Title
The role of feedback in speech motor learning: insights from healthy speakers and applications to the treatment of apraxia of speech

Permalink
https://escholarship.org/uc/item/2j35v8vg

Author
Austermann-Hula, Shannon Noelle

Publication Date
2008

Peer reviewed|Thesis/dissertation
The Role of Feedback in Speech Motor Learning: Insights from Healthy Speakers and Applications to the Treatment of Apraxia of Speech

A Dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Language and Communicative Disorders

by

Shannon Noelle Austermann Hula

Committee in charge:

Professor Jeff Elman, Chair, University of California, San Diego
Professor Donald Robin, Co-Chair, University of Texas Health Sciences Center, San Antonio
Professor Jessica A. Barlow, San Diego State University
Professor Michael Caligiuri, University of California, San Diego
Professor Victor Ferreira, University of California, San Diego
Professor Malcolm McNeil, University of Pittsburgh
Professor Gabriele Wulf, University of Nevada, Las Vegas

2008
The Dissertation of Shannon Noelle Austermann Hula is approved, and it is acceptable in quality and form for publication on microfilm:

Co-Chair

Chair

University of California, San Diego
San Diego State University
2008
For Diane, Joseph, Janine,
Evan and Matt
Knowing what
Thou knowest not
Is in a sense
Omniscience

-Piet Hein
# TABLE OF CONTENTS

Signature Page .................................................................................................................... iii

Dedication ........................................................................................................................ iv

Epigraph ............................................................................................................................ v

Table of Contents ................................................................................................................. vi

List of Abbreviations ......................................................................................................... xi

List of Figures .................................................................................................................... xiii

List of Tables ...................................................................................................................... xvii

Acknowledgements .......................................................................................................... xviii

Vita ..................................................................................................................................... xxi

Abstract ............................................................................................................................ xxvii

1.0. Introduction ................................................................................................................... 1

1.1. Overview ....................................................................................................................... 5

2.0. Motor Control and Learning ....................................................................................... 7

2.1. Schema Theory ........................................................................................................... 8

2.2. Two-Stage Model of Motor Programming ................................................................ 10

2.3 Motor Learning ............................................................................................................. 13

2.4. Practice Variables ..................................................................................................... 16

2.5. Feedback Variables .................................................................................................. 23

3.0. Speech Production ..................................................................................................... 33

3.1. Psycholinguistic Models ............................................................................................ 34

3.2. Four-Level Framework ............................................................................................. 41
3.3. DIVA ............................................................................................................ 42

4.0. Apraxia of Speech ........................................................................................ 48
   4.1. Defining AOS ............................................................................................ 48
   4.2. Behavioral Definition ................................................................................ 50
       4.2.1. Speech Behaviors ............................................................................ 50
       4.2.2. Nonspeech behaviors ...................................................................... 53
   4.3. Neuroanatomic Substrates ....................................................................... 56
   4.4. Cognitive Definition .................................................................................. 60
       4.4.1. Psycholinguistic Theories ................................................................. 60
       4.4.2. Dual Route Theory .......................................................................... 62
       4.4.3. Reduced Buffer Capacity Hypothesis ............................................. 63
       4.4.4. Four-Level Framework .................................................................... 65
       4.4.5. Coalitional Breakdown ..................................................................... 66
       4.4.6. Schema Theory ............................................................................... 67
       4.4.7. Two-Stage Model ........................................................................... 69
   4.5. Application of Motor Learning Concepts to Treatment of AOS .............. 73

5.0. Effects of Feedback Frequency and Timing on Acquisition, Retention, and Transfer of Speech Skills in Acquired Apraxia of Speech .................................................. 77
   5.0.1. Theoretical Framework ....................................................................... 77
   5.0.2. Application of Principles of Motor Learning to Treatment for Apraxia of Speech ............................................................................................................. 83
   5.1. Experiment 1: High versus Low Frequency Feedback .............................. 85
7.1.3. Tasks and Procedures ................................................................. 234
7.1.4. Design and Analysis ................................................................. 239
7.2. Results ......................................................................................... 243
  7.2.1 AOS1 .................................................................................... 243
  7.2.2 AOS2 .................................................................................... 257
  7.2.3 AOS3 .................................................................................... 271
  7.2.4 AOS4 .................................................................................... 285
7.3. Discussion .................................................................................. 298
  7.3.1. Feedback Requests ................................................................. 300
  7.3.2. Response Execution ............................................................... 303
  7.3.3. Motor Programming ............................................................... 310
  7.3.4. Conclusions ......................................................................... 315
Appendix B ....................................................................................... 345
8.0. General Discussion and Conclusions ............................................ 319
References ....................................................................................... 332
LIST OF ABBREVIATIONS

ABA-2 = Apraxia Battery for Adults-2 (Dabul, 2000)

ABCD = Arizona Battery for Communication Disorders in Dementia (Bayles & Tomoeda, 1993)

AE = absolute error

ANCDS = Academy of Neurologic Communication Disorders and Sciences

AOS = Apraxia of Speech

ATD = alternating treatments design

BDAE = Boston Diagnostic Aphasia Examination (Goodglass, Kaplan, & Barresi, 2001)

DF = delayed feedback

DIVA = Directions Into Velocities of Articulators

E = total variability, also root mean square error

EMG = electromyography

fMRI = functional magnetic resonance imaging

GMP = generalized motor program (Schmidt, 1975)

HFF = high frequency feedback

IF = immediate feedback

INT = response preparation process in motor programming (Klapp, 1995)

KP = knowledge of performance

KR = knowledge of results

LCA = learner-controlled after
LCA+E = learner-controlled after with self-evaluation

LCB = learner-controlled before

LCB+E = learner-controlled before with self-evaluation

LCF = learner-controlled feedback

LFF = low frequency feedback

MCA = middle cerebral artery

NVOA = non-verbal oral apraxia

PPT = phonetic placement therapy (Van Riper & Irwin, 1958)

PROMPT = Prompts for Restructuring Oral Muscular Phonetic Targets (Square et al., 1985)

RE = relative error

RMS = root mean square

RT = reaction time

SEQ = sequencing process in motor programming (Klapp, 1995)

SPT = sound production treatment (Wambaugh, West, & Doyle, 1998)

ST = study time

VMT = visuomotor tracking

WAB = Western Aphasia Battery (Kertesz, 1982)

WAB AQ = Western Aphasia Battery aphasia quotient

WAB CQ = Western Aphasia Battery cognitive quotient

YF = yoked feedback
LIST OF FIGURES

Figure 5-1. P1 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 1. ......................................................... 99

Figure 5-2. P2 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 1. ..................................................... 103

Figure 5-3. P3 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 1. ..................................................... 106

Figure 5-4. P4 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 1. ..................................................... 111

Figure 5-5. P1 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 2. ..................................................... 120

Figure 5-6. P2 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 2 ..................................................... 122

Figure 6-1. Events on each practice trial in the self-selection paradigm for each group in Experiment 3…………………………………………………………. 149

Figure 6-2. Mean accuracy for each group during practice, retention, and transfer blocks in Experiment 3……………………………………………………….. 160

Figure 6-3. Study Time for groups across practice, retention, and transfer phases of Experiment 3…………………………………………………………. 163

Figure 6-4. Reaction Time for groups across acquisition, retention, and transfer phases of Experiment 3 .............................. 166

Figure 6-5. Absolute Error for groups across acquisition, retention, and transfer phases of Experiment 3 .................................................................. 169

Figure 6-6. Relative Error for groups across acquisition, retention, and transfer phases of Experiment 3 .................................................................. 171

Figure 6-7. Study Time for groups and response types across practice and retention phases of Experiment 3.......................................................... 174

Figure 6-8. Reaction Time for groups and response types across acquisition and retention phases of Experiment 3................................. 176
Figure 6-9. Absolute Error for groups and response types across acquisition and retention phases of Experiment 3 .................................................. 178

Figure 6-10. Relative Error for groups and response types across acquisition and retention phases in Experiment 3 .................................................. 179

Figure 6-11. Study Time for groups and response types in the four transfer probes in Experiment 3 ................................................................. 183

Figure 6-12. Reaction Time for groups and response types across transfer probes in Experiment 3 ................................................................. 187

Figure 6-13. Absolute Error for groups and response types across transfer probes in Experiment 3 ................................................................. 190

Figure 6-14. Relative Error for groups and response types across transfer probes in Experiment 3 ................................................................. 192

Figure 6-15. Stimuli target durations and participant total response durations across three blocks of practice for each group in Experiment 3 ............... 201

Figure 7-1. Events during each practice trial in the self-selection paradigm for both conditions of Experiment 4 ......................................................... 237

Figure 7-2. Stimuli target durations and actual response durations for AOS1 for each practice block during acquisition in Experiment 4 ................. 246

Figure 7-3. Stimuli target durations and actual response durations for AOS1 for random transfer probes during each probe session in Experiment 4 .......... 249

Figure 7-4. Participant AOS1 average absolute error (AE) for responses during treatment sessions and retention and transfer probes in Experiment 4 ...... 250

Figure 7-5. Participant AOS1 average relative error (RE) for responses during treatment sessions and retention and transfer probes in Experiment 4 ...... 251

Figure 7-6. Participant AOS1 study time and reaction time for treated responses elicited as random-order transfer probes during baseline, treatment, and retention phases in Experiment 4 ................................................. 256

Figure 7-7. Stimuli target durations and actual response durations for AOS2 for each practice block during acquisition in Experiment 4 ................. 260
Figure 7-8. Stimuli target durations and actual response durations for AOS2 for random transfer probes during each probe session in Experiment 4 ..........264

Figure 7-9. Participant AOS2 average absolute error (AE) for responses during treatment sessions and retention and transfer probes in Experiment 4 ........265

Figure 7-10. Participant AOS2 average relative error (RE) for responses during treatment sessions and retention and transfer probes in Experiment 4 ........266

Figure 7-11. Participant AOS2 study time and reaction time for treated responses elicited as random-order probes during baseline and each treatment phase in Experiment 4 ..........................................................270

Figure 7-12. Stimuli target durations and actual response durations for AOS3 for each practice block during acquisition in Experiment 4 .........................274

Figure 7-13. Stimuli target durations and actual response durations for AOS3 for each probe session in Experiment 4 ..............................................278

Figure 7-14. Participant AOS3 average relative error (RE) for responses during treatment sessions and retention and transfer probes in Experiment 4 ........279

Figure 7-15. Participant AOS3 average relative error (RE) for responses during treatment sessions and retention and transfer probes in Experiment 4 ........280

Figure 7-16. Participant AOS3 study time and reaction time for treated responses elicited as random-order probes during baseline, treatment, and retention phases in Experiment 4 ..........................................................286

Figure 7-17. Stimuli target durations and actual response durations for AOS4 for each practice block during acquisition in Experiment 4 .........................288

Figure 7-18. Stimuli target durations and actual response durations for AOS4 for each random transfer probe in Experiment 4 .................................291

Figure 7-19. Participant AOS4 average absolute error (AE) for responses during treatment sessions and retention and transfer probes in Experiment 4 ........292

Figure 7-20. Participant AOS4 average relative error (RE) for responses during treatment sessions and retention and transfer probes in Experiment 4 ........293

Figure 7-21. Participant AOS4 study time and reaction time for treated responses elicited as random-order probes during baseline, treatment, and retention phases in Experiment 4 ..........................................................297
Figure B-1. Percent differences of ST, RT, AE, and RE between feedback and no-feedback trials for each participant in the Learner-Controlled Feedback (LCF) treatment condition and the Yoked Feedback (YF) treatment condition in Experiment 4………………………………………………………………317

Figure B-2. Study time and Reaction time standard deviations for random transfer probe “pa” (top panels) and “chee” (bottom panels) responses for each participant in Experiment 4……………………………………………………………..318
LIST OF TABLES

Table 5-1. Participant characteristics for Experiments 1 and 2. .............................. 88
Table 5-2. Design and orthographic stimuli for Experiment 1................................. 90
Table 5-3. Design and orthographic stimuli for Experiment 2............................... 117
Table 5-4. Summary of Results for Experiments 1 and 2. ................................. 125
Table A-1. Carrier phrases used in Experiment 1.................................................. 133
Table 6-1. Order of experimental tasks and orthographic target cues with associated temporal patterns for Experiment 3. ................................................................. 151
Table 6-2. Questionnaire questions and number of responses for each group in Experiment 3........................................................................................................... 154
Table 6-3. Overview of Experiment 3 predictions and results.................................. 195
Table 7-1. Participant information for Experiment 4............................................. 231
Table 7-2. Treatment stimulus and independent variable pairings for each participant in Experiment 4............................................................................................... 233
Table 7-3. Example target cues and associated patterns for Experiment 4.............. 234
Table 7-4. Four different effect size calculations used in Experiment 4, and the random transfer probe data used to compute them............................................. 261
Table 7-5. Overview of Experiment 4 response execution results for each participant in each experimental task................................................................. 309
Table 7-6. Overview of key Experiment 4 motor programming results for the random transfer probe for each participant...................................................... 310
ACKNOWLEDGEMENTS

Chapter 5 is in press with the *Journal of Speech, Language, and Hearing Research* as a manuscript entitled “Effects of Feedback Frequency and Timing on Acquisition, Retention, and Transfer of Speech Skills in Acquired Apraxia of Speech,” by Shannon N. Austermann Hula, Donald A. Robin, Edwin Maas, Kirrie J. Ballard, and Richard A. Schmidt. Permission for its use in this dissertation was granted by the publisher. The dissertation author is the primary investigator and author of the manuscript. The experiments presented in Chapters 6 and 7 were supported by a fellowship from the VA Associated Health Rehabilitation Research Program.

The completion of this dissertation would not have been possible without the help of a great many people. First and foremost, I would like to extend my deepest gratitude to my advisor, Don Robin. Eight years ago when I was a hapless clinical master’s program student thirsting for research experience, he volunteered to mentor me, and has made my education a top priority of his ever since. No matter what else he has had on his plate, Don has always made ample time for me and provided thoughtful discussion, stimulating challenges, steadfast encouragement, and unwavering support. I am honored to have had such a top-notch scientist as an advisor and mentor. I would also like to thank the rest of my committee-- Jeff Elman, Mike Caligiuri, Vic Ferreira, Jessica Barlow, Mick McNeil, and Gabriele Wulf-- for their flexibility, constructive input, responsiveness, and enthusiasm. It is an honor for me to have worked with such
an amazing group of talented scientists, and I value all of the advice and insights that they have shared with me.

I owe a debt of gratitude to Pat Doyle and Mick McNeil for giving me a home in their labs at the Pittsburgh VA. They have generously shared their resources with me and served as inspirational role models for careers in clinical research. Thank you also to my colleagues at the VA and in the Joint Doctoral Program, especially MaryBeth Ventura for her help with subject recruitment and Christine Matthews, Mike and Laura Dickey, and Josée Poirier for their stimulating collegiality and constant selfless support and friendship. I am also grateful for the assistance of Courtney Heckman, Katy Schmidt, Jennifer George, Mandy Case, Gabriella Perez, and Marie Winther, who spent countless hours helping with data collection and acoustic analyses. I am deeply appreciative of all of the participants who endured the challenges of these learning studies, especially the participants with AOS and their partners, who understood the value of research and each dedicated many months to these projects.

Edwin Maas has been an invaluable source of intellectual and personal support and guidance. I am indebted to him for his generous responsiveness to my seemingly neverending questions and for his helpful comments on Chapter 6 of this document. This work stems in large part from stimulating debates and conversations with Edwin, and I am grateful to him for leading the way and showing me how it’s done.
I am also deeply appreciative to my family for their encouragement, and in particular my parents and my Aunt Kerstin for supporting – personally and financially – my education. I would not be where I am today without them.

To Will, I am profoundly humbled by the depth of your patience, devotion, and thoughtfulness, as well as by the sacrifices that you have made to support me in this endeavor. I honestly could not have made it through this without you, and this accomplishment is as much yours as it is mine. We have survived two dissertations in the last year, and I anticipate that the coming year will be pleasantly boring in contrast!
VITA

Education

2008  Doctor of Philosophy, University of California, San Diego and San Diego State University (Language and Communicative Disorders)
2002  Master of Science, California State University, Hayward (Speech Pathology and Audiology)
1997  Bachelor of Arts, University of California, Berkeley (Molecular Cell Biology/Neurobiology and Philosophy)

Professional Experience

Present  Clinical Research Scientist, Veteran’s Research Foundation of Pittsburgh; Pittsburgh, PA. Supervisor: Patrick Doyle, Ph.D.
2006-2008 Research Fellow, VA Associated Health Rehabilitation Research Fellowship; Pittsburgh, PA. Supervisors: Patrick Doyle, Ph.D., and Malcolm McNeil, Ph.D.
2006  Research Assistant, Speech Motor Aphasia Cognition Laboratory, VA Pittsburgh Healthcare System; Pittsburgh, PA. Supervisor: Malcolm McNeil, Ph.D.
2005  Teaching Assistant, Design and Analysis of Experiments, University of California, San Diego; La Jolla, CA. Supervisor: Rafael Núñez, Ph.D.
2005  Teaching Assistant, Cognitive Development and Education, University of California, San Diego; La Jolla, CA. Supervisor: Gedeon Deák, Ph.D.
2004-2005 Research Assistant, Kutas Cognitive Electrophysiology Lab, University of California, San Diego; La Jolla, CA. Supervisor: Tom Urbach, Ph.D.
2004  Clinical Fellow in Speech-Language Pathology, VA Pittsburgh Healthcare System; Pittsburgh, PA. Supervisors: Thomas Marshall, Ph.D. and Joan Olsiewski, M.S.
2003-2004 Clinical Fellow in Speech-Language Pathology, San Diego State University Communications Clinic; San Diego, CA. Supervisors: Donald Robin, Ph.D., Amy Ramage, Ph.D., and Charlotte Lopes, M.S.
2002-2006 Research Assistant, Speech and Language Neuroscience Laboratory, San Diego State University; San Diego, CA. Supervisor: Donald Robin, Ph.D.
2002  Clinical Intern in Speech-Language Pathology, Stanford University Hospital; Stanford, CA. Supervisor: Deborah Kutch, M.S.
2002  Clinical Intern in Speech-Language Pathology, Marin Elementary School; Albany, CA. Supervisor: Marilyn Shumacher, M.S.
2000-2002 Research Assistant, Fluency Treatment Laboratory, California State University, Hayward; Hayward, CA. Supervisor: William Rosenthal, Ph.D.
1999-2002  Research Assistant, Aphasia Treatment Laboratory, California State University, Hayward; Hayward, CA. Supervisor: Jan Avent, Ph.D.
1996-1997  Research Assistant, Cognition and Action Laboratory, University of California, Berkeley; Berkeley CA. Supervisors: Richard Ivry, Ph.D. and Susan Ravizza, Ph.D.

Teaching Experience

Teaching Assistantships

Design and Analysis of Experiments (undergraduate)  2005
Department of Cognitive Science (UCSD)

Cognitive Development and Education (undergraduate)  2005
Teacher Education Program (UCSD)

Guest Lectures

“Treatment of the Respiratory System in Acquired Neuromotor Speech Disorders” (graduate)  2006
School of Speech, Language, and Hearing Sciences (SDSU)

“Single Subject Design” (graduate)  2006
School of Speech, Language, and Hearing Sciences (SDSU)

“Estimation and Confidence Intervals” (undergraduate)  2005
Department of Cognitive Science (UCSD)

“Language Development” (undergraduate)  2005
Teacher Education Program (UCSD)

“Neuromotor Disorders” (graduate)  2004
School of Speech, Language, and Hearing Sciences (SDSU)

“Cranial Nerve Examination” (graduate)  2004
School of Speech, Language, and Hearing Sciences (SDSU)

“Communication and Aging” (graduate)  2003
School of Speech, Language, and Hearing Sciences (SDSU)

“Aphasia” (undergraduate)  2002
Department of Psychology (SDSU)

“Attention” (undergraduate)  2002
Department of Psychology (SDSU)

“Right Hemisphere Dysfunction” (professional in-service)  2002
Stanford University Hospital
**Honors and Awards**

UCSD Department of Cognitive Science “Superior Teaching” Award, 2005
UCSD Dean of Social Science Travel Award to attend the 12th Biennial Conference on Motor Speech, 2004
Travel Fellowship to attend the NIDCD-supported Student Research Symposium at 34th Annual Clinical Aphasiology Conference, 2004
Travel Fellowship to attend the NIDCD-supported Student Research Symposium at 33rd Annual Clinical Aphasiology Conference, 2003
ASHA Research in Higher Education Mentorship Program Award, 2000

**Clinical Service Experience**

VA Pittsburgh Healthcare System, Clinical Fellow in Speech-Language Pathology 2004
San Diego State University, Clinical Fellow in Speech-Language Pathology 2003-2004
Stanford University Hospital, Clinical Intern in Speech-Language Pathology 2002
Marin Elementary School, Clinical Intern in Speech-Language Pathology 2002
California State University, Hayward Speech, Language, and Hearing Clinic, Graduate Student Clinician in Speech-Language Pathology 2001-2002

**Certification and Licensure**

Licensure in Speech-Language Pathology, Pennsylvania 2006-present
Certificate of Clinical Competence, American Speech, Language, and Hearing Association 2005-present
Licensure in Speech-Language Pathology, California 2004-present
*Porch Index of Communication Ability* Administrator Certification 2004

**Memberships**

California Speech, Language and Hearing Association 2001-present
American Speech, Language, and Hearing Association 1999-present

**Research Grants and Contracts**

VA Pre-Doctoral Associated Health Rehabilitation Research Fellowship 2006-2007
National Institutes of Health (NIDCD), Pre-Doctoral Training Grant 2005-2006
National Institutes of Health (NIDCD), Pre-Doctoral Training Grant 2002-2003
Publications

Book Chapters

Journal Articles (Refereed)


Conference Presentations
Functioning Item Pool. Paper accepted for presentation at the 2008 Clinical Aphasiology Conference, Jackson Hole, WY.


ABSTRACT OF THE DISSERTATION

The Role of Feedback in Speech Motor Learning: Insights from Healthy Speakers and Applications to the Treatment of Apraxia of Speech

by

Shannon Noelle Austermann Hula

Doctor of Philosophy in Language and Communicative Disorders

University of California, San Diego, 2008
San Diego State University, 2008

Professor Jeff Elman, Chair
Professor Donald A. Robin, Co-chair

Feedback is considered one of the most important variables in motor learning. The method of its application, however, can determine whether it exerts positive or
negative influences on different aspects of motor control and learning. Thus, a principled approach must govern its use, particularly in the treatment of motor impairments. This dissertation examined the effects of theory-driven feedback manipulations on treatment of the speech motor programming disorder apraxia of speech (AOS).

