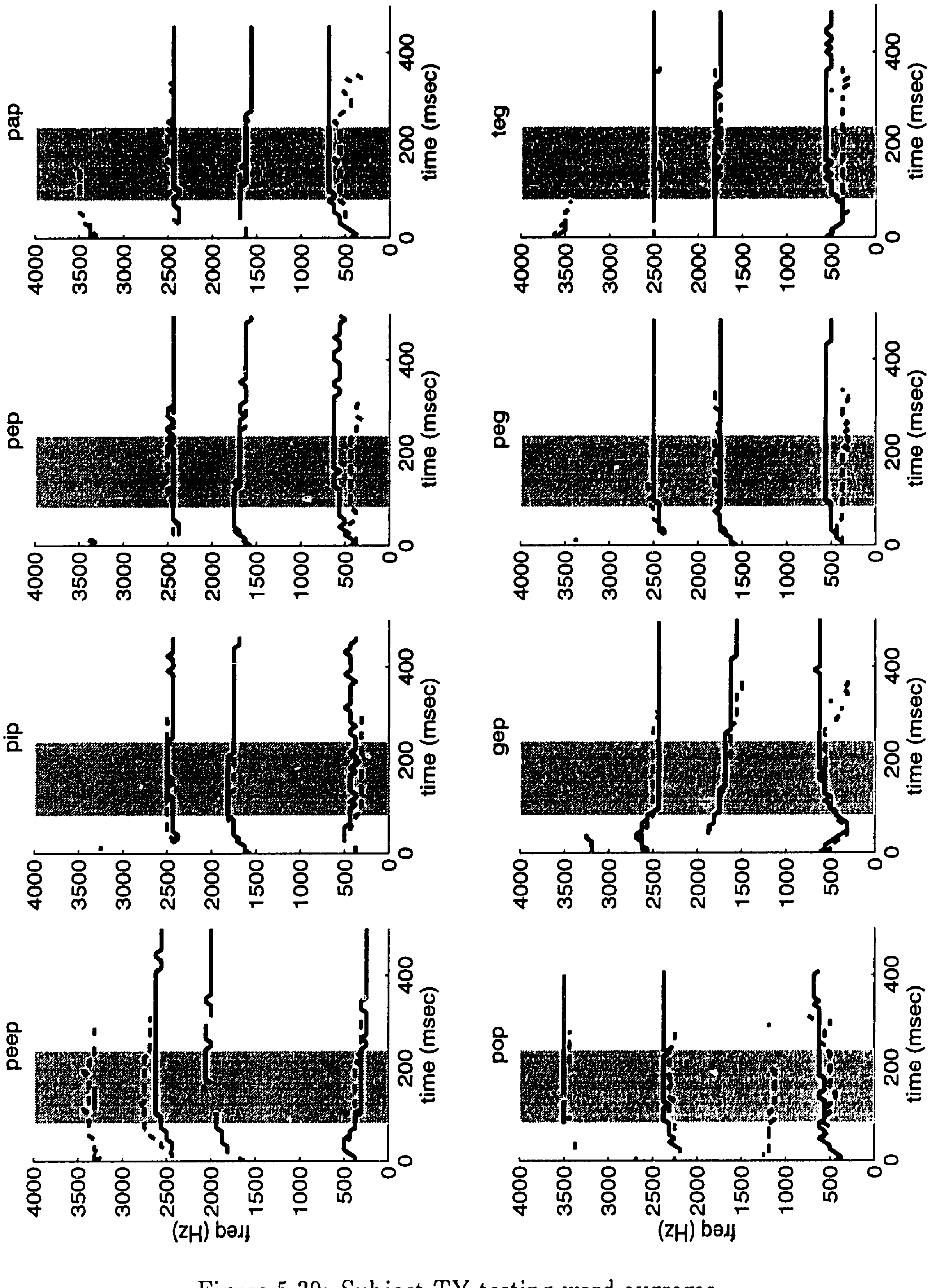
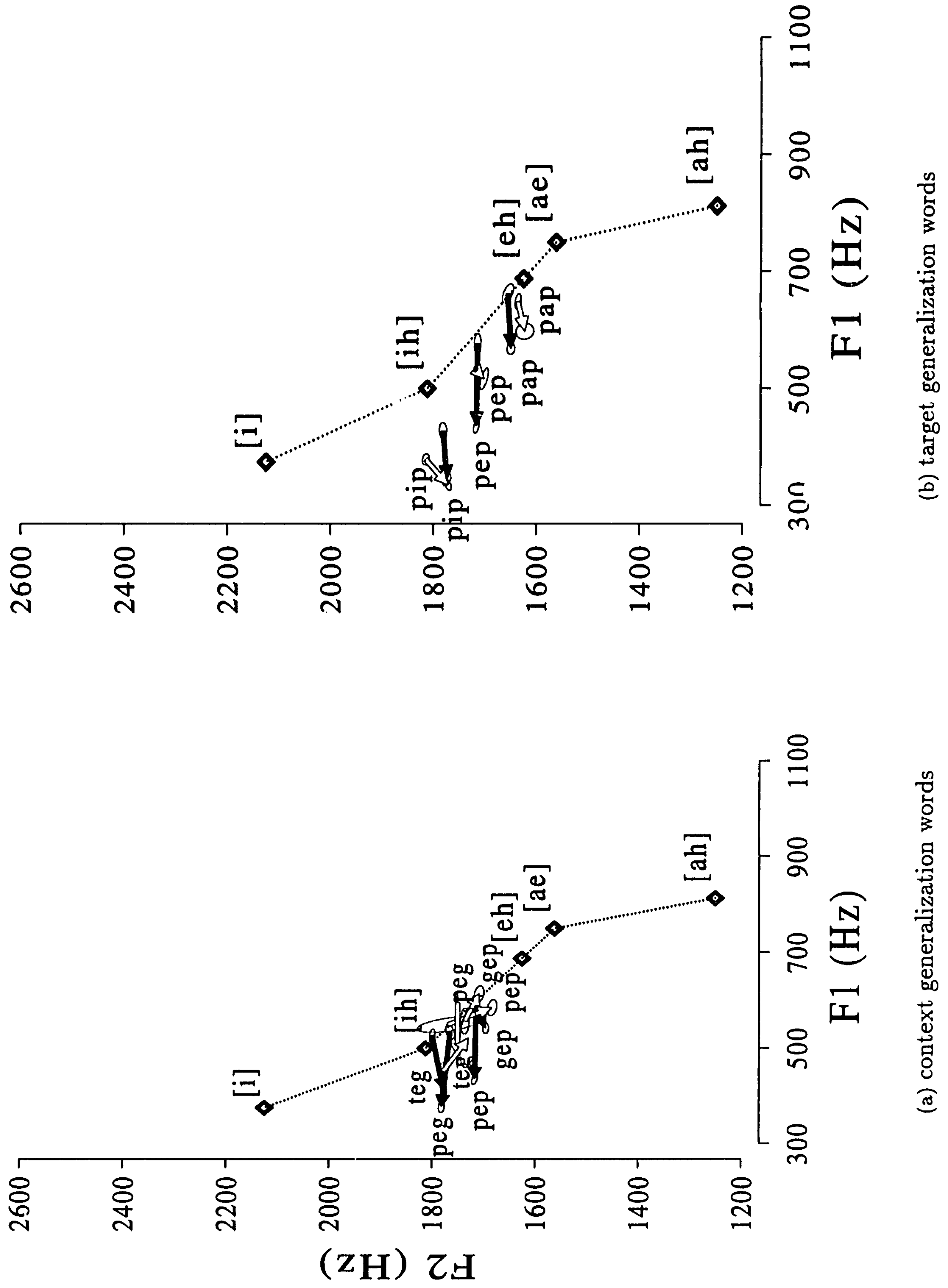


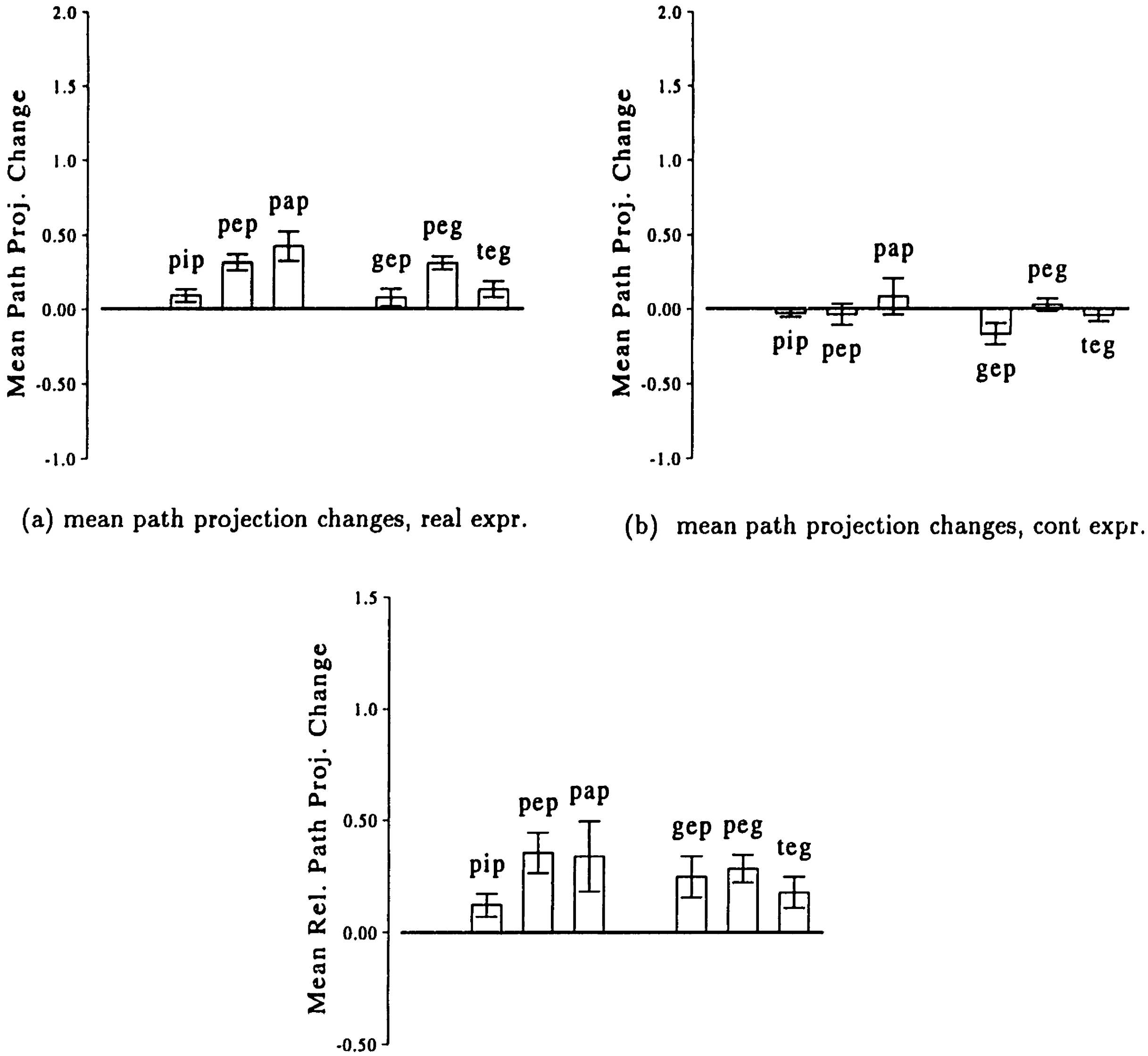
Figure 5-39: Subject TY testing word avgrams.



(a) context generalization words

(b) target generalization words

Figure 5-40: Subject TY testing word vowel plots.



(a) mean path projection changes, real expr.

(b) mean path projection changes, cont expr.

(c) mean relative path projection changes

Figure 5-41: Subject TY testing word path projection changes.

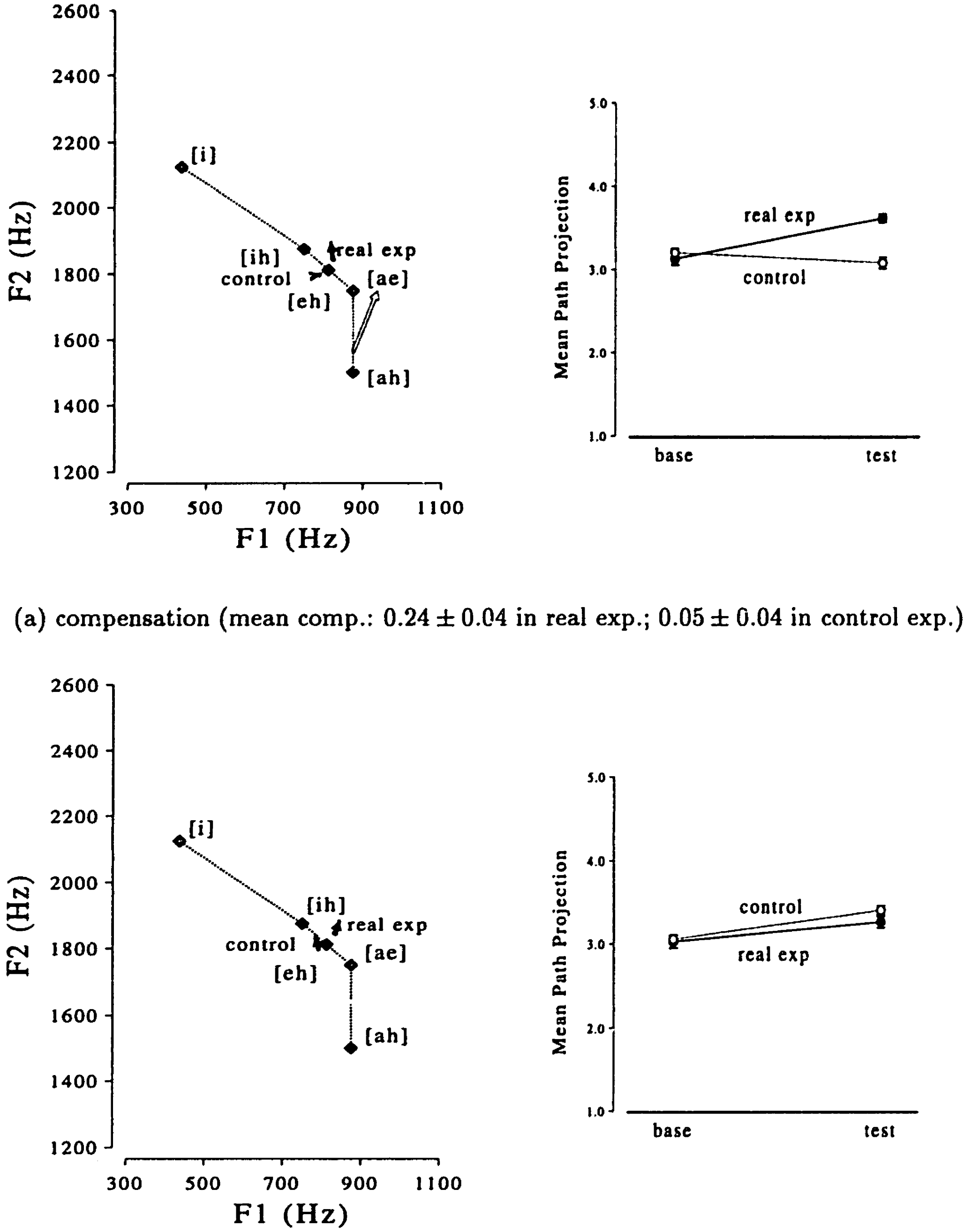
5.4.7 Subject VS

Figures 5-42 through 5-43 show the results for subject VS. For this subject, the real experiment was run on 4/13/96, at 1:16 PM using the +2.0 feedback transform. The control experiment was run 33 days later, on 5/16/96 at 9:33 AM.

Compensation and Adaptation Results Figure 5-42 shows VS's overall compensation and adaptation results. These results show evidence of some compensation but very little evidence of any adaptation.

Because VS did not appear to exhibit any adaptation, no analysis of his generalization results was performed.

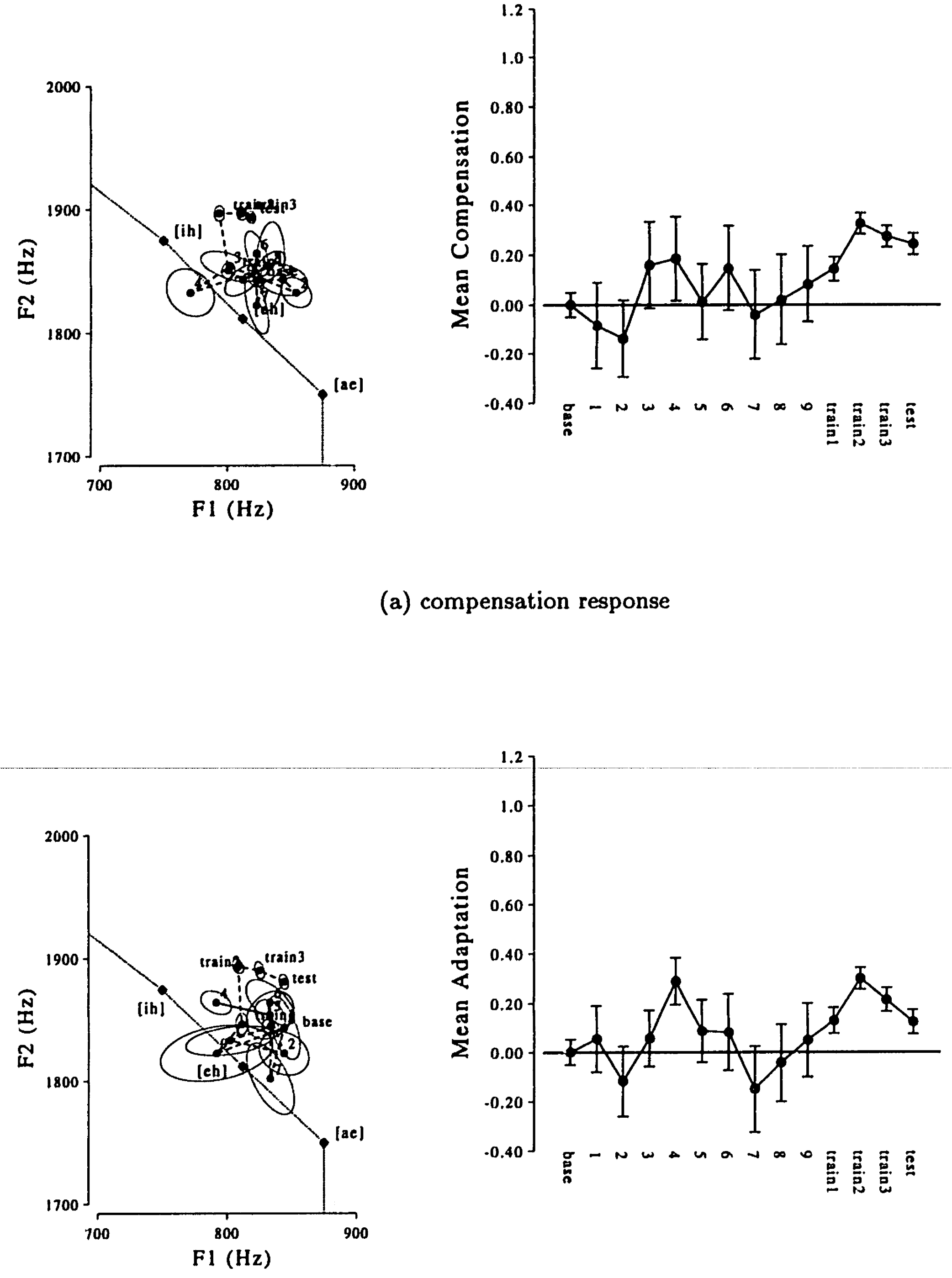
Timecourse Results Figure 5-43 shows VS's compensation and adaptation timecourse results. The mean compensation plot shows some evidence of VS's compensation reaching a stable limit midway through the train phase. The mean adaptation plot shows VS's adaptation reaches a maximum midway trough the train phase but then noticeably decreases. VS's ramp phase timecourse results exhibit such large variation that it is difficult to see any definite features.



(a) compensation (mean comp.: 0.24 ± 0.04 in real exp.; 0.05 ± 0.04 in control exp.)

(b) adaptation (mean adapt.: 0.12 ± 0.05 in real exp.; 0.18 ± 0.04 in control exp.)

Figure 5-42: Subject VS overall compensation and adaptation responses.



(a) compensation response

(b) adaptation response

Figure 5-43: Subject VS compensation and adaptation timecourses.