Motor learning research shows that delaying or reducing the frequency of feedback promotes retention and transfer of skills. In contrast, immediate or frequent feedback may temporarily enhance performance during acquisition, yet interfere with learning. The first two experiments in this dissertation tested these predictions in the context of a common treatment for AOS. The results provided evidence that reducing or delaying feedback provision promotes learning in some individuals with AOS, but also raised questions about how to determine which feedback manipulations provide optimal learning conditions for particular individuals. The general motor learning literature has shown that allowing learners to control their own feedback schedules facilitates learning, suggesting that it enables learners to receive feedback when they need it. Experiments 3 and 4 extended these notions to speech motor programming and learning in healthy speakers and those with AOS. Results from healthy speakers indicated that learner-controlled feedback manipulations affected different stages and aspects of motor programming (i.e., the translation of language units to speech motor commands), and that the benefits of learner-controlled feedback may be related to promotion of self-evaluation skills. Results from speakers with AOS provided
qualified support for the application of learner-controlled feedback to the treatment of this disorder and provide impetus for further investigation.

Overall, this research has demonstrated that individuals with AOS benefit from structured intervention, and that it is important to consider the influence of feedback variables when implementing AOS treatment. It has also demonstrated that the speech motor control system may share properties in common with general motor control, and that by extending general motor principles to the speech domain, we may increase our understanding of normal and disordered speech motor control and learning.
1.0. Introduction

Since Apraxia of Speech (AOS) was first described by Frederick Darley and his colleagues in the late 1960s (Darley, 1968 in Duffy, 2005), the nature of the disorder has been the subject of extensive debate, with far-reaching consequences in both theoretical and applied domains. As described by McNeil (2002:221),

Although clinically accepted, theoretical controversy over the term “apraxia”—its cognitive, linguistic, or neurological theoretical explanation; its diagnostic criteria; and its most effective treatment—has surrounded this category of motor speech disorder since its origin.

…With few exceptions,…interest has focused on various theoretical accounts for the disorder, and only a small handful of clinical researchers have continued to work tenaciously on addressing diagnostic and treatment issues.

The purpose of this dissertation is to examine the contributions that theories of motor control and learning may make to our understanding of the impairments and management of speech motor control disorders, namely, apraxia of speech. The experiments described herein comprise the foundation for a research program focused on improving the management of this disorder through the adoption of a theoretically-grounded motor control perspective. The long-term objectives of this research program are to (a) investigate the extent to which speech motor learning processes share properties in common with general motor learning and to (b) extend the application of a principled approach, grounded in motor learning theory, to the treatment of AOS. The work described in this dissertation focuses on the role of external feedback in speech motor learning.
The current state of clinical affairs in speech-language pathology is not optimistic with regard to AOS management. The prevailing sentiment amongst clinicians faced with treating persons with this disorder, particularly in the chronic phase, is that it is difficult to diagnose and very resistant to the various extant therapeutic approaches. Secondary to Rosenbek and colleagues’ early suggestion that treatments for AOS “should emphasize regaining points of articulation and sequencing,” (Rosenbek, Lemme, Ahern, Harris, & Wertz, 1973) most treatments that have been applied to this disorder have targeted improving the accuracy of these skills or facilitating sound production. These treatments have involved techniques such as articulatory placement and cueing (e.g., Bose, Square, Schlosser, & van Lieshout, 2001; Holtzapple & Marshall, 1977; Square, Chumpelik, & Adams, 1985; Wertz, LaPointe, & Rosenbek 1984), sound contrast practice (Wambaugh, Doyle, Kalinyak, & West, 1996; Wambaugh, Kalinyak-Fliszar, West, & Doyle, 1998; Wambaugh, Martinez, McNeil, & Rogers, 1999), prosodic training and singing (e.g., Keith & Aronson, 1975; Southwood, 1987), vibrotactile stimulation (e.g., Rubow, Rosenbek, Collins, & Longstreth, 1982), pacing and finger counting (Dworkin & Abkarian, 1996; Dworkin, Abkarian, & Johns, 1988; Simmons, 1978; Wambaugh & Martinez, 2000), biofeedback (e.g., Fossett et al., 2008; Howard & Varley, 1995; Katz, Bharadwau, & Carstens, 1999; McNeil et al, 2007; McNeil, Prescott, & Lemme, 1976), and rapid sequence repetition (e.g., Dabul & Bollier 1976). Yet, as Wambaugh (2002) points out, there are limited data (which include some from poorly-controlled case studies) available on the efficacy of these techniques in the management of AOS, and a
particular dearth of replication and extension data to support their use. Further, since most of these findings were based on only a few participants, and since the diagnostic criteria for AOS have evolved over the years (McNeil, Pratt, & Fossett, 2004; McNeil, Robin, & Schmidt, 1997), it is possible, and probably likely, that some of these limited treatment data were derived from investigations that relied on or included subjects whose impairments were due to another disorder, and not AOS, at least as it is currently defined.

As Wambaugh (2002) discusses in a recent summary of AOS treatment approaches, repeated investigations have only been attempted with four basic therapeutic methods: Rosenbek’s 8-step continuum (Rosenbek et al., 1973), metronomic pacing (e.g., Dworkin et al., 1988), Prompts for Restructuring Oral Muscular Phonetic Targets (PROMPT; e.g., Square et al. 1985), and Sound Production Treatment (SPT; e.g., Wambaugh, West, & Doyle, 1998). Although investigations of these therapies have generated promising preliminary results, efficacy has only been partially-established for two of these treatments (PROMPT and SPT). Evidence oriented toward implementation specifics of these treatments is still in demand.

Taking a cue from the limb motor control literature, however, researchers in speech motor learning have recently begun to assess the roles of different general practice variables, “principles of motor learning,” inherent in implementation of these treatments for motor speech disorders. These principles specify conditions of practice and augmented feedback that are thought to enhance motor skill learning. Many of these variables, such as order of practice, complexity of stimuli, type of feedback, and
schedule of feedback delivery, are inherent factors in any treatment protocol, although they have received relatively little attention in the speech domain until recently. It can be speculated that poor control and underspecification of some of these factors may be responsible for some of the equivocal or irreproducible results found in the AOS treatment literature. Indeed, current clinical practice in speech-language pathology often runs contradictory to some of these principles.

In addition to the obvious clinical significance of developing efficacious treatments for speech motor control impairments, the questions posed by the present research program also have important theoretical implications. For instance, there has been an ongoing debate in the literature regarding the uniqueness of motor control systems and processes for speech versus nonspeech functions (e.g., Ballard, Robin, & Folkins, 2003; Folkins et al., 1995; Robin, Solomon, Moon, & Folkins, 1997; Weismer, 2006; Weismer & Liss, 1991; Ziegler, 2002, 2003a, b). Moreover, this debate has highlighted the lack of consensus amongst researchers about the very definition of speech (e.g., Ballard et al., 2003) and the units that comprise it. Investigating AOS from a general motor control perspective has the potential to bring important evidence to bear on these fundamental issues. First, the degree to which predictions that are generated by general motor control models and supported by research in the extremities can be productively applied to the speech domain suggests how likely it is for speech and nonspeech tasks to be subserved by separate control systems. Second, observations of transfer between more and less putatively “speech-like” tasks (e.g., word production and syllable repetition; Austermann Hula, Robin,
Maas, Ballard, & Schmidt, in press) or between speech and nonspeech tasks (e.g., Shaiman, McNeil, Szuminsky, Meigh, & Botler, 2006) also provide evidence about the neuropsychological overlap between these control systems and the nature of the fundamental units upon which they act.

1.1. Overview

This dissertation takes the position that AOS is most appropriately considered a disorder of motor control, and as such, is best-understood in terms of models of general motor control that can be integrated with current psycholinguistic models of language production. Chapter 2 presents two prominent theories of motor control that are productively applied to speech. The predictions that these models make with regard to motor learning will be discussed, and the evidence supporting the use of several practice variables which enhance nonspeech motor learning will be reviewed. Next, the general psycholinguistic approach to the modeling of speech production is reviewed in Chapter 3, and possible ways in which speech production can be understood in concert with the general models of motor control are suggested. The current understanding of AOS at behavioral, neuroanatomic, and cognitive levels of description is reviewed in Chapter 4, and the disorder is considered in terms of the motor and psycholinguistic models presented in the previous two chapters. Chapter 4 also presents the arguments that applying principles of motor learning that have been found effective for enhancing limb motor learning is a well-motivated next step in improving the treatment outcomes in disorders of speech motor control, such as AOS,
and that AOS is a logical place to begin in testing the generalizability of principles of motor learning to the speech domain.

The four experiments that comprise this dissertation focus on the use of a subset of motor learning principles, that is, variables that govern the provision of external feedback, in the learning of speech motor skills. Experiments 1 and 2, presented in Chapter 5, investigate the effects of feedback timing variables on short-term performance changes and long-term learning in the context of a common speech treatment for AOS. Specifically, the frequency and post-response timing of external feedback are manipulated in an effort to enhance retention of treatment gains and transfer of treated speech skills to different contexts. Experiments 3 and 4 apply a chronometric approach to examine the effects of a different feedback variable, learner-controlled feedback, on pre-movement response preparation as well as response execution accuracy. In Experiment 3 (Chapter 6), the mechanism of the learner-controlled feedback effect (i.e., learning is enhanced when individuals have control over their feedback schedules) is investigated in healthy speakers in the context of a two-process motor programming model. The learner-controlled feedback manipulation is then applied to speech motor programming and learning in individuals with AOS in Experiment 4 (Chapter 7). The current experiments represent the first attempts to apply these motor learning principles to AOS (and the first attempt to apply the learner-controlled feedback manipulation to normal speech motor programming processes), and thus constitute an important step in the translation of a general conceptual framework of motor control to speech control and (re) learning.
2.0. Motor Control and Learning

Theories of motor control and learning provide guidance about how people acquire and use skilled actions. These theories have driven approaches to understand such diverse activities as athletic skills, playing musical instruments, and speech production. While some have considered motor control for speech unique (e.g., Ziegler, 2003a) given its dependence on the language system, others have recognized that it is a highly integrative motor activity that likely shares functional components with other motor control systems (e.g., Ballard et al., 2003; Folkins et al., 1995; Robin et al., 1997). Therefore, there is potentially a great deal to be learned from exploring the contributions that general models of motor control and learning can make to the speech domain.

One major approach to the study of motor control is based on the assumption that the time that intervenes between a stimulus and a response in a basic reaction time (RT) paradigm represents certain kinds of information processing related to preparation of the response. This chronometric view has posited that the serial processes that occur during RT include stimulus identification (external or internal), response selection, and response programming. These three basic tasks constitute the process of advance specification of movement goals before commands are sent to the effectors to initiate the execution of the response. In terms of speech production (in the

---

1 While some models of speech production make a distinction between motor planning and programming (e.g., Van der Merwe, 1997), there is a lack of consensus about the specification of these terms (Ballard, Granier, & Robin, 2000; McNeil et al. 1997). Therefore, the use of the term motor programming that is consistent with the more-established limb literature is adopted to refer to “organizing a selected action” (Klapp, 1995, p. 1015), or more specifically, to the set of processes that specify “a representation that, when initiated, results in the production of a coordinated movement sequence” (Schmidt & Lee, 2005, p.466).
context of picture naming, for example), stimulus identification and response selection may be thought of in terms of lexical retrieval and phonemic activation and will be discussed in the context of language production models considered in the next chapter. Response programming then, for a speech production task, involves assembling (e.g., Ferreira & Pashler, 2002) and/or mapping of these language units (phonemes, syllables) onto spatially and temporally coordinated effectors (articulator musculature), thus translating the discrete and invariant language units into continuous, context-dependent, fluent speech. A disruption of this process, AOS, will be discussed in detail in Chapter 4. The current chapter discusses two prominent models of general motor control and empirical evidence for factors that influence motor learning in terms of these models.

2.1. Schema Theory

A key assumption of theories of response programming is that programming organizes units of movement that are used in a given skill. In other words, actions are comprised of a series of units that are activated and then serially ordered during a skilled activity. A prominent conceptualization of this notion was developed under the framework of Schema Theory (Schmidt, 1975) as a “generalized motor program” (GMP). The Schema Theory of motor control and learning assumes that the brain stores “generalized” motor programs representing the relative timing and force of muscle commands necessary for carrying out discrete movements. Given particular response specifications, including desired movement outcomes and initial conditions,
the movement parameters are selected, specifying the absolute timing and force of muscle contractions in the chosen effectors (Shea & Wulf, 2005).²

Although Schema Theory is often contrasted with a dynamic systems or task dynamics view (e.g., Saltzman & Munhall, 1989; Thelen & Smith, 1994), on a superficial level, these approaches share similar features. Briefly, a central tenet of dynamic systems theory is that actions must be regarded as inextricably related to the contextual framework in which they are performed. Like Schema Theory, dynamic systems theory also proposes an organization of motor control that utilizes units of action/movement to constrain the degrees of freedom of the system (e.g., Kelso & Tuller, 1981). Whereas the units in Schema Theory (GMPs) represent the invariant relative timing and force patterns for a class of movements, the functional units in dynamic systems are referred to as coordinative structures or coalitions, and represent synergistic relationships between neural and muscular groups. The invariant temporal pattern, instead of being a stored memory structure as in Schema Theory, emerges as a result of the interaction between the person and the environment, which constrains or parameterizes the motor system (Kelso & Tuller, 1981).

Schema Theory provides the theoretical foundation for the current work because of its substantial advantages over dynamic systems in terms of explanatory and predictive power regarding motor control and learning phenomena (e.g., Schmidt, 2003; Schmidt & Lee, 2005). Evidence for generalized motor programs as

² It is acknowledged that other researchers (Shea & Wulf, 2005) have recently advocated a reconceptualization of the GMP as a “scalable response structure” (SRS), and emphasize processing mechanisms instead of schemata, but the terms ‘GMP’ and ‘schemata’ are used here to maintain terminological consistency with the literature from which this paper draws.
conceptualized by Schema Theory has been derived from electromyography (EMG) studies of muscle activation patterns in the presence of unexpected environmental perturbations (e.g., Wadman, Denier van der Gon, Geuze, & Mol, 1979) and different relative temporal scalings of movements (e.g., Carter & Shapiro, 1984), as well as from studies that have demonstrated the cohesion of relative movement characteristics within, but not across, putative units of movement (e.g., Clark, Robin, McCullagh, & Schmidt, 2001; Schneider & Schmidt, 1995). Further, Schema Theory makes explicit predictions about the effects of various practice and feedback conditions on motor learning (see section 2.3), which is particularly relevant to the focus of this dissertation. These predictions can be integrated with programming models that have compatible views about motor learning, such as the two-stage model discussed in the next section. Although these predictions are not necessarily incompatible with a dynamic systems view, they have been given more systematic and explicit consideration within the framework of Schema Theory.

2.2. Two-Stage Model of Motor Programming

Because movements in daily activities such as speech usually entail the serial ordering of successive putative units of action (or GMPs in Schmidt’s (1975) terminology), a two-process model of response programming that specifies both single motor program assembly as well as multiple program sequencing has been proposed (e.g., Klapp, 1995; 2003). Under this framework, response programming is divided into two components: one that can be preprogrammed, and one that is undertaken only
after the movement is initiated. Incorporating decades of research involving variables that affect RT in different experimental paradigms, the hypothesized functions of these two stages of processing (pre- and post- movement initiation) have been described (e.g., Klapp, 1995; 2003; Sternberg, Monsell, Knoll, & Wright, 1978). The pre-initiation stage is sensitive to the complexity of individual units to be performed, and is thus thought to specify the internal structure of an individual unit of movement. This preprogrammable stage is called “INT.” Movement duration is thought to be one important dimension of complexity that affects INT processing. To put it in terms of Schema Theory, this process may be thought of as entailing the activation and parameterization of a GMP. The second stage, which begins with movement initiation, is sensitive to the number of units being performed, and is thus thought to entail the retrieval of units of action (or GMPs) from a motor buffer and sequencing into the correct serial order. This post-initiation process is called “SEQ.”

Evidence for this two-stage model of response programming has come from chronometric studies in which simple and choice reaction times were measured (e.g., Klapp 1995; 2003). In a simple reaction time paradigm, a participant is instructed by a pre-cue to prepare a certain response, and then the latency between an imperative “go” signal and the subject’s response is measured. This measure varies as a function of the number of units that are performed, and not their complexity, and serves as an index of the SEQ process. In a choice reaction time paradigm, on the other hand, subjects are not told which of several responses to prepare until the imperative signal is given, so choice reaction time (latency between imperative signal containing the response cue
and response) can be used to index the preprogramming process of INT as well. Although INT and SEQ theoretically can occur in parallel during the choice RT interval, it is assumed that the INT process takes longer, and thus, choice reaction time varies as a function of unit complexity when the number of units is constant (Klapp, 1995).

One major limitation of this reaction time approach for detailing these two processing stages has been that choice and simple RT data are necessarily collected on different trials, usually on different subjects in separate experiments, and thus conclusions about influences on INT and SEQ processing have typically relied on cross-experiment comparisons. Recently, a new method, the self-selection paradigm, was introduced for measuring both INT and SEQ in a single trial. In the self-selection paradigm (Immink & Wright, 1998; 2001), INT is indexed by “Study Time” (ST), or the amount of time that it takes a person to press a button to indicate that he has prepared a response after a cue indicates the specific response to be performed (e.g., single movement, multiple movements, varying durations). Since the movement is not initiated during ST, and only the pre-programmable portion of the response can be processed, ST appears to be a reasonable index of the INT process. Then, within the same trial, after a variable delay, an imperative “GO” signal is given, and the subject’s simple reaction time is measured as the latency between this signal and the onset of

---

3 Although INT is conceptualized in terms of processes of response programming, it should be noted that in a choice RT measure, response selection is also necessarily processed during the RT interval. However, in the evidence reviewed here concerning response programming, the independent variables that can affect response selection (e.g., response uncertainty, stimulus-response compatibility) have been controlled (see Klapp, 1995).
the preprogrammed response. As in the earlier-described approach, this simple reaction time (RT) is an index of SEQ processing.

### 2.3 Motor Learning

Both the INT/SEQ model and Schema Theory generate testable predictions about motor learning. Motor learning refers to a “set of internal processes associated with practice or experience leading to relatively permanent changes in the capability for movement” (Schmidt & Lee, 2005:302). According to the two-stage model of INT and SEQ, learning complex motor skills entails the “chunking,” or integration, of multiple units into larger units (e.g., Klapp, 1995) thus reducing the load on the sequencing process. The notion of concatenating smaller units into a larger single unit with practice is described in other models as well (Schmidt & Lee, 2005). Klapp (1995) provided initial evidence demonstrating that with practice of a 4-part button press sequence, simple reaction time (an index of SEQ) decreases, while choice reaction time (an index of INT) increases, suggesting that the 4 serially-ordered units become integrated into one larger, more complex unit.

Schema Theory assumes the refinement of the GMP and the selection of parameters to be critical to motor learning (e.g., Schmidt & Lee, 2005). According to Schema Theory, four types of information are stored after a movement is executed, including the initial conditions (task and environment conditions prior to movement production), the parameters that were assigned to the GMP, the outcome of the movement in terms of the environmental goal, and the sensory consequences of the
movement. Learning new skills involves developing abstract relationships between these pieces of information with practice of the movement. One of these relationships, the recall schema, involves activation of past parameters, initial conditions, and movement outcomes produced by their combinations. On future trials, when the initial conditions and desired outcomes are noted, the recall schema serves to select the parameters most appropriate for achieving the movement goal, and applies them to the GMP. The second relationship is the recognition schema, which is the relationship between the initial conditions, the environmental outcomes, and the sensory consequences of a particular movement. When a performer notes the initial conditions and desired outcome before a movement, the sensory consequences of that movement can then be estimated. These expected sensory consequences are then used to evaluate the movement after its execution. Under the framework of Schema Theory, learning consists of strengthening these two schemas through experience with the four stored elements.

Schema Theory makes several predictions about motor learning (see Schmidt & Lee, 2005; Shea & Wulf, 2005) and the effects that certain practice variables have on it. Collectively, these variables have become known as “principles of motor learning,” and specify conditions of practice structure and augmented feedback that can be manipulated to enhance learning of motor skills. Most of the evidence regarding the application of these principles to motor learning has been derived from studies of limb movement learning, although increased attention has been given in recent years to the effects that they might have on rehabilitation of limb control.
(Goodgold-Edwards & Cermak, 1990; Hanlon, 1996; Jarus, 1994; Landers, Wulf, Wallmann, & Guadagnoli, 2005; Sabari, 1991; Stevans & Hall, 1998; VanVliet & Wulf, 2006), speech motor learning, and re-learning of speech skills in the context of neurologic injury or disease. The evidence concerning the application of these principles to speech will be presented in the final chapter of this paper.

If learning is to be conceived of as entailing the relatively permanent development and refinement of abstract movement representations, then informative measures of learning should include assessing retention of the trained skill (i.e., persistence of performance), transfer (i.e., change in the performance capability of related, but untrained skills), and long-term retention of improved performance. Unfortunately, researchers in motor learning (and also clinical researchers in speech-language pathology, as will be discussed later) have not always recognized the differential effects that some practice variables have on performance during acquisition versus long-term learning. Therefore, an important consideration when measuring the degree of motor learning is to differentiate between variables which temporarily enhance performance and those that actually enhance learning in a relatively permanent way (i.e., when these conditions are removed). The review of literature below supporting principles of motor learning, when possible, excludes early studies that failed to examine long-term learning effects (i.e., retention and transfer).

When classifying extrinsic practice variables (environmental conditions present during practice) that affect motor learning, researchers have primarily been interested in conditions that have to do with either the structure of practice or the
augmented feedback that is provided to the learner. Within each of these domains, several influential variables have been studied.

2.4. Practice Variables

It is generally agreed that the most influential variable in motor skill learning is the amount of practice (Schmidt & Lee, 2005). Clearly, a large number of repetitions of a skill provides the learner with more opportunities to extract relevant movement data (e.g., initial conditions, movement outcomes, sensory consequences) and associate them to build stable, coherent schemas. High amounts of practice may also lead to more “automatic” (i.e., impervious to interference) processing of the task (e.g., Schmidt, 1987; Schmidt & Lee, 2005). However, researchers have also discovered conditions in which learning is degraded by too much practice. For example, massive amounts of practice with one member of a class of movements may lead to the emergence of an “especial” skill (e.g., basketball shots from the foul line) that does not transfer to related skills (e.g., nearby shots) (Keetch, Schmidt, Lee, & Young, 2005). Similarly, while a high amount of movement practice in one effector enhances retention of that skill within that effector system, it limits transfer of the skill to other effectors (Park & Shea, 2003). Although effector assignment is typically considered a movement parameter, these results suggest that after many trials in one effector system, this parameter becomes represented as part of the movement structure (GMP). A large number of repetitions also appears to be detrimental under conditions of constant practice, and will be discussed in that context in greater detail below.
Related to amount of practice, the intensity of practice is another concern when attempting to optimize the efficiency of motor learning. Over the years, the concepts of “massed” and “distributed” practice have shifted from referring to the time between trials (usually in terms of seconds and minutes) to referring to the time between practice sessions (usually in terms of days). Under both meanings of the terms, distributing practice seems to exert a positive influence on learning (see Baddeley & Longman, 1978; Lee & Genovese, 1989; Shea, Lai, Black, & Park, 2000). Allowing a greater amount of time between practice sessions (on the order of 1-2 days) is thought to support memory consolidation processes and strengthen memory representations required for long-term learning (Robertson, Pascual-Leone, & Miall, 2004).

Another important condition of practice that has garnered much attention in the limb motor learning literature and demonstrates differential effects on acquisition and retention is the order in which skills are practiced. Robust effects have been found when comparing blocked (i.e., a single behavior is practiced repeatedly during a single block of trials) versus random (i.e., multiple behaviors are practiced in random order within a block of trials) practice schedules, demonstrating a benefit of blocked practice on acquisition (i.e., temporary performance enhancement during practice), but a benefit of random or serial practice on retention (Lee & Magill, 1983; Lee, Magill, & Weeks, 1985; Li & Wright, 2000; Shea & Morgan, 1979). These effects seem to interact with task difficulty, such that random practice may not provide as great a benefit when applied to difficult tasks (Wulf & Shea, 2002).
Schema Theory does not offer explicit predictions about the differential effects of random and blocked practice on acquisition performance and retention (Shea & Wulf, 2005), but other hypotheses have been proposed. The elaboration hypothesis (Shea & Morgan, 1979; Shea & Zinny, 1983) forwarded the notion that the different cognitive demands placed on learners by the different practice orders influence skill acquisition or retention. Blocked practice is thought to only engage learners in intratask processing, since it relies on single task analysis without reference to other sources of information. Random practice, on the other hand, supports intertask processing, or associative and referential analysis amongst different tasks, and thereby helps the learner form a more detailed representation of the behavior by incorporating new task information with existing knowledge. These different processing requirements are thought to be responsible for the observed benefit of blocked practice on acquisition and the benefit of random practice on retention. A similar logic is suggested by the reconstruction hypothesis (Lee & Magill, 1983; Lee et al., 1985), which argues that blocked practice is beneficial in early skill acquisition, when cognitive demands are high, because it allows a single “action plan” to be held in a buffer and repeatedly deployed without interference. Random practice, on the other hand, which requires repeated retrieval and reconstruction of responses, enhances retention because it requires greater engagement in movement planning processes, which more closely approximates the increased processing demands when practice conditions are removed.
This pattern of results has been tested in the context of the two-stage response programming model discussed above. Immink and Wright (1998) found the same pattern of practice order results (i.e., blocked benefits acquisition; random benefits retention) on the INT stage of response programming, but demonstrated that random practice subjects could perform as well as blocked practice subjects in acquisition when they were given sufficient time. These authors attributed this finding to what they called the *contextual interference effect*, that the trial-to-trial interference inherent in random practice conditions was likely associated with extra planning processes that disrupted acquisition performance but facilitated retention and transfer. This group also found that the effect of number of units on SEQ processing decreased under conditions of random practice, suggesting that random practice facilitated chunking and resulted in a more stable and coherent representation of the learned movement sequences (Wright, Black, Immink, Brueckner, & Magnuson, 2004).

Recent studies have also tested the differential effects of random and blocked practice on the learning of GMPs and parameterization, as conceptualized under the framework of Schema Theory. The learning of absolute timing, which can be thought of as a movement parameter, is enhanced under random practice conditions, when instructions specify either segment durations or ratios thereof. Learning relative timing, on the other hand (specified by GMPs), is facilitated by blocked practice when instructions are given in terms of relative movement durations, and facilitated by random practice when instructions specify movement duration ratios (Shea, Lai, Wright, Immink, & Black 2001). To account for these findings, the *stability*
A related factor in the structure of practice has to do with the variability of the behaviors that are practiced. Schema Theory predicts that increasing variability by practicing one movement structure (GMP) under a wide variety of initial conditions and outcomes enhances learning (e.g., Schmidt & Bjork, 1992). The argument is that variable practice involves scaling a single movement pattern across different superficial dimensions and observing the associated outcomes, and thus provides the learner with opportunities to develop and strengthen the schemas for that movement class, and thereby enhance the proficiency of assigning novel parameters to it in the future. The empirical support for this manipulation is strong (e.g., Lee et al. 1985; Shapiro & Schmidt, 1982; see van Rossum, 1987; Wulf and Schmidt, 1997), although the effects may be mitigated by interactions with other practice conditions such as high skill proficiency (Chamberlin & Lee, 1993; Shapiro & Schmidt, 1982; but see van Rossum, 1987) or blocked practice schedule (Lee et al., 1985; Shea et al., 2001). Of note, there is also evidence to suggest that constant practice of one variant of a movement over a high number of trials may actually deter learning processes (Shea & Kohl, 1991). This finding is compatible with the reconstruction hypothesis (Lee &
Magill, 1983; Lee et al., 1985), which would argue that since constant practice limits experience with retrieval and reconstruction processes (by keeping only one representation in a buffer), a large amount of this kind of practice may result in a decrement in the ability to engage those operations in the future.

Just as random and blocked practice schedules of multiple behaviors have been demonstrated to differentially affect GMP learning and parameterization, a congruent effect has been shown with constant versus variable practice of a single movement structure. Lai, Shea, Wulf, and Wright (2000) demonstrated in a sequenced key-press task that constant practice facilitated acquisition of relative timing patterns (GMPs), but that variable practice enhanced learning of absolute timing (parameterization), as measured by a transfer test. Further, by combining different permutations of constant and variable practice across two phases of practice, these authors demonstrated that providing variable practice after experience with constant practice optimized the learning of both relative and absolute timing.

Another important factor that is involved in the optimization of motor learning is movement complexity. Many studies have suggested that complexity, and a related construct, difficulty, may interact with other principles of motor learning (e.g., Beilock, Bertenthal, McCoy, & Carr, 2004; Chamberlin & Lee, 1993; Shapiro & Schmidt, 1982; Swinnen, Lee, Verschueren, Serrien, & Bogaerds, 1997; cf. Wulf, Lauterbach, & Toole, 1999), however a satisfactory operationalization of these terms has not been established. While some have defined complexity in terms of reaction time or movement time, others have considered it in relation to the number of degrees
of freedom to control in a given movement (Guadagnoli & Lee, 2004; Wulf & Shea, 2002). In some types of tasks, it is possible to characterize complexity in terms of the subset/superset notion, in which whole movements can be broken down into composite parts. Conventional wisdom, particularly in the rehabilitation sciences, holds that decomposing a skill and independently training its various components facilitates learning by reducing the cognitive demands of processing the whole task. According to Schema Theory (e.g., Schmidt & Lee, 2005), however, when a “whole” movement is governed by a single GMP, whole-practice should promote the abstraction of the movement structure and enhance transfer to simpler, or “part,” movements. Part-training, on the other hand, may not transfer to the “whole” movement pattern when it is governed by a single GMP. When part-movements are characterized as serial elements (each a putative single GMP) of a “whole” sequential movement, part-practice is beneficial to learning (Hansen, Tremblay, and Elliott 2005) and promotes greater flexibility in breaking down large sequences into parts (Park, Wilde, & Shea 2004).

It appears that in order to effectively apply the strategies of part- or whole-practice, trainers need an accurate count of how many GMPs are involved in a given movement. However, GMP size likely varies as a function of context, as well as individual learner skill and experience (e.g., “chunking,” Klapp, 1995; Wright et al., 2004), making it impossible to determine standard movement complexities for all learners in all conditions. A recently-proposed framework has attempted to model this interaction of learner- and task-dependent factors contributing to complexity. The
Challenge Point Framework (Guadagnoli & Lee, 2004) argues that learners must be challenged at an optimal level in order for efficient learning to occur. That optimal level, or “challenge point,” is a function of the nominal task difficulty, functional task difficulty, and the background skill level of the learner. In much the same way, collectively, principles of motor learning are thought to operate on the assumption that certain practice variables influence the effort required to undertake cognitive processes that play an important role in early skill acquisition. Manipulating these practice variables enhances motor learning by promoting an optimal level of cognitive effort by the learner (Lee, Swinnen, & Serrien, 1994).

2.5. Feedback Variables

In addition to considerations related to the structure of practice, manipulations in the provision of augmented feedback during practice represent another major domain of principles of motor learning and have been the subject of extensive investigation in the limb motor learning literature (see Swinnen, 1996; Wulf & Shea 2004 for reviews). As indicated above in the discussion of schema development, the availability of outcome information after a movement is critical to motor learning and the development of memory representations (e.g., Bilodeau, Bilodeau, & Shumsky, 1959). Due to the difficulty of studying the effects of intrinsic sources of outcome information in humans, researchers have developed paradigms for manipulating extrinsic, “augmented” feedback as a means of deducing the operations of intrinsic feedback in naturalistic contexts. Further, as augmented feedback is often used in
training or rehabilitative circumstances, its informational, motivational, and associational effects on learning need to be understood in order to provide optimal learning conditions.

Augmented feedback typically assumes one of two forms: knowledge of performance or knowledge of results (Schmidt & Lee, 2005). Knowledge of performance (KP) feedback provides the learner with information about the movement pattern, and usually inherently contains some information about how to correct the incorrect movement pattern (e.g., “Your knees were bent”). Knowledge of results (KR), on the other hand, provides information concerning only the movement outcome, in terms of the environmental goal. KR can be general (e.g., “correct” or “incorrect”), or specific (e.g., “You overshot the target by 12 cm”), and might also provide information for error correction on a subsequent trial. In general, KP is thought to be most beneficial early in learning when the movement goal is unknown (e.g., Newell, Carlton, & Antoniou, 1990), or when a more flexible reference criterion for accuracy is employed (Schmidt & Lee, 2005). Conversely, KR may be more facilitative later in skill acquisition when a single optimal movement goal is clear (e.g., Hodges & Franks 2001; Swinnen, Walter, Lee, & Serrien 1993).

Early studies investigating augmented feedback supported the provision of frequent, immediate, and precise feedback, although, as described earlier, these studies largely suffered a common flaw; few examined the effects of these manipulations on multiple indices of performance and learning (i.e., acquisition, retention, and transfer). These studies often only assessed acquisition effects, which may have been transitory
More recently, investigators recognized the potentially differential effects that manipulations in feedback intensity and timing may exert on acquisition versus retention and transfer of motor skills.

In investigating the effect that altering the temporal locus of feedback has on motor performance and learning, researchers have been concerned with when, in the time course of a practice trial or inter-trial interval, KR or KP is delivered. For instance, feedback can either be provided at the end of a movement (terminal feedback), or online during the movement production (concurrent feedback). Concurrent feedback, while it may enhance online performance, has generally been found to disrupt learning as measured by retention and transfer (Park, Shea, & Wright, 2000; Schmidt & Wulf, 1997; Vander Linden, Cauraugh, & Greene, 1993). One exception to this conclusion has been found when studying the interaction between concurrent feedback and a learner’s focus of attention. Although results are not uncontested (e.g., Beilock, et al. 2004; Wulf et al. 1999), in general, inducing an external focus of attention in the learner (i.e., on something outside of the body) via instructions or feedback seems to be more effective in promoting accurate and stable performance and learning than an internal focus of attention (i.e., on inside the body) (e.g., Wulf, McNevin, & Shea 2001; see Wulf & Prinz, 2001 for review). The constrained action hypothesis explains this result by arguing that adopting an internal focus of attention causes learners to attempt to consciously control, or “freeze,” their otherwise automatic motor control processes, inducing an abnormal processing mode (Wulf et al., 2001). Of note, learners who have not been given attentional focus
instructions perform as poorly as those who have been instructed to adopt an internal focus of attention (Wulf, Höß, & Prinz, 1998). Compared to a no-feedback condition, concurrent feedback that induces an external focus of attention can be beneficial to learning by alleviating the internal-focus tendency (Shea & Wulf 1999). In other words, the detrimental effect that concurrent feedback typically has on learning is overridden when it causes learners to focus their attention outside of their bodies.

Based on animal studies suggesting that delaying a reward after response to a stimulus has a detrimental effect on conditioning (e.g., Perin, 1943; Skinner, 1936), researchers in human performance became interested in evaluating the effects of delaying terminal feedback in human motor skill learning (see Salmoni et al., 1984, for review). Most early studies revealed null or inconsistent effects of feedback delay, but failed to distinguish between acquisition performance and learning (e.g., Becker, Mussina, & Persons, 1963; Koch & Dorfman, 1979; Mulder & Hulstijn, 1985; Weltens & de Bot, 1984). Swinnen, Schmidt, Nicholson and Shapiro (1990) examined the effect of a short feedback delay (3.2 or 8 seconds) on acquisition and retention of a motor skill, and concluded that instantaneous feedback initially supported acquisition of the behavior, but at a certain point began to impede the continued improvement during acquisition and also interfered with retention of that skill tested 10 minutes and, more dramatically, 2 days later. This trend persisted on a 4-month retention test, although the group difference was no longer significant. More recently, Anderson, Magill, Sekiya, and Ryan (2005) reported that trials-delayed feedback (i.e., feedback given after two intervening trials) resulted in less accurate acquisition of an unfamiliar
aiming behavior, but stronger retention, particularly when tested after a delay (24 hours vs. 1 minute). The size of this difference was moderate, although it did not reach significance. Additionally, the decline in performance between acquisition and retention was smaller for the delayed feedback group. These subjects also reported using a greater number and variety of intrinsic feedback sources during practice, suggesting that this type of processing may have been blocked in subjects receiving immediate feedback, contributing to their poor retention.

A related feedback manipulation that also delays the availability of outcome information is known as summary feedback. Under this condition, researchers suspend the delivery of KR after a trial until a specified number of trials (X) have been completed. Then, feedback is delivered which summarizes the learner’s performance over the past X trials. Although this paradigm introduces additional factors (e.g., intervening trials between a movement and feedback can create interference and increase memory demands), it has produced results congruent with those found for straightforward feedback delay, suggesting that although summary KR appears least effective for acquisition, it is more effective than immediate KR for learning (e.g., Lavery, 1962; Salconi et al., 1984; Schmidt, 1991; Schmidt, Young, Swinnen, & Shapiro, 1989). Furthermore, there is an inverse relationship between summary length (the number of trials summarized) and acquisition performance (Gable, Shea, & Wright, 1991; Schmidt, Lange, & Young, 1990; Yao, Fischman, & Wang; 1994). A positive relationship between summary length (up to 16 trials) and long-term learning effectiveness has also been demonstrated (e.g., Gable et al., 1991; Schmidt, et al.,
1989), although for more complex tasks, the optimal summary length appears to be about 5 trials (Yao et al., 1994; Schmidt et al., 1990).

Another important factor in augmented feedback that potentially confounds a straightforward interpretation of summary KR in terms of feedback delay has to do with the frequency with which a learner’s practice is interrupted for delivery of feedback. Summary KR presents an interesting situation in which feedback may be given about 100% of the productions, but may only be delivered after 20% of trials (i.e., a summary length of 5). Although this particular interaction has not been addressed specifically in motor learning research, much is known about the effects of the relative frequency of feedback. Like several other variables discussed above, feedback frequency may exert differential effects on acquisition performance and long-term learning (e.g., Weinstein & Schmidt, 1990; Wulf & Schmidt, 1989). According to the guidance hypothesis, providing more frequent feedback likely serves to enhance performance during training by providing increased guidance and heightening energy and motivation (Lee, White, & Carnahan, 1990; Salamone, et al., 1984). However, these performance effects are temporary, and are dependent on the continued delivery of the feedback. Presenting KR feedback with lower frequency is hypothesized to enhance long-term learning by facilitating the development of self-evaluation and error-detection skills that the learner can apply to situations in which

---

4 Although studies of feedback frequency typically discuss manipulations in terms of relative frequency of feedback (e.g., on 50% of trials versus 100%), since the number of trials is usually held constant in these comparisons, this manipulation actually results in a changed absolute frequency of feedback as well. Purely relative feedback frequency manipulations are not commonly used because of the potentially large and confounding effect that the greater number of trials (in a reduced relative feedback condition) could have (Schmidt & Lee, 2005).
external feedback is not available (Bruechert, Lai, & Shea, 2003). This hypothesis provides a similar explanation for the results of KR-delay studies (above), because delaying feedback would also promote the development of these self-evaluation skills by allowing the learner sufficient time to process and build relationships (schemas) based on the four kinds of movement information discussed above. Indeed, participants in Anderson and colleagues’ (2005) novel aiming task study were more likely to report using a greater variety of intrinsic feedback sources under conditions of feedback delay than with immediate or no feedback. Alternatively, it has been suggested that reducing the frequency of augmented feedback may serve to relieve the performer of a state of internal attentional focus induced by constant feedback, and to allow the performer to assume an external attentional focus guided by automatic motor control processes (Wulf, McConnel, Gärtner, & Schwarz, 2002). Although not all experimental conditions have produced a benefit of reduced KR relative to 100% KR (e.g., Lee et al., 1990; Sparrow & Summers, 1992; Winstein & Schmidt, 1990), a superior effect of 100% KR on learning has not been found (i.e., effects are equal or enhanced under reduced KR conditions).

Although Schema Theory predicts a positive relationship between availability of outcome information and degree of learning, differential effects of reduced-frequency KR on GMPs and parameters have been demonstrated. While increased frequency of external knowledge-of-results (KR) feedback promotes parameter learning (e.g., Wulf, Schmidt, & Deubel, 1993), the provision of too much external KR feedback appears to degrade GMP learning, suggesting that reduced availability of
external outcome information is important for promoting performers’ own abstractions of the invariant features of a movement pattern (GMP) (Wulf, Lee, & Schmidt, 1994). The effect of feedback frequency on parameter learning is predicted by Schema Theory, in which stable and accurate movement representations (schemata) are built by associating parameters with movement outcomes. Interestingly, on the view that more efficient learning may involve first developing GMPs (see discussion of constant and variable practice order effects above), it would seem that beginning with relatively infrequent KR and then *increasing* feedback over the course of learning would optimize the learning of GMPs and parameters. This is the opposite of what typically happens in dynamic feedback paradigms, such as in bandwidth feedback (e.g., Lee & Carnahan, 1990; Sherwood, 1988), in which explicit KR is only provided for errors outside a predetermined level of tolerance. Thus, as a learner’s skill increases, the likelihood of receiving explicit KR decreases.

Like many principles of motor learning, feedback frequency interacts with other factors, and the negative effects of highly-frequent feedback can be mitigated in some circumstances. For example, highly-frequent feedback is not as detrimental when it induces an external focus of attention (e.g., Shea & Wulf, 1999), and greater feedback frequency is beneficial for learning more complex movements (e.g., Swinnen, et al., 1997; see Shea & Wulf, 2005, for review). Further, the detrimental effects of high-frequency feedback can be alleviated when subjects are asked to estimate their own errors (Guadagnoli & Kohl, 2001), lending further support to the guidance hypothesis.
Given that many of these feedback manipulations are aimed at maximizing the likelihood that learners will receive the right kind and amount of feedback when they need it, another logical manipulation is to allow learners to determine the schedule with which they receive feedback. Giving learners some control over the practice conditions is predicted to enhance learning by heightening motivation, encouraging learners to take charge of the learning process and engage in deeper task processing, and by providing conditions that are more tailored to individual learners’ needs. Earlier studies on this nascent topic (i.e., in the motor domain) have confirmed the benefit of self-controlled feedback (e.g., Janelle, Barba, Frehlich, Tennant, & Cauraugh, 1997; Janelle, Kim, & Singer, 1995), and these benefits extend to learner-control of other task conditions as well (Wulf, Clauss, Shea, & Whitacre, 2001; Wulf, Raupach, & Pfeiffer, 2005). Furthermore, recent studies exploring the reason why self-controlled feedback works have produced results that support the notion that it is the tailoring of the feedback to the learners’ needs, and not merely motivation or perception of self-control, that benefits learning (Chiviacowsky & Wulf, 2002; 2005). Interestingly, this manipulation also appears to have differential effects on absolute and relative aspects of movement structures, although the pattern of results is not consistent (e.g., Chiviacowsky & Wulf, 2002; 2005). More research is needed to understand the mechanisms and interactions of this potentially influential learning variable.

Given the vast amount of research that has been conducted testing these principles of motor learning on nonspeech skills and activities, it is not surprising that
researchers are beginning to recognize the potential importance and implications of these findings to the field of speech-language pathology. As alluded to earlier, clinicians treating motor speech disorders often employ tactics that are contrary to these findings, such as structuring practice of skills in homogenous blocks and providing immediate feedback with high frequency. Chapters in clinical management textbooks (e.g., Duffy, 1995; 2005; McNeil et al., 1997) have recently advocated for the application of these principles in the speech domain, and have called for further investigation. The early outcomes of these endeavors will be presented in Chapter 4.
3.0. **Speech Production**

Control of the rapid articulatory movements underlying speech represents one of the most complex and uniquely human motor feats. Despite its role as the fundamental modality of communication between humans and the research interest that its complexities have generated, the neural organization of and interactions between linguistic, motoric, and sensory information that underlie fluent speech are not well understood. Reconciling what is known about the speech-language production system with what motor researchers have learned about limb control will be essential in fully explicating a theory of speech motor control and learning.