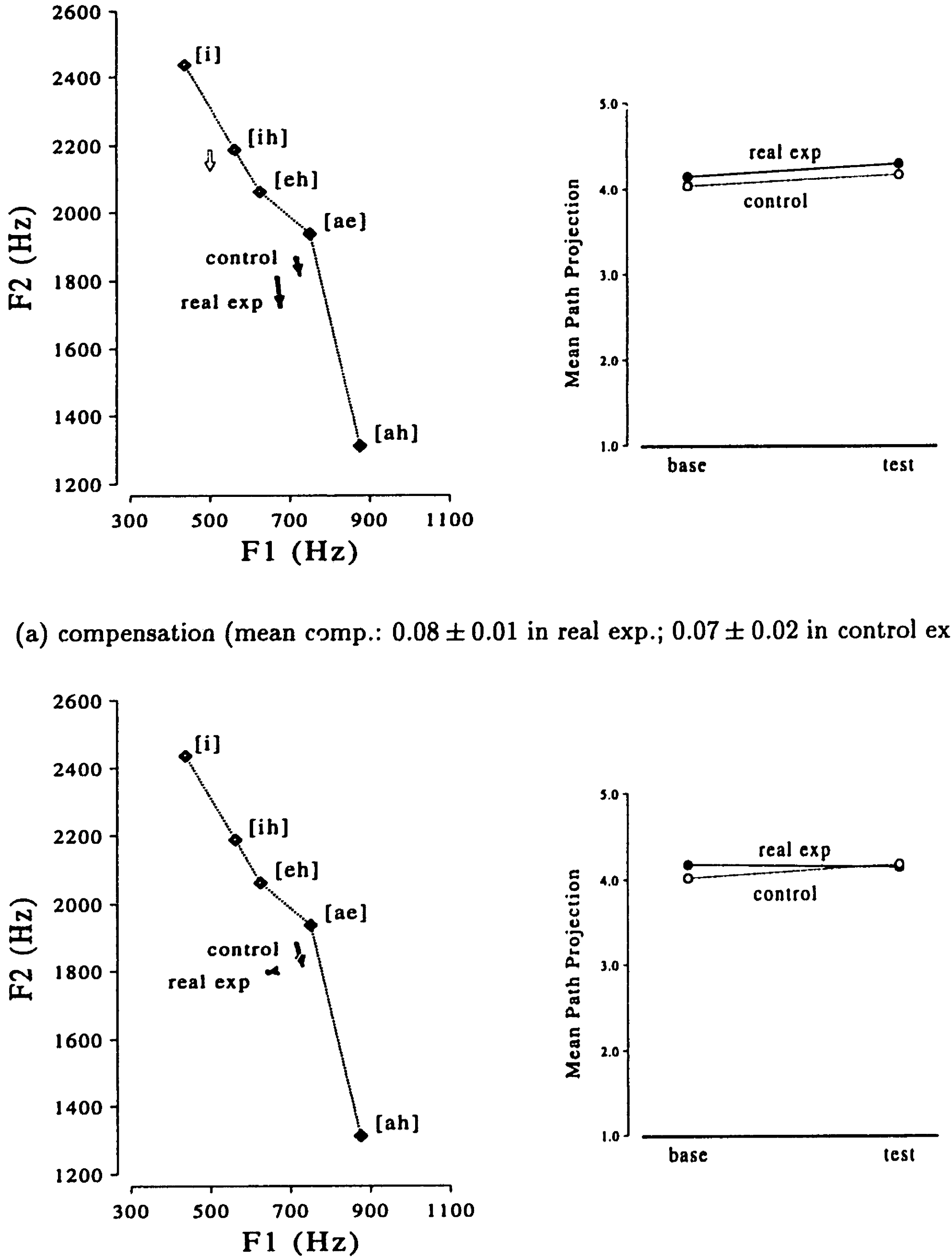
5.4.8 Subject AH

Figures 5-44 through 5-45 show the results for subject AH. For this subject, the real experiment was run on 4/04/96, at 3:36 PM using the -2.0 feedback transform. The control experiment was run 40 days later, on 5/14/96 at 12:59 PM.

Compensation and Adaptation Results Figure 5-44 shows AH's overall compensation and adaptation results. These results show evidence of slight compensation but no evidence of any adaptation.

Because AH did not appear to exhibit any adaptation, no analysis of his generalization results was performed.

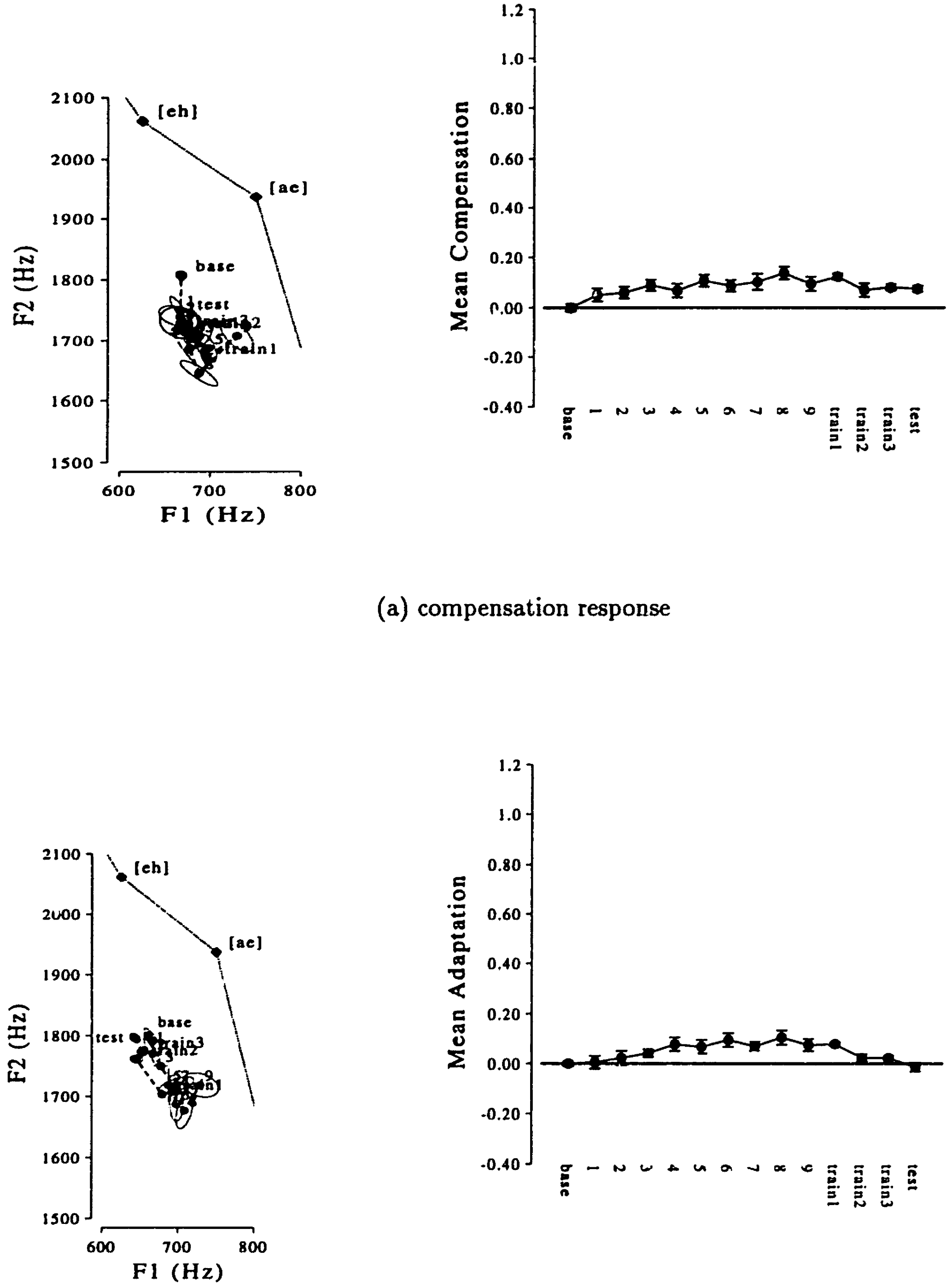
Timecourse Results Figure 5-45 shows AH's compensation and adaptation timecourse results. The mean compensation plot shows some evidence of AH's compensation reaching a stable limit. However, AH's overall production changes are so slight that it is difficult to determine where this stable limit is reached or whether its magnitude is significant. The mean adaptation plot shows an interesting rise and fall in mean compensation over the course of the experiment. Again, however, AH's overall production changes are so slight that it is difficult to assess whether his changes mean adaptation are significant.



(a) compensation (mean comp.: 0.08 ± 0.01 in real exp.; 0.07 ± 0.02 in control exp.)

(b) adaptation (mean adapt.: −0.01 ± 0.02 in real exp.; 0.08 ± 0.02 in control exp.)

Figure 5-44: Subject AH overall compensation and adaptation responses.



(a) compensation response

(b) adaptation response

Figure 5-45: Subject AH compensation and adaptation timecourses.

Chapter 6

Discussion

The primary objective of this thesis was to examine the hypothesis that vowel production, like reaching, exhibits sensorimotor adaptation (SA). A second objective was to exhibit the potential of speech SA for examining the phonetic structure of speech production.

6.1 Existence of Speech SA

The hypothesis that vowel production exhibits SA can be factored into two subhypotheses:

1. Speakers will adjust vowel productions to compensate for perceived alterations of their auditory feedback.

2. This compensation will be sufficiently permanent to be partly retained in the absence of auditory feedback. Such retained compensation is called adaptation.

Both Study 1 and Study 2 tested this hypothesis. Both did so by altering formants in subjects' auditory feedback and analyzing how this affected their productions of whispered [ε]. Both studies found that exposure to this altered feedback caused subjects' productions to exhibit the following characteristics:

1. *Compensation:* Their production of [ε] was changed in ways that compensated for the effects of the feedback alteration. The amount of compensation varied

widely across subjects but overall was highly significant.

2. *Adaptation:* The subjects retained much of this compensation when producing [ε] while prevented from hearing it. This retained compensation, or adaptation, was also significant across subjects.

Study 2 included two additional aspects in its design to provide confirmation and elaboration of these results. First, Study 2 included a comparison of subject's responses in the real experiment with those in a control version (no feedback alteration). Compensation and adaptation were found to be significantly greater in the real experiment.

Second, in Study 2 the feedback transformation was introduced gradually to minimize subjects' awareness of it. This awareness was assessed by post-experiment interviews. In these interviews, no subject reported noticing the alteration of his feedback.

These results provide strong support for the hypothesized existence of speech SA and reveal several important characteristics of it. These characteristics concern subjects' perception of the feedback alteration, and their compensation and adaptation responses.

6.1.1 Perception of the Altered Feedback

In Study 2, no subject reported noticing feedback alterations, yet many subjects showed considerable compensation. This suggests that compensation does not require conscious perception of unusual or unexpected feedback.

On the other hand, some subjects showed very little compensation. For these subjects, an explanation is needed for why they did not compensate more and why they didn't report noticing the altered feedback. In Section 5.3.1, several possible explanations were discussed.

Possible methodological problems were considered. Problems with generating the altered feedback are possible but unlikely since subject pretesting screened out subjects whose formants transformed poorly. Also, post-experiment interviews may have unreliably gauged subjects' awareness of the altered feedback. This too was rated as unlikely: subjects' exhibited noticeable curiosity about the experiment's purpose and thus seemed likely to remember any of its unusual aspects.

Other explanations concerned speech perception. The first is that the poor compensators could be insensitive to the whispered vowel sound differences created by the altered feedback. Given that these differences amount to complete changes of vowel phonetic identity, this explanation also seems unlikely.

A more likely explanation for the poor compensators is that the altered feedback induced adaptation of their speech perception, not their speech production. That speech perception can adapt has already been shown in other types of experiments: subjects' voiced/voiceless feature perception will shift after repeatedly hearing only stop consonants with one of these features [Cooper, 1979].

This explanation can account for the range of compensations observed across subjects. Each subject could have a different capacity to adapt his perception of the altered feedback. This perceptual adaptation reduces his perception of the true amount of feedback alteration. He then produces compensations only for the perceived amount of feedback alteration.

By this account, subjects who produced large compensations have small capacities to adapt perception, while subjects who produced small compensations have large capacities to adapt perception.

6.1.2 Compensation

For all subjects, compensation was greater than (or, occasionally, equal to) adaptation. One possible explanation for this difference is that the presence of masking noise somehow causes subjects to whisper differently. This account, however, does not specify why compensation would be greater than adaptation.

Another possible explanation is that some portion of each subject's compensation was accomplished by some temporary correction mechanism, active only in the presence of the altered feedback. This explanation suggests that vowel production may be partly under auditory feedback control.

Control of speech production based on auditory feedback was first proposed by Grant Fairbanks in 1954 [Fairbanks, 1954]. However, since then, several arguments have been made that auditory feedback plays no such direct role in the control of speech.

The first argument is based on minimality: it isn't necessary to suppose auditory feedback control, since speech is producible without auditory feedback. Speakers deafened in adult life retain intelligible speech [Cowie and Douglas-Cowie, 1983, Lane and Webster, 1991]. Many other experiments (including those of this thesis) have shown that speech remains intelligible even when hearing is blocked by masking noise [Lombard, 1911, Lane and Tranel, 1971].

However, this argument does not rule out the possibility that, when available, auditory feedback control is used in speech production. In fact, evidence for auditory feedback control has been found in other aspects of speech. Kawahara and others have found evidence of fast pitch corrections in response to sudden perturbations of pitch feedback [Kawahara, 1993]. These experiments found a compensating response within 100-200ms of the onset of perturbation.

The other major argument against auditory feedback control is that it is too slow: the neural delays in processing auditory feedback probably make it unusable for the control of fast speech movements [Perkell, 1996]. But not all speech tasks require fast movements. Maintaining a pitch frequency or a steady-state vowel are examples of speech tasks not requiring fast movements. For these tasks, the speech production system may take advantage of the feasibility of using control based on auditory feedback, when it's available.

6.1.3 Adaptation

The study results concerning adaptation show that significant production changes of some permanence can be induced by relatively brief (e.g. 1 hour) exposure to altered feedback. These results are surprising, given the stability of speech control in the absence of feedback. Such stability can be seen in speakers deafened in adult life: their speech remains intelligible for years after deafness [Cowie and Douglas-Cowie, 1983, Lane and Webster, 1991]. Speech thus appears to be both stable in the absence of feedback and yet easily affected by altered feedback.

One explanation of these observations is the speech-equivalent of the "reafference hypothesis" proposed by Held to explain reaching SA [Hein and Held, 1962]. This explanation assumes a speaker retains an *expected outcome* of any speech motor commands. Any sensory reports of the actual outcome (e.g., auditory or proprioceptive) are compared with the expected outcome. A mismatch drives the speaker to make some corrective response that minimizes the mismatch.

In this account, if a sensory report is not available (as when the subject's hearing was blocked by noise), no comparison is made, no mismatch is generated, and the speaker is not driven to make a corrective response. If a sensory report is available and feedback is unaltered, the outcome reported matches expectations, and there is no mismatch to correct. However, when feedback is altered, the sensed outcome no longer matches expectations. Only in this case is the speaker driven to correct the mismatch.

As discussed above, there are several possible ways the speaker could correct the mismatch. Perceptual adaptation could occur, causing the sensed outcome better match the expected outcome. Auditory feedback control mechanisms could also cause some temporary production correction. Finally, the speaker could respond by making long-term adjustments to his speech control.

Recently, the reafference hypothesis has been incorporated into computational models of motor learning [Jordan and Rumelhart, 1992]. In these models, minimizing the mismatch between actual and expected outcomes is a basic process in learning control of directed movements. These models would predict that speech SA is simply a more limited and controlled version of the initial process of learning speech motor control.

6.1.4 Summary

In sum, the characteristics of speech SA observed in studies 1 and 2 are explained by the following theory:

1. Perception of altered speech feedback is partially offset by perceptual adaptation. The capacity to adapt perception is limited and speaker-specific.
2. Perceived feedback alterations are compensated for.
3. Compensation is preferentially achieved by a temporary correction mechanism (active only while exposed to the altered feedback). The capacity of the temporary correction mechanism to compensate is limited and speaker-specific.
4. When required compensation exceeds the capacity of the temporary correction mechanism, long-term speech control adjustments must be made – i.e., adaptation occurs.

This theory is preliminary, and further investigations will be needed to confirm its assertions.

6.2 Using Speech SA to Address Phonetic Structure Questions

Chapter 1 suggested that a principal value of speech SA is in its potential for examining questions concerning phonetic structure in speech production. To illustrate this, a hypothetical experiment was described showing how speech SA could be used to determine (1) if words specify their productions via shared intermediate production units, and (2) whether these production units have independent representations.

Study 2 implemented this experiment for the vowel [ɛ]. It looked at how adaptation of [ɛ] in a bilabial CVC word context affected [ɛ]'s production in other CVC words. It also looked at how other vowels' productions were affected. The results showed that adaptation of [ɛ] in the bilabial context caused:

1. Similar production changes in [ɛ] in other word contexts. This is called *context*

generalization.

2. Similar production changes in other vowels. This is called *target generalization.*

6.2.1 Context Generalization

The context generalization results showed that [ε] adapted in one word context caused [ε] to be produced in this adapted manner in other word contexts. These words all appeared to access the same adapted production of [ε]. This result rules out the possibility that these words each have independent production mechanisms. Instead, it appears that production of a word involves specifying representations of intermediate production units that are shared by many words. This is consistent with most phonetic theories, which suppose word productions to be represented as sequences of elemental production units like phonemes or syllables [Halle, 1990, Levelt, 1989, Meyer, 1991].

The above result suggests that speech SA experiments could be designed to investigate a number of issues related to word production. Several of these issues are discussed below.

6.2.1.1 Word Frequency Differences

It is possible that high-frequency words are not produced in the same way as low-frequency words and non-words. Because there are arbitrarily many low-frequency words and non-words, it seems unlikely that their productions are not constructed from smaller shared production units. However it does seem plausible that frequently-used words might start to function as production units themselves. If this were the case, adapting a vowel's production in a high-frequency word might not alter its production in other words.

6.2.1.2 Syllables Versus Phonemes

Study 2 did not have sufficient power to determine if generalization was affected by syllable context. This is an important issue because the syllable has been suggested as a possible intermediate production unit [Levelt, 1989]. If this were the case, context generalization might be restricted to words sharing the same syllable.

6.2.2 Target Generalization

The target generalization results showed that [ε] adapted in one word context caused similar production changes in other vowels. These results demonstrate that vowel production representations are not independent. This suggests that vowel representations may share common features.

The results suggest ways in which future speech SA experiments could be used to examine the representations of speech production units. Consider vowels as an example. More detailed investigations of target generalization could be conducted to reveal its pattern across many vowels. Different vowel representations could then be examined by evaluating how easily they explain the target generalization pattern. For example, the pattern might be easily explained by assuming adjustment of a single articulatory parameter (e.g., tongue height): this would be evidence for articulatory vowel representations. On the other hand, the pattern might be more easily explained as a function of distance in formant space (e.g., nearby vowels show similar production changes): this would be evidence for acoustic vowel representations.

6.3 Conclusions

To summarize, the principal thesis results are the following:

1. Experience with auditory feedback induces long-term adaptation of parameters controlling vowel production.

2. Adapting production of a vowel in one word context affects production of this vowel in different word contexts and the production of other vowels.

Result 1 shows that speech, like reaching, exhibits sensorimotor adaptation (SA). Result 2 shows that words share common intermediate production representations, and that vowel representations are not independent. These results show that speech

SA exists and provides a new tool for examining fundamental questions concerning phonetic representations in speech production.

Appendix A

Formant Estimation, Analysis, and Subject Pretesting

Accurate formant estimation was central to all aspects of the studies discussed in this thesis. During an experiment, the feedback transformation was based on formant estimates produced by the digital signal processor (DSP). Post-experiment, the results analyses assessed compensation by examining changes in mean formant values of the recorded formant data.

In the first section of this appendix, the key procedures relating to formant estimation are described. First, the formant estimation method used by the DSP in transforming a subject's whispered speech is considered. As discussed in chapter 3, the DSP estimated formants from the magnitude spectrum of a frame of input speech data. These formant estimates were then used in the transformation and resynthesis of feedback for a subject. They were also the data recorded for that input speech frame.

In the second section, the method of deriving mean formant values from this recorded formant data is discussed.

Finally, in the third section, the procedures used in pretesting subjects are described. Pretesting of subjects was needed for a number of reasons, but primarily because the formant estimation and transformation methods required parameters which were measured in these pretest procedures.

A.1 Formant Estimation

As described in Section 3.2, the feedback transformation process consisted of:

1. Acquiring a frame of whispered speech from the subject.

2. Producing a magnitude spectrum of it

3. Estimating formants from the spectrum

4. Altering the formants

5. Synthesizing from the altered formants a frame of speech fed back to the subject.

All steps of this process are considered in more detail in appendix B. Here, we focus on the method of estimating formants (step 3), which was complicated by certain spectral characteristics of whispered speech. We begin by defining what we mean by the term "formant".

A.1.1 What are Formants?