The ultimate goal in developing a complete account of language and speech production is to frame these systems in relation to other cognitive processes such as memory, attention, perception, and motor programming to reveal how abstract communication goals are realized as coordinated muscle commands for normal and disordered speech. Psycholinguistic models (e.g., Dell, 1986; Garret, 1984; Levelt, Roelofs, & Meyer, 1999) have approached this goal by conceptualizing word production as a multistage process that proceeds from more (e.g., conceptual-semantic) to less (e.g., phonologic, phonetic) abstract. Although these models have been clinically relevant to the modeling of language production disorders, their underspecification of processes that occur after the phonological level have limited their utility in the speech domain. In contrast, models of speech production have typically not addressed pre-motor linguistic processing (e.g., Guenther & Perkell, 2004). A complete understanding of normal speech production, as well as higher-level
motor speech disorders such as AOS, will require a detailed model that encompasses not only the language system or the speech motor system, but also their junction. It is precisely this translation of phonologic information into kinematic patterns for speech that AOS is thought to disrupt (McNeil et al., 1997), underscoring the importance of modeling this connection. This chapter will briefly describe word production from psycholinguistic and motor speech perspectives, and will suggest potential points of convergence between these theoretical frameworks and the models of motor control discussed in the previous chapter.

3.1. Psycholinguistic Models

Current psycholinguistic models of word production (e.g., Bock, 1995; Cutting & Ferreira, 1999; Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, 1989; Levelt et al., 1999; Roelofs, 1992; Roelofs, 1997; Roelofs, Meyer, & Levelt, 1996) are typically organized in hierarchical stages, with spreading activation between levels and different linguistic representations operated on at each level. These linguistic representations (e.g., phonemes, syllables, morphemes) are input to, but theoretically distinct from, the motor system that instantiates them physiologically. While input to the motor system may be possible at multiple levels of the linguistic processing hierarchy (e.g., metrical and segmental plans), most theories posit a primary distinction between speech and language stages of processing. Support for this notion has been derived from picture naming studies of word production showing greater interference from semantically-related distracters at short stimulus onset
asynchronies, and greater influence of phonologically-related primes at longer stimulus onset asynchronies, taken to suggest that phonological encoding follows word retrieval in a hierarchical manner that has been conceptualized by some models as feedforward and independent (e.g., Levelt et al., 1991), and by others as cascaded and interactive (e.g., Dell & O’Seaghdha, 1991). Under either interpretation of the priming experiment data, articulatory planning and motor execution are hypothesized to follow these linguistic processes, creating a primary separation of speech and language processes.

There is general consensus among the many psycholinguistic theories of word production that linguistic rules serve to build frames at different levels of processing, which are filled by the representations activated at those levels. Lemmas, or syntactically and semantically specified word forms without phonological content, and lexemes, containing phonologic structure, are representations that are activated early in the multistage process. Phonemes, the fundamental units of lexemes, are then assembled and processed for motor programming and execution. Debate about the precise nature of these notions has stimulated a vast amount of research, which has given rise to variations on this general hierarchical model of word production.

Sometimes considered the most prominent and comprehensive of these models, the Nijmegen model developed by Levelt and colleagues (e.g., Levelt, 1989; Levelt et al., 1999) is a feedforward model of discrete activation between serially-processed levels. In this model, lexical retrieval forms the first level of processing, the output of which, in terms of morphemes, is transformed into phonological words
through the process of phonological encoding. During phonological encoding, phonemes and metrical frames are simultaneously retrieved, and through the serial process of prosodification (Levelt et al., 1999) are integrated to form phonological syllables and then words. Next, the process of phonetic encoding occurs, beginning what Levelt considered to be the motor programming stage (Levelt, 1989). Abstract phonological words are translated into articulatory gestures, either by retrieval of phonetic syllable representations from a memory store of articulatory primitives called the “syllabary” (for frequently-used syllables), or by assembling individual phoneme segments (for infrequent syllables). The gestural scores specifying the movement goals for each articulator are then stored in a buffer until all of the gestural scores for a phonological word are computed, and then they are deployed to the effectors to be implemented as continuous, context-dependent speech movements.

Despite the prominence and comprehensiveness of the Nijmegen model, its assumptions and explanatory power are not without criticism. For instance, the difficulty that the model has accounting for certain patterns of production errors has been a common topic of debate. Essentially, a model such as this with strictly feedforward connections and discrete activation of one level at a time cannot easily accommodate errors that appear to result from interactions between different levels of representation. These error patterns include category constraint on whole-word exchange errors which usually involve members of the same syntactic category (Dell & Reich, 1981; Fay & Cutler, 1977; Garret, 1980; MacKay, 1982), lexical bias in which phonological errors are more likely to result in real words instead of nonwords
(Dell, 1986; 1990; Dell & Reich, 1981), semantic bias in which phonological errors are more likely to result in words that are related to the surrounding linguistic context, and similarity effects, in which there is a tendency for word errors to involve similar-sounding and -meaning words (Dell, 1986). In order to accommodate these error patterns, Levelt and colleagues (1999) invoked verification processes such as the binding-by-checking mechanism and the post-lexical self-monitoring system, that, when malfunctioning, allow these types of errors to occur. In contrast, these errors can easily be explained by a cascaded, interactive model in which feedback between different levels of representation (e.g., phonemes and lexemes) is permitted (e.g., Dell, 1986). While Dell’s model can accommodate these phenomena, it is also limited by its focus on pre-motor linguistic processing and interactions at the expense of including other potential constraints such as those from the motor system.

A further issue that has been raised with regard to the Nijmegen model is that the evidence for the existence of the syllabary is not unequivocal (Levelt et al., 1999; Nijland et al., 2003). However, the notion of the syllable as a fundamental unit of speech programming is nevertheless appealing, as it is plausible and accepted under several other accounts (e.g. Aichert & Ziegler, 2004; Klapp, 2003; Stetson, 1951; see Ziegler & Maassen, 2004 for a review). For example, in Klapp’s (2003) two-stage model of motor programming, syllable-sized units are preprogrammed and fed into a buffer, where they are later sequenced online after movement initiation, reminiscent of the Nijmegen model’s syllable-by-syllable encoding and articulatory buffering. Klapp’s notion of “chunking” more practiced sequences into concatenated units is also
compatible with the Nijmegen model’s notion of frequently-used phoneme sequences being unitized into syllables stored in the syllabary, although Klapp’s work demonstrated syllables being chunked into multisyllabic units, whereas in the syllabary, phoneme segments are unitized into syllables.

Other units of production have been proposed or assumed, from phonetic features (Perkell, 1980) to phonemes (Guenther, Hampson, & Johnson, 1998; Shattuck-Hufnagel, 1987; van der Merwe, 1997), and even phrases (Kozhevnikov & Chistovich, 1966, in Smith & Goffman, 2004). Further, Klapp (2003) found evidence for flexibility of speech unit size when he demonstrated that a sequence of up to three syllables could be programmed as a single-unit pseudoword, or that unitization could be discouraged by certain practice conditions. The two-stage model of motor programming that Klapp developed was used to show that when chunking (i.e., unitization) was allowed, adding more syllables to a pseudoword affected the process of INT, as measured by choice RT, indicating that the syllables had been concatenated into a single more complex unit. However, when chunking was discouraged, the number of syllables affected SEQ, measured by simple RT, indicating that multiple individual units were being sequenced. This paradigm offers a promising way to measure the size of the units used for speech, which has practical as well as theoretical relevance, since unitization is considered reflective of motor skill learning (e.g., Klapp, 1995; Wright et al., 2004).

Other researchers have suggested that speech unit size varies as a function of frequency of use, from syllables to words to phrases (Varley, Whiteside, Windsor, &
Fisher, 2006). In children learning to speak, changes in organizational unit size may occur in the opposite direction, beginning with broader unit boundaries (whole words) before motor constraints mature to allow contrasts to be made at the segmental level (e.g., Ferguson & Farwell, 1975; Vihman, 1996). Further, it is likely that the linguistic and motor systems interface at multiple levels, with different linguistic representations (e.g., grammatical encoding, metrical and segmental plans), and speech control processes (e.g., temporal and spatial frames of various sizes) interacting at different levels. The unresolved issue of what constitutes units of speech production may be key to the successful convergence of psycholinguistic and speech production models into a “parallel, multiple unit view of the language/motor interface” (Smith & Goffman, 2004), since a consistent mapping between the discrete, invariant linguistic units and the continuous, context-dependent, co-produced units of speech production has been, and will likely continue to be, elusive.

The Schema Theory of motor control (e.g., Schmidt, 1975) offers a parsimonious account that may be flexible enough to accommodate some of these concerns. Recall that this theory posits the existence of memory representations called generalized motor programs (GMPs), which can be adaptively deployed in different contexts through the process of parameterization. Recently, researchers in the speech sciences have begun to consider how language units such as those discussed above might translate to GMPs and parameters in the motor speech domain. One plausible idea is that syllables, which are approximately the same duration as other putative units of action under this theory (i.e., about 200 ms, Schmidt & Lee, 2005), constitute
or access motor programs (e.g., Aichert & Ziegler, 2004; Ballard, Maas, & Robin, 2007). Working under this hypothesis, the question then becomes one of how general these motor programs are and what variables are parameterized. For example, if effector specification (which muscles are to carry out a movement) is conceived of as a parameter, then very general syllable patterns such as opening and closing may be GMPs that are parameterized to different manners of production and points of constriction along the vocal tract. Alternatively, GMPs may be less general, specifying distinctive features (e.g., place of articulation, manner, voicing) for each syllable, in which case they would need only to be parameterized in absolute amplitude and timing. Another possibility is that one distinctive feature (e.g., manner) may be classified by GMPs, while another feature (place of articulation) may be parameterized. Evidence for this was recently uncovered (Ballard et al., 2007) by applying the principle that transfer of skill learning is expected between members of the same GMP (e.g., Schmidt & Lee, 2005). However, there is also evidence that when a given action is produced repeatedly with a particular structure, that structure’s biomechanical properties (parameters) become part of the movement representation (GMP) (Park & Shea, 2003), highlighting the importance of flexibility and adaptability in a theory about units of production.

A major shortcoming of the Nijmegen model, and one that Levelt acknowledged (Levelt et al., 1999), is that it refrains from specifying how the highly complex motor processing of the articulatory system proceeds from the phonological stages. The abstract gestural scores that are fed into the articulatory system represent
movement goals, but do not explicate how articulatory trajectories or positions for achieving these goals in various contexts are computed. As a result, this and other psycholinguistic models are agnostic about the etiologies of any higher-level speech motor programming disorders (e.g., AOS) or any disruptions that arise downstream from the phonological processing level.

3.2. Four-Level Framework

In an attempt to further explicate the speech motor programming process neglected by these models, van der Merwe (1997) proposed a four-level framework that expanded on prior three-stage (i.e., linguistic encoding, articulatory programming, and execution) notions by including a motor planning stage. In the four-level framework, motor planning comprises the second stage, in which motor goals for the articulators in the production of the utterance are specified. In the planning stage, invariant phonemic units from the linguistic system are transformed into core motor plans, which specify temporal and spatial characteristics for each phoneme. These core motor plans are accessed from a sensorimotor memory store and then modified in spatial and temporal dimensions, depending on the phonetic (e.g., coarticulation, rate) or linguistic context. The process of phonetic adaptation of core motor plans in van der Merwe’s (1997) framework invokes the notion of internal feedback or “predictive stimulation” that takes the initial conditions of the system into account when setting spatial and temporal parameters, consistent with ideas developed in dynamic systems (e.g., Kelso & Tuller, 1981) and Schema Theory (e.g., Schmidt, 1975). Once these
modifications of core motor plans are computed, subroutines comprising adapted motor plans for each articulator (e.g., lip rounding, glottal closure) are temporally-arranged and deployed to the motor programming system, which controls the third stage in van der Merwe’s (1997) framework.

Whereas motor planning in the four-level framework operates to set spatial and temporal goals for the articulators, the motor programming stage selects and sequences motor programs for the muscles of each articulator. These muscle-specific motor programs represent the spatio-temporal and force parameters for individual muscles, and have also been thought of in terms of Schema Theory’s GMPs (e.g., McNeil et al., 2004), although Schema Theory does not specifically frame these concepts in terms of language production. McNeil and colleagues (2004) also suggest that whereas motor plans can be construed as strategies, motor programs are analogous to tactics for fluent, synchronized speech movements. Once motor programs are sent to the muscles as motor commands, the movement is initiated, and sensory feedback can be used (but is not required) to execute the complete movement (van der Merwe, 1997).

3.3. DIVA

Although van der Merwe’s (1997) four-level framework is better-specified at relevant levels for the discussion of the motor control of speech production, a shortcoming that it shares with psycholinguistic models of production is that it has little to say about learning. A recent neural network model of speech production,
Directions Into Velocities of Articulators (DIVA; e.g., Guenther, 1995; Guenther, Ghosh, & Tourville, 2006; Guenther & Perkell, 2004), not only models acquisition and learning, but provides a detailed account of the speech-language junction, incorporating both feedforward and feedback control. In addition, it accounts for imaging results and many long-observed speech phenomena, including articulation kinematics, contextual variability, motor equivalence, coarticulation, and rate effects. Further, it has been used to model several different speech patterns including those induced by articulator growth, constraint and perturbation, those related to change in hearing status, and stuttering (e.g., Guenther et al., 1998). The central tenet of the DIVA model is that movement goals are represented in auditory temporal space (Guenther & Perkell, 2004). This view comports well with observations of speech changes related to hearing status, as well as neurophysiological evidence for activation of the auditory cortex slightly before articulatory initiation (e.g., Levelt, et al., 1998) and for motor to sensory discharges induced by auditory masking of a speaker’s own voice (e.g., Bookheimer, Zeffiro, Blaxton, Gaillard, & Theodore, 1995; Hickok et al., 2000; Paus, Perry, Zatorre, Worsley, & Evans, 1996). Speech motor planning in the DIVA model involves mapping between the articulators and acoustic/auditory consequences or goals. In a computer simulation of this neural network, model parameters (i.e., synaptic weights encoding general sensory-motor transformations) are tuned during a “babbling” phase by various types of feedback, including tactile, proprioceptive, and auditory information. Phoneme-, syllable-, or word-specific mappings are then learned and maintained primarily via auditory feedback. During
repeated production attempts, the model develops feedforward computations that eventually subserve the production of fluent speech.

A revolutionary aspect of the DIVA model is that its representations and computational vectors (maps) each correspond to a specific set of neurons in the brain, based on neurophysiological and neuroanatomical evidence. Outputs and activation levels of these maps correspond to neural firing rates and postsynaptic potentials, and are modulated by synaptic weights that are tuned by experience to create three mappings (transformations between maps). The first, called “convex region targets,” maps phoneme string (or syllable) inputs from a speech sound map to phonemic targets coded in multidimensional orosensory (specified as vocal tract constrictions) and acoustic (specified as formant ratios) space. The speech sound map that encodes each native or frequently-occurring phoneme, syllable, word, or short phrase is reminiscent of Levelt’s mental syllabary, and its projections to primary motor cortex have been described as motor programs (e.g., GMPs in Schema Theory; Schmidt, 1975) or gestural scores (e.g., Levelt et al., 1999; Guenther et al., 2006). This speech sound map is hypothesized to lie in the left ventral lateral premotor cortex and functionally correspond to mirror neurons, which have been suggested by recent research to be involved in movement learning (e.g., Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). The “convex” region refers to the one-to-many multidimensional mapping of phonemes/syllables to vocal tract configurations and acoustic dimensions, and is a key feature of this model, accounting for motor equivalence and the other speech phenomena listed above.
According to the DIVA model, the projections from the speech sound map to higher-order auditory cortical areas in the superior temporal gyrus (representing anticipated sound patterns of the speaker’s own voice) and the orosensory areas of the somatosensory cortex (representing anticipated somatosensory stimulation patterns) and supramarginal gyrus (implicated in processing of phonological perception (e.g., Caplan, Gow, & Makris, 1995; Celsis, Boulanouar, Ranjeva, et al., 1999) and production (e.g., Damasio & Damasio, 1980; Geschwind, 1965)) form the basis of the auditory and somatosensory feedback control subsystems. These target representations are compared to somatosensory and auditory feedback, and error signals are computed and then mapped to the motor cortex as corrective motor commands. Although this feedback mechanism is too slow for online correction of natural speech, its error signals, if generated regularly enough (e.g., due to articulator growth, constraint, or persistent perturbation), become integrated into the feedforward control system (e.g., Guenther & Perkell, 2004).

The second mapping, called the “directional mapping,” computes the movement directions in articulatory space from the auditory and orosensory space goals. This mapping would best coincide with the recall schema notion in Schema Theory (e.g., Schmidt, 1975), which associates past parameters and outcomes with initial conditions to appropriately parameterize the current movement. Based on patterns learned during the babbling phase (Guenther & Perkell, 2004), this one-to-many mapping of a desired auditory consequence to compatible articulator movements is computed such that articulatory movements between adjacent segments are
minimized, automatically controlling for constraints and perturbations and providing a further mechanism to account for speech phenomena such as articulatory variability, anticipatory coarticulation, and carryover coarticulation. This directional mapping is encoded by projections from the premotor cortex (speech sound map) to the motor cortex (articulatory velocity and position maps) directly and via the cerebellum, forming part of the basis of the feedforward control subsystem.

The third mapping in the DIVA model, called the “forward model,” estimates the sensory consequences of movements by mapping orosensory feedback from the current state of the vocal tract (via the subcortical nuclei) to a somatosensory state map in the somatosensory cortex and an efference copy of the motor command (via the premotor cortex) to an auditory state map in the superior temporal cortex. These representations are transmitted via the pontine nuclei to the cerebellum, which in turn projects to the motor cortex where they are used to compute appropriate motor command sequences. This forward model system offers a potential description of the cortical mechanism underlying the recognition schema described by Schema Theory (e.g., Schmidt, 1975), which estimates sensory consequences based on initial conditions and previous outcomes. Once learned, these feedforward command sequences for each unit of speech obviate the need for sensory feedback during rapid speech, unless an unexpected constraint or perturbation occurs.

From this brief discussion of prominent models of speech production, it is evident that the current direction of research is aimed at integration of theories of motor control and cognition to develop new models that are informative to
developmental, clinical, neurophysiological, and behavioral concerns. The current emphasis on integration across multiple parallel processes involving cognitive, linguistic, motoric, and sensory systems requires an understanding of how these complex functions interact, not only in the speech domain, but in other activities of motor control and learning as well. Applying these models to dysfunctional systems, as will be discussed in the next chapter, has the potential to illuminate our understanding of normal functioning and further hone these models.
4.0. Apraxia of Speech

Since AOS was first described by Darley almost four decades ago (Darley, 1968, in Duffy, 2005), the nature of the disorder and its diagnosis and treatment have been the subject of active debate. The prevailing view, and one that is compatible with many theories, is that AOS disrupts the specification and/or control of spatiotemporal parameters of articulator movements for speech. However, there continues to be controversy over precisely what this view entails. As McNeil and colleagues have recently pointed out (McNeil, et al., 2004: 389),

…it is not a lack of theory or the inability to select the correct theory from the known alternatives that limits understanding of AOS…Neither is it the inability to construct critical experiments, nor the inability to select the appropriate level of description or contrast with the appropriate comparison group that limits understanding of AOS. It is, likewise, not the lack of neurologic or anatomic instantiation that limits AOS understanding. The most important impediment to theoretical and clinical advancement in AOS is, however, the lack of a comprehensive and clear definition that leads to an agreed-upon set of criteria for subject selection.

4.1. Defining AOS

AOS has undergone numerous definitions, from purely linguistic to purely motoric, depending on the perspective of the clinician and/or researcher. Of note, early definitions of AOS (e.g., Johns & Darley, 1970; Wertz et al., 1984) were problematic in that although they recognized AOS as a motor programming disorder and thus different from its clinical neighbors (i.e., the language disorder aphasia and the speech motor execution disorder dysarthria), ambiguous speech symptoms and a lack of appropriate models led to sets of subject classification criteria that lacked sensitivity
and specificity in diagnosing the disorder. Further, AOS shares clinical features with and is usually accompanied by aphasia and/or dysarthria, and is extremely rare in its pure form (without deficits in motor physiology, perception, or language), making the study of its unique characteristics difficult. Although an improved definition has been developed based on well-tested and instantiated models (see below), the underspecificity of earlier attempts at defining AOS continues to influence current work in ways that are both clinically and theoretically unfortunate. For example, the most widely used assessment tool that speech-language pathologists in the United States use to diagnose AOS, the Apraxia Battery for Adults-2 (Dabul, 2000), characterizes AOS by behaviors that are also consistent with phonemic paraphasia, a phonologic impairment feature of aphasia (McNeil et al., 2004). As McNeil and colleagues have also noted (1997; 2004), it is likely that the AOS database from which much of the popular understanding of AOS and its management has been derived is greatly confounded because many studies (e.g., Dronkers 1996; Halpern, Keith, & Darley, 1976; Rochon, Caplan, & Waters, 1991) have employed diagnostic criteria for AOS that “not only include subjects with phonemic paraphasias, but …perhaps guarantee their admission” (McNeil et al., 2004:392). The currently accepted, comprehensive and clear definition of AOS is as follows:

AOS is a phonetic-motoric disorder of speech production caused by inefficiencies in the translation of a well-formed and filled phonologic frame to previously learned kinematic parameters assembled for carrying out the intended movement, resulting in intra- and interarticulator temporal and spatial segmental and prosodic distortions. It is characterized by distortions of segment and intersegment
transitionalization resulting in extended durations of consonants, vowels, and time between sounds, syllables, and words. These distortions are often perceived as sound substitutions and as the misassignment of stress and other phrasal and sentence-level prosodic abnormalities. Errors are relatively consistent in location within the utterance and invariable in type. It is not attributable to deficits of muscle tone or reflexes, nor to deficits in the processing of auditory, tactile, kinesthetic, proprioceptive, or language information. In its extremely infrequently occurring “pure” form, it is not accompanied by the above listed deficits of motor physiology, perception, or language.

This chapter will review the evidence for the phonetic-motoric basis of AOS by discussing the current behavioral, neuroanatomic, and cognitive understanding of the disorder.