Specifying all the spectral features that define formants is the subject of continued research that is beyond the scope of this thesis. Here, we give a functional definition of formants that served the purposes of the experiments.

If spectral analysis is done on some time interval of voiced speech, the envelope of the resulting magnitude spectrum will exhibit peaks. These peaks are usually labeled in order of increasing frequency as F1, F2,...FN. They have the following properties:

1. From a reasonable pitch function and F1, F2, F3, and F4, it is possible to synthesize speech that is perceptually similar to the original speech from which F1, F2, F2, and F4 were derived [Klatt, 1980, O'Shaughnessy, 1987].

2. If the speech interval is from a sustained vowel sound, then the frequencies of F1 and F2 determine the vowel's identity. In particular, as the vowel is changed from [i] to [ɩ] tc [ɛ] to [æ] to [ɑ], F1 increases in frequency and F2 decreases [Peterson and Barney, 1952].

The feedback transformation process depended on extracting from the whispered speech spectrum quantities which had these same properties:

- Property 1 was needed for synthesizing speech feedback that the subject considered an adequate substitute for his real feedback.

- Property 2 was needed because the feedback alteration was based on the premise that shifting the frequencies of F1 and F2 altered perceived vowel identity. In particular, it was critical that changing vowel identity between [i], [ι], [ɛ], [æ], and [ɑ]could be achieved by shifting the frequencies of F1 and F2.

Quantities which have these two properties are generally referred to as formants. For the purposes of the experiments, these two properties will define what we consider to be formants.

Unlike voiced speech, however, the spectral envelope peaks of whispered speech sounds did not always have these properties, and thus could not be used directly as formant estimates.

A.1.2 Formant Estimation Problems

It was found that the spectral peaks of whispered speech had property 1: if peak frequencies were left unaltered, whispered speech synthesized from these peaks was perceptually similar to the input whispered speech.

However, these peaks did not always have property 2. In particular, the two lowest frequency spectral peaks did not always distinguish the vowels [i], [ι], [ɛ], [æ], and [ɑ]. In normal voiced speech, this progression of vowels causes the lowest peak (F1) to increase in frequency and the next highest peak (F2) to decrease in frequency.

In many subjects' whispered speech, this same vowel progression instead exhibited the following peak pattern (peaks considered in order of increasing frequency):

- The first peak remained fixed in frequency but decreased in amplitude.

- The second peak varied little in frequency and increased in amplitude.

- The third peak decreased in frequency, increased in amplitude, but often disappeared in the vowel [ɑ].

Figure A-1 illustrates these observed spectral differences. The figure shows a comparison of the author's voiced and whispered productions of the vowels [i], [ɩ], [ɛ], [æ], and [ɑ]. As indicated by the labeling, each row of the figure displays the magnitude spectra of a given vowel, with the spectrum of voiced production on the left and that of the whispered production on the right. The gray streaks in the figure highlight how corresponding peaks change frequency across the vowel spectra.

The data for these spectra were 3-5 sec. productions of voiced or whispered steady-state vowels. Each vowel's spectrum was calculated from a 1.7 sec. analysis window positioned within the waveform data corresponding to the production of the vowel. The *ESPS* system's *xspectrum* tool was used to calculate the spectra.[1] This tool windowed the analyzed data with a Hanning function and used Cepstral smoothing (low-pass liftering) of the calculated spectrum.

Considering only the spectra of voiced vowel productions (the left column of the figure), it can be seen that the two lowest-frequency peaks exhibit property 2: moving down the column, as the vowel is varied from [i] to [ɑ], the voiced F1 peak (vF1) generally increases in frequency while the voiced F2 peak (vF2) decreases in frequency.

On the other hand, the spectra of whispered vowel productions (the right column of the figure) clearly do not exhibit property 2. Consider first the spectrum of whispered [i]: its peaks do not match up with those seen in the spectrum of voiced [i]. In the voiced spectrum, the lowest-frequency peak, vF1, appears at 300 Hz, while the second-lowest-frequency peak, vF2, appears at 2000 Hz. In the whispered spectrum, the vF1 peak may still be present (albeit at a higher frequency of 400 Hz). But here, vF1 is almost overshadowed by two flanking peaks: wF1 and wF2. The wF1 peak frequency (300 Hz) roughly matches that of the vF1 peak, but the wF2 peak frequency (800 Hz) does not come close to that of the vF2 peak. Note, however, that the whispered spectrum's wF3 peak (2100 Hz), does approximate the vF2 peak.

[1] The ESPS system is a collection of UNIX/X-windows speech analysis tools from Entropic Research Laboratory, Inc.

Now consider what happens to these peaks in the whispered spectra as the vowel is changed from [i] to [ɩ] to [ɛ] to [æ] to [ɑ]. Over this vowel progression, the wF1 and wF2 peaks do not appear to change in frequency. Instead, their amplitudes change in complementary fashion: wF1's amplitude decreases while wF2's increases. The wF3 peak continues to match vF2 over the progression, increasing in frequency until [ɑ], where it curiously splits into two smaller peaks. The higher-frequency peaks of the whispered spectra (wF4 and wF5) generally also match peaks seen in the voiced spectra (vF3 and vF4).

In sum, this discussion has highlighted two differences between the voiced and whispered spectra of the vowels shown:

1. The F1 peak seen in the voiced spectra, which increases in frequency from [i] to [ɑ], is overshadowed by two peaks (wF1 and wF2) in the whispered spectra. These peaks appear fixed in frequency, but vary their amplitudes in complementary fashion as the vowel changes from [i] to [ɑ].

2. The F2 peak seen in the voiced spectra is also seen in the whispered spectra (as wF3). However, in whispered [ɑ], it splits into two small peaks (that, in fact, in many subjects, are so small as to disappear).

We refer to both these observed spectral differences as *splitting phenomena*. The F2 splitting phenomenon (item 2) could be avoided by concentrating the experiments on vowels other than [ɑ]. The F1 splitting phenomenon, however, was a more pervasive problem, and required development of a new estimation approach to mitigate it.

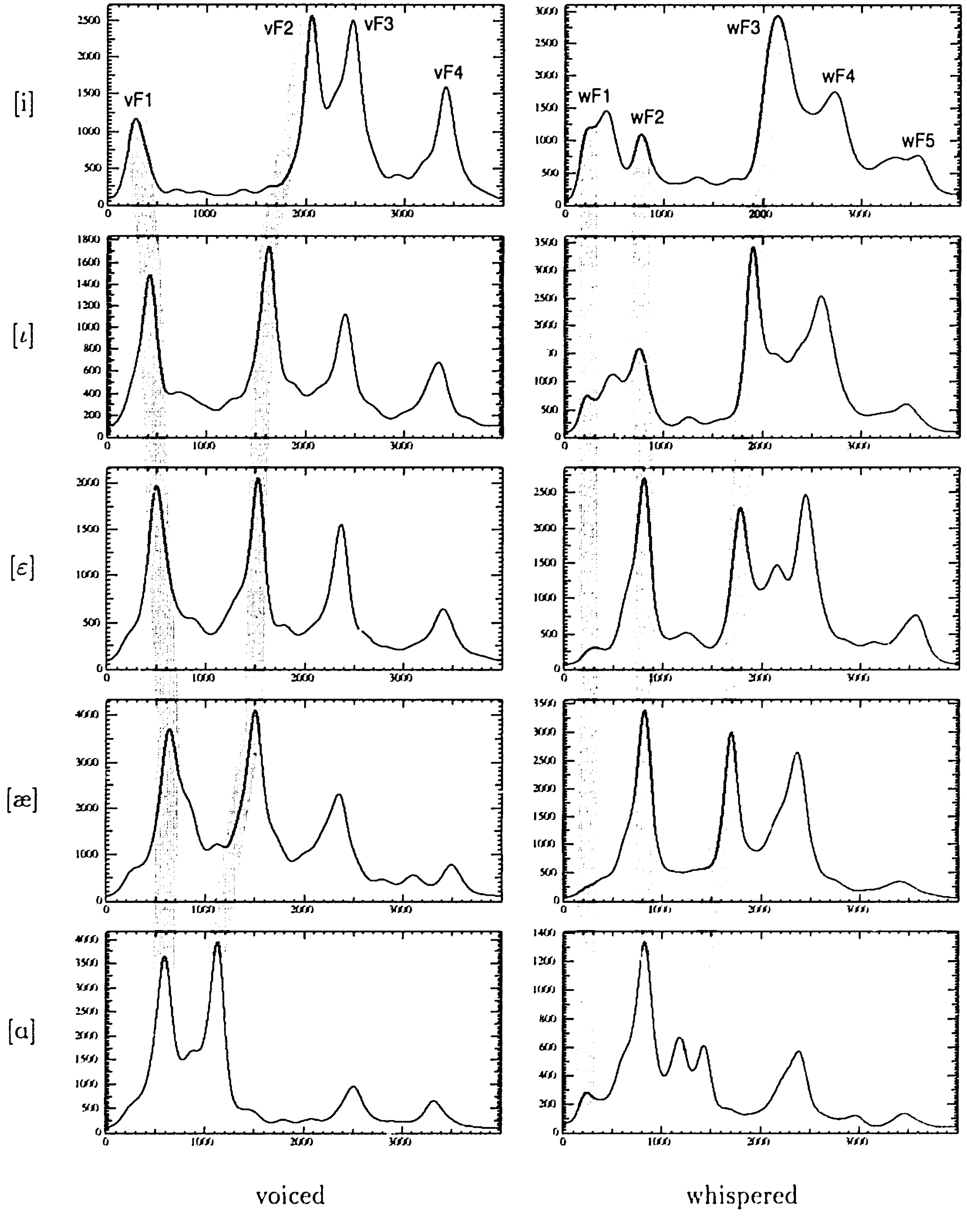


Figure A-1: Comparison of voiced and whispered vowels.

A.1.3 The Chosen Estimation Procedure

The most troubling aspect of the F1 splitting phenomenon was its inconsistency: the degree to which the flanking peaks wF1 and wF2 overshadowed vF1 depended on the subject. For some subjects, their whispered vowel spectra looked just like those seen in Figure A-1. For others, the splitting phenomena were absent, and the spectra of their whispered vowels looked like the voiced vowel spectra of Figure A-1. Many subjects' whispered vowel spectra exhibited characteristics between these two extremes.

Because of this, the chosen F1 estimation procedure did not use peaks of the whispered spectrum, but instead calculated the centroid of a limited region of it. It was done in the following way:

1. For each subject, a region from 0Hz up to some maximum frequency was chosen as the subject's *F1 range*. This frequency range was chosen to contain the peaks pertaining to F1:

    - If the subject exhibited the splitting phenomenon, this range was chosen to contain wF1 and wF2. (For example, in the whispered vowel spectra of Figure A-1 wF1 and wF2 appear confined to the 0-1000 Hz frequency range.)

    - If the subject did not exhibit significant splitting phenomena, this range was chosen to contain the frequency of F1 over the range of vowels shown in Figure A-1.

2. Within this region, the frequency of F1 was estimated as the centroid of the distribution of spectral amplitudes, while the amplitude of F1 was estimated as the average of these spectral amplitudes.

The complete formant estimation procedure thus relied on having determined a subject's F1 region:

- Within this region, F1 was estimated via the centroid method just discussed.

- Above this region, F2, F3, and F4 were estimated in the conventional way from the spectral envelope peaks:
  - F2: the lowest frequency peak above the F1 region.
  - F3: the next highest peak
  - F4: the next highest peak beyond F3.

This estimation procedure is illustrated in Figure A-2 (which is a repeat of Figure 3-3 in chapter 3).

The advantage of this formant estimation procedure was that it produced usable F1 estimates regardless of the existence of splitting phenomena in the spectrum. If a subject's spectra exhibited F1 splitting, then the complementary amplitude variations of wF1 and wF2 caused the F1 range centroid to behave as desired. As Figure A-1 shows, from [i] to [ɑ] wF1's amplitude decreases while wF2's increases. This causes the F1 range centroid to increase in frequency over this vowel progression. If a subject's spectra did not exhibit F1 splitting, then the F1 range centroid simply tracked the F1 peak. Thus, in the vowel progression from [i] to [ɑ], the F1 range centroid again increased in frequency.

Because of this, the frequencies of the F1 and F2 estimates sufficed to distinguish the whispered vowels [i], [ɩ], [ɛ], [æ], and [ɑ]. Moreover, they varied in the same way that F1 and F2 vary in the voiced versions of these vowels.

Figure A-3 illustrates this. It shows spectra of resynthesized versions of the whispered vowels shown in Figure A-1. This resynthesis was accomplished using the DSP to perform a 0.0 transformation (no formant alteration) of each whispered vowel's recording. For each vowel, the spectrum of the DSP's output was again calculated using the ESPS system in the same fashion described above.

In the figure, the left column shows the spectra of the original whispered vowels (the same shown in the right column of Figure A-1). The right column shows the spectra of the resynthesized versions of these vowels.

The peaks of the resynthesized spectra result from the formant estimates used by the DSP in its analysis and resynthesis of the original whispered vowels. Since

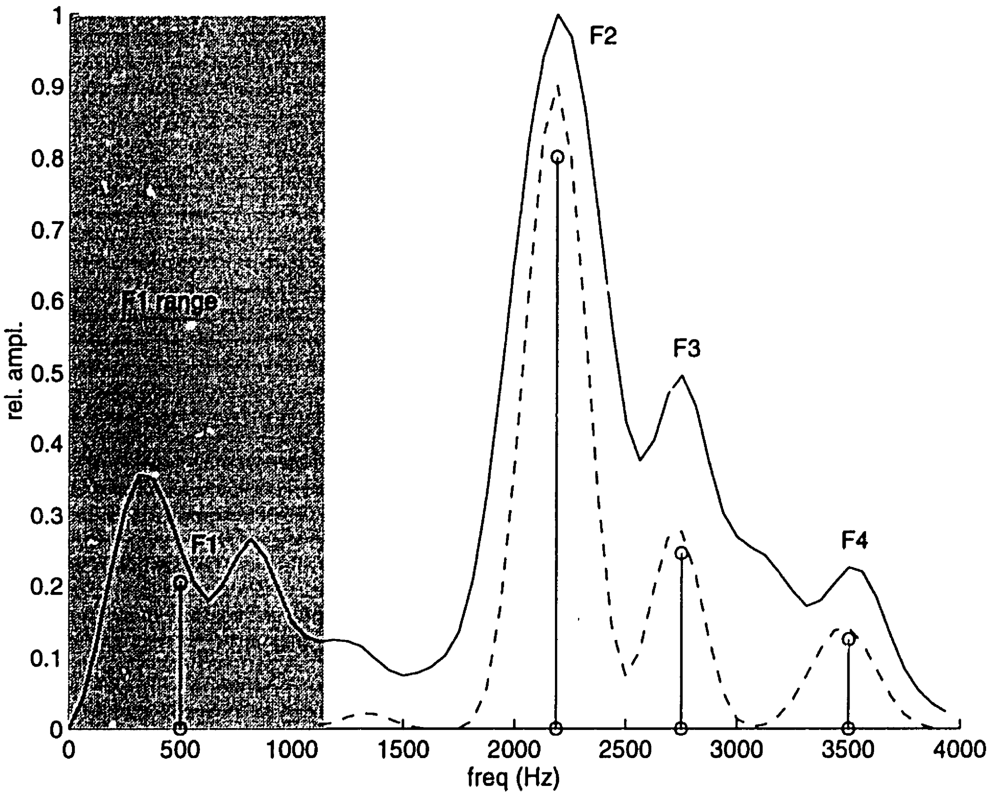


Figure A-2: An illustration of the formant estimation procedure. The solid line shows the spectrum of one frame of the author's whispering of [i]. The gray region highlights the F1 range, within which the two peaks of wF1 and wF2 can clearly be seen. The circle-terminated vertical lines display the formants estimated from this spectrum. As these lines indicate, F1 is estimated as the centroid of spectral amplitudes within the F1 range, while outside of this range, F2, F3, and F4 are estimated from the spectrum's peaks, just as they would be in a voiced spectrum. (The dashed line is a peak-enhanced version of the spectrum used to facilitate peak finding. Note also that, for display purposes, the spectra and estimated formant amplitudes have been offset from each other by scaling. Further details of this process can be found in appendix B.)

these estimates were created using the above-described formant estimation procedure, they exhibit the desired behavior for F1 and F2. This is seen in the spectra of the resynthesized vowels as changes in peak frequencies. In the progression from [i] to [ɑ], the lowest-frequency peak (resulting from the F1 estimate) increases in frequency, while the second-lowest-frequency peak (resulting from the F2 estimate) decreases in frequency. Because of this, the spectra of the resynthesized vowels look more like the voiced vowel spectra shown in Figure A-1.[2]

Interestingly, in spite of the obvious spectral differences, each vowel's resynthesized version sounded perceptually similar to the original version. This was confirmed by all subjects in feedback tests preceding each experiment.

Thus, the formants estimated by this method possessed both properties of the formant definition given above, and therefore served as usable formant estimates.

[2] In the resynthesized [i] and [ɩ] spectra, the amplitude of F2 is, for reasons to be investigated, unusually high. This makes the F1 amplitude appear lower than that of wF1 or wF2 in the original spectra. Checking the amplitude scales shows this is not the case.

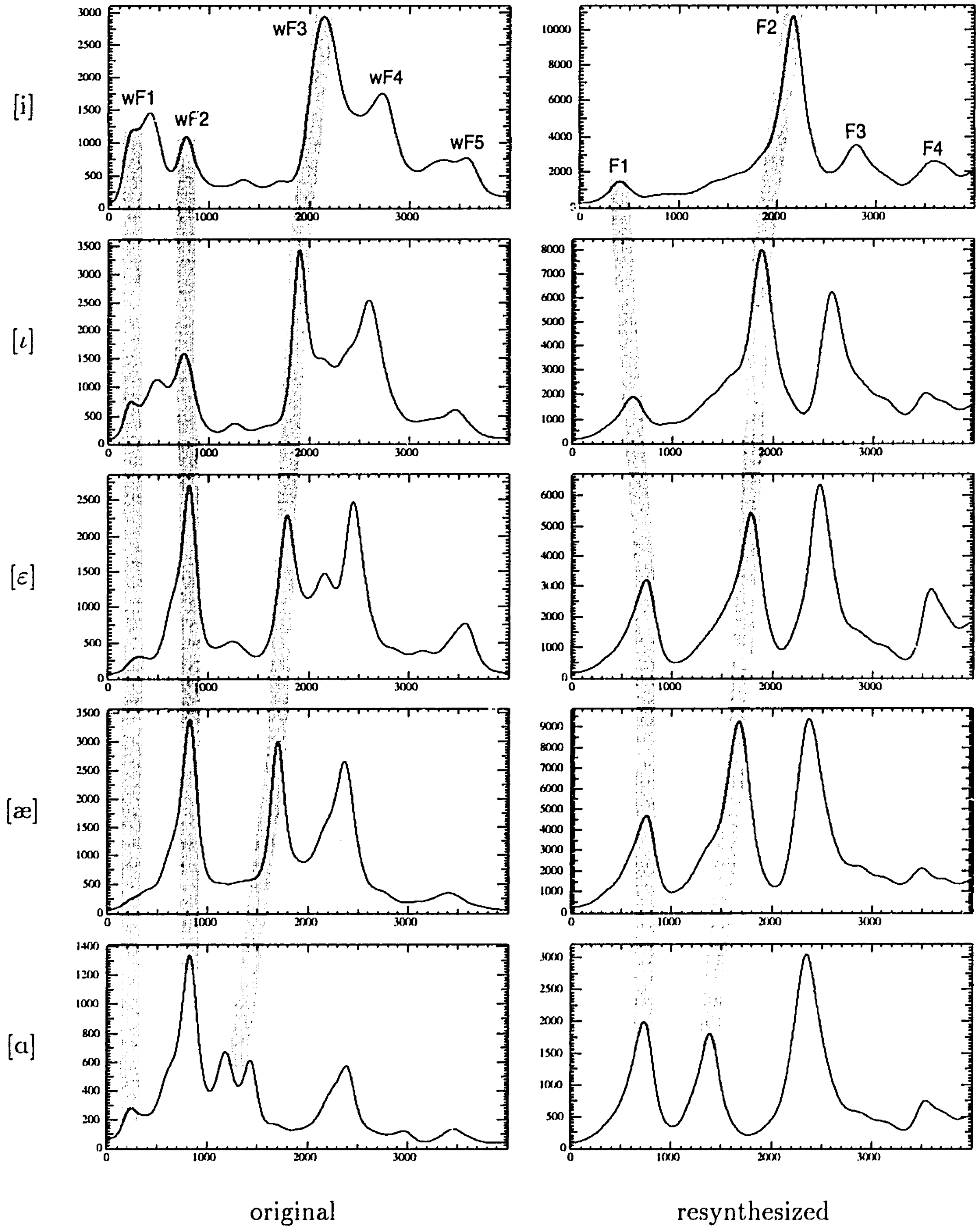


Figure A-3: Comparison of original and resynthesized whispered vowels.