4.2. Behavioral Definition

4.2.1. Speech Behaviors

Behaviorally, AOS can be measured in terms of perceptual, acoustic, and physiological evidence. Although perceptual measures are perhaps the most relied-upon for everyday clinical purposes, as the definition above indicates, they are not always reliable for distinguishing errors of paraphasic versus apraxic origin. For instance, because of categorical perception, sound distortions characteristic of AOS are commonly perceived as substitutions. When the two disorders are overlaid, as they often are in patients with stroke, the reliability of perceptual classification becomes particularly tenuous. Evidence from careful narrow transcription and acoustic and kinematic measures can augment perceptual evidence, although it is generally recognized that no single behavioral symptom is unique to AOS and not shared by
another disorder. Differential diagnosis is achieved via observation of a constellation of behaviors.

The key necessary and sufficient acoustic/physiologic speech features of AOS, and those that are not found in phonemic paraphasia, are lengthened segment and inter-segment durations and sound distortions (McNeil et al., 1997; 2004). These disruptions manifest in both “on-” and “off-target” utterances of any length (consonants and vowels, syllables, words, phrases, and sentences). For example, sound substitutions, a prominent feature of phonemic paraphasia, which is often concomitant with AOS, are also subject to distortion in AOS. Devoicing of voiced phonemes is a common type of distortion that may be perceived as sound substitution. The perceptual effects attributable to lengthened durations in AOS include perceived overall slow rate, distorted speech, and prosodic abnormalities including stress neutralization and sound, syllable, and word segregation. Reduced anticipatory coarticulation secondary to lengthened inter-segment duration contributes to the perception of speech distortion in AOS (e.g., Ziegler & von Cramon, 1985).

In addition to these cardinal features of AOS, a number of other symptoms commonly occur with the disorder, although they do not differentiate it from phonologic impairments in aphasia (McNeil et al., 2004). These include difficulty increasing speech rate while maintaining sound integrity (e.g., Kent & McNeil, 1987), so-called “islands” of less-impaired speech (usually in short nonpropositional utterances), relative consistency of errors with regard to location and type across repetitions of the same utterance (McNeil, Odell, Miller, & Hunter, 1995), increased
errors with longer or more phonetically difficult utterances, unsuccessful self-initiated repair attempts resulting in audible or visible “groping” or searching behavior, difficulty initiating speech, and the impression of increased effort or struggle. While AOS shares many perceptual features with phonologic production errors in aphasia, it is important to realize that the underlying causes of these behaviors may be quite different, and will be discussed below in the context of models of speech production.

A recent report from the Academy of Neurologic Communication Disorders and Sciences (ANCDS) (Wambaugh, Duffy, McNeil, Robin, & Rogers, 2006:xvii) summarized the primary clinical characteristics of AOS as “1) slow rate of speech, typified by lengthened sound segments and intersegment durations, 2) sound distortions, 3) distorted, perceived sound substitutions, 4) errors that are relatively consistent in terms of type and invariable in terms of location, and 5) prosodic anomalies.”

Another clinical neighbor that is often confused with AOS is dysarthria, a category of heterogeneous disorders that affect speech motor execution. In the case of a single unilateral lesion, the type of dysarthria that would be expected, and that may be mistaken for AOS, is unilateral upper motor neuron dysarthria (e.g., Duffy, 2005). Whereas dysarthric errors of timing, range, rate, force, steadiness, strength, and direction of movement are attributable to deficits of muscle tone and reflexes that affect multiple speech systems (phonation, resonance, articulation, and prosody) (e.g., Duffy, 2005), timing and kinematic disturbances in AOS originate at the motor planning/programming level of speech production. Physiological evidence of
abnormal temporal and kinematic features during speech, in the presence of otherwise normal muscle tone and reflexes, provide evidence for its underlying phonetic-motoric nature, discernable from purely linguistic or motor execution disorders. For instance, persons with pure AOS can achieve normal average or peak velocities of articulator movements in the context of a speech segment that is nonetheless of abnormal (increased) duration, although velocity-displacement profiles are not always intact (Robin, Bean, & Folkins, 1989). Increased segment durations of vowels (e.g., Collins, Rosenbek, & Wertz, 1983; Freeman, Sands, & Harris, 1978; Seddoh et al., 1996), consonants (e.g., Kent & Rosenbek, 1983), and transitions (e.g., Kent & Rosenbek, 1983; McNeil, Caligiuri, & Rosenbek, 1989), as well as longer and more variable stop gap and vowel durations (Seddoh et al., 1996), disturbances in voice onset time (Ballard, et al., 1997; Freeman et al., 1978; Itoh, Sasunuma, Tatsumi, & Kobayashi, 1979; Kent & Rosenbek, 1983) and lingual-velar timing (Itoh, Sasunuma, & Ushijima, 1979) also demonstrate disruption of spatiotemporal coordination in AOS and help differentiate it from aphasia.

4.2.2. Nonspeech behaviors

Unlike the dysarthrias, it has been argued that AOS can theoretically occur without clinically apparent consequence to nonspeech oromotor skills (Duffy, 2005; Ziegler, 2003). However, because of measurement issues, concomitance of other disorders, and the confounded subject selection criteria in some studies of AOS, it is unclear whether these cases exist. It has been estimated that in 48% to 85% of reported
cases of AOS, apraxic symptoms (i.e., nonverbal oral apraxia, NVOA) are also manifest in the nonverbal domain (McNeil, Doyle, & Wambaugh, 2000). It is also possible that in patients without NVOA, impairments of AOS may extend to motor control of some nonspeech tasks, particularly when they approximate the complexity of speech movements (Ballard et al., 2000). On an integrative view of oromotor control, speech and nonspeech activities share principles and substrates in common as parts of a more general motor control system, so it is logical to expect an interaction of some skills in each domain (Ballard et al., 2003; Robin et al., 1997). Indeed, many nonspeech sequelae have been associated with AOS, and nonspeech assessment can be used to disambiguate the contributions of motoric and linguistic impairments to disrupted speech (Ballard et al., 2000).

Visuomotor tracking (VMT) is a paradigm that has been applied to assessment of nonspeech motor control in AOS. In VMT, a subject is required to track a moving visual target on an oscilloscope or computer screen by moving a transducer with the articulators (e.g., lip, jaw, voice) Hageman and colleagues (1994) had subjects track predictable (sinusoidal) and unpredictable (random) waveforms. In unimpaired subjects, tracking error is expected to be reduced in the predictable condition (vs. unpredictable), in which development of a motor plan/program is possible. Results indicated that although subjects with AOS were impaired at tracking the predictable waveform, they performed no differently than unimpaired controls when tracking an unpredictable waveform. These results support the notion that AOS entails a deficit in the development and/or implementation of motor programs/plans. These data have
been replicated and extended to other tracking conditions, tasks, and patient groups (e.g., Ballard & Robin, 2007, Clark & Robin, 1998), demonstrating that subjects with conduction aphasia perform no differently than controls on this task (Robin et al., submitted, but see one subject in Clark & Robin, 1998), and that subjects with ataxic dysarthria are impaired in tracking both predictable and unpredictable waveforms (Hageman et al., 1994). Furthermore, a correlation (r = 0.68 - 0.84) between tracking performance and three perceptual measures of speech intelligibility (overall articulatory precision, speech defectiveness, and overall intelligibility) has been shown in AOS (Robin et al., submitted), supportive of the argument for interdependence of control mechanisms for speech and nonspeech oral movements.

Patients with pure AOS were also compared to those with ataxic dysarthria and conduction aphasia, as well as to normal controls, in a study of isometric force and static position control of the articulators (McNeil, Weismer, Adams, & Mulligan, 1990). While subjects with AOS were significantly less accurate and more variable than normal controls in force and posture control, they were not differentiated from those with ataxic dysarthria on these measures. Many authors (e.g., McNeil et al., 2004; Ballard et al., 2000) have noted that it is not surprising that the deficits of AOS sometimes resemble those of ataxic dysarthria, given that ataxic dysarthria typically occurs as a result of damage to the cerebellum, which is heavily involved in movement timing and sensorimotor control. These disorders are, however, clinically differentiable, and illustrate how similar symptoms may arise from different etiologies. An additional and unexpected finding from the study by McNeil and
colleagues (1990) was that subjects with conduction aphasia performed between normal controls and the apraxic and dysarthric groups (i.e., not differently than any group), raising the possibility of the presence of motor involvement in conduction aphasia, or the concomitance of some degree of AOS in these subjects. Other data from EMG have failed to differentiate these groups on the basis of measures such as antagonistic muscle co-contraction, continuous undifferentiated muscle activity, and EMG shutdown of the orbicularis oris muscle (e.g., Forrest et al., 1991), although contradictory findings exist (Fromm, Abbs, McNeil, & Rosenbek, 1982, Hough & Klich, 1987). Taken as a whole, the data from nonspeech assessment supports the view that while AOS predominantly affects speech skills, it exerts measurable influence on other motor control tasks as well.

4.3. Neuroanatomic Substrates

AOS most commonly occurs as a result of focal brain injury (stroke), although other etiologies (e.g., neurodegenerative disease) have been reported (Duffy, 2005; Gorno-Tempini et al., 2004; Nestor et al., 2003). Despite debate over the precise localization of lesions responsible for AOS, there has been relatively broad agreement that left posterior inferior frontal gyrus (Broca’s area) and anterior insula are important for speech articulation. The symptoms of AOS are commonly associated with lesions in these cortical areas secondary to left middle cerebral artery (MCA) occlusion, although attempts at more precise localization (i.e., dissociating brain regions necessary and sufficient for AOS versus aphasia) have yielded ambiguous data. An
influential lesion overlap study by Dronkers (1996) reported that all patients with AOS who were studied (N=25) had a common infarction to the superior tip of the left precentral gyrus of the anterior insula. In addition, in 19 left-hemisphere stroke patients without AOS who were studied, this region was completely spared. Dronkers (1996) concluded that these findings were evidence for a robust double dissociation, and that the common lesion site shared by all patients with AOS must be specialized for speech motor planning.

Although Dronkers’ (1996) finding is compatible with other reports of this specific brain-behavior relationship in AOS (single case studies; Marien, Pickut, Engelborghs, Martin, & De Deyn, 2001; Nagao, Takeda, Komori, Isozaki, & Hirai, 1999), nonfleunt progressive aphasia (Gorno-Tempini et al., 2004; Nestor et al., 2003), and functional magnetic resonance imaging (fMRI) studies of normal speech production (e.g., Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000; Wise, Green, Buchel, & Scott, 1999), this relationship has not been borne out in all studies of normal speech (e.g., Riecker et al., 2000) and AOS. For example, of the four “pure” apraxic subjects studied by McNeil and colleagues (1990), two did not have any insular damage. A more recent study of 11 patients with AOS (2 “pure,” 7 with concomitant Broca’s aphasia, and 2 with concomitant mixed nonfluent aphasia) found insular involvement in only three patients (27%) (Jones, Peach, & Schneck, 2003). A larger-scale study of 80 acute left MCA stroke sufferers (40 with insular damage and 40 without) failed to find an association between AOS and insular damage, although it
did find a strong association between AOS and stroke or hypoperfusion to Broca’s area (Hillis, et al., 2004).

If the anterior insula must be damaged for AOS to occur (Dronkers, 1996), how can these reports of AOS in the absence of insular damage be reconciled? At least three explanations suggest that although the association between anterior insular damage and chronic articulatory deficits may be reliable, Dronkers’ strong causative conclusion may not be completely supported. First, as suggested by Hillis and colleagues (2004), apraxia of speech and insular lesions may simply be coincident but independent results of large MCA lesions. The patients selected by Dronkers (1996) had chronic speech articulation deficits (duration over one year), indicative of relatively large lesions, since rapid resolution of AOS is observed after relatively small ischemia (e.g., Marien et al., 2001). The common lesion site found by overlap methodology by Dronkers may have been a reflection of a particularly vulnerable region to large MCA strokes. Indeed, 94% of cases of acute MCA occlusion have insular involvement (Finley et al., 2003). Relatedly, an important part of the logical argument that would need to be made if the causative role of the anterior insula in AOS were to be upheld is that damage to this area always causes AOS and without damage to this area, AOS cannot occur. In other words, whereas Dronkers assessed the probability that AOS was related to a specific lesion site, the probability that a lesion at that site was related to AOS was not addressed. These were precisely the arguments tested by Hillis and colleagues (2004), who found that in the presence of left insula damage (in general, in the anterior portion, or in the superior precentral
gyrus of the anterior insula), AOS was absent more often than it was present. Further, when the insula was not damaged or hypoperfused, AOS was present in 37-42% (depending on size of the area studied within the insula) of stroke patients. A much stronger association was found between Broca’s area abnormalities and acute AOS (87%) (Hillis et al., 2004).

Finally, these studies must also be interpreted with caution because of the issues raised above with regard to subject selection criteria. The subjects in Dronkers’ (1996) study were selected if they met four criteria for AOS, based on characteristics described by Wertz, LaPointe, and Rosenbek (1991): “(1) effortful, trial-and-error, groping articulatory movements and attempts at self-correction, (2) dysprosody unrelieved by extended periods of normal rhythm, stress, and intonation, (3), articulatory inconsistency on repeated productions of the same utterance, (4) obvious difficulty initiating utterances” (Dronkers, 1996:161). Similar criteria were employed in other recent studies (e.g., Hillis et al., 2004; Jones et al., 2004); however, according to the current understanding of AOS and neighboring disorders, these features do not differentiate AOS from phonemic paraphasia (McNeil et al., 1997; 2004). Therefore, it is likely, and perhaps guaranteed (McNeil et al., 2004), that the samples studied were contaminated with subjects who did not have AOS. What can be gleaned from these contradictory neuroimaging results supports what theorists (e.g., Guenther & Perkell, 2004; van der Merwe, 1997) have suggested about the organization of planning and programming of speech articulation, namely, that these functions are likely subserved by a distributed network of processes and neural resources, and that functional
inactivation of any of them may disrupt the chain of events required for fluent speech. Further study is clearly required to understand the lesion-deficit relationships along this chain.

4.4. Cognitive Definition

Considering all of the controversial behavioral and neuroanatomical findings, how can AOS be unambiguously defined? Differential diagnosis of AOS and its clinical neighbors must invoke inferences about the underlying disturbance in the cognitive architecture subserving fluent speech. Although theories of speech and language production abound (previous chapter), few explicitly model AOS; yet most can accommodate the distinction between this disorder and phonological or motor execution disorders. In addition, theories of motor control and programming can also be productively applied to developing a cognitive definition of AOS.

4.4.1. Psycholinguistic Theories

As discussed in Chapter 2, psycholinguistic models of speech production are underspecified at the levels relevant for modeling AOS. However, they can be productively applied to detailing phonological errors, and thus help disambiguate phonological paraphasias from errors arising further downstream in the translation of intact phonological representations to motor commands. Earlier definitions of AOS (e.g., Wertz et al., 1984) proposed sound substitution and sequencing errors to be features of AOS. However, those perceptual features were likely aspects of
concomitant aphasia or, as suggested by McNeil and colleagues’ (1997) definition, they were misperceptions of AOS-driven distortions in phoneme duration and voice onset time. The Nijmegen model and others similar to it (e.g., Roelofs, 1997; Dell, 1986) illustrate that sound sequencing and substitution errors occur at different levels of processing than distortions. For example, phonological errors are possible in the processes of phonological encoding, wherein a word’s phonological form may not be properly retrieved, or in the process of phonetic encoding, wherein phonological words may not bind to or activate the appropriate syllables. Disruptions in phonetic assembly, for less-frequent representations not retrieved from the syllabary, may also lead to phonological speech errors. In addition, phonetic codes may fail to index the appropriate gestural scores, or buffers at any stage may fail. All of these disruptions involve discrete, static, context-independent linguistic representations, and therefore result in phonemic errors such as sound or syllable perseveration, anticipation, metathesis, or substitution. These stages are disrupted in instances of phonemic paraphasia.

Although the psycholinguistic models mentioned here were not developed to account for phonetic-motoric production disorders such as AOS, this disorder would be attributed to disturbance further downstream, in the unpacking or implementation of gestural scores. In an attempt to explicate the phonologic-motoric translation process further, Roelofs (1999) suggested that contextual adaptation occurs by overlapping gestural scores, and that relative levels of activation of competing motor programs determine which will be implemented. It still remains unclear, however,
how the motor system can exert bottom-up influences on these processes (Ziegler, 2002). Although not addressed by Roelofs, others (McNeil et al., 2004) have speculated that a slowing in the process of motor program selection might result in symptoms characteristic of AOS, such as prolonged segments, transitions, and intervals. These hypotheses await experimental verification.

4.4.2. Dual Route Theory

Another possibility for the etiology of AOS hypothesizes that AOS results from a disruption in the ability to retrieve phonetic codes from the syllabary or similar memory stores (the “direct” route), forcing phoneme-by-phoneme assembly of each syllable or word (the “indirect” route). This “dual route” theory (Varely & Whiteside, 2001; Whiteside & Varley, 1998) assumes that frequently-used syllables and words are stored as verbo-motor patterns in a manner that is similar to the syllabary. These verbo-motor patterns specify relative force and timing components of movement representations for coordinative structures, and are described as movement synergies or “learned links between muscle commands” (White & Varley, 1998:223), and are also reminiscent of GMPs. According to the dual route theory, AOS results from a breakdown in the ability to store and/or access verbo-motor patterns. Thus, the person with AOS must rely on the indirect route for assembling each utterance phoneme-by-phoneme. This process is assumed to be more resource-demanding and inefficient, and particularly in persons with brain damage who may have reduced resources or allocation inefficiencies (e.g., Ballard, Robin, Woodworth, & Zimba, 2001), reliance
on this indirect route for speech production may result in the types of speech characteristics exhibited by those with AOS. The dual route theory has been criticized on several points, including that it lacks falsifiability and explanatory power over other models (Ballard et al., 2001), it does not accommodate programming units that may be smaller than the word or syllable (e.g., Rogers & Spencer, 2001), it does not explain nonspeech symptoms of AOS (Ballard et al., 2001), and it cannot explain, if the direct route is inaccessible, how persons with AOS can often produce “automatic” speech (i.e., overlearned utterances, short, non-propositional words and phrases) better than other utterances (Rogers & Spencer, 2001).

4.4.3. Reduced Buffer Capacity Hypothesis

Another theory of AOS, the reduced buffer capacity account (Rogers & Storkel, 1999) posits a reduction in processing capacity, specifically in the phonological output buffer that holds motor programs prior to execution. According to this theory, AOS represents an inability to effectively hold more than one syllable in the buffer, resulting in the need for these speakers to program one unit at a time. According to Rogers and Storkel, this accounts for the segmentalized nature of speech production in AOS.

This hypothesis was derived based on a series of experiments conducted with normal speakers that demonstrated the “phonological similarity effect” (Rogers & Storkel, 1998). This effect refers to the observation that in a speeded word-reading task, when initial phonemes of monosyllabic word pairs share articulatory features, the
onset latency for producing the second word is longer than when the words are dissimilar. These data were taken as evidence that sublexical units comprising monosyllabic words needed to be assembled for each production, and that temporary inactivation of a recently-used program required greater processing resources to overcome, resulting in an increased latency when that unit needed to be reprogrammed (Rogers & Storkel, 1998). Rogers and Storkel (1999) predicted that when unimpaired speakers were allowed to program both words at one time (i.e., the words were presented simultaneously and sufficient time was allowed), the phonological similarity effect would disappear because the relevant programs could be maintained in the buffer long enough for the production of both words (i.e., reprogramming of the similar element would not have to occur). However, if the buffer capacity were reduced and could not accommodate both words in the pair, each word would have to be programmed and buffered separately and sequentially. This would result in the phonological similarity effect, since the same unit would have to be programmed twice, and it would require greater processing to overcome the temporary inactivation and program it a second time. Indeed, this result was obtained for subjects with AOS, but not those with aphasia or normal speech and language skills (Rogers & Storkel, 1999), supportive of the claim that persons with AOS had a reduced-capacity phonological buffer.

The reduced buffer capacity hypothesis is consistent with other accounts of specific short-term memory/phonologic loop deficits in AOS (e.g., Waters, Rochon, & Caplan, 1992), and resonates with others who have suggested the influence of reduced
resource capacity or allocation in AOS (e.g., Kent & McNeil, 1987; Whiteside & Varley, 1998). Clark and Robin (1998) have also invoked the notion of a resource allocation deficit to explain why some subjects with AOS can program or parameterize well in the context of a VMT task, but cannot successfully do both at once. While Rogers and Storkel’s (1998; 1999) view of AOS is appealing because of the perceptual features that it accounts for (e.g., segmentation, reduced rate, reduced coarticulation, length effects), it has been criticized on the grounds that it does not explain the nonspeech deficits evidenced in AOS, or the predominant type of speech error, namely, distortions. Further, the presence of a phonologic similarity effect has not been demonstrated in all studies of persons with AOS (e.g., Rochon et al., 1991).

### 4.4.4. Four-Level Framework

The four-level framework developed by van der Merwe (1997) allows specific assignment of different disorders to different stages of the framework, although van der Merwe does acknowledge that since different levels do not necessarily correspond to non-overlapping neural resources, symptoms may be shared by multiple disorders. In this framework, linguistic-symbolic deficits such as those seen in aphasia, including phonemic paraphasias, are attributable to disturbances in the first level of processing. Impairment in the second level of processing, the motor planning stage, are assumed to result in difficulty identifying, recalling and indexing core motor plans to corresponding phonemes, difficulty sequencing movements within motor plans for each phoneme and across phonemes, difficulty adapting core motor plans within the
boundaries of motor equivalence to the spatiotemporal dimensions compatible with the context in which a phoneme is to be produced, and difficulty transferring adapted motor plans to the motor programming system. According to van der Merwe, any of these difficulties would manifest as the kinds of behaviors evidenced in persons with AOS, such as slow, effortful, and distorted speech (van der Merwe, 1997).

Impairments at the next two levels of speech production, motor programming and execution, would cause disruptions that would manifest as various types of dysarthria (van der Merwe, 1997).

4.4.5. Coalitional Breakdown

According to Kelso and Tuller’s (1981) theory of coalitional control as mentioned in Chapter 2, speech, like any other motor behavior, is organized in functional units, or coalitions, that emerge out of an individual’s interaction with the environment. According to Kelso and Tuller, “apractic deficits may be more properly viewed as a breakdown in the syngergistic relationship between the individual and the environment as defined by the behavioral goal” (1981:233). Breakdowns in these syngergistic relationships may result in instability of neuromotor and myomotor coalitions or attractor states, and reduce the consistency of the temporal invariance that emerges from these states. Spatiotemporal inconsistency and variability, if pathological, would manifest as movement distortions. Although these notions appear compatible with AOS, this model does not further elaborate the interface of the
language and motor systems, nor does it provide a framework for considering different error patterns which may be related to different linguistic and/or motoric pathologies.