A.1.4 Discussion

For this thesis, the spectral characteristics of whispered vowels served only as a technical problem to overcome. However, they are sufficiently interesting in their own right to warrant further discussion. In particular, we consider here a possible articulatory explanation for them and their relevance to theories of vowel perception. This discussion will conclude this section on formant estimation.

A.1.4.1 An Articulatory Explanation

The hypothesized articulatory explanation of whispered vowel's spectral characteristics comes from [Stevens, 1996], and concerns how a speaker positions his vocal folds when whispering.

Vocal fold position determines two aspects of a speakers speech. First, it determines what type of speech is produced: if the folds are held in one particular position, they vibrate and voiced speech is produced; if they are held in another position, air from the lungs passing through them produces turbulence that results in whispered speech. Intermediate positionings result in breathy speech, which is a mixture of voiced and whispered speech.

The other aspect this positioning determines is the amount of acoustic coupling to the sub-glottal cavities. In voiced speech, the timing of vocal fold vibration is such that, acoustically, these cavities are largely isolated from the rest of the vocal tract. This is because, within each cycle of glottal vibration, maximal excitation of the vocal tract occurs when the vocal folds are closed. In whispered speech, however, the vocal folds are constantly open, resulting in a constant amount of sub-glottal coupling.

The effect of this coupling is to introduce pole-zero pairs in the spectrum of the whispered speech produced. Each pole-zero pair occurs at a resonant frequency of the sub-glottal cavities. For an adult speaker, the lowest two of these resonant frequencies have been observed to occur at 600 and 1400 Hz [Fant and Ishizaka, 1972]. The amount of separation between the pole and zero in the pole-zero pairs is a function of the amount of coupling. With no coupling, as occurs with a closed glottis, there is zero separation between the poles and zeros. In this case, the poles and zeros cancel each other's effects, resulting in no net effect on the spectrum. With significant coupling, as occurs with a more open glottis, the poles and zeros are separated. In this case, the spectrum is amplified near each pole's frequency and attenuated near each zero's frequency.

Thus, spectral distortion produced by sub-glottal coupling occurs around the resonant frequencies of the sub-glottal cavities, and is a function of the amount of coupling. Since this coupling is controlled by the size of the glottal opening, the amount of spectral distortion is a function of the glottal opening.

It is hypothesized that the observed splitting phenomena in whispered speech spectra arise from this mechanism. Moreover, the variability seen across subjects in the amount of splitting is thought to result from differences in whispering *styles*:

- Some subjects are thought to whisper with an *open-glottis* whispering style. These subjects whisper with a large glottal opening, which results in spectral distortions due to sub-glottal coupling. These distortions occur around 600 and 1400 Hz. The distortions at 1400 Hz affect F2 when it is nearby, as is the case for [ɑ]. The distortions at 600 Hz affect F1 when it is nearby, as is the case for [i], [ɩ], [ɛ], and, to a lesser extent, [æ], and [ɑ]. This results in whispered vowel spectra like those seen in the right column of Figure A-1.

- Other subjects are thought to whisper with a *closed-glottis* whispering style. These subjects whisper with a small glottal opening, producing little sub-glottal coupling and little spectral distortion. These subjects' whispered vowel spectra thus differ only minimally from their voiced spectra; their whispered spectra would look like those seen in the left column of Figure A-1.

Subjects with whispering styles between these two extremes would show intermediate amounts of distortion in their whispered vowel spectra.

Confirming evidence for this explanation was provided by a pilot investigation. In it, the author produced vowels in both a closed-glottis and open-glottis whispering style and compared spectra of the results. To make the comparisons more direct, the compared whispering styles were from the same vowel production. This was accomplished by beginning each vowel's production in the closed-glottis whispering style, and abruptly changing to the open-glottis style midway through the production.

The results of this comparison for the vowel [ɛ] are shown in figures A-4 and A-5.

Figure A-4 shows the waveform of the recorded whispering of [ɛ]. The labels above the waveform indicate the closed and open-glottis portions of the utterance.

The figure shows the waveform amplitude is greatly affected by whispering style. The initial amplitude is consistent and low, corresponding to the closed-glottis portion of the utterance. This amplitude abruptly increases at the point where the glottis was opened to the open-glottis whispering style. The amplitude then gradually decreases as the speaker's lung capacity is expended.

The labeled gray regions in the figure show the waveform data time windows from which spectra of the whispering styles were calculated. The closed-glottis spectrum was calculated from time window (a); the open-glottis spectrum, from time window (b). These time windows had the same size (0.4 sec), and were positioned to contain approximately equal-amplitude waveform samples. Figure A-5 shows plots of the resulting magnitude spectra. Each was calculated using the same *xspectrum* settings used to create the spectra of figures A-1 and A-3.

Figure A-5(a) shows the spectrum of the closed-glottis whispering of [ɛ]. This spectrum looks similar to that of a voiced production of [ɛ]: the peaks look like the normal formants of voiced [ɛ]. These peaks are therefore labeled as F1, F2, F3, and F4. Only minimal spectral distortion is evident: the slight peak below F1 is the only apparent deviation from the normal voiced spectrum.[3]

Figure A-5(b) shows the spectrum of the open-glottis whispering of [ɛ]. The distortions seen in this spectrum look similar to those seen in the whispered vowel spectra of Figure A-1. F1 of the closed-glottis spectrum has been largely replaced in this spectrum by peaks wF1 and wF2.[4] These two peaks are centered roughly around

[3] In discussing these spectra, we will use the word "below" to mean lower in frequency, and "above" to mean higher in frequency.

[4] This spectrum is visibly different from the whispered [ɛ] spectrum in Figure A-1. This is perhaps due to glottal opening size differences: the whispered vowels of Figure A-1 were produced with no conscious manipulation of glottal opening, while in Figure A-5(b) glottal opening was deliberately maximized.

600 Hz – the predicted frequency of the first sub-glottal resonance.

Above wF2, the largest spectral peaks occur approximately where the voiced formants would be: wF3 corresponds to F2, wF4 corresponds to F3, and wF5 corresponds to F4. However, additional spectral distortions can be seen between these peaks. A small peak is seen between wF2 and wF3. The frequency of this peak – 1350 Hz – is near the predicted frequency of the second sub-glottal resonance. Another peak is seen between wF4 and wF5.

In sum, the observed differences between the spectra of Figure A-5 are consistent with the hypothesized distortion mechanism:

- Figure A-5(a) shows whispering [ɛ] with a nearly-closed glottis (minimizing sub-glottal coupling) produces minimal spectral distortion.

- Figure A-5(b) shows whispering [ɛ] with a wide-open glottis (maximizing sub-glottal coupling) produces significant spectral distortion. Most of these distortions occur near frequencies of the predicted sub-glottal resonances.

Furthermore, these two spectra represent the extremes of the range of spectral distortions seen in the subject data. Some subjects produced undistorted spectra like Figure A-5(a), while others produced spectra with the same kinds of distortions seen in Figure A-5(b).

Thus, the pilot study results for [ɛ] show that sub-glottal coupling produces the kinds of spectral distortions seen in the subject data. Similar results have been found for [ɩ] and [æ].

To confirm the trends seen in the pilot study, a more complete investigation of sub-glottal coupling in whispering is necessary. Such a study, however, is beyond the scope of this thesis and will be reported elsewhere.

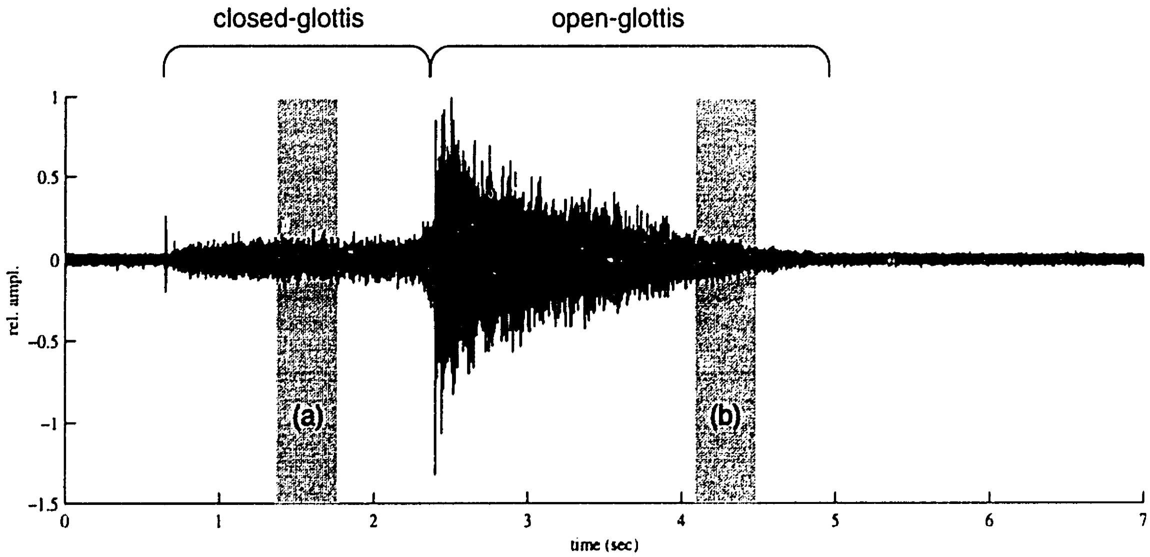
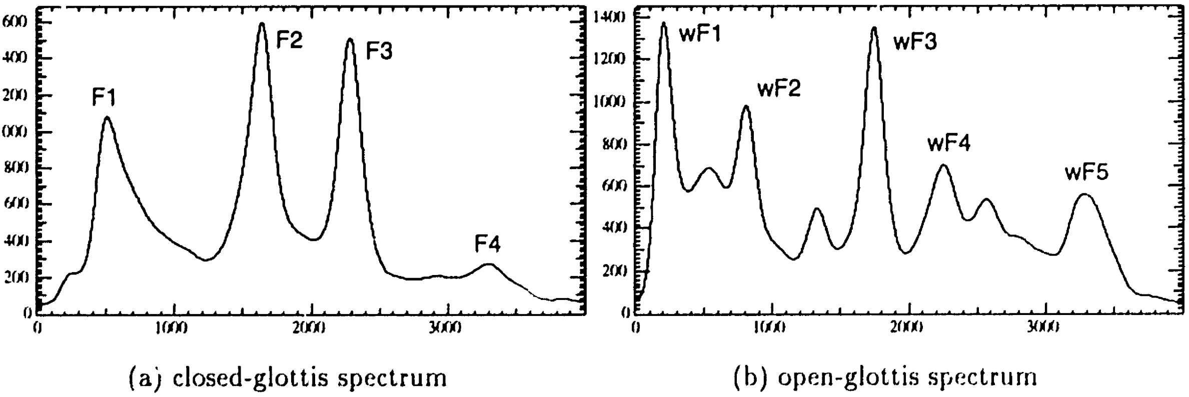


Figure A-4: The time waveform of [ɛ] from which spectra of the whispering styles were calculated. As indicated by the brackets above the waveform, [ɛ] was initially whispered closed-glottis style, but was subsequently whispered open-glottis style. The gray regions marked (a) and (b) indicate the time windows from which spectra were calculated.



(a) closed-glottis spectrum

(b) open-glottis spectrum

Figure A-5: Spectra of the closed and open-glottis portions of the whispered [ɛ] waveform shown in Figure A-4. These spectra were calculated from time windows indicated by gray regions in the waveform figure: spectrum (a) was calculated from time window (a); spectrum (b) was calculated from time window (b).

A.1.4.2 Implications for Vowel Perception Theories

To conclude discussion of this topic, we consider briefly the perceptual implications of the spectral characteristics of whispered vowels.

Since peak frequencies distinguish voiced vowels, it is natural to suppose listeners also use peak frequencies to distinguish whispered vowels.

A study done by Klatt supports this hypothesis [Klatt, 1982]. In it, subjects rated phonetic similarity of synthesized versions of a vowel (either [æ] or [ɑ]). The versions differed in some acoustic parameter value (e.g., formant frequency, amplitude, or bandwidth). The study found versions differing in formant frequency were rated as most phonetically dissimilar. Since, in voiced vowels, formants are spectral peaks, the results offer evidence that listeners use peak frequencies to distinguish vowels.

However, as we have shown above, peak frequencies are often insufficient to distinguish whispered vowels. In particular, frequencies of the first two spectral peaks – wF1 and wF2 – often do not change across whisperings of the [i]–[ɑ] path vowels ([i], [ɩ], [ɛ], [æ], and [ɑ]). How, then, do listeners distinguish whispered vowels?

Several possibilities are apparent:

1. Listeners do not use F1 to distinguish vowels. For the range of whispered vowels just mentioned, the spectral peaks above wF2 appear to be the same seen in voiced vowels. The lowest of these – wF3 – corresponds to the normal voiced F2 peak. This peak's frequency decreases over the [i]–[ɑ] path vowels and is therefore sufficient to distinguish them. However, the distribution of vowel positions seen in formant-space plots of the complete vowel triangle suggest F1 is necessary for some vowel distinctions (see Figure 2-2).

2. Listeners compute F1 as a centroid of the spectral region they expect F1 to occur in. Evidence for this comes from subjects' judgments in the feedback tests beginning each SA experiment. In these tests, all subjects judged the DSP output to be perceptually similar to their actual whispered speech. As shown by Figure A-3, for subjects exhibiting F1 splitting, spectra of their actual whispering and the DSP output can differ significantly. Within the F1 range,

the spectral peaks were often completely different; only the centroid of the F1 range was preserved by the DSP processing. Lack of sensitivity to this difference may indicate subjects also compute a centroid to estimate F1.

3. Listeners have independent representations of voiced and whispered vowels. This would obviate the need to suppose auditory processing invariant to the spectral distortions of whispered speech. Listeners could be supposed to use one criterion for judging voiced vowels (e.g., peak frequencies), and another for judging whispered vowels (e.g., peak frequencies and amplitudes).

   Results of the adaptation experiments indirectly support this theory. Study 2 found subjects retained measurable whispering adaptation over month-long intervals. During a month, each subject should have had ample experience producing voiced speech with unaltered feedback. It was expected that this would restore original whispered vowel productions. That this did not occur suggests that voiced and whispered vowels may have independent representations in the speech production system.

Investigations that address these hypotheses will be conducted in the future, but are beyond the scope of this thesis.

At this point, we resume discussion of issues central to the thesis by considering formant data analysis.

A.2 Formant Data Analysis

All the data analysis methods used in this thesis involved estimating mean formant values of formant data sets. For each data set, these estimates were made from a *formant histogram* of the data. This section describes how these histograms were created and used to estimate mean formants.

A.2.1 Word Production Data Records

Formant data sets come from data records of a subject's word productions. These data records contain formant estimates – by-products of the feedback transformation process.

During an experiment, the DSP estimates F1, F2, F3, and F4 for each input speech frame. To effect the feedback transformation, these formant estimates are then altered and used to synthesize output speech. In addition, these estimates were the data recorded for the input speech frame.

Thus, a word produced by a subject was converted into a sequence of input speech frames. The (F1,F2,F3,F4) estimates for each of these frames constituted the data record of the word production. (For this reason, we will also call each (F1,F2,F3,F4) estimate a "frame".) This process is illustrated in Figure A-6.

A.2.2 Calculating Mean Formants of Formant Data

Figure A-7 illustrates the steps in creating a formant data set and calculating mean formant values from it.

Formant data sets consist of frames collected from word production data records. Depending on the type of data analysis, these frames might be from the same frame position in different data records, or they might be from a range of frames in a single word production's data record.

Once a formant data set is created, the four-step process illustrated in the figure is used to calculate the set's mean formant values. These steps were:

1. Splitting the formant data into individual formant data sets.

   To do this, the individual formant estimates in each frame of the data set are collated into separate data sets for each formant. Thus, an F1 data set is created from the F1 estimates in each frame, an F2 data set is created from the F2 estimates in each frame, etc.

2. Creating an amplitude-weighted histogram of the data in each formant's data set.

As described in Section A.1, a formant estimate consists of two numbers: the formant frequency and amplitude. Each formant's data set is a set of these formant estimates: histograms were made of the formant estimate frequencies, weighted by their amplitudes.

The standard procedure for creating a data set's histogram involves:

(a) Creating bins that cover the range of data values.

(b) For each data value, incrementing the count in the bin representing that data value.

For a formant data set, the data values are the frequencies of the formant estimates. A histogram is created for these data values using only one departure from the standard procedure: for each formant estimate, the count in the bin representing the estimate's frequency is incremented, not by one, but by the amplitude of the formant estimate.

This amplitude weighting of each formant estimate was done to minimize the effect of spurious low-amplitude formant estimates.

After creation of the individual formant histograms, the final two steps in estimating mean formants are:

3. Adding the individual formant histograms to create the combined histogram, which is called the *formant histogram.*

4. Estimating mean F1, F2, F3, and F4 from the highest four peaks of the formant histogram.

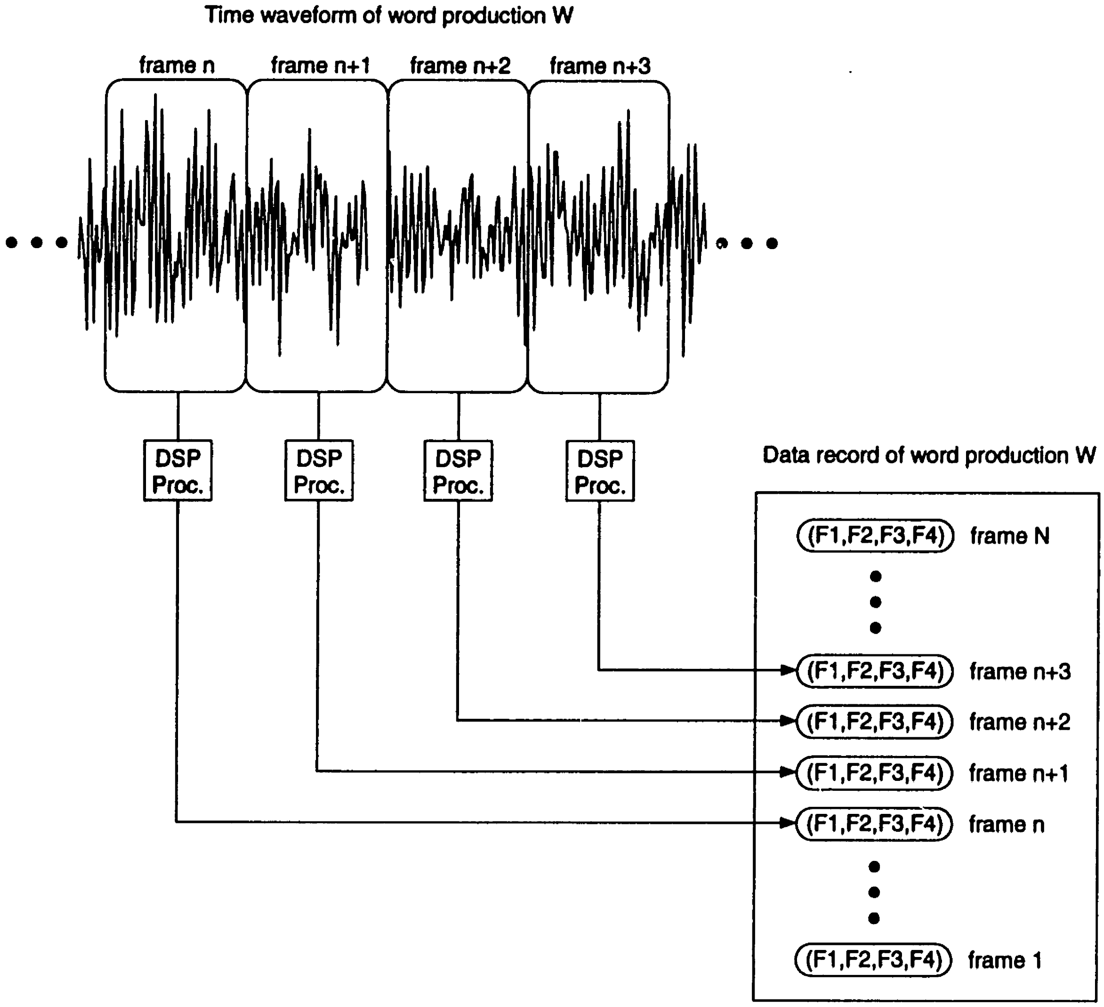


Figure A-6: Illustrating how word data records were created. The upper right of the figure shows a portion of the time waveform of a subject's word production W. To create transformed feedback, the word production is chunked into input speech frames, as indicated by the oval boxes labeled "frame n" through "frame n+3". As indicated by the boxes labeled "DSP Proc.", An intermediate step in this processing is the estimation of F1, F2, F3, and F4 from a magnitude spectrum of the frame. These (F1,F2,F3,F4) estimates are used to synthesize the subject's feedback. In addition, as indicated by the arrows below the "DSP Proc." boxes, these formant estimates are the data stored for each frame. The data record of word production W is thus a sequence of (F1,F2,F3,F4) estimates that we also call frames.