4.4.6. Schema Theory

Schmidt’s (1975) Schema Theory has recently been used to develop a perspective on AOS that is grounded in an influential and well-tested model of motor control. Framed in terms of Schema Theory, AOS may be thought of as a disruption in the ability to select or activate a generalized motor program and/or set the correct parameters for the execution of movements required for speech production (Clark & Robin, 1998; Ballard et al., 2000). This hypothesis is supported by the nonspeech VMT results presented above, demonstrating that subjects with AOS are impaired in tracking tasks which require development and/or implementation of a movement plan/program. In another study, Clark and Robin (1998) examined the differential learning of GMPs and parameterization in the context of learning a tracking skill. By subtracting the tracking error related to the absolute timing and amplitude parameters, they used the residual root-mean-square (RMS) tracking error as an index of GMP error. Results indicated that, relative to controls, some subjects with AOS appeared to show disruption in the GMP but normal parameterization, whereas others demonstrated the opposite pattern. These results, although not derived from the speech domain, are nonetheless compatible with McNeil and colleagues’ (1997) definition of AOS. As stated by Robin et al. (in press):
While this idea may appear inconsistent with the proposal by McNeil et al. (1997) that AOS represents a disruption of the mapping between linguistic and motor levels of processing, this account does not necessarily claim that the mapping between other motor goals and motor levels of processing is intact. Rather, AOS can be viewed as the manifestation in speech of a more widespread problem with accessing, selecting, or parameterizing GMPs for oral movements.

The precise nature of the breakdown of GMPs and/or parameterization in AOS is still unclear. Citing evidence of predominating sound distortions in AOS, some researchers have suggested that the structural integrity of GMPs is damaged, and not their retrieval or selection (Aichert & Ziegler, 2004). Alternatively, Kent and Rosenbek (1983) proposed that unavailability of information about initial conditions and movement consequences constituted the programming disruption in AOS. As discussed in Chapter 2, Schema Theory regards these pieces of information as critical for updating and utilizing the recall and recognition schemas for movement programming and self-evaluation, respectively. Without these bits of information, the recall schema would not function to select the appropriate parameters and apply them to a GMP for movement production. Additionally, the recognition schema would not be supplied with information necessary for the self-evaluation process. Since the ability to update these schemas is considered critical for motor learning to occur, this proposal would entail the prediction that persons with AOS would be impaired in aspects of motor learning. Indeed, the VMT data discussed above suggest that persons with AOS are impaired in the ability to develop new motor programs (Clark & Robin,
Clinical evidence citing limited transfer of trained skills in therapy also lends support to this view (e.g., Wambaugh, 2002).

4.4.7. Two-Stage Model

Even more recently, the two-stage model of motor programming explicated by Klapp (1995; 2003) has been invoked to develop further hypotheses about the nature of the motor programming disturbance in AOS. Recall that the preprogramming stage of the model, INT, is responsible for specification of the internal structure of a unit of movement. If this process were disrupted in speech, the expected outcome would be distortion, disrupted prosody, and slowed rate. The second stage of the model, SEQ, is responsible for the serial ordering of preprogrammed units of movements. A disruption in SEQ would produce errors in the sequencing of units. Following these predictions, AOS has been hypothesized to reflect a disruption in the INT process, but not SEQ. Because parameterization of a GMP is consistent with the INT process, this account is compatible with others postulating impaired selection, activation, integrity, or parameterization of GMPs (e.g., Aichert & Ziegler, 2004; Ballard et al., 2001; Clark & Robin, 1998; McNeil et al., 1997).

Evidence for this hypothesis comes from several recent findings. For instance, in VMT, subjects with AOS tend to resort to a preferred tracking frequency when the target is removed, although it may not be the one required by the task (Ballard & Robin, 2007). This suggests that when subjects are able to activate and implement a
stable GMP (albeit perhaps a task-inappropriate one), they do not make errors that would suggest a sequencing disturbance.

Several recent studies have explicitly tested the performance of persons with AOS under conditions which affect INT and SEQ. For instance, Deger and Ziegler (2002) found that, unlike persons with aphasia only or normal controls, persons with AOS had particular difficulty assembling motor programs for non-repetitive syllables (e.g., /daba/ versus /dada/). Wright, Magnuson, Robin, Maas, and Ballard (2005; in Magnuson, Robin, & Wright, 2006) applied the self-selection paradigm to the programming of Morse-code-like finger tapping patterns to individuals with AOS. As predicted by the Klapp model (1995), a sequence length effect (in terms of number of button presses) was exerted on simple RT, the index of SEQ. Further, this length effect was present in both normal controls and subjects with AOS, confirming that the SEQ process was not disrupted in AOS. However, study time (ST) evidence emerged that was consistent with the hypothesized INT impairment in AOS. First, ST was significantly longer overall in subjects with AOS compared to those without, indicating a disruption localized to the INT (as opposed to SEQ) process. Second, the complexity effect (i.e., the difference in preprogramming (INT) demands of long-versus short-duration units) was significantly greater in subjects with AOS than normal controls. Converging evidence for this general pattern of results was reported by Maas, Robin, Wright, and Ballard (in press, Experiment 1), although the complexity effect was not significant for either group in this study.
Similar results have been obtained in using the self-selection paradigm to test the INT/SEQ model on speech motor programming in AOS. Maas and colleagues (in press, Experiment 2) had participants with AOS, participants with aphasia only, and normal controls produce the syllable “ba” with different durations (long versus short) and different numbers of repetitions (1 versus 4), similar to the Morse-code-like finger movement studies conducted earlier. A key result provided evidence supporting the localization of the deficit in AOS to the INT process: Overall, STs were significantly longer for speakers with AOS (but not for speakers with aphasia only) than for normal controls. Also consistent with finger movement findings (Maas, Robin, Wright et al., in press, Experiment 1; Wright et al., 2005), there was no significant difference in RT between the two groups, suggesting that SEQ was intact for serial ordering of units of speech. The absence of a sequence length effect on RT demonstrated that speakers from both groups were able to integrate the syllable sequences into single units. However, longer STs for the apraxic speakers indicated that they required more time for organizing these integrated movement patterns.

Although Deger and Ziegler (2002) did not measure the time required for preprogramming (INT), their simple RT finding that participants with AOS were impaired in organizing alternating-syllable patterns into single units is consistent with an INT deficit. The discrepancy between this report and that of Maas, Robin, Wright and colleagues (in press, Experiment 2) which did provide evidence for “chunking” in AOS might be attributable to differences in severity between the two samples of patients (Maas, Robin, Wright et al., in press). A potential conclusion based on these
studies is that persons with AOS exhibit a deficit in the response preparation stage of motor programming (INT) that, in its milder forms is responsible for inefficient movement unit organization (e.g., Maas, Robin, Wright et al., in press), but in its more severe forms causes the deconstruction of movements into separate smaller units (e.g., Deger & Ziegler, 2002).

These findings of a disruption of the INT subprocess of motor programming, for both speech and nonspeech activities, provides compelling evidence that AOS may best be considered a central, not a speech-specific, disorder. This perspective can thus accommodate findings of nonspeech disturbances in AOS (e.g., Clark & Robin, 1998; Hageman et al., 1994; McNeil et al., 1990). In the brief discussion of the models, theories, and frameworks of AOS presented here, it is evident that the most powerful explanations are those that can accommodate the greatest number of empirical observations of speech and nonspeech behaviors in AOS. Currently, models of general motor control such as Schema Theory and the INT/SEQ model of motor programming, when considered in conjunction with specific models of the speech-language interface, offer this advantage. Although an account of AOS has not yet been developed in the context of the DIVA model, the detailed theory of speech motor control discussed in Chapter 2 (e.g., Guenther & Perkell, 2004), the neurophysiological and neuroanatomical grounding of DIVA, as well as its ability to account for acquisition patterns and various speech phenomena, hold promise for a forthcoming application to AOS, potentially in a manner compatible with general models of motor control.
4.5. Application of Motor Learning Concepts to Treatment of AOS

Current research, as outlined in this chapter, strongly supports the notion that AOS is a disorder of motor control. Therefore, it is logical to consider applying principles that have been found effective for motor learning to the clinical management of this disorder. The set of principles of motor learning discussed in detail in Chapter 2 have guided our understanding of how humans learn skilled actions, primarily in the limb domain. However, recently, researchers have begun to acknowledge that speech is also a skilled action that shares many features in common with other skills, in terms of motoric, neural, and cognitive architecture. Principles of motor learning that have been studied extensively in the learning of limb movements and actions have great potential to enhance the efficiency and/or effectiveness of learning when applied to the speech domain. Indeed, the little evidence that has been collected to date supports this assertion and offers promising directions for future research.

As described in Chapter 2, principles of motor learning can be divided into variables which determine the conditions of practice in which a new skill is learned, and variables which specify dimensions of the augmented feedback that is given to the learner. In addition, some researchers have identified important pre-practice conditions, or learning variables which are present before practice begins, that can influence learning. Specifically, pre-practice considerations typically include motivational factors such as deciding whether to include the patient in aspects of
treatment planning such as target selection (e.g., Strand & Debertine, 2000) and setting specific goals (e.g., McNeil et al., 1997). These practices are generally regarded as conducive to learning because of the motivational effects that they impart (Maas, Robin, Austermann Hula et al., in press). Other pre-practice considerations in terms of how the learner is instructed to perform the task are also influential variables. For example, learning appears to be enhanced when participants are instructed to adopt an external focus of attention (i.e., focus on something outside of one’s body) (e.g., Wulf et al., 1998; 2001). Although it is not completely certain what this would entail in the speech domain (perhaps focusing on the listener or on communicative goals or environmental outcomes), this phenomenon was recently applied to the oromotor domain in a tongue force learning task (Freedman, Maas, Caligiuri, Wulf, & Robin, 2007). The results of this study suggest that Wulf and colleagues’ (2001) constrained action hypothesis, developed to explain the learning benefit that an external focus of attention confers on limb movement skill-learning, applies to the oromotor domain as well. Specifically, participants learning to produce a particular movement force with their tongues were more accurate and stable when instructed to adopt an external, rather than an internal, focus of attention. More research is required to determine if these hypotheses can be confirmed for speech production as well.

Several principles of motor learning relevant to practice conditions have also recently been applied to the speech domain. For example, in the treatment of AOS, it has been shown that training more complex skills first (i.e., consonant clusters) improved transfer to less complex skills (i.e., singletons) (Maas, Barlow, Robin, &
Enhanced retention of speech production skills has been demonstrated under conditions of random practice (i.e., mixed target sequences) instead of blocked practice (i.e., homogenous blocks of targets) in normal speakers (Adams & Page, 2000) and speakers with AOS (Knock, Ballard, Robin, & Schmidt, 2000). Ballard and colleagues (2007) have also provided evidence for applying the principle of variable practice (which has been shown to enhance limb learning relative to constant practice, e.g., Wulf & Schmidt, 1997) to the training of voiced phonemes in patients with AOS. Although most of these studies (cf. Adams & Page, 2000) have involved single subject experimental designs with small numbers of subjects (i.e., 2-4), the results that they have produced offer justification for the application of principles of practice conditions to the speech domain and call for more research to replicate and extend these findings.

In the realm of principles of augmented feedback, Clark & Robin (1998) first provided evidence that reduced feedback frequency facilitates retention of a new oral motor skill in healthy speakers. Steinhauer and Grayhack (2000) subsequently applied the principle of reduced feedback frequency to the motor learning of a vowel nasalance task in unimpaired speakers, and found an inverse relationship between the frequency of feedback (0%, 50%, or 100%) and measures of performance and learning of the skill. Similarly, Adams, Page, and Jog (2002) demonstrated that providing summary feedback after 5 trials as opposed to after every trial enhanced retention of a novel speech skill in normal subjects and speakers with Parkinson’s Disease. Overall, the results of these studies are supportive of the role of principles of motor learning in
speech learning and re-learning in the context of neurogenic disorders, and provide justification for further research.

This dissertation represents a series of studies examining the application of augmented feedback variables to normal speech skill learning and the re-learning of speech skills in adults with AOS. As advocated above, this program of research highlights the distinction between acquisition, or temporary performance changes during practice, and true learning, the relatively permanent changes in the capability for movement. Speech-language pathologists naturally tend to employ techniques that facilitate performance in the treatment setting, but critically, motor learning studies (discussed in Chapter 2) have demonstrated that variables that enhance performance during training may actually impede progress toward the more important goals of long-term retention and generalized improvement of other related skills. Differences in measures of acquisition, retention, and transfer will be explored in each of the following experiments. The results of these experiments are expected to make important contributions to the clinical and theoretical literatures, as they represent the first studies in which the effects of three feedback variables (i.e., feedback frequency, feedback timing, and feedback control) have been investigated in speech skill learning in AOS, as well as the first investigations of the effects of a feedback variable (i.e., learner-controlled feedback) on different motor programming subprocesses.
5.0. Effects of Feedback Frequency and Timing on Acquisition, Retention, and Transfer of Speech Skills in Acquired Apraxia of Speech

The purpose of the current experiments\(^5\) was to explore the effect of feedback frequency and timing on the acquisition and retention of speech skills in persons with apraxia of speech (AOS). These data extend work on principles of motor learning in limb motor learning to the treatment of speech, as well as provide more data on treatment efficacy in AOS. The theoretical motivation for this work has its genesis in the Schema Theory of motor control and learning (Schmidt, 1975). However, other views of motor learning and programming (e.g., Li & Wright, 2000; Shea & Wulf, 2005) that can be integrated with Schema Theory are also compatible with our theoretical approach. AOS is best considered a disorder of motor control and, in particular, one of motor programming (McNeil et al., 1997). We use the term *motor programming* to refer to the set of processes that specify “a representation that, when initiated, results in the production of a coordinated movement sequence” (Schmidt & Lee, 2005:466). Since principles of motor learning are thought to operate, in part, at the programming level, this group of patients is ideal to test the application of these principles to the treatment of speech. The theoretical framework will first be outlined, followed by discussion of its recent application to motor speech learning and AOS.

5.0.1. Theoretical Framework

\(^5\) Specifically, Experiments 1 and 2 of this dissertation. Chapter 5 is a verbatim reproduction of a manuscript in press, and thus is not framed in the context of the larger work presented in this dissertation.
Motor learning refers to a “set of internal processes associated with practice or experience leading to relatively permanent changes in the capability for movement” (Schmidt & Lee, 2005:302). The Schema Theory of motor control and learning proposed by Schmidt (1975) assumes that the brain stores “generalized” motor programs (GMPs) that represent the relative timing and relative force of muscle commands necessary for carrying out members of a class of movements. Given particular response specifications, the processing mechanism selects the parameters, or details, of the movement to be executed (Shea & Wulf, 2005). The parameters assigned to a given GMP specify the absolute timing and absolute force of muscle contractions in the chosen effectors. It is assumed that the specification of GMPs and the selection of parameters are shaped and refined during motor learning.

While other theories of motor control exist (e.g., Kelso & Tuller, 1981; Saltzman & Munhall, 1989; Thelen & Smith, 1994), Schema Theory provides the theoretical framework for this program of research because of its emphasis on motor learning and its influence on the development of specific principles of motor learning such as those addressed in this paper. Other authors (Shea & Wulf, 2005) have recently advocated a reconceptualization of the GMP as a “scalable response structure” (SRS), and emphasize processing mechanisms instead of schemata, but the terms “GMP” and “schemata” have been used here to maintain terminological consistency with the literature from which this paper draws.

According to Schema Theory, four types of information are stored after a movement is executed. This information includes the initial conditions (task and
environment conditions prior to movement production), the parameters that are
assigned to the GMP, the outcome of the movement in terms of the environmental
goal, and the sensory consequences of the movement. In order to learn new skills,
reorganize older skills to be performed at more challenging levels, and presumably to
re-learn skills that have been lost, the performer must develop abstract relationships
between these pieces of information upon completion of the movement. The first of
these relationships, the recall schema, involves the relation between past parameters,
the initial conditions, and the movement outcomes produced by their combinations. In
future trials, when the initial conditions and desired outcomes are noted, the recall
schema serves to select the parameters most appropriate for achieving the movement
goal, and applies them to the GMP. The second abstract relationship is the recognition
schema. This is the relationship between the past initial conditions, the past
environmental outcomes, and the past sensory consequences of those movements.
When a performer notes the initial conditions and desired outcome before a
movement, s/he can then estimate (anticipate) the sensory consequences of that
movement. These expected sensory consequences are then compared to the actual
feedback produced in order to evaluate the movement after its execution to detect
error. Under the framework of Schema Theory, learning consists of
refining/strengthening these two schemata through experience.

Schema Theory makes several predictions about how development of GMPs
and schemata might be affected by specific conditions present during practice (see
Schmidt & Lee, 2005; Shea & Wulf, 2005). These conditions define how new skills
are practiced and the type and frequency of external, or augmented, feedback that is provided to the learner. Through numerous studies of motor learning in the limbs, a set of motor learning principles has been identified that differentiates between variables that enhance performance temporarily and those that bring about robust long-term learning (see Maas, Robin, Austermann Hula et al., in press; Schmidt & Lee, 2005, for reviews). This distinction between performance and long-term learning then requires that one measures short-term changes in performance, from training session to training session, as well as long-term retention of trained skills after termination of training and transfer of trained skills to related, but untrained skills.

While there are a number of principles of motor learning, we focus here on those involving provision of augmented feedback, a ubiquitous component of motor speech treatments. Due to the difficulty of studying the effects of intrinsic sources of outcome information in humans, researchers have developed paradigms for manipulating extrinsic, “augmented” feedback as a means of deducing the operations of intrinsic feedback in naturalistic contexts. Studies in limb motor learning have shown that increased frequency of external knowledge of results (KR) feedback promotes parameter learning (e.g., Wulf et al., 1993). However, the provision of too much external KR feedback appears to degrade GMP learning, suggesting that reduced availability of external outcome information is important for promoting performers’ learning of the core features of a movement pattern, or GMP (Wulf et al., 1994). To date, studies have found that reduced KR is either equally or more effective in promoting learning (e.g., Lee et al., 1990; Sparrow & Summers, 1992; Winstein &
Schmidt, 1990). No studies have reported a superior effect of 100% KR over reduced KR on learning.

In addition to frequency of KR feedback, the temporal locus of feedback is an important determinant of the availability of external outcome information. Researchers have examined three intervals: feedback delay interval (time between participant’s production and provision of feedback), post-feedback delay interval (time elapsed between feedback and next stimulus presentation), and inter-trial interval (time between two successive trials) (Schmidt & Lee, 2005). Animal studies have suggested that delaying a reward after response to a stimulus has a detrimental effect on conditioning (e.g., Perin, 1943; Skinner, 1936). Based on this, researchers became interested in evaluating the effects of feedback delay on human motor learning (see Salmoni, Schmidt, & Walter, 1984, for review). Although most studies have revealed null or inconsistent effects of feedback delay on human motor skills learning (e.g., Becker, Mussina, & Persons, 1963; Koch & Dorfman, 1979; Mulder & Hulstijn, 1985; Weltens & de Bot, 1984), most have focused on temporary changes in performance rather than long-term retention and transfer of motor skills. Swinnen, Schmidt, Nicholson, and Shapiro (1990) examined the effect of a short feedback delay on acquisition and retention of a bat-swing motor skill. They concluded that instantaneous feedback initially supported acquisition of the behavior, but at a certain point began to impede the continued improvement during acquisition. Furthermore, it interfered with retention of the trained skill at ten minutes and, more dramatically, two days after training. This trend persisted on a 4-month retention test, although the group
difference was no longer significant. More recently, Anderson, Magill, Sekiya, and Ryan (2005) reported that delayed feedback (i.e., feedback given after two intervening trials) resulted in less accurate acquisition performance of an unfamiliar aiming behavior, but stronger retention after a 24-hour delay. The size of this difference was moderate, although it did not reach significance. Additionally, the decline in performance from acquisition to retention (one minute and 24 hours) was smaller for the delayed feedback group than the immediate feedback group and the delayed group reported using a greater number and variety of intrinsic feedback sources during practice.

The vast majority of limb motor learning studies have tested healthy individuals. Several studies have extended the work on principles of motor learning to patients with neurological disease in re-learning limb control (Goodgold-Edwards & Cermak, 1990; Hanlon, 1996; Jarus, 1994; Sabari, 1991; Stevans & Hall, 1998). In general, the principles appear to apply similarly in the intact and the neurologically impaired system. It is reasonable then to hypothesize that the same principles which enhance limb motor learning will also apply to speech motor learning in both healthy and neurologically impaired individuals. Apraxia of Speech (AOS) is a logical starting point, as it is widely considered a disorder of motor programming. Therefore, clear predictions based on Schema Theory can be made regarding its response to specific principles of motor learning.
5.0.2. Application of Principles of Motor Learning to Treatment for Apraxia of Speech

Acquired AOS is a motor speech disorder that has been estimated to account for 4% of all acquired neurologic communication disorders (Duffy, 2005). Current research indicates that AOS is a disorder of motor programming (Ballard & Robin, 2007; Clark & Robin, 1998; Deger & Ziegler, 2002; Hageman, Robin, Moon, & Folkins, 1994; Itoh & Sasanuma, 1987; Maas, Robin, Magnuson, Wright, & Ballard, 2005; McNeil, Weismer, Adams, & Mulligan, 1990; see Ballard, Granier, & Robin, 2000, and McNeil et al., 1997, for reviews) that affects programming the kinematic patterns used during speech production (McNeil et al., 1997). Within a motor-programming framework, AOS is a disruption in the ability to select or activate a GMP, and/or select correct parameter values for the execution of movements required for speech production. Motor-learning theory, which models the programming of skilled actions, provides an organizing framework that can be applied to the relearning of speech skills in persons with AOS.