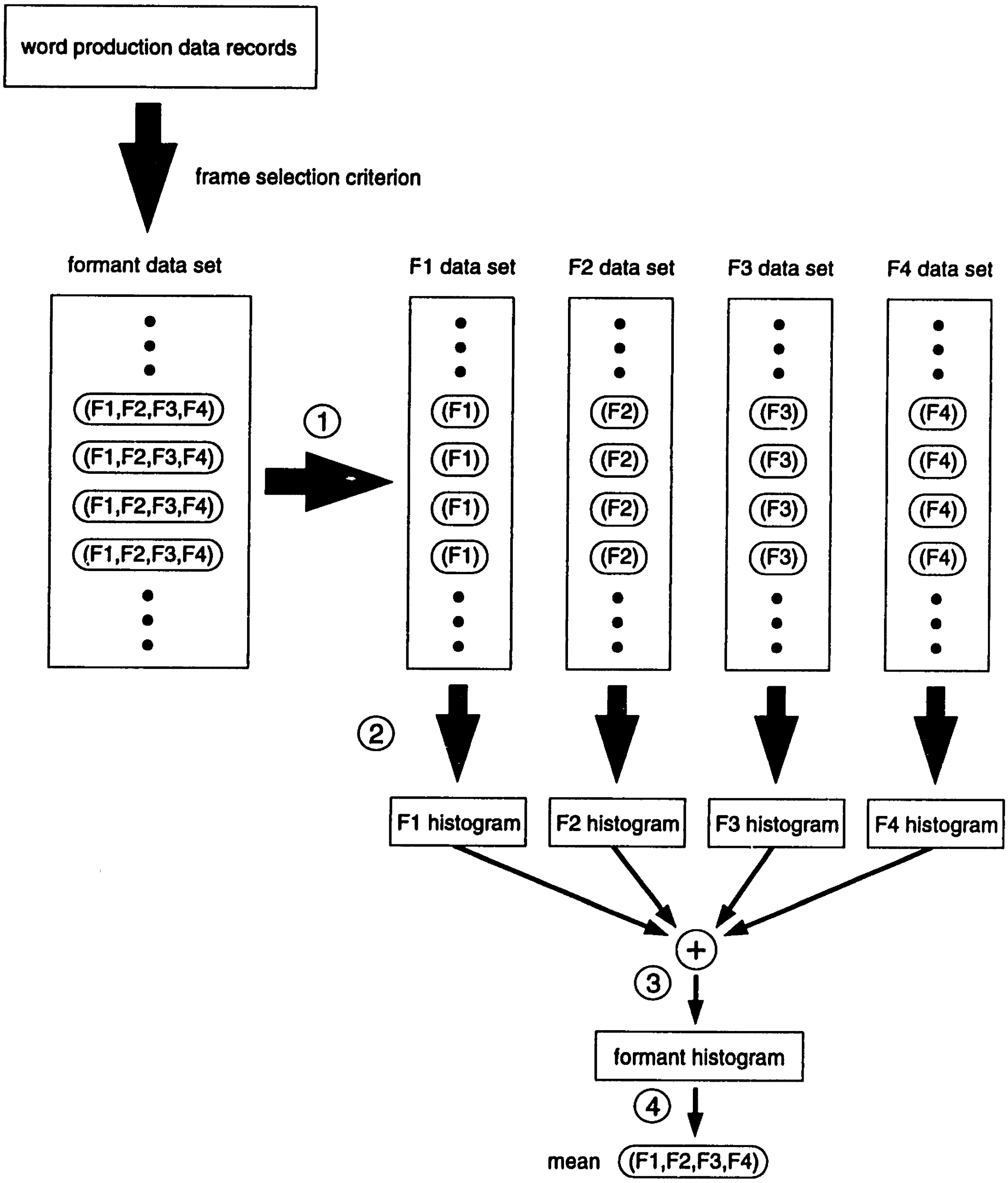


Figure A-7: Steps in the calculation of mean formants of a set of formant data. By some selection criteria, frames from word production data records are collected into a formant data set. Then, via a four-step process, mean formants of this data set are estimated. These steps are: (1) splitting the formant data into individual formant data sets; (2) creating an amplitude-weighted histogram of the data in each formant's data set; (3) adding the individual formant histograms to create the combined *formant histogram*; (4) estimating mean F1, F2, F3, and F4 from the formant histogram.

A.2.3 Analysis of the Method

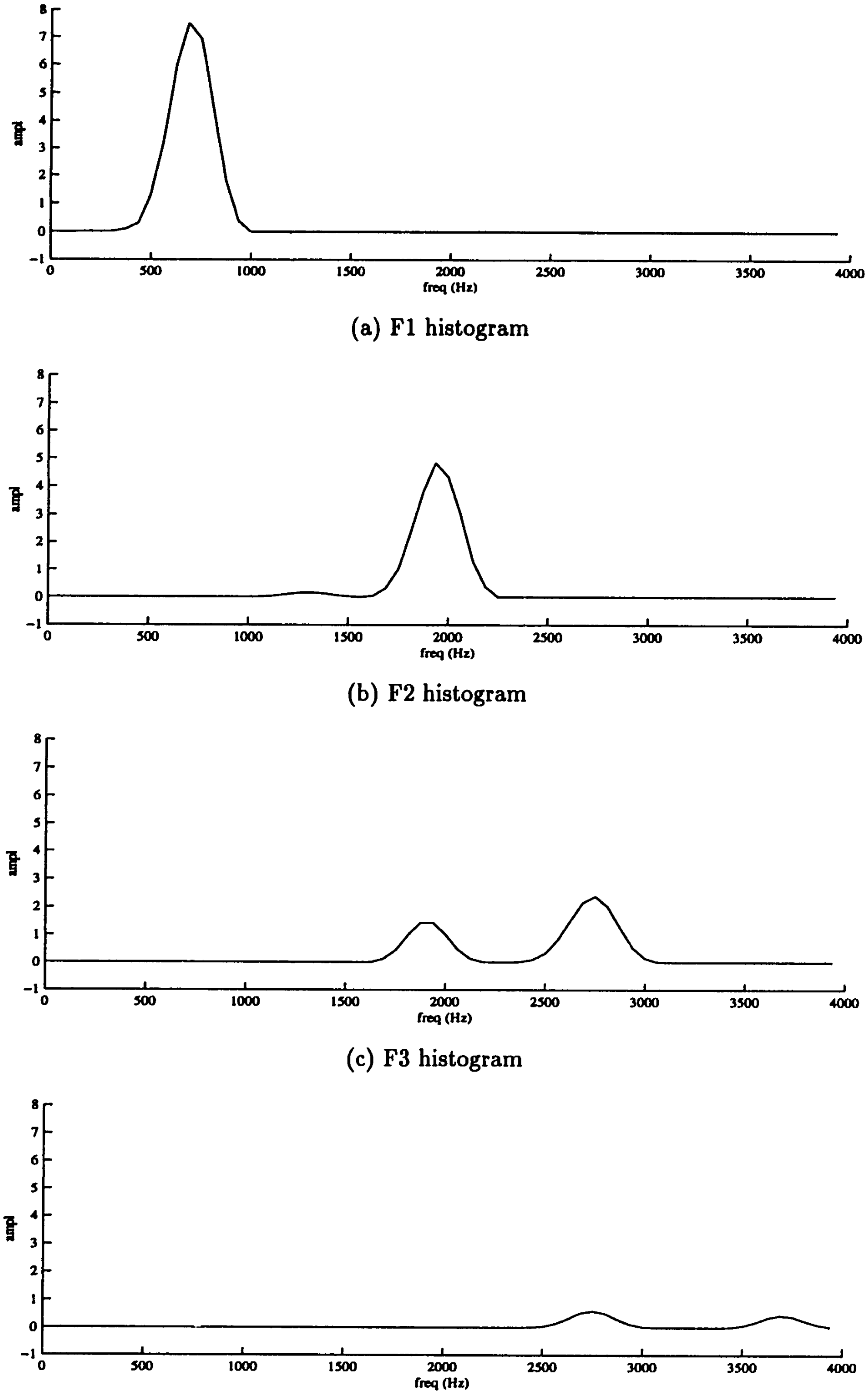
To see why mean formants are estimated from the combined formant histogram, instead of directly from the individual formant histograms, we consider here an example of this process.

Figure A-8 shows an example of individual formant histograms calculated from an actual formant data set. In this case, the set's frames are from a word produced by subject BK in study 1. This figure illustrates why it was unreliable to estimate mean formant values directly from the individual formant histograms. Figures A-8(a) and A-8(b) show the F1 and F2 histograms to be suitable for estimating mean formant values: these histograms both exhibit a single-peaked distribution. Such a distribution is the predicted outcome of a subject intending to produce a single formant frequency.

However, figures A-8(c) and A-8(d) show the F3 and F4 histograms are unsuitable for estimating mean formant values: their double-peaked distributions are not the likely result of a single intended formant frequency. Indeed, the mean of each distribution is a formant frequency value never produced by the subject.

The cause of these double-peaked distributions can be seen in how their peaks line up with those of the other formants' distributions. In the F3 distribution, the lower frequency peak has the same frequency as the peak in the F2 distribution. In the F4 distribution, the lower frequency peak is aligned with the higher F3 distribution peak. This alignment of peaks suggests the double-peaked distributions are due to mislabeling in the formant estimation process: as a result of noise, an insignificant spectral peak (whose frequency is lower than the true F2 peak) is occasionally getting labeled as F2. This results in the true F2 peak being labeled as F3 and the true F3 peak being labeled as F4.

The effects of this mislabeling on the F2 histogram are minimized by amplitude weighting the data used in creating it, as described above. Because the spectrum peaks occasionally mislabeled as F2 have small amplitudes, they contribute little to the overall F2 histogram. However, the F2 peaks mislabeled as F3 are likely to have high amplitudes and contribute significantly to the overall F3 histogram. The same



(a) F1 histogram

(b) F2 histogram

(c) F3 histogram

(d) F4 histogram

Figure A-8: Showing the separate histograms for each formant's data from subject BK.

holds true for F3 peaks mislabeled as F4.

To mitigate the effects of this mislabeling, mean formant values are instead estimated from the combined formant histogram shown in Figure A-9. From this histogram, mean formants are estimated as the frequencies and amplitudes of the four largest peaks.

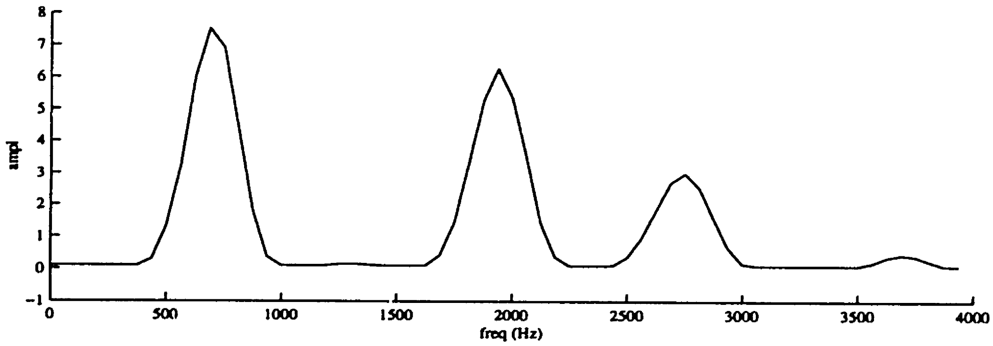


Figure A-9: The complete formant histogram made by adding together the individual formant histograms of Figure A-8.

By estimating mean formants in this fashion, we have, in effect, removed the peak labels in the individual formant histograms. This makes the approach robust to formant mislabeling: it does not matter which individual formant histogram a formant peak initially belongs to, since all its occurrences will sum to a single peak in the combined histogram.

Having described how formant histograms are created and used to estimate mean formants, we can now describe the subject pretest procedure. This procedure is based on examining formant histograms.

A.3 Subject Pretesting

The formant estimation procedure described in the first section was based on knowing the F1 range of a subject. This necessitated its measurement in a subject pretest procedure. Pretesting of subjects was also necessary for several other reasons:

- Since feedback transformations were based on shifts along a subject's [i]–[ɑ] path, formant values of subjects' path vowels ([i], [ι], [ε], [æ], and [ɑ]) needed measurement.

- Subjects with weak spectral peaks, or who otherwise had trouble whispering had to be screened out.

- Subjects whose vowels sounded unnatural when heard through the feedback transformation also had to be screened out.

These requirements led to the a three-step subject pretest procedure consisting of:

1. Measuring the subject's F1 range.

2. Measuring his path vowel formants.

3. Assessing the fidelity of his transformed vowel sounds.

A.3.1 Measuring the F1 Range

Measuring a subject's F1 range began by first acquiring vowel data in a brief experiment. Data from this experiment was then analyzed using formant histograms to estimate the range of F1.

An important aspect of this experiment was how the DSP estimated formants in it. Since no F1 range was known at this point, the normal formant estimation method could not be used. Instead, all formants were estimated as peaks in the magnitude spectrum, just as is normally done for voiced speech. This will be called the DSP's *peak mode* of formant estimation, to distinguish it from the DSP's *normal mode* of F1-range-dependent formant estimation.

As described above, whispered formants estimated as peaks are not a suitable basis for the feedback transformation procedures. But, if left unaltered, they can be used to produce reasonable synthesized whispered speech. The subject's feedback was left unaltered in this experiment, because its purpose was only to record normal whispered vowel productions of a subject. The DSP was therefore able to provide adequate feedback using peak mode formant estimation during the experiment.

A.3.1.1 Vowel Data Acquisition

The experiment collected data on the subject's productions of the five different path vowels.

The time course and subject interface of this experiment were identical to those described in chapter 3. The experiment consisted of 10 epochs, each of which consisted of two stages: a feedback stage and a noise stage. These two stages both consisted of prompting, in random order, of all of the words from the set {"beed", "bid", "bed", "bad", "bod"}.[5] In the feedback stage, the DSP output was in mixed mode, allowing the subject to hear his whispering. In the noise stage, the DSP output was in noise mode, which blocked the subject's hearing. The feedback heard by the subject in the mixed mode state was unaltered.

In order for the utterance data to primarily reflect the steady vowel portion, the prompted-for utterance duration was 500ms (63 analysis frames). This was done in an effort to make the constant transition times small compared to the overall utterance duration. For the initial [b] and final [d], these transition times were, in general, sufficiently fast to be largely completed in 2–3 analysis frames.

A.3.1.2 Vowel Data Analysis

Formant histograms were used to analyze the results of the data collection experiment and determine the subject's F1 range.

For each word and each feedback condition, a histogram was made of the formant data from all productions of that word whispered in that feedback condition. These histograms were then combined to make two different composite histograms: one for words whispered while the subject heard feedback (a *feedback* composite histogram) , and one for words whispered while he heard only noise (a *noise* composite histogram). A subject's F1 range was estimated from the feedback composite histogram and verified in the noise composite histogram.

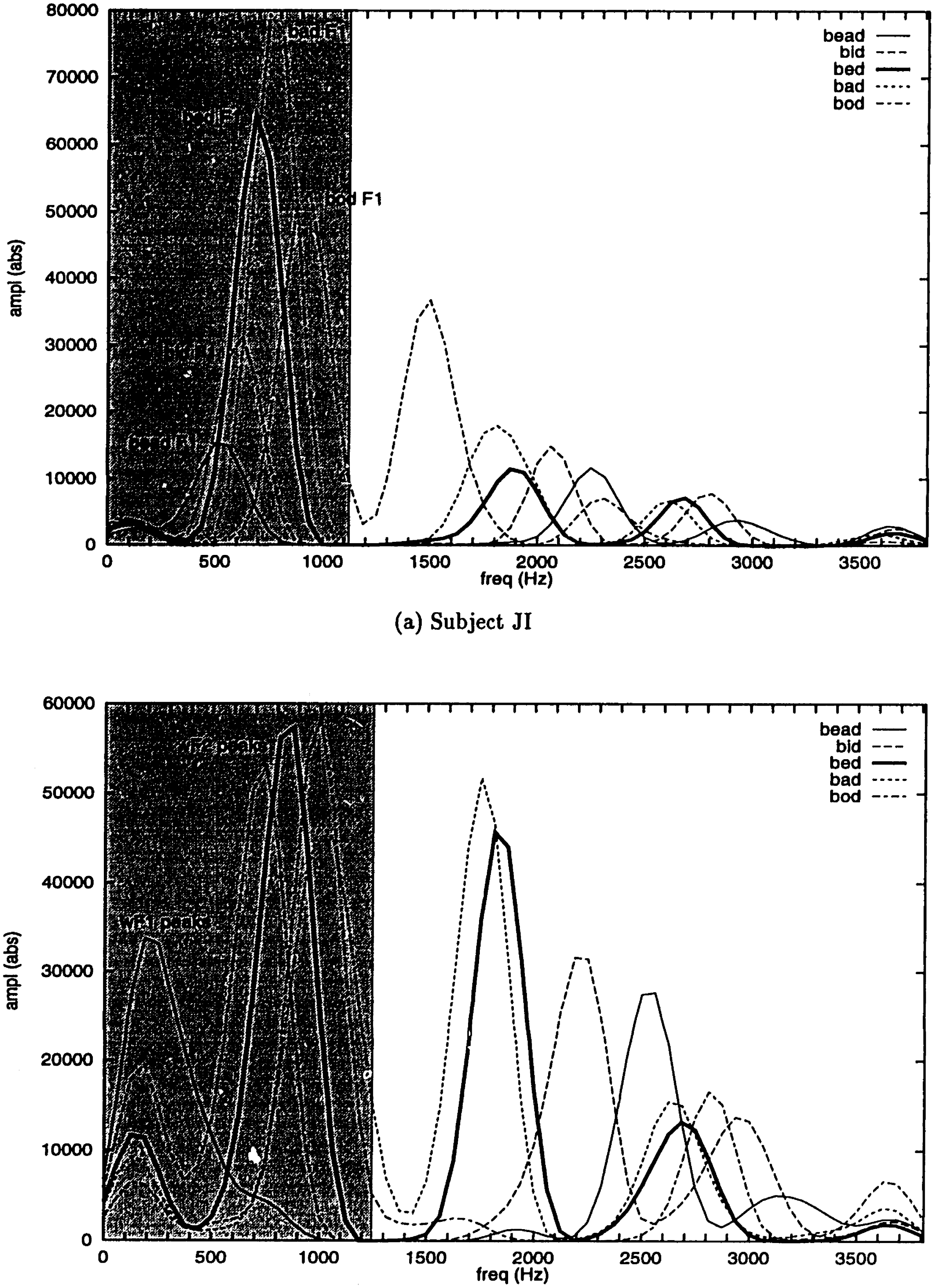
[5] For early versions of the procedure other vowel sounds were also tested for, and the set included the word "bawd" or the word "bahd". However, nearly all subjects pronunciation of these words were the same as "bod", and data from these pronunciations were never used.

In the feedback composite histogram, estimation of the F1 range depended on the whispering style of the subject. If the subject whispered closed-glottis style, his F1 range was the frequency range over which his F1 peak varied across the individual word histograms. If the subject whispered open-glottis style, his F1 range was the frequency range containing the wF1 and wF2 peaks of the individual word histograms.

Feedback composite histograms of two different subjects are shown in Figure A-10.

Figure A-10(a) shows the composite histogram of subject JI's word productions. This subject whispered closed-glottis style. As indicated by the figure, his F1 range was determined from the range containing the F1 peak of all his individual word histograms.

Figure A-10(b) shows the composite histogram of subject GL's word productions. This subject whispered open-glottis style, and indicated by the figure, his F1 range was determined from the positions of the wF1 and wF2 peaks of his individual word histograms.



(a) Subject JI

(b) Subject GL

Figure A-10: Composite histograms used to measure the F1 range (shown in gray) of two subjects. In these, each line style shows a different word's formant histogram.