Remediation of AOS has been studied for many years, although long-term retention has rarely been reported in clinical research literature (Wambaugh, Duffy, McNeil, Robin, & Rogers, 2006a, b). A theory-based approach, incorporating principles that enhance learning of motor skills, is lacking in current clinical practice, where anecdotal evidence suggests that speech pathologists tend to employ variables that lead to better performance during the therapy session and rarely measure long-term retention and transfer.
Much of the data supporting the application of principles of motor learning to training of speech skills has been derived from healthy populations or limited numbers of speakers with motor speech disorders (e.g., Adams & Page, 2000; Ballard, Maas, & Robin, 2007; Knock, Ballard, Robin, & Schmidt, 2000; Maas, Barlow, Robin, & Shapiro, 2002). Clark and Robin (1996) first provided evidence that reduced feedback (KR) frequency facilitates retention of a new oral motor skill in healthy speakers. Steinhauer and Grayhack (2000) subsequently applied the principle of reduced KR frequency to the motor learning of a vowel nasalence task in unimpaired speakers, and found an inverse relationship between the frequency of feedback (0%, 50%, or 100%) and measures of performance and learning of the skill. Similarly, Adams and Page (2000) and Adams, Page, and Jog (2002) demonstrated that providing summary feedback (feedback about every trial, presented after a number of intervening trials) after five trials as opposed to feedback after every trial enhanced retention of a novel speech skill in normal speakers and speakers with Parkinson’s Disease. No studies have yet examined the effect of immediate versus delayed provision of KR feedback on motor speech learning. Overall, the results of early studies suggest that continued work on the influence of principles of motor learning on the speech motor system is warranted.

We designed a series of treatment studies in an effort to understand how feedback (KR) affects acquisition, retention, and transfer of motor speech skills in speakers with AOS. The first two experiments in this series are reported here and examined the effects of two feedback variables on the treatment of AOS: frequency of
feedback and temporal locus of feedback. Although it can be argued that targeting functionally relevant words may be more appropriate in treatment contexts, it was necessary to use nonwords in this study in order to examine the effects of these principles while avoiding other potentially confounding factors (e.g., concreteness, frequency, familiarity, phonological structure). In other words, this was not a clinical outcome study; it was a research study to examine the influence of these variables on speech skill learning.

5.1. Experiment 1: High versus Low Frequency Feedback

The purpose of Experiment 1 was to examine the effect of frequency of feedback on the learning of speech skills by adults with AOS. Based on evidence from the limb literature (see Schmidt & Lee, 2005, for review), two predictions were made. First, we predicted that high-frequency feedback (HFF) would promote temporary performance enhancement but interfere with long-term retention and transfer of speech skills. The second prediction was that lower frequency feedback (LFF) would best-promote the long-term retention of treated speech sounds and facilitate transfer of treated skills to similar but untreated stimuli. These feedback conditions were compared using single-subject design in a common treatment method for AOS.

5.1.1. Method

5.1.1.1. Participants
Four participants (three male, one female; mean age = 70.3, sd=3.0 years) with AOS (mean time post-onset 13.3 months; range 6-20 months; sd=5.9 months) subsequent to left-hemisphere middle cerebral artery stroke participated in the study. Participants were recruited from the San Diego State University Communications Clinic. Brain scans and detailed lesion information were not available. Three of the participants were right-handed monolingual English speakers. Participant 2 was a left-handed simultaneous bilingual (English and Spanish) with a background in foreign language teaching. He reported that he considered English to be his primary language. Although Participant 1 was six months post-onset, the possibility of spontaneous recovery did not pose a threat to the validity of his inclusion in the study for two reasons. First, the application of principles of motor learning should impact learning during the subacute stage as well as the chronic stage. Second, the use of the chosen single subject design (i.e., alternating treatments design, see below) allows the separation of the effects of the experimental variables from those related to potential spontaneous recovery. Further, since individuals with AOS receive most of their treatment in the early stages, it is important to study how these variables affect learning at these stages as well.

We conducted formal testing approximately two weeks prior to commencement of the study (Table 5-1), and assessed language skills with the *Boston Diagnostic Aphasia Battery* (BDAE; Goodglass, Kaplan, & Barresi, 2001). Language Competency Indices revealed a wide range of degree of impairment between participants in the language comprehension and expression domains (range = 15th
percentile to 81st percentile). To evaluate praxis skills, the *Apraxia Battery for Adults-2* (ABA-2; Dabul, 2000) was administered to all participants. Performance on six ABA-2 subtests was analogous to each participant’s BDAE performance, ranging from mild to severe. All participants were diagnosed with AOS by two speech-language pathologists with expertise in motor speech disorders, and produced speech characterized by the following cardinal features of the disorder: increased segmental duration, increased inter-segmental duration, errors consisting primarily of distortions, substitutions distorted, consistent error types, and prosodic anomalies (McNeil et al., 1997; Wambaugh et al., 2006). These criteria are typical of those used to characterize the perceptual features of AOS in other treatment studies (e.g., Wambaugh et al., 1998; Wambaugh et al., 1999; Wambaugh & Nessler, 2004, Ballard et al., 2007). An oral mechanism examination revealed the probable concomitance of unilateral upper motor neuron dysarthria in Participant 2.
Table 5-1. Participant characteristics. Note that age and time post-onset are given at commencement of Experiment 1. Experiment 2 began 4 months later.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Handedness</th>
<th>Age (yrs.)</th>
<th>Time post-onset (mos.)</th>
<th>BDAE Percentiles</th>
<th>ABA-2 Levels of Impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Expression</td>
<td>Comprehension</td>
</tr>
<tr>
<td>P1</td>
<td>M</td>
<td>R</td>
<td>74.3</td>
<td>6</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>P2</td>
<td>M</td>
<td>L</td>
<td>67.3</td>
<td>12</td>
<td>82</td>
<td>80</td>
</tr>
<tr>
<td>P3</td>
<td>M</td>
<td>R</td>
<td>69.0</td>
<td>20</td>
<td>50</td>
<td>93</td>
</tr>
<tr>
<td>P4</td>
<td>F</td>
<td>R</td>
<td>70.6</td>
<td>15</td>
<td>95</td>
<td>87</td>
</tr>
</tbody>
</table>

m=70.3 m=13.3
sd= 3.0 sd = 5.9

5.1.1.2. Experimental Design

A single-subject alternating treatments design (ATD; McReynolds & Kearns, 1983) was utilized with related but untrained behaviors probed throughout the study to assess transfer. An ATD involves administering all treatment conditions (HFF and LFF, in this case) in parallel to all participants. In this way, each participant serves as his/her own control. Each treatment condition must be paired with a different, independent set of behaviors in order to isolate its effects (e.g., Knock et al., 2000). Treatment condition – behavior set pairing was counterbalanced within participant across two phases of treatment. Counterbalancing of conditions across participants was not possible due to the range of severities and speech impairment profiles represented by the participants.
Since reducing the frequency of feedback is thought to enhance the learning of motor programs as opposed to parameters (Wulf et al., 1993), speech behaviors that were based on different manner classes (e.g., fricatives and plosives; Ballard et al., 2007; Knock et al., 2000; Rubow et al., 1982) were chosen for participants 1, 2, and 4 based on their stimulability and profile of impairment (Table 5-2). Since differences in manner of production reflect differences in the relative force and timing of muscle contractions, speech behaviors in different manner classes are presumably governed by different GMPs (see Ballard et al., 2007), and therefore were considered sufficiently independent to preclude cross-condition contamination. In contrast, place of articulation can be considered a parameter that selects the appropriate muscle groups (or effectors, in Schema Theory) to execute the program. For participant 4, stress assignment was also varied to add complexity due to this participant’s high level of functioning. Participant 3, instead of receiving training for different manner classes, was trained to produce the /ɹ/ sound in the context of already-established front versus back consonants, a skill area of relative weakness for him. Each participant underwent two phases of treatment. In Phase I, one set of behaviors (e.g., /ʃ/ /s/ /v/ for the fricative set) was trained with treatment type 1 (e.g., HFF) and the other behavior set (e.g., /p/ /b/ /t/ for the plosive set) with treatment type 2 (e.g., LFF) and in Phase II the behavior set – treatment condition pairing was reversed.
<table>
<thead>
<tr>
<th>Phase I Conditions</th>
<th>Phase I Targets</th>
<th>Phase I Probes</th>
<th>Phase II Conditions</th>
<th>Phase II Targets</th>
<th>Phase II Probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFF</td>
<td>Fricative</td>
<td>suh, fuh, vuh</td>
<td>see, us, seam, fee, uf, fate, vee, uv, vain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td></td>
<td></td>
<td>Fricative</td>
<td></td>
<td>suh, ees, cuss</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>fuh, eef, huff</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>vuh, eev, love</td>
</tr>
<tr>
<td>P2</td>
<td>Plosive</td>
<td>tuh, puh, buh</td>
<td>tee, ut, ton, pee, up, pat, bee, ub, beam</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tuh, eet, hut</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>puh, eep, cup</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>buh, eeb, rub</td>
</tr>
<tr>
<td>P3</td>
<td>Fricative /affricate</td>
<td>chuh-chuh, thuh-thuh, zuh-zuh</td>
<td>cheechee, uchuh, chill, theethee, uothuh, thumb, zeezee, uhrh, zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>echee, chuhchuh, achieve, eethee, rhuithuh, athena</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>eeece, uhrh, azores</td>
</tr>
<tr>
<td>HFF</td>
<td>L-blends</td>
<td>fluh-fluh, pluh-pluh, bluh-blah</td>
<td>fleaflie, uhflu, flame, plueoplee, uhpluh, plague, bleeltee, uhbluh, blade</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>eeflee, fluh-fluh, afloat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>eeplee, pluhpluh, aplomb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>eeblee, bluhbluh, ablate</td>
</tr>
<tr>
<td>P4</td>
<td>Front-initial.</td>
<td>MER-nuh, PER-tuh, BER-duh</td>
<td>merNUH, NUHmer, morning perFUH, TUHper, parton berDUH, DUHber, beared</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MER-nuH, nuH-MER, NO-mare</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PER-tuh, tuh-PER, temper</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BER-duH, duh-BER, dubber</td>
</tr>
<tr>
<td>P5</td>
<td>Back-initial.</td>
<td>HER-nuh, KER-tuh, GER-duh</td>
<td>herNUH, NUHher, homet kerFUH, TUHker, carton gerDUH, DUHger, guarded</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HER-nuH, nuH-HER, NO-hair</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KER-tuh, tuh-KER, tanker</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GER-duH, duh-GER, dagger</td>
</tr>
<tr>
<td>LFF</td>
<td>S-cluster -initial</td>
<td>struh-MUH-nuh, scruh-MUH-nuh, spruh-MUH-nuh</td>
<td>STRUHmuhnuh, muhnUHstruh, stratify SCRUHmuhnuh, muhnUHscruh, scrutinize SPRUHmuhnuh, muhnUHspruh, sprucify</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>STRUHmuhnuh, muhnUHSTRUH, tapestry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SCRUHmuhnuh, muhnUHSCRUH, redescribe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SPRUHmuhnuh, muhnUHSPRUH, overspread</td>
</tr>
<tr>
<td>HFF</td>
<td>S-cluster medial</td>
<td>Nuh-MUH-struh, Nuh-MUH-scruh, Nuh-MUH-spruh</td>
<td>nuh-MUH-fluh, nuh-MUH-bluh, nuh-MUH-gluh, glamorize</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>nufMUHnuH, nuhnMUHFLUH, megaflop</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>bluhMUHnuH, nuhnMUHBLUH, notably</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gluhMUHnuH, nuhnGLUH, polyglot</td>
</tr>
</tbody>
</table>
Each participant demonstrated stable performance on three baseline probe sessions before treatment began. Each treatment phase was four weeks in length (with approximately four treatment sessions per week) with a four week maintenance period following (including three to four probes, depending on participant availability). Weekly probe sessions were administered throughout the 16 weeks; these assessed retention of trained behaviors when training conditions were removed, and transfer of the trained behaviors to related but untrained targets. In addition, long-term retention data for Phase I were collected on Participants 1 and 4 eight and seven months, respectively, following the end of Phase I treatment. Participant 2 was unavailable for long-term retention testing, and Participant 3 did not participate because he had suffered a second stroke in the interim.

5.1.1.3. Baseline and Probe Testing Procedures

Baseline and weekly probe sessions consisted of the random elicitation of ten each of six trained nonword behaviors, 12 related but untrained nonword transfer items, and six related but untrained real word transfer items, for a total of 240 items for each baseline or probe session. Productions were elicited using orthographic prompts only, and no feedback was given. Rate of stimulus presentation varied as a function of participants’ response times, with the average time between response and presentation of the next stimulus being approximately two seconds. Probes during treatment phases were administered on a day during which no treatment was received.
5.1.1.4. Treatment Procedure

Ninety-minute treatment sessions took place four times per week for a total of 14 to 16 sessions per participant per treatment phase. The sessions were divided into two periods, with one of the two behavior-condition pairings presented in one period and the other behavior-condition pairing in the other period. Order of treatment conditions within session was counterbalanced across sessions within subject within each treatment phase. Each period began with a pre-practice component, usually 5-15 minutes in length, involving the use of phonetic placement strategies to elicit at least five correct productions of each of the targets for one behavior set before practice began. The Phonetic Placement Therapy (PPT; Van Riper & Irwin, 1958) involved using orthographic stimuli, pictures, diagrams, verbal descriptions of articulatory features, and/or modeling to shape correct target productions by the participants. For example, if working on /pʌ/, the clinician might model the target, instruct the participant to look at the clinician and listen carefully to the sound, “puh” would be shown on a card orthographically, and the clinician would say (for example) “for this sound, start with the lips pressed together.” Feedback involving Knowledge of Results (KR; whether the sounds were correct or incorrect) and Knowledge of Performance (KP; how the sounds were produced, e.g., “your lips were apart”) were given during PPT pre-practice. Consistent with previous studies examining motor learning in AOS (e.g., Knock et al., 2000), PPT was selected because it is part of almost all treatments for AOS across a range of severities. Moreover, PPT results in clear acquisition effects under controlled experimental conditions (Wambaugh & Doyle, 1994, for review).
This pre-practice was followed by a practice component in which 30 productions of each target in a set were elicited in random order with orthographic prompts only, for a total of 90 productions. General KR feedback (i.e., “correct” or “incorrect”) was provided on 60% (LFF) or 100% (HFF) of productions. During both pre-practice and practice, the feedback interval and post-feedback interval each approximated two seconds. For trials in which feedback was not given, inter-trial intervals were approximately two seconds. For the LFF condition, feedback schedules were constructed beforehand and used on-line by the clinician to ensure reliability of the independent measure. Three different LFF schedules were used, in order to avoid the same trials receiving feedback during each session.

5.1.1.5. Stimuli and Materials

Due to the wide range of speech and language skills observed in these participants, each received individually-tailored stimulus sets appropriate to his or her profile of impairment and stimulability (Table 5-2). For all participants, three nonword syllables or syllable sequences were trained under each feedback condition (e.g., “suh” or “MER-nuh;” see Table 5-2). Transfer was tested to untrained stimuli of similar complexity (e.g., “see” and “us” or “mer-NUH” or “NUH-mer,” respectively) and to more complex real words related to the trained behaviors (e.g., “seam” or “morning,” respectively). Word frequency was balanced across both sets of complex real words in each treatment phase (Francis & Kučera, 1982) for participants 1, 2, and 4. Complex real-word stimuli for Participant 4 were verbs (for homogeneity and conformity with
carrier phrase restrictions) with very low frequency (1 or less; Francis & Kučera, 1982); however, since a three-syllable English verb beginning with the /spr/ cluster was not available, a pseudoword was constructed and defined for the participant (“sprucify: to spruce or clean up the appearance of something”). Since the participant was unfamiliar with most of the other real words, this decision seemed justified. For Participant 3, it was not possible to balance the sets of real word stimuli for word frequency due to the need to control for other, potentially more influential factors such as stimulability. However, there was no difference in baseline performance on these sets of real word stimuli in Phase I, and the real word stimuli chosen for Phase II were all highly infrequent (Francis & Kučera, 1982).

The use of nonwords is motivated by several points. First, the consensus is that AOS is a disorder of speech motor programming, and nonwords (here, syllables licensed in English) also involve the relevant processes, from phonological encoding through motor programming through articulation. Also, in order to remove possible confounds arising at prior (semantic and/or lexical) stages of processing as a result of concomitant aphasia, it was decided to isolate the relevant processes via the use of nonwords. In addition, there is evidence to support the use of nonwords in treatment of AOS, in that generalization has been observed to real words (e.g., Mass, Robin, Wright et al., 2002; Schneider & Frens, 2005). However, it is possible that using real words would lead to greater learning. No studies are available to our knowledge that have directly compared the efficacy of using real words vs. nonwords in this population, and as such it remains an open question as to which is better. For example,
Tjaden (2000) included both nonwords and words; however, these were not counterbalanced (nonwords always preceded words) and this study was not designed to compare the two.

All stimuli were presented to participants orthographically. Prior to the beginning of the study, all participants were screened for their ability to read the stimuli by demonstrating reliability in matching examiner-produced tokens to the appropriate written stimulus in a field of six. For Participant 1, who was severely impaired, place-of-articulation photographs were also used during the practice component of the treatment sessions and during a “modified” probe (see below). For Participant 2, target behaviors were embedded in easy or difficult carrier phrases (e.g., “It’s a ___ a day” or “The hyperthyroidism will ____ the ray”; see Appendix A) in Phases I and II of the study, respectively. Targets for Participant 4 were embedded in difficult carrier phrases in both phases (Appendix A).

5.1.1.6. Scoring and Reliability

All productions were scored online as correct or incorrect by the examiner and up to two other clinicians for purposes of determining inter-rater reliability. Only the examiner’s judgment counted for purposes of providing feedback. Responses were considered correct if all phonemes were accurately produced (without distortion), the appropriate stress pattern was applied (in multisyllablic productions), and the rate of the production was deemed “normal.” Error-free carrier phrase production was not required for a “correct” score on a target item. For Participant 2, coarticulation across
sylables was not required for a score of “correct” during Phase I, as he was encouraged to think of the targets as reduplicated consonant-initial nonwords for the early goal of achieving correct articulation of these consonants in the initial position of syllables (for which he was most stimulable). The second goal, of producing these consonants in the medial position of a two-syllable nonword without segmentation, was encouraged during the second phase, when coarticulation across syllables was required for a score of “correct.” Participant responses were also recorded via a DAT recorder for future analyses.

Inter-rater reliability was measured for a randomly selected 25% (Participant 1: 26%, Participant 2: 29%, Participant 3: 22%, Participant 4: 24%) of treatment and probe sessions. Point-to-point agreement on “correct” and “incorrect” binary judgments was 95.3% (fricatives: 95.5%, plosives: 95.1%) for Participant 1, 83.6% (fricatives/affricates: 85.5%, L-blends: 81.7%) for Participant 2, 86.2% (front phonemes: 85.8%, back phonemes: 86.8%) for Participant 3, and 88.9% (S-clusters: 91.9%, L-blends 85.8%) for Participant 4.

5.1.2. Results and Discussion

5.1.2.1. Participant 1

Phase 1

Baseline. Participant 1 (P1) demonstrated stable baseline performance on fricative and plosive CV behaviors (e.g., “fuhi” and “tuh”) before Phase 1 treatment began. Accuracy ranged from 0 to 5% for items to be trained (Figure 5-1, Panel A,
Acquisition. Improvements in performance during treatment sessions (Figure 5-1, Panel A, filled shapes) were observed for trained behaviors for both treatment conditions (LFF and HFF). The rates of improvement and accuracy levels achieved did not differ for LFF and HFF items, which reached 69% and 67%, respectively.

Retention and Transfer. P1’s probe performance in Phase I revealed no changes in accuracy of production for retention of trained items (Figure 5-1, Panel A, unfilled shapes) and transfer to untrained items (Figure 5-1, Panels B, C, and D, Phase I) in either feedback condition (see below for note on word transfer items in Panel D).

Phase II

Baseline. P1 demonstrated stable baseline performance on fricative and plosive VC behaviors (e.g., “uf” and “ut”) prior to treatment. Accuracy rates were similar across conditions, and ranged from 0 to 5% for items to be trained (Figure 5-1, Panel B, sessions 1-21).

Acquisition. P1 acquired the HFF-fricative behaviors at a faster rate and to a slightly higher level (80%) than the LFF-plosive behaviors (58%) (Figure 5-1, Panel B, filled shapes), supporting our first prediction, that HFF would enhance initial skill acquisition. Interestingly, after only one session of Phase II treatment, P1 improved from 0 to 50% accuracy on the HFF-fricative behaviors (Figure 5-1, Panel B, sessions 24-25). This could be due to a very early benefit of HFF for acquisition or,
alternatively, a longer-term effect of LFF-fricative training carried over from Phase I. HFF throughout Phase II treatment likely compounded this effect.

Retention and Transfer. Probe performance in Phase II suggests that P1 was unable to transfer trained behaviors to untrained items in either condition (Figure 5-1, Panels C-D, Phase II). Small fluctuations in performance on word transfer items (Figure 5-1, Panel D) are not considered reflective of feedback condition influence since their baseline was not perfectly stable, and also since they consistently favor plosive items across both phases (i.e., both feedback conditions). However, examination of the retention of trained VC targets in regular probe sessions (Figure 5-1, Panel B, Phase II, unfilled shapes) suggests that beginning in Phase II treatment and continuing throughout maintenance, this participant was able to demonstrate learning in the difficult context of the probe session format (consisting of the random elicitation of 24 different targets ten times each), and his performance consistently favored the LFF condition (LFF range 0-43%; HFF range 0-23%).
Figure 5-1. P1 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 1. Filled shapes in panels A and B represent performance during High and Low Frequency Feedback practice trials in treatment sessions. Unfilled shapes in panels A and B represent baseline and retention of trained behaviors in probe sessions. Panels C and D represent transfer to related but untrained stimuli. (Note that different transfer stimuli were probed in each of the two phases.) Panel E depicts retention of trained behaviors in modified probe sessions. Session “LT” represents a long-term retention probe that occurred 8 months following the end of Phase I treatment.
Modified Probe

Due to P1’s extreme difficulty producing trained behaviors in the large probe stimulus set, and his inconsistent and perseverative error patterns, it was felt that his retention of trained speech skills was not reflected by probe performance. Therefore, a limited set of trained items only, elicited five times each with photographic support, was presented every week after Phase I treatment as a “modified probe” (Figure 5-1, Panel E). As with the regular probes, no feedback was given during this modified probe. Due to P1 traveling, only two modified probes were administered during the maintenance stage of Phase I. The first maintenance probe was administered the day after P1 returned, and may reflect his reported jetlag. However, a modified probe one week later (four weeks after Phase I treatment, session 23) illustrated strong retention of both sets of trained behaviors, but no evident difference in retention between the two sets of behaviors under different feedback conditions in Phase I (LFF = 74%; HFF = 72%). When long-term retention was assessed with the modified probe task eight months after the termination of Phase I treatment (with Phase II and Experiment 2 intervening), a small difference in retention (LFF = 54%; HFF = 38%) between the two conditions was detected (Figure 5-1, Panel E, session “LT”). Speech targets that were trained under reduced feedback conditions were retained better at 8 months post-treatment. The other therapies that the patient received in the intervening months were not likely contributors to this effect, as they did not involve production of Phase I targets, and there is no evidence that this speaker was able to transfer skills between
related items. Time constraints did not permit the collection of long-term retention
data 8 months after Phase II.

Consistent with the postulated carry-over effect from LFF conditions in Phase I
(see Phase II acquisition above) is P1’s performance on the first session of the
modified probe testing in Phase II (Figure 5-1, Panel E, session 28), in which LFF-
plosives (formerly HFF) were 0% accurate, and HFF-fricatives (formerly LFF) were
almost 70% accurate. The carry-over hypothesis for these results seems tenable,
especially after examining the time-course of modified probe performance throughout
Phase II. By the end of Phase II maintenance, retention of HFF behaviors, as measured
via modified probes, was roughly unchanged from this first probe session, while
retention of LFF behaviors had increased. This result, along with the outcome of the
long-term retention probe in Phase I (Session “LT”) and the maintenance of
performance in Phase II (Figure 5-1, Panel B, unfilled symbols), supports the second
prediction that LFF leads to enhanced learning.

5.1.2.2. Participant 2

Phase I

Baseline. Participant 2 (P2) demonstrated stable baseline performance on L-
blend and fricative/affricate C(c)VC(c)V behaviors (e.g., “pluhpluh” and “zuhzuh”) before Phase I treatment began (Figure 5-2, Panel A, sessions 1-3), correct responses ranging from 0 to 10% for items to be trained (Figure 5-2, Panel A, sessions 1-3) and
from 0 to 17% for untrained transfer items (Panels B, C, and D, sessions 1-3), with no substantial differences across conditions.

**Acquisition.** Contrary to the first prediction, P2’s acquisition performance on trained behaviors was better for LFF-fricative/affricate items (range 5%-97%) than for HFF-L-blend items (range 2%-71%) (Figure 5-2, Panel A, filled shapes).
Figure 5-2. P2 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 1. Filled shapes in panels A and B represent performance during High and Low Frequency Feedback practice trials in treatment sessions. Unfilled shapes in panels A and B represent baseline and retention of trained behaviors in probe sessions. Panels C and D represent transfer to related but untrained stimuli. (Note that different transfer stimuli were probed in each of the two phases.) More difficult carrier phrases were introduced at the onset of Phase II.
Retention and Transfer. Retention (Figure 5-2, Panel A, unfilled shapes) and transfer (Figure 5-2, Panels B-D, Phase I) were greater and more stable for items related to trained LFF-fricative/affricate behaviors than for items related to trained HFF-L-blend behaviors, appearing to lend support to the second experimental prediction. Accuracy of responses to retention probes of trained stimuli during treatment and maintenance periods of Phase I ranged from 3% to 97% for LFF-fricative/affricate behaviors, and from 0 to 43% for HFF-L-blend behaviors. Transfer to untrained productions involving different vowels (LFF range = 2%-80%; HFF range = 0-40%; Panel C) and to untrained words (LFF range = 42%-83%; HFF range = 6%-63%; Panel D) demonstrated a similar pattern of results. A large degree of transfer was observed to the VCV stimuli that were going to be used as treatment targets for Phase II (Figure 5-2, Panel B, Phase I). Because this improvement continued to grow and fluctuate drastically throughout Phase I Maintenance, an additional probe was administered before Phase II Baseline began. These items were returned to near-zero accuracy levels by increasing their difficulty for Phase II (see sections 5.1.15. Stimuli and Materials and 5.1.1.6. Scoring and Reliability above).

Phase II

Baseline. After implementing the above-mentioned difficulty manipulations, P2 demonstrated stable baseline performance on L-blend and fricative/affricate VC(c)V behaviors (e.g., “uhpluh” and “uhzuh”) before Phase II treatment began (Figure 5-2, Panel B, sessions 28-30). Response accuracy ranged from 0 to 3% for items to be trained in the LFF condition and from 0 to 7% in the HFF condition.
Baseline accuracy for untrained transfer behaviors ranged from 0 to 10% for LFF items and from 0 to 20% for HFF items.

**Acquisition.** Performance on targets given HFF (now fricatives/affricates) was better during acquisition than those given LFF (now L-blends) (Figure 5-2, Panel B, filled shapes). Performance accuracy ranged from 3% to 87% for HFF targets and from 0 to 61% for LFF targets.

**Retention and Transfer.** Retention (Figure 5-2, Panel B, Phase II, unfilled shapes) and transfer to untrained items (Figure 5-2, Panels C-D, Phase II) were also slightly better in the HFF condition (HFF retention range = 50%-87%; LFF retention range = 23%-33%). This pattern of results across both phases of the study, wherein fricative/affricate behaviors fairly consistently outperformed L-blend behaviors regardless of feedback conditions, suggests a probable stimulus effect. It is likely that the additional phonological and motoric complexity of L-blends as compared to fricative/affricates influenced accuracy enough to conceal any underlying effect of feedback frequency in this participant.

**5.1.1.3. Participant 3**  
**Phase I**

**Baseline.** Participant 3 (P3) demonstrated stable baseline performance with zero accuracy on back-initial and front-initial behaviors (e.g., “KERtuh” and “MERnuh”) before Phase I treatment began (Figure 5-3, Panel A, sessions 1-3).

**Acquisition.** P3’s performance on both sets of behaviors improved with treatment (Figure 5-3, Panel A, filled shapes), but contrary to our first prediction, HFF
did not appear to promote substantially better performance in acquisition of speech skills than LFF did (HFF range = 0-82%; LFF range = 0-76%).

**Retention and Transfer.** Despite the absence of marked acquisition differences, retention (Figure 5-3, Panel A, unfilled shapes) and transfer (Figure 5-3, Panels C-D, Phase I) of trained behaviors were enhanced in the LFF condition, as predicted. Ranges of accuracy for trained items (LFF = 0-53%; HFF = 0-23%), untrained stress-transfer items (LFF = 0-63%; HFF = 0-27%), and untrained word-transfer items (LFF = 0-63%; HFF = 0-33%) revealed obvious overall enhanced learning in the LFF condition.

**Phase II**

**Baseline.** Stable zero-accuracy baseline performance on back-medial and front-medial behaviors (e.g., “TUHker” and “NUHmer”) was demonstrated before Phase II treatment began (Figure 5-3, Panel B, sessions 1-26).

**Acquisition.** P3’s performance of both behaviors improved markedly with Phase II treatment (Figure 5-3, Panel B, filled shapes). However, as in Phase I, HFF did not enhance performance during acquisition of speech skills relative to LFF, contrary to our first prediction (HFF range = 0-79%; LFF range = 6%-89%).
Figure 5-3. P3 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 1. Filled shapes in panels A and B represent performance during High and Low Frequency Feedback practice trials in treatment sessions. Unfilled shapes in panels A and B represent baseline and retention of trained behaviors in probe sessions. Panels C and D represent transfer to related but untrained stimuli. (Note that different transfer stimuli were probed in each of the two phases.)
Retention and Transfer. Although overall enhanced learning under conditions of LFF was more obvious in Phase I, Phase II did provide evidence of earlier retention in probe sessions for trained targets given LFF (Figure 5-3, Panel B, sessions 31 and 35). Also, despite the retention of trained HFF behaviors reaching a higher peak right after treatment (HFF = 70%; LFF = 33%; Figure 5-3, Panel B, session 47), it soon began to decline (sessions 48-49; final probe HFF = 27%), whereas behaviors trained in the LFF condition in Phase II were retained at a level similar to that achieved during the treatment phase (final probe LFF = 43%). An apparent contradiction to this pattern was the real word transfer set in Phase II, which appeared to be enhanced under HFF conditions (Figure 5-3, Panel D, Phase II). These data, however, should be interpreted with caution, since the baseline level was higher for real words related to the HFF than the LFF condition (Figure 5-3, Panel D, session 27). When pre-treatment and post-maintenance endpoints for these Phase II real word data sets were compared (LFF = 0, 10%; HFF = 26%, 33%), the LFF condition indeed was associated with as much change as the HFF condition.

5.1.1.4. Participant 4

Phase I

Baseline. Participant 4 (P4) demonstrated stable zero-accuracy baseline performance on S-cluster and L-blend behaviors (e.g., “struh-MUH-nuh” and “fluh-muh-NUH”) before Phase I treatment began (Figure 5-4, Panel A, sessions 1-3). Baseline accuracy for untrained order and stress transfer behaviors (Figure 5-4, Panels
C-D, sessions 1-3) was also zero, while untrained word transfer items (Panel E) ranged from 7% to 27% for HFF items and from 7% to 8% for LFF items.

**Acquisition.** In the treatment setting, P4 acquired L-blend behaviors, given HFF, at a faster rate and to a higher level than S-cluster behaviors given LFF, in accord with our first prediction (Figure 5-4, Panel A, filled shapes). Accuracy rates during acquisition ranged from 3% to 92% for HFF items, and from 0 to 49% for LFF items.

**Retention and Transfer.** Contrary to our second prediction, retention (Figure 5-4, Panel A, unfilled shapes) and transfer (Figure 5-4, Panels C-E, Phase I) were stronger for the L-blend behaviors given HFF than for the S-cluster behaviors given LFF. It is noteworthy, however, that after the termination of treatment, retention of behaviors given HFF steadily declined from 93% to 63% across the four maintenance probes, whereas retention of behaviors given LFF remained stable or increased slightly (from 33% to 47%), relative to the level achieved during treatment (Figure 5-4, Panel A, sessions 23-26). A long-term follow-up probe session indicated that P4 retained much of the improvement in trained skills (HFF = 60%; LFF = 40%) seven months after the termination of Phase I treatment (Figure 5-4, Panel A, session “LT”). Performance levels on transfer to untrained stress and word transfer behaviors (Figure 5-4, panels C-E, session “LT”) were also maintained to varying degrees, with HFF L-blend skills continuing to surpass LFF S-cluster skills. The only other treatment that this participant received in the intervening months was Experiment 1 Phase II
treatment, which did not involve production of Phase I targets. Time constraints did not permit the collection of long-term retention data seven months after Phase II.

**Phase II**

**Baseline.** A stable baseline with zero accuracy was demonstrated on syllable-order reversed S-cluster and L-blend behaviors (e.g., “NUH-muh-struh” and “nuh-MUH-fluh”) before Phase II treatment commenced (Figure 5-4, Panel B, sessions 27-29). Untrained transfer items were also stable at or near zero accuracy (Figure 5-4, Panels C-E, sessions 27-29).

**Acquisition.** While P4 made gains in both behavior sets during Phase II treatment, HFF applied to the training of S-cluster behaviors did not result in better performance, contrary to prediction. A greater level of acquisition was achieved for L-blend behaviors given LFF (Figure 5-4, Panel B, filled shapes), with correct responses ranging from 1% to 56% for HFF items and from 0 to 83% for LFF items.
Figure 5-4. P4 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 1. Filled shapes in panels A and B represent performance during High and Low Frequency Feedback practice trials in treatment sessions. Unfilled shapes in panels A and B represent baseline and retention of trained behaviors in probe sessions. Panels C, D, and E represent transfer to related but untrained stimuli. (Note that different transfer stimuli were probed in each of the two phases.) Session “LT” represents a long-term retention probe that occurred 7 months following the end of Phase I treatment.
Retention and Transfer. After the termination of treatment, probe performance of many trained and untrained items continued to improve throughout the maintenance phase, but a difference between HFF and LFF conditions was not evident in retention of trained targets (Figure 5-4, Panel B, unfilled shapes) or transfer to untrained real words (Figure 5-4, Panel E, Phase II). Transfer to different syllable order (HFF range = 0-25%; LFF range = 0-38%) and stress pattern (HFF range = 3%-8%; LFF range = 7%-57%) was better for LFF-L-blend behaviors (Figure 5-4, Panels C-D, Phase II), apparently supportive of our second hypothesis. However, as was the case with the other mildly-impaired participant, P2, the results for P4 are also suggestive of a stimulus effect. In general, across both phases of the study, P4 acquired (Figure 5-4, Panels A-B), retained (Figure 5-4, Panels A-B), and transferred (Figure 5-4, Panels C-E) best the L-blend as opposed to the S-cluster behaviors regardless of feedback condition, in accord with hypothesized differences in phonologic and motoric complexity.

5.2. Experiment 2: Immediate versus Delayed Feedback

Experiment 2 examined the effect of feedback timing on speech skill learning in adults with AOS by comparing the effects of instantaneous feedback versus delayed feedback on acquisition and learning. Based on recent evidence in the limb literature (e.g., Anderson et al., 2005), two predictions were made. First, we predicted that immediate feedback (IF) would promote temporary performance enhancement but interfere with retention and transfer of speech skills. Our second prediction was that
delayed feedback (DF) would best-promote retention of treated speech sounds and transfer of treated skills to similar but untreated stimuli. These feedback conditions were also compared using a single-subject design in a common treatment method for AOS.

5.2.1. Method

5.2.1.1. Participants

Two participants who had completed participation in Experiment 1 (P1, age 74.7, and P2, age 67.7) with chronic AOS (time post-onset 10 and 16 months, respectively) subsequent to left-hemisphere stroke participated in the second study, which commenced approximately 1 week later. The third participant from Experiment 1 began Experiment 2, but withdrew when he suffered another stroke. The fourth participant from Experiment 1 did not participate in Experiment 2 because it was not possible to construct a new set of stimuli that were challenging enough given her mild impairments, yet could also be reliably judged online by the examiner. Formal testing was conducted approximately four months prior to commencement of the study. Participant details and test results were reported for Experiment 1 (Table 5-1).

5.2.1.2. Experimental Design

The design of Experiment 2 comparing immediate and delayed feedback was identical to that of Experiment 1. Stimuli were again chosen based on the assumption that different production manners represented different motor programs (Ballard et al.,
2007), and therefore could presumably be trained concurrently under different feedback conditions without introducing cross-condition contamination.

5.2.1.3. Baseline and Probe Testing Procedures

Baseline and weekly probe sessions were similar to those described above in Experiment 1, except that due to the differences in participant ability and stimulus set makeup, composition of probes differed between participants and between elicitation of 180 (P1) or 240 (P2) items for each baseline or probe session. Productions were elicited via orthographic prompts only.

5.2.1.4. Treatment Procedure

Treatment procedures were similar to those of Experiment 1. General knowledge-of-results feedback, instead of being given verbally as in Experiment 1, was provided visually by showing the participant a red or green signal (for “incorrect” or “correct”), and was provided on 100% of trials either immediately [IF; instantaneously (within 1 second) upon completion of the trial] or after a 5-second delay (DF) following the participant’s production. The post-feedback delay interval was a constant five seconds in each of these conditions. Since the inter-trial interval includes both of these segments, it was also necessarily different between conditions (five sec. for IF, 10 sec. for DF).
5.2.1.5. Stimuli and Materials

Participants in Experiment 2 continued to present with different levels of speech and language skills, and therefore stimuli were individually-tailored for appropriateness to each participant’s profile of impairment and stimulability (Table 3). Participants were given training on production of three nonword syllables or syllable sequences under each feedback condition (e.g., “shee” or “VAYmuhnay;” see Table 5-3). Transfer to similar but untrained behaviors (e.g., “eesh” or “muhNAYvay” and “FAYmuhnay,” respectively) and to real words related to the trained behaviors (e.g., “sheep” or “vaporize,” respectively) was assessed in weekly probe sessions. It was not possible to control for word frequency of these real word stimuli due to the need to control for other, potentially more influential factors such as a participant’s baseline stimulability for these words (as in P1), or the unavailability of English single words related closely enough to the trained nonwords (as in P2). Real word stimuli used for P2 were either of very low frequency (4 or less; Francis & Kuçera, 1982), or were “composite” words made of real words and/or morphemes and defined for the participant (e.g., “misinvoke”). P2 was encouraged to think of and produce “composite” word stimuli made of more than one real word (e.g., “visacard”) as single words. As before, speech behaviors were elicited by orthographic printing of stimuli on cards.

As in Experiment 1, due to P1’s difficulty demonstrating retention of familiar trained items during standard probe sessions, a “modified probe” was administered for each treatment set during the regular probe session each week after the termination of
Phase I treatment (starting with Session 23 and continuing throughout Phase II). This modified probe essentially segmented the probe task into two smaller retention tests (one for each stimulus set) consisting of trained items only (elicited randomly ten times each), followed by a transfer test consisting of the random elicitation of the remaining 12 untrained items (elicited randomly ten times each). As in the standard probe protocol, feedback was withheld during modified probes.
Table 5-3. Design and orthographic stimuli for Experiment 2. Note for P2 stimuli: Capital letters indicate stress.

<table>
<thead>
<tr>
<th>Phase I Conditions</th>
<th>P1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IF</td>
<td>DF</td>
</tr>
<tr>
<td></td>
<td>fricative</td>
<td>plosive</td>
</tr>
<tr>
<td><strong>Phase I Targets</strong></td>
<td>shay</td>
<td>kay</td>
</tr>
<tr>
<td></td>
<td>shee</td>
<td>kee</td>
</tr>
<tr>
<td></td>
<td>shoo</td>
<td>koo</td>
</tr>
<tr>
<td><strong>Phase I Probes</strong></td>
<td>aysh</td>
<td>eek</td>
</tr>
<tr>
<td></td>
<td>eesh</td>
<td>eek</td>
</tr>
<tr>
<td></td>
<td>oosh</td>
<td>ook</td>
</tr>
<tr>
<td></td>
<td>shame</td>
<td>came</td>
</tr>
<tr>
<td></td>
<td>sheep</td>
<td>keep</td>
</tr>
<tr>
<td></td>
<td>shut</td>
<td>cut</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase II Conditions</th>
<th>IF</th>
<th>DF</th>
<th>IF</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>plosive</td>
<td>fricative</td>
<td>R-blend</td>
<td>Fricative</td>
</tr>
<tr>
<td><strong>Phase II Targets</strong></td>
<td>ayk</td>
<td>aysh</td>
<td>muhNAbray</td>
<td>muhNAyvay</td>
</tr>
<tr>
<td></td>
<td>eek</td>
<td>eesh</td>
<td>muhNEE bree</td>
<td>muhNEEvee</td>
</tr>
<tr>
<td></td>
<td>ook</td>
<td>oosh</td>
<td>muhNO bro</td>
<td>muhNO vo</td>
</tr>
<tr>
<td><strong>Phase II Probes</strong></td>
<td>kay</td>
<td>shay</td>
<td>VAYmuhnay</td>
<td>muhNAyvay</td>
</tr>
<tr>
<td></td>
<td>kee</td>
<td>shee</td>
<td>VEEmuhnee</td>
<td>BREEmuhnee</td>
</tr>
<tr>
<td></td>
<td>koo</td>
<td>shoo</td>
<td>VOmuhno</td>
<td>BROmuhno</td>
</tr>
<tr>
<td></td>
<td>back</td>
<td>lash</td>
<td>celebrate</td>
<td>salivate</td>
</tr>
<tr>
<td></td>
<td>walk</td>
<td>wash</td>
<td>gentle-breeze</td>
<td>old-TVs</td>
</tr>
<tr>
<td></td>
<td>look</td>
<td>bush</td>
<td>go-for-broke</td>
<td>misinvoke</td>
</tr>
</tbody>
</table>

5.2.1.6. Scoring and Reliability

All productions were again scored online as correct or incorrect following the same guidelines as established and described in Experiment 1. Inter-rater reliability
was calculated on a randomly selected 13% (P1: 12%, P2: 14%) of treatment and probe sessions. Point-to-point agreement for “correct” and “incorrect” perceptual judgments was 94.4% (plosives: 94.8%, fricatives: 94.1%) for P1 and 97.8% (r-blends: 98.1%, fricatives: 97.5%) for P2.

5.2.2. Results and Discussion

5.2.2.1 Participant 1

Phase I

Baseline. P1 demonstrated stable baseline performance on trained fricative and plosive CV targets (e.g., “shoo” and “koo”) before Phase I treatment began (Figure 5-5, Panel A, sessions 1-3). Accuracy ranged from 0 to 3% for items to be trained and from 0 to 7% for untrained transfer items (Figure 5-5, Panels B-C, sessions 1-3), with no substantial differences across conditions.

Acquisition. During Phase I treatment, acquisition of trained behaviors was better under conditions of DF than IF (Figure 5-5, Panel A, filled shapes), contrary to the first prediction. Correct responses ranged from 20% to 77% for DF items and from 2% to 57% for IF items.

Retention and Transfer. P1’s probe performance in Phase I revealed enhanced retention of trained skills in the DF condition (DF range = 0-90%; IF range = 0-67%; Figure 5-5, Panel A, unfilled shapes), in accord with our second prediction, but mixed results for untrained transfer probes (Figure 5-5, Panels B-C, Phase I). Specifically, effects of DF treatment transferred to the opposite phoneme order (VC) (DF range = 0-23%; IF range = 0; Figure 5-5, Panel B, Phase I), whereas the effects of IF treatment
showed greater transfer to the real word stimuli (DF range = 0-27%; IF range = 3%-50%; Figure 5-5, Panel C, Phase I). P1’s performance on real word probes might be explained by differences in his personal experience with the specific words (he associated one of the fricative words with an important political figure at the time, and relished every opportunity to practice it), so it is perhaps more telling that for “phoneme order” non-word probes (Figure 5, Panel B, Phase I), P1 was never able to transfer to this behavior in the IF condition, but was able to do so (although not to a great extent) in the DF condition.

**Phase II**

**Baseline.** A stable baseline was demonstrated on fricative and plosive VC targets (e.g., “oosh” and “ook”) prior to the commencement of the study (accuracy range 0-13%), although slight instability was induced before Phase