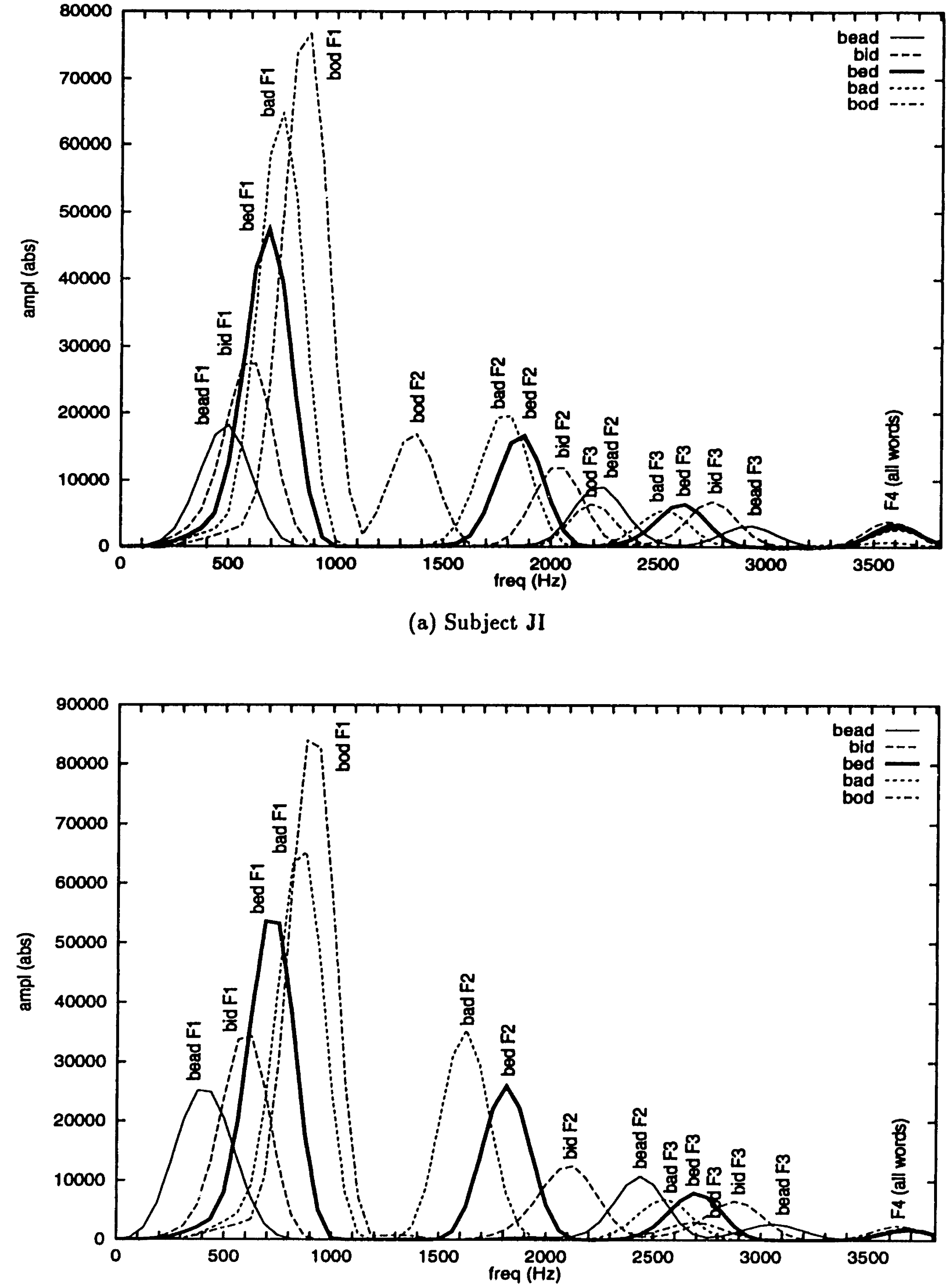
A.3.2 Measuring Path Vowel Formants

Following measurement of a subject's F1 range, a second procedure was used to measure his path vowel formant frequencies. This procedure was almost identical to the F1 range measurement procedure. The only differences were:

- The data collection experiment was 20 epochs long.
- In the experiment, the DSP's normal mode of formant estimation was used.
- In analyzing the results, vowel formant frequencies were estimated from the feedback composite histogram.

Examples of feedback composite histograms from this measurement procedure are shown in Figure A-11. The histograms are from the same subjects shown in the previous figure. Each histogram shows the formant labeling given the peaks of the individual word histograms. Note that because of the centroid-based F1 estimate used by the DSP, the previously seen wF1 and wF2 peaks are no longer present in subject GL's histogram.

Since the words used in the data collection experiment contained all the path vowels, formants estimated from their histograms could be used as path vowel formant estimates. Table A.1 shows the path vowel formants estimated for subjects JI and GL from the word histograms in Figure A-11.



(a) Subject JI

(b) Subject GL

Figure A-11: Composite histograms used to measure path vowel formants of two subjects (the same shown in Figure A-10).

	Subject JI				Subject GL			
path vowel	F1	F2	F3	F4	F1	F2	F3	F4
[i] (in "beed")	500	2250	2937	3625	375	2437	3062	3687
[ι] (in "bid")	625	2062	2750	3562	625	2125	2875	3687
[ε] (in "bed")	687	1875	2625	3625	687	1812	2687	3687
[æ] (in "bad")	750	1812	2500	3625	875	1625	2562	3687
[ɑ] (in "bod")	875	1375	2187	3562	875	1312	2687	3625

Table A.1: Path vowel formant frequencies (Hz) for subjects JI and GL, estimated from peaks of the word histograms in Figure A-11.

A.3.3 Assessing Feedback Transformation Fidelity

Following determination of a subject's path vowel formants, a brief assessment procedure was used to evaluate transformed versions of his path vowels. This procedure consisted of:

1. Using the path vowel formant estimates to generate the lookup tables for the 2p and 2m feedback transformations.

2. Listening to how the subject's path vowels sounded under each feedback transformation.

The listening test of step 2 was a subjective evaluation by the experimenter. The subject did not hear the DSP output during this test; he removed the insert earphones and heard his whispering in the normal acoustic fashion. The experimenter listened to the DSP output and made requests for the subject to whisper different words. These words were the same as those used in the previous measurement procedure: they therefore contained all the path vowels. For each word whispered, the experimenter judged phonetic identity of the vowel heard in the DSP output.

These phonetic judgments were recorded for each path vowel under each feedback transformation. The recorded judgments were then compared with the predicted action of each feedback transformation. For example, [ɛ] under the 2p transformation should be heard as [ɑ]and under the 2m transformation as [i].

For a given transformation, if recorded phonetic judgments matched predictions for all path vowels except [ɑ] (which was often missing F2), the subject was rated as suitable for participating in experiments using that feedback transformation.

This assessment of feedback transformation fidelity concluded the pretest done on each subject.

Appendix B

Signal Processing

In Chapter 3, Section 3.2 summarized the feedback transformation signal processing as consisting of five steps:

- **Acquisition:** An 8ms (64 sample) frame of the subject's whispered speech is acquired.
- **Spectral analysis:** This frame is analyzed into a magnitude spectrum, which is further processed before formats are estimated from it.
- **Formant estimation:** From the processed spectrum, the frequencies and amplitudes of F1, F2, F3, and F4 were estimated.
- **Formant alteration:** To implement the feedback transformation, a lookup table was used to shift the frequencies of F1, F2, and F3.
- **Synthesis:** Whispered speech is synthesized from the altered formant estimates and output as feedback to the subject.

The implementation of these steps was based on a number of sources. Design of the acquisition step was based on example programs provided by the manufacturer of the DSP system used (see Appendix D). Design of the spectral analysis and synthesis steps were based on [Houde, 1994]. Finally, design of the formant estimation step was based on observed characteristics of whispered speech, as described in Appendix A.

These processing steps will now be discussed in more detail. In this discussion, some knowledge of speech and linear systems theory is assumed.

B.1 Acquisition

Two issues were important in the acquisition of the subject's speech: sampling rate and frame size.

B.1.1 Sampling Rate

The digital nature of the signal processing required that the continuous-time speech signal be sampled. The rate at which these samples were acquired (the *sampling rate*) was determined from two competing considerations: (1) higher sampling rates incurred greater computational cost; (2) lower sampling rates decreased system bandwidth.[1]

For typical adult male speakers, F4, the highest formant likely to contain significant phonetic information [Stevens, 1989], ranges as high as 4KHz (with female and child speakers, F4 will typically be much higher). Thus, by restricting the choice of subjects to adult male speakers, the bandwidth of the system could be as small as 4KHz, which set the sampling rate at 8KHz.

B.1.2 Frame Size

To maximize the veridicality of the subject's feedback, there must be little delay and no interruptions in its generation. The requirements this places on the signal processing are that the input speech must be continuously sampled without interruptions, and that, simultaneous to this and in synchrony with it, synthesized speech feedback samples must be generated.

In order to accomplish this, the signal processing employed a system of two input buffers and two output buffers. At any given time, one set of the buffers is active

[1] According to sampling theory [Oppenheim and Willsky, 1983], system bandwidth is equal to half the sampling rate.

while the other is idle. The active input buffer is in the process of being filled (with incoming samples of the subject's speech), while the active output buffer is in the process of being emptied (as its contents are sequentially output to the subject as his feedback). Both buffers fill and empty at the the sampling rate, and both hold a frame-sized amount of data.

While the active buffers are being filled/emptied, the idle buffers are being processed: the frame of data in the idle input buffer is added to a larger data buffer called the analysis window, the contents of which are then then analyzed into a magnitude spectrum. From this spectrum, formants are estimated, altered, and used to generate synthesized speech that fills the idle output buffer.

Once the the active buffers have finished filling/emptying, the buffers must immediately change roles before the next input speech sample arrives (and the next output speech sample must be delivered). When this happens:

- the current idle buffers become the new active buffers (ready for input/output), and

- the active buffers become the new idle buffers (ready to be processed).

This double-buffering scheme works because the DSP system hardware allows the active buffers' input/output to proceed simultaneous to the processing of the idle buffers. Thus, the main system time constraint is that the processes filling the idle output buffer must complete before the active buffers have finished filling/emptying.

Since the sampling rate is fixed, the time it takes for the active input buffer to fill (or, equivalently, the active output buffer to empty) is determined solely by the frame size. It was found that this could be set as small as 64 samples and still leave enough time for the processing of the idle buffers to finish.

In this double-buffering scheme, feedback delay is the time between two events:

- Acquisition of an incoming speech frame's first data sample.

- Output of the first synthesized speech sample generated from processing this incoming speech frame.

Thus, from the above description it can be seen that:

$$\text{feedback delay} = \begin{pmatrix}\text{time it takes to ac-}\\ \text{quire the input frame}\\ \text{of data}\end{pmatrix} + \begin{pmatrix}\text{time it takes to pro-}\\ \text{cess this frame of}\\ \text{data}\end{pmatrix}$$

But since, after the idle buffers are processed, the system must still wait for the active buffers to finish filling/emptying, the above equation reduces to:

$$\begin{aligned}\text{feedback delay} &= \begin{pmatrix}\text{time it takes to ac-}\\ \text{quire the input frame}\\ \text{of data}\end{pmatrix} + \begin{pmatrix}\text{time it takes to out-}\\ \text{put a frame of data}\end{pmatrix}\\ &= 2\begin{pmatrix}\text{time it takes to ac-}\\ \text{quire the input frame}\\ \text{of data}\end{pmatrix}\\ &= 2\,(\text{frame size})\,(\text{time per sample})\\ &= 2\,(\text{frame size})\,(1/\text{sampling rate})\\ &= 2(64)(0.125ms)\\ &= 16ms\end{aligned}$$

This value is well below the 30ms delay at which speakers begin to notice and be disturbed by the delay in feedback [Lee, 1950, Yates, 1963].

B.2 Spectral analysis

The goal of the spectral analysis processing step was creation of spectral representations suitable for formant estimation. This involved creating a magnitude spectrum that was then further processed into smoothed and peak-enhanced versions.

B.2.1 Creating the Magnitude Spectrum

The first step in processing the new frame of data in the idle input buffer is calculation of its magnitude spectrum. This magnitude spectrum is calculated from a buffer of data called an *analysis window.* Thus, to calculate the magnitude spectrum of the new frame of data, the oldest frame of data in the analysis window is shifted out of the window and the new frame shifted in. This process is illustrated in Figure B-1.

The size of the analysis window is determined by the resolution of the spectral analysis done. In this case, it was found that there was sufficient processing time to do a 64-channel magnitude spectrum analysis. This meant that a 128-point FFT (fast Fourier Transform [Cooley and Tukey, 1965, Oppenheim and Schafer, 1975]) was done on the data, which required the analysis window to hold 128 data samples, or 2 frame's worth of data.

The FFT is not immediately done on the analysis window after the newest frame of data had been added: first, its contents are multiplied by a windowing function (a hamming window) to minimize spectral distortion in the FFT processing. After this, The FFT processing is done, resulting in the initial magnitude spectrum.

B.2.2 Processing the Magnitude Spectrum

Several operations are performed on the magnitude spectrum before its formants are estimated.

First, the ambient spectrum is subtracted from it. This ambient spectrum is the magnitude spectrum of background sound sources existing in the room where the experimental setup exists, plus any noise sources in the electronics of the microphone and microphone amplifier.

Next, the average magnitude of the spectrum is calculated and saved for later use in controlling amplitude of the speech feedback synthesized from the spectrum. This average spectrum magnitude therefore controls the amplitude of the feedback heard by the subject. As a safeguard for the subject's hearing, this average magnitude is limited to be less than a fixed threshold value.

After that, the spectrum is smoothed in time by averaging it with the previous frame's spectrum. The resulting time-smoothed spectrum is saved for further processing, and is later used for estimating F1.

Finally, the peaks of the time-smoothed spectrum are enhanced by a series of operations. First, a running average (the average magnitude of the n surrounding channels) is subtracted from each channel. This removes overall trends in the spectrum and thus tends to enhance localized variations which are usually peaks. Any small peaks in the valleys (which, at this point, have negative magnitude values) are then removed by keeping only the positive channel amplitude values. This peak-enhanced spectrum is then smoothed in frequency (by convolution with an appropriate kernel) and then in time (by weighted average with the sum of past frames' peak-enhanced spectra), to make the final peak-enhanced spectrum. F2, F3, and F4 were estimated from this spectrum.

Examples of the time-smoothed and peak-enhanced spectra are illustrated in Figure B-2.

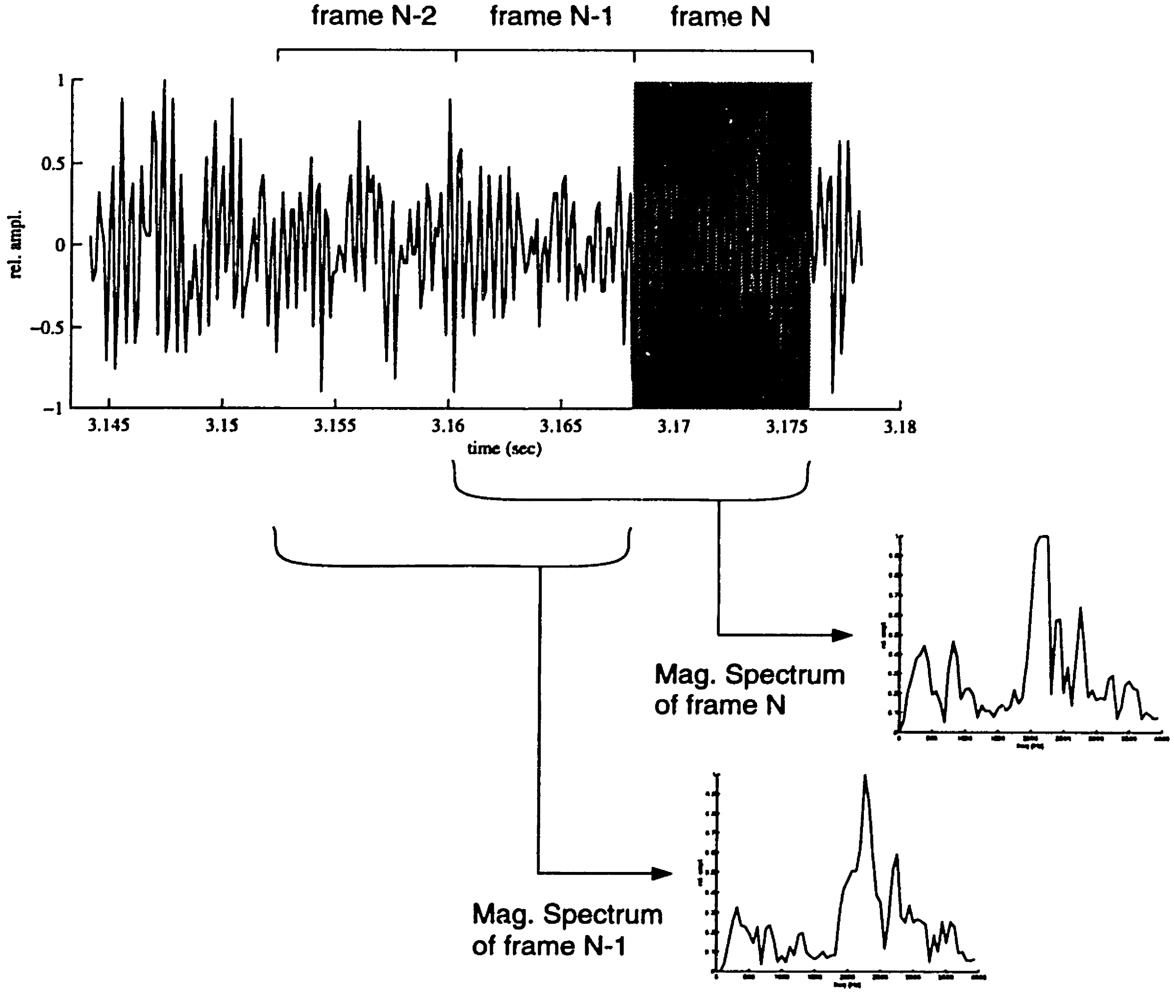


Figure B-1: Showing how the incoming speech data was processed into magnitude spectra. The time waveform shown is from the author whispering the vowel [ɛ] in the word "beed". Input speech data was processed in 8ms (64-sample) frames. As soon as a complete frame of data was acquired (the gray box labeled "frame N"), its data and the previous frame's data were processed into a magnitude spectrum. This is indicated below the time waveform by the brace spanning frames N and N-1 (representing the analysis window), and by the arrow leading from this brace to the magnitude spectrum of frame N. During the calculation and subsequent processing of this spectrum, two other processes were simultaneously occurring: (1) new input speech data (the time waveform shown extending beyond frame N) was being acquired and (2) the synthesized speech generated from the previous processing of frame N-1 was being output.

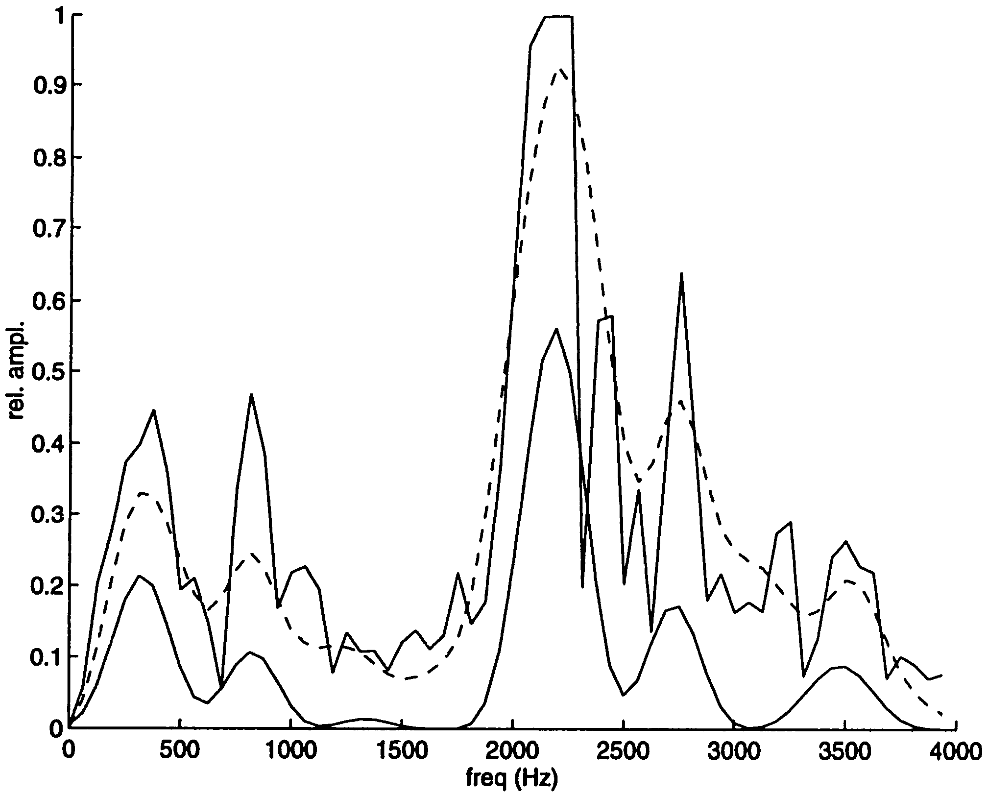


Figure B-2: Key steps in processing the magnitude spectrum of frame N from the previous figure. The highest solid-line spectrum is the unprocessed FFT magnitude spectrum of frame N. The superimposed dashed-line spectrum is the time-smoothed version of it from which F1 was estimated. The lower smooth solid-line spectrum is the peak-enhanced spectrum. This was derived from the time-smoothed spectrum, and was used to estimate F2, F3, and F4. (For display purposes, these spectra have been offset from each other by scaling. In the actual signal processing, each spectra is normalized to have the same average amplitude.)

B.3 Formant Estimation

Because of problems with determining F1 in whispered speech (see Appendix A), formants could not be estimated simply from the peaks of the peak-enhanced spectrum.

The method used required measuring the frequency region over which a subject's F1 ranges:[2]

- **Within the F1 range**, the time-smoothed spectrum was used to estimate F1. The frequency of F1 was estimated as the centroid frequency of the spectral amplitude distribution within the F1 range. The amplitude of F1 was estimated as the average spectral amplitude within this range.

  As discussed in Appendix A, this produces a robust F1 estimate with the desired characteristics: it increases in frequency as the whispered vowel changes from [i] to [ɑ], just as F1 does for the voiced version of these vowels.

- **Outside of the F1 range**, the peak-enhanced spectrum was used to estimate F2, F3, and F4. This was done using a standard, hill-climbing-based technique to find the peaks of the spectrum. F2, F3, and F4 were then estimated as the highest three of these peaks.

This method is illustrated in Figure B-3, which is a repeat of Figure A-2.

Because F1 estimation is based on calculating a center of mass, it is desirable to have the spectrum which more closely represents the true distribution of channel amplitudes in the F1 range. The peak-enhanced spectrum is not a good representation of this since many overall amplitudes trends have been removed from it. For this reason, F1 is estimated from the time-smoothed spectrum rather than the peak-enhanced spectrum

Following this, the formant amplitudes are rescaled so their average equals the original (previously calculated) average amplitude of the original magnitude spectrum.

[2] This was done in the subject pretest procedure described in Section A.3.

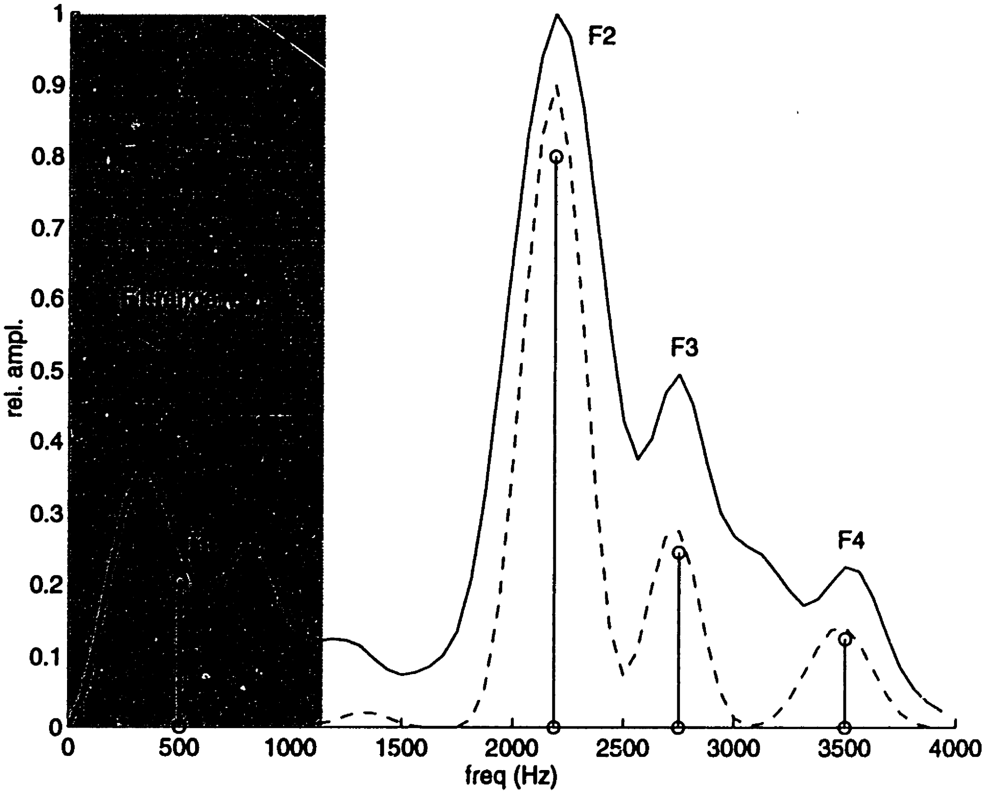


Figure B-3: Showing how formants were estimated from the spectra in Figure B-2. In this figure, the solid line is the time-smoothed spectrum of frame N (shown in Figure B-2 as a dashed line) and the dashed line is the peak-enhanced spectrum (shown in Figure B-2 as the lower solid line). F1 is estimated as the centroid of the time-smoothed spectrum within the frequencies of the F1 range (shown in gray). Outside of this region, the peak enhanced spectrum is used: F2, F3, and F4 are estimated as its three highest peaks. The vertical lines terminated by circles indicate the frequencies and relative amplitudes of the formant estimates. (Note again: for display purposes, the spectra and estimated formant amplitudes have been offset from each other by scaling. In the actual signal processing, all are normalized to have the same average amplitude.)

B.4 Formant Alteration

To implement the feedback transformation, the frequencies of F1, F2, and F3 were shifted via a lookup table. Use of this table and its construction are described in detail in Chapter 3.

B.5 Synthesis

Synthesizing a new frame of output speech from the altered formants was based on an approach called *formant synthesis*, which is a method used in many current speech synthesis systems [Klatt, 1980, O'Shaughnessy, 1987]. This approach is in turn based on the *source-filter theory* of speech production [Fant, 1960, Flanagan, 1972, O'Shaughnessy, 1987, Titze, 1994] that supposes the vocal tract can be modeled as a linear time-varying filter, characterized by an impulse response function. Speech is then modeled as the convolution of this impulse response with the glottal source function.

The synthesis process based on this theory consisted of three steps:

(1) Creating a vocal tract impulse response function corresponding to the formant estimates.

(2) Generating a whispered pitch glottal source function.

(3) Simulating the response of a vocal tract with an impulse response calculated in (1) to the pitch function generated in (2).

B.5.1 Creating the Vocal Tract Impulse Response

The vocal tract's impulse response function was calculated by assuming it could be modeled as the parallel combination of four resonances, each corresponding to a different formant.

Each resonance was modeled as a second-order linear filter with a damped-sinusoid impulse response. All resonances were assumed to have the same fixed damping factor controlling their impulse response decay rate.

The resonant frequencies of these filters were specified by the estimated formant frequencies. The filter outputs were weighted by the estimated formant amplitudes. In this way, the complete vocal tract impulse response was calculated as the weighted sum of the impulse responses corresponding to each formant.

Illustrations of these weighted formant impulse responses are shown in Figure B-4. Their sum made the vocal tract impulse response function shown in Figure B-5(a).

B.5.2 Generating a Glottal Source Function

Because subjects were restricted to whispered speech, their glottal source functions were assumed to be random. Thus, a synthesized random source function could be generated which functioned as an adequate substitute for subjects' true glottal source function. This allowed for significant computational savings since no source analysis then needed.

The actual random source function used was a random series of impulses. These impulses all had the same magnitude but alternated in sign. The intervals between the impulses were random numbers generated from a uniform distribution between some minimum and maximum limit values. It was found that this source function generated synthesized whispered speech that was perceptually nearly indistinguishable from the input whispered speech.

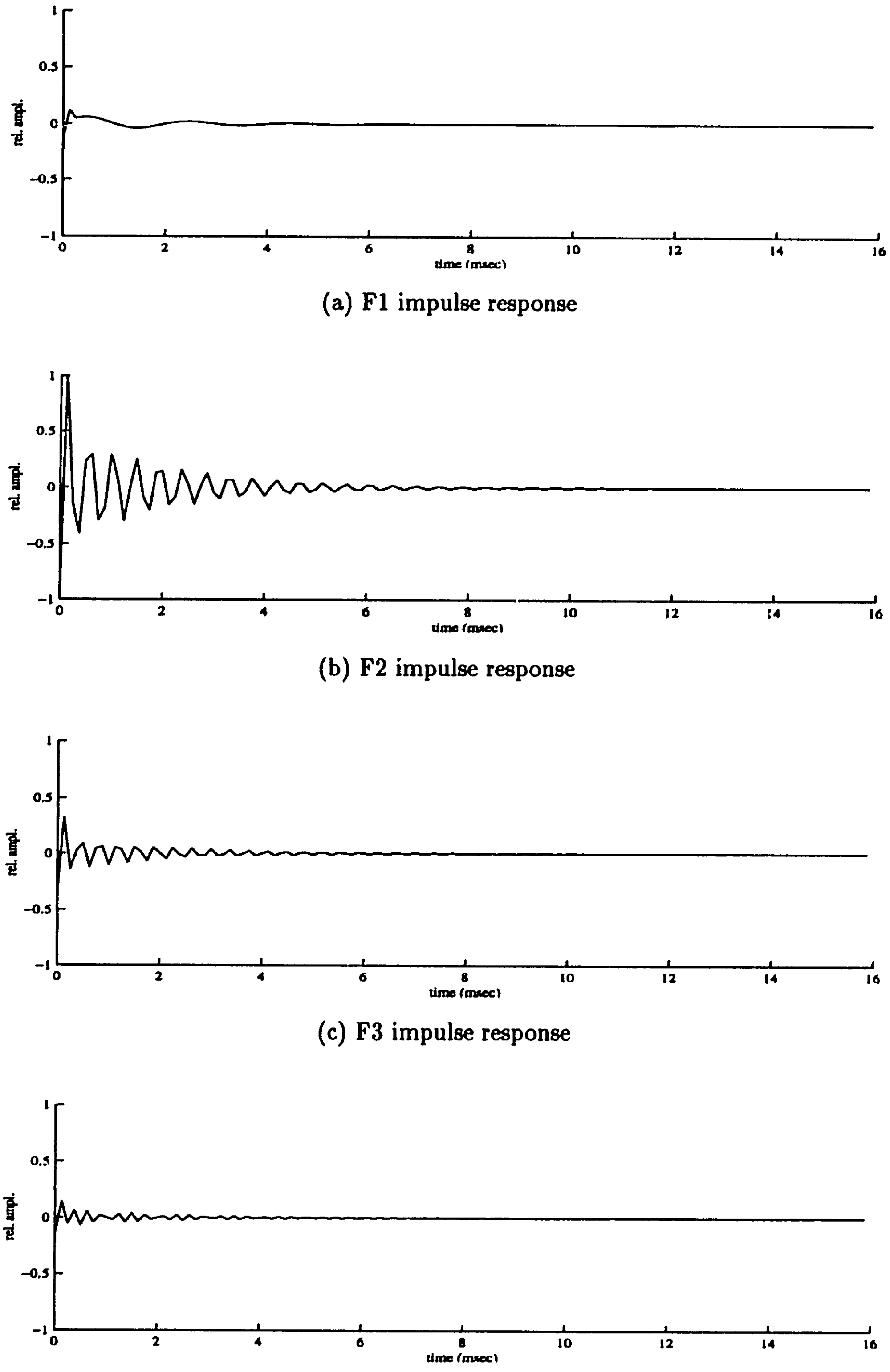
An example of this type of glottal source function is shown in Figure B-5(b).

B.5.3 Simulating the Response of the Vocal Tract: Generating Whispered Speech

At this point in the synthesis process, the impulse response of the vocal tract has been created, as has the glottal source function. All that remains is to determine the response of a vocal tract with this impulse response to this glottal source function. Since the vocal tract is assumed to be an linear, time-invariant system over the frame, computing this response involves merely convolving the vocal tract impulse response with the glottal source function.

The result of this final step is a frame of synthesized whispered speech corresponding to the peak representation derived from the subject's input speech. An example is shown in Figure B-5(c). This frame is synthesized in the current idle output buffer. With the completion of the synthesis process, this buffer is now ready to become the next active output buffer. When this happens, the buffer's contents are delivered to the subject as feedback of his whispered speech.

This completes the description of the signal processing used in the experimental apparatus.



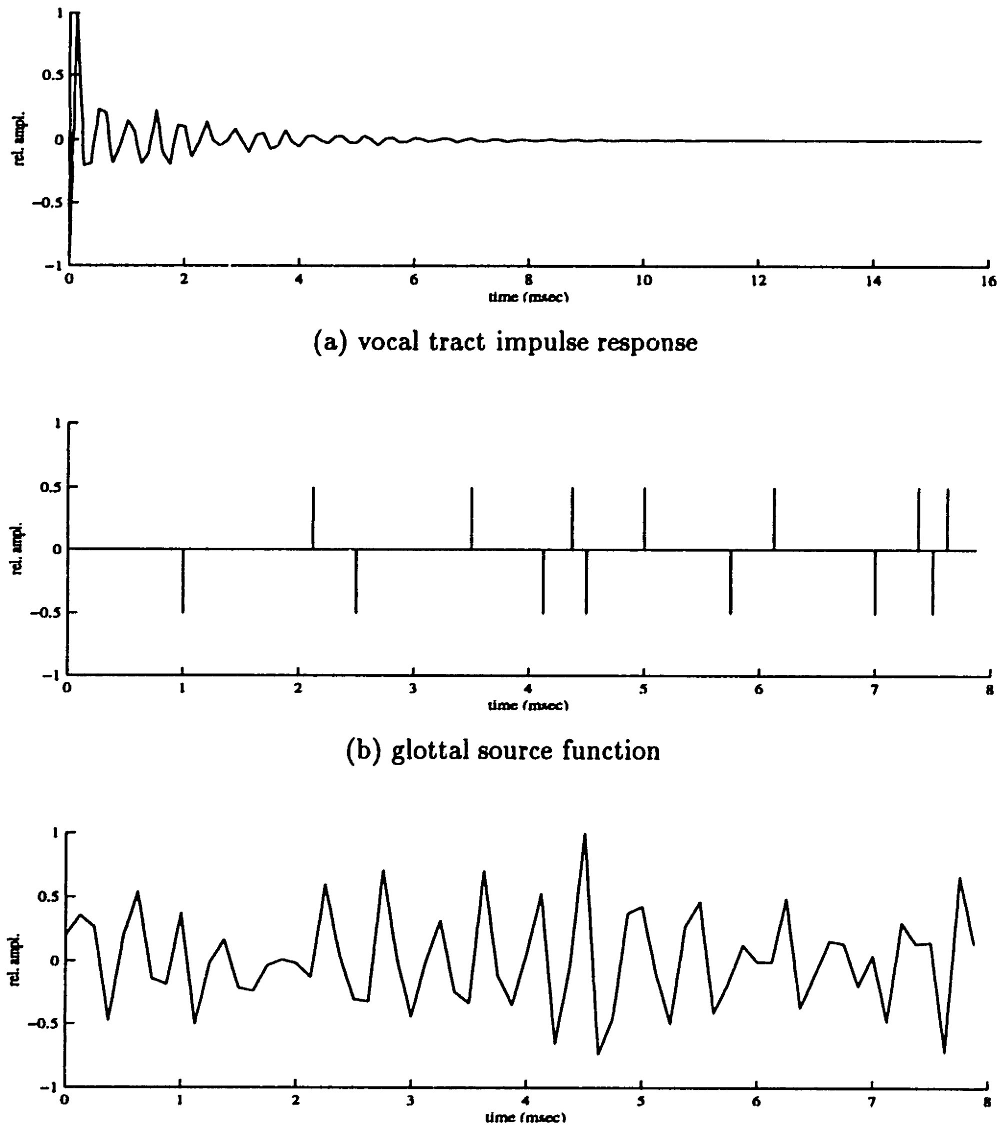
(a) F1 impulse response

(b) F2 impulse response

(c) F3 impulse response

(d) F4 impulse response

Figure B-4: The impulse response functions created from the formant estimates of Figure B-3. These response functions were then added together to make the complete vocal tract impulse response shown in Figure B-5(a).



(a) vocal tract impulse response

(b) glottal source function

(c) output speech

Figure B-5: Showing how the next frame of output synthesized speech was constructed from the convolution of the vocal tract impulse response and one frame of glottal source function. (a) shows the vocal tract impulse response function, which was created by summing the separate formant impulse response functions shown in Figure B-4. (b) shows the frame of glottal sourre function, which was calculated as a random sequence of alternating impulse functions. (c) shows the frame of output synthesized speech resulting from the convolution of the functions in (a) and (b). This frame was then output to the subject while the spectrum of the next input frame was being processed (see Figure B-1).

Appendix C

Resolution and Stability of Path Projections

To construct the feedback transformation tables, a line-segment definition of the [i]–[ɑ] path was used. In this definition, the [i]–[ɑ] path is formed from line segments joining the the path reference points.[1] Subsequent to running the experiments, however, several problems were discovered with using this path definition to calculate path projections.

C.1 Cubic-Spline [i]–[ɑ] Path Definitions

These problems were mitigated by changing to a cubic-spline [i]–[ɑ] path definition. In this definition, the [i]–[ɑ] path is formed by fitting a 3rd-order spline curve to the path reference points. Figure C-1 shows line-segment and cubic-spline [i]–[ɑ] paths made from the same path reference points. As the figure shows, both paths pass through the path reference points. The main difference between the two paths is smoothness: the line-segment path makes abrupt direction changes **at** path reference points, while the cubic spline path makes smooth direction changes **between** the path reference points.

[1] Path reference points are the path vowels plus the endpoint extensions – see Section 3.3.1.

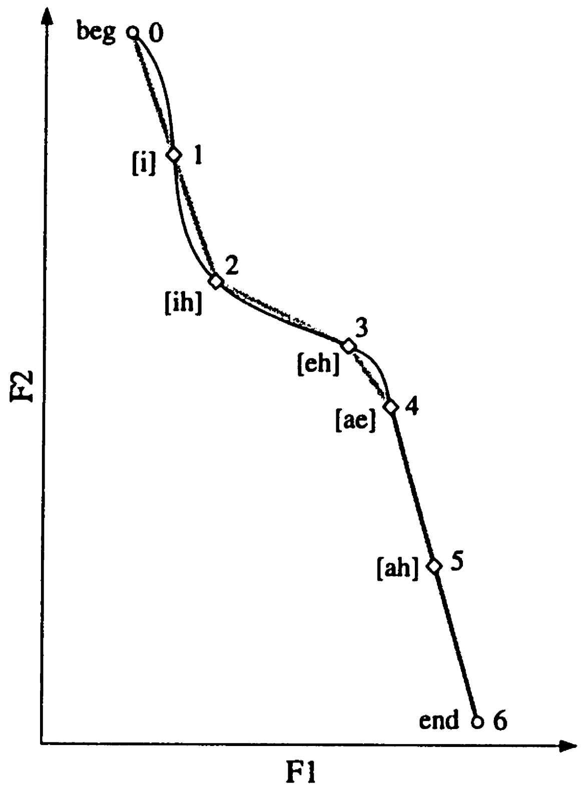


Figure C-1: Comparison of line-segment and cubic-spline [i]–[ɑ] paths made from the same path reference points. In this figure, the line-segment path is the thick gray line while the cubic-spline path is the thin dark line.

C.2 Resolution and Stability Issues

The problems mitigated by using the cubic-spline path concerned the resolution and stability of path projections. The problems are explained in more detail in figures C-2 and C-2. The figures show the problems are caused by the line-segment path's abrupt direction changes at path reference points. Because of this, the problems worsen with sharper direction changes at path reference points. The problems are also worse for vowel sounds more distant from the [i]–[ɑ] path.

Consider, now, the impact of these problems on quality of the feedback transformations and on data analysis.

C.3 Impact on the Feedback Transformations

Recall that the feedback transformations were defined as path projections shifts. They shift perceived formants of a vowel sound by shifting its path projection without altering its path deviation. Maintaining path deviations insures that all formant shifts follow the contour of the [i]–[ɑ] path. In this sense, all vowel sounds are shifted are in the same direction.

Thus, bad path deviation estimates would affect vowel sound shift directions, whereas bad path projection estimates would affect shift magnitudes. Therefore, the line-segment path projection problems discussed above only affected formant shift magnitudes, not their directions. The resolution problem would affect continuity of the shift magnitudes, while the stability problem would affect consistency of the shift magnitudes.

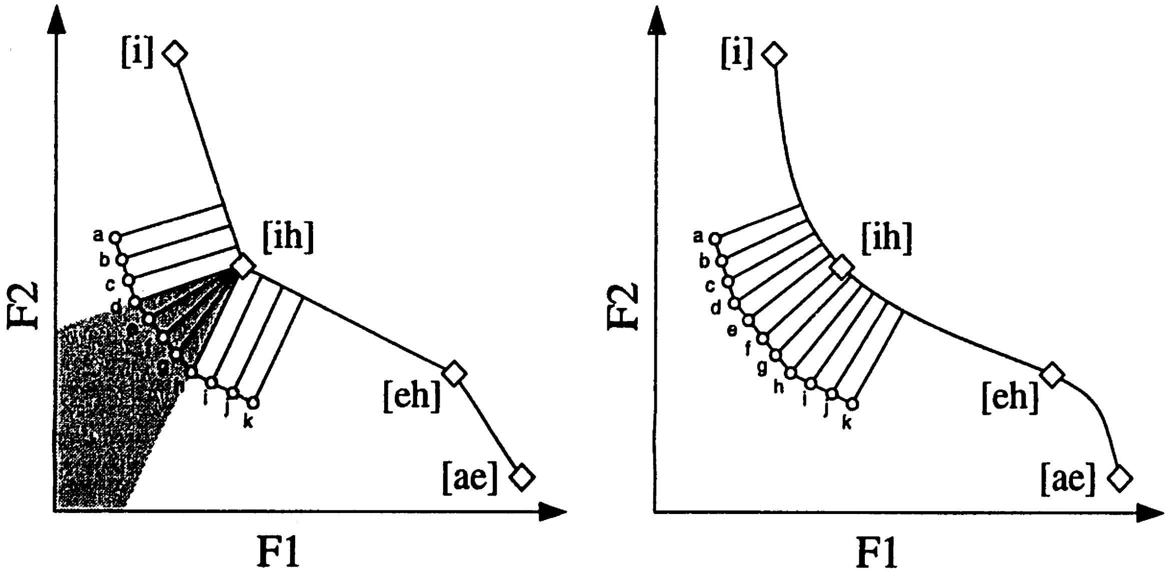
The path projection problems would be most severe for subjects whose [i]–[ɑ] paths exhibit sharp direction changes near close-together path reference points. These subjects may therefore have experienced discontinuous and inconsistent shifts of some of their vowel sounds. However, any subjects whose transformed vowels sounded noticeably bad were screened out in the subject pretest.

Thus, via subject screening, the impact of the line-segment path projection problems on feedback transformations was minimized. For this reason, plots of accepted subjects' feedback transformations (such as Figure 3-10) looked generally consistent in magnitude and direction. And, as the experiment data show (chapters 4 and 5), these transformations were consistent enough to cause subjects to compensate.

C.4 Impact on Data Analysis

Although the resolution and stability problems had only limited overall effects on the feedback transformations, they introduced avoidable inaccuracies in the data analysis.

For this reason, all path projection and deviation analysis of utterance data was done using the cubic-spline [i]–[ɑ] path definitions.



(a) resolution, line-segment path

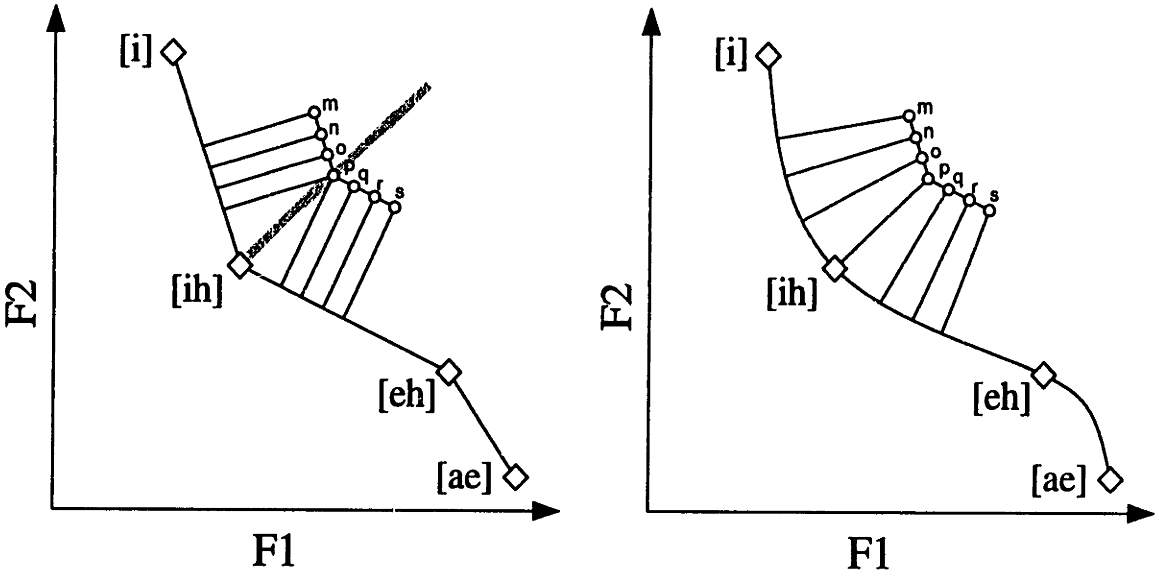
(b) resolution, cubic-spline path

Figure C-2: Resolution of path projections. (a) and (b) show different [i]–[ɑ] path definitions based on the same path reference points (shown as diamonds). In both subfigures, lines show path projections of vowel sounds a thru k.

(a) shows path projections based on the line-segment path. a-b-c-d and h-i-j-k are parallel to the [i]-[ι] and [ι]-[ε] path segments, respectively. Because of this, a, b, c, i, j, and k all have different path projections. However, all vowel sounds in the gray sector (e.g., sounds d, e, f, g, and h) have [ι] as their closest path point, so they all have the same path projection. (The sector's size is determined by the angle between the [i]-[ι] and [ι]-[ε] path segments: the sharper the angle, the larger the sector.) Now consider vowel sound f. Perturbations of f's formants to d or h are still in the sector and have the same path projection. (Note however, if f were closer to [ι], the same size perturbations of f would be outside of the sector. These f perturbations would have different path projections.)

Path projection insensitivity to formant variations in the gray sector is called the resolution problem. As shown above, the problem is worse at greater distances from the path. Sharper path segment angles also make it worse.

(b) shows path projections based on the cubic-spline path. Here, because there is no angle between path segments, there is no sector of vowel sounds with the same path projections. Thus, in this case, vowel sounds a thru k all have different path projections.



(a) stability, line-segment path

(b) stability, cubic-spline path

Figure C-3: Stability of path projections. (a) and (b) show the same [i]–[ɑ] path definitions seen in Figure C-2. In both subfigures, lines show path projections of vowel sounds m thru s.

(a) shows path projections based on the line-segment path. m-n-o-p and p-q-r-s are parallel to the [i]-[ι] and [ι]-[ε] path segments, respectively. m thru s all have the same path deviation. Thus p is on the bisector (shown in gray) of the angle between the [i]-[ι] and [ι]-[ε] path segments.

The two lines emanating from p illustrate the stability problem: p is equidistant from either path segment, so it is ambiguous which segment p projects to. This projection ambiguity amplifies any variation in p's formants: p will project to one segment or the other, depending on which it is ınfinitesimally closer to. Thus, small variations in p's formants make larger variations in its path projection. All points on the bisector of an acute path segment angle have this problem: their projection ambiguities amplify formant variations. The ambiguity gets worse for points more distant from the path. It also gets worse for smaller path segment angles.

(b) shows path projections based on the cubic-spline path. Here, there is no projection ambiguity: vowel sound p projects to only to [ι].

Appendix D

Equipment

Section 3.1 discussed the apparatus used in the experiments. Here, more detailed specifications of the equipment comprising this apparatus are described.

D.1 Detailed Apparatus Description

Figure D-1 shows a diagram of the equipment used in the experimental setup. In this diagram, the main pathway involved in processing the subject's feedback is highlighted.

The subject whispered into a head-mounted, noise-cancellation microphone that was connected to a mixer that functioned as a pre-amplifier. The auxòutput of the mixer was fed to both the DSP system's input A and input 1 of a four-track tape deck. The DSP system transformed the speech signal; the tape deck was used to record both the subject's speech and his feedback. The DSP system produced two outputs:

- Output A of the DSP system was the feedback signal sent to the subject. This signal was either pure noise to block his hearing, or a mixture of mild noise and transformed feedback of his whispering. This signal when to input 2 of the tape deck as well as to the Tuner L input of the Amplifier. The Phones output of the Amplifier when to the insert earphones that actually delivered the transformed feedback to the subject.

- Output B of the DSP system was a copy of the subject's transformed feedback without the noise mixed in. This signal when to input 3 of the tape deck.

In this way, the subject's actual whispered speech, what he heard as feedback, and the transformed version of his speech could be simultaneously recorded and monitored throughout an experiment.

The DSP system resided on a board installed in a PC computer. Words for the subject to whisper were presented visually on the PC's video monitor, and the PC's mouse was used by the subject to control the pace of the experiment.

D.2 Equipment Specifications

The specifications of the equipment used in the experimental apparatus were the following:

Microphone:	*Shure* model SM10A (noise cancellation, headmounted). *Shure Brothers, Inc.*
Mixer:	*Shure* model M268. *Shure Brothers, Inc.*
DSP system:	*Ariel* DSP-96. *Ariel Corp.*
PC:	*Alcom* 486DX/33 Mhz VESA Local Bus System. *Alcom Technology Corp.*
Tape Deck:	*TEAC* model A-3440 (four-track, reel-to-reel). *Teac Corp.*
Amplifier:	*Realistic* SA-150 integrated stereo amplifier. *Radio Shack Div., Tandy Corp.*
Earphones:	*E-A-RTone* 3A insert earphones. *E-A-R Auditory Systems Div., Cabot Safety Corp.*

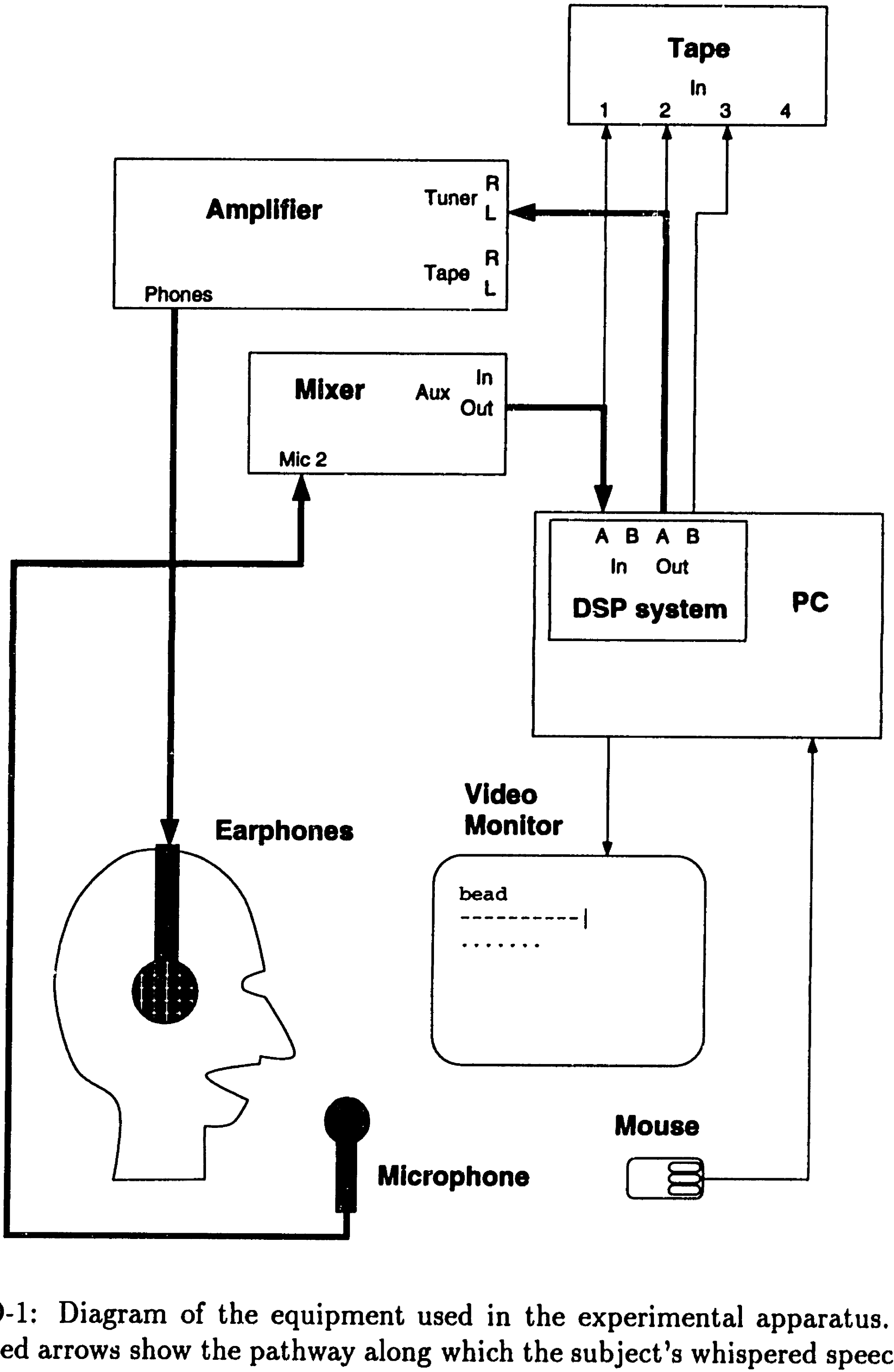


Figure D-1: Diagram of the equipment used in the experimental apparatus. The highlighted arrows show the pathway along which the subject's whispered speech was intercepted, processed, and fed back to him.

Bibliography

[Békésy, 1949] Békésy, G. V. (1949). The structure of the middle ear and the hearing of one's own voice by bone conduction. *Journal of the Acoustical Society of America*, 21(3):217–232.

[Borden et al., 1994] Borden, G. J., Harris, K. S., and Raphael, L. J. (1994). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*, chapter 5: Speech Production II: The Finished Products – The Articulation and Acoustics of Speech Sounds, pages 90–173. Williams and Wilkins, Baltimore, MD, 3rd edition.

[Cooley and Tukey, 1965] Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Math. Computation*, 19:297–301.

[Cooper, 1979] Cooper, W. E. (1979). *Speech Perception and Production: Studies in Selective Adaptation.* Language and Being. Ablex Publishing Corp., Norwood, NJ.

[Cowie and Douglas-Cowie, 1983] Cowie, R. and Douglas-Cowie, E. (1983). Speech production in profound post-lingual deafness. In Lutman, M. and Haggard, M., editors, *Hearing Science and Hearing Disorders*, pages 183–231. Academic Press, New York.

[Fairbanks, 1954] Fairbanks, G. (1954). Systematic research in experimental phonetics: 1. a theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders*, 19:133–139.

[Fant and Ishizaka, 1972] Fant and Ishizaka (1972). Stl qpsr 1. Technical report, KTH.

[Fant, 1960] Fant, G. (1960). *Acoustic Theory of Speech Production.* Mouton and Co., 's-Gravenhage.

[Flanagan, 1972] Flanagan, J. L. (1972). *Speech Analysis Synthesis and Peception*, volume 3 of *Kommunikation und Kybernetik in Einzeldarstellungen.* Springer-Verlag, New York, NY, 2 edition.

[Halle, 1990] Halle, M. (1990). Phonology. In Osherson, D. N. and Lasnik, H., editors, *Language*, volume 1 of *An Invitation to Cognitive Science*, chapter 3, pages 43–68. MIT Press, Cambridge, MA, 1 edition.

[Hein and Held, 1962] Hein, A. V. and Held, R. (1962). A neural model for labile sensorimotor coordination. In Bernard, E. and Hare, M., editors, *Biological Prototypes and Synthetic Systems*, volume 1. Plenum Press, New York.

[Held, 1996] Held, R. (1996). Personal Correspondence.

[Held and Durlach, 1991] Held, R. and Durlach, N. (1991). Telepresence, time delay and adaptation. In Ellis, S. R., editor, *Pictorial communication in virtual and real environments*, chapter 14, pages 232–246. Taylor and Francis Ltd., London, UK.

[Held and Gottlieb, 1958] Held, R. and Gottlieb, N. (1958). Technique for studying adaptation to disarranged hand-eye coordination. *Perceptual and Motor Skills*, 8:83–86.

[Houde, 1994] Houde, R. A. (1994). Personal Correspondence.

[Howard, 1968] Howard, I. P. (1968). Displacing the optical array. In Freedman, S., editor, *The Neuropsychology of Spatially Oriented Behavior.* Dorsey Press, Homewood, IL.

[Jordan and Rumelhart, 1992] Jordan, M. I. and Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16(3):307–354.

[Kawahara, 1993] Kawahara, H. (1993). Transformed auditory feedback: Effects of fundamental frequency perturbation. *Journal of the Acoustical Society of America*, 94(3 Part 2):1883. Proceedings of the 126th Meeting, Denver, CO. Oct 4-8, 1993.

[Klatt, 1980] Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67:971–995.

[Klatt, 1982] Klatt, D. H. (1982). Prediction of perceived phonetic distance frcm critical-band spectra: A first step. In *Proceedings of ICASSP-82: The IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1278–1281, Piscataway, NJ. IEEE, IEEE.

[Kornheiser, 1976] Kornheiser, A. S. (1976). Adaptation to laterally displaced vision: A review. *Psychological Bulletin*, 83(5):783–816.

[Kuhl, 1991] Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50:93–107.

[Ladefoged, 1982] Ladefoged, P. (1982). *A Course in Phonetics.* Harcourt-Brace Jovanovich, San Diego, CA., 2nd edition.

[Lane and Tranel, 1971] Lane, H. and Tranel, B. (1971). The lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14:677–709.

[Lane and Webster, 1991] Lane, H. and Webster, J. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America*, 89:859–856.

[Lee, 1950] Lee, B. S. (1950). Some effects of side-tone delay. *Journal of the Acoustical Society of America*, 22(5):639–640.

[Levelt, 1989] Levelt, W. J. (1989). *Speaking: From Intention to Articulation*, chapter 9: Generating Phonetic Plans for Words, pages 318–363. ACL-MIT Press Series in Natural-Language Processing. MIT Press, Cambridge, MA.

[Lombard, 1911] Lombard, E. (1911). Le signe de lelevation de la voix. *Ann. maladies oreille larynx nez pharynx*, 37:101–119.

[Meyer, 1991] Meyer, A. S. (1991). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. In Levelt, W. J., editor, *Lexical Access in Speech Production*, chapter 6, pages 181–211. Blackwell Publishers, Cambridge, MA.

[Oppenheim and Schafer, 1975] Oppenheim, A. V. and Schafer, R. W. (1975). *Digital Signal Processing*, chapter 6: Computation of the Discrete Fourier Transform, pages 581–661. Prentice-Hall Signal Processing Series. Prentice-Hall, Englewood Cliffs, NJ.

[Oppenheim and Willsky, 1983] Oppenheim, A. V. and Willsky, A. S. (1983). *Signals and Systems*, chapter 8: Sampling, pages 513–572. Prentice-Hall Signal Processing Series. Prentice-Hall, Englewood Cliffs, NJ.

[O'Shaughnessy, 1987] O'Shaughnessy, D. (1987). *Speech Communication: Human and Machine.* Addison-Wesley Series in Electrical Engineering: Digital Signal Processing. Addison-Wesley, Reading, MA.

[Perkell, 1996] Perkell, J. S. (1996). Articulatory processes. In *The Handbook of Phonetic Sciences*, pages 333–370. Blackwell, London.

[Peterson and Barney, 1952] Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24:175–184.

[Peterson and Lehiste, 1960] Peterson, G. E. and Lehiste, I. (1960). Duration of syllable nuclei in english. *Journal of the Acoustical Society of America*, 32:693–703.

[Rabiner and Levinson, 1981] Rabiner, L. R. and Levinson, S. E. (1981). Isolated and connected word recognition - theory and selected applications. *IEEE Transactions on Communications*, COM-29(5):621–659.

[Schacter, 1995] Schacter, D. L. (1995). Implicit memory: A new frontier for cognitive neuroscience. In Gazzaniga, M. S., editor, *The Cognitive Neurosciences*, chapter 52, pages 815–824. MIT Press, Cambridge, MA.

[Schwartz, 1970] Schwartz, M. F. (1970). Power spectral density measurements of oral and whispered speech. *Journal of Speech and Hearing Research*, 13:438–448.

[Stevens, 1989] Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17:3–45.

[Stevens, 1996] Stevens, K. N. (in preparation as of 1996). *Acoustic Phonetics*, chapter 3, section 3.6.4.

[Titze, 1994] Titze, I. R. (1994). *Principles of Voice Production.* Prentice-Hall, Inc., Englewood Cliffs, NJ.

[Welch, 1972] Welch, R. B. (1972). The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception and Psychophysics*, 12(6):453–456.

[Welch, 1978] Welch, R. B. (1978). *Perceptual Modification: Adapting to Altered Sensory Environments.* Academic Press Series in Cognition and Perception. Academic Press, New York.

[Welch, 1986] Welch, R. B. (1986). Adaptation of space perception. In Boff, K. R., Kaufman, L., and Thomas, J. P., editors, *Handbook of Perception and Human Performance*, volume 1: Sensory Processes and Perception, chapter 24, pages 24–1 – 24–45. John Wiley and Sons, New York, NY.

[Yates, 1963] Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60(3):213–232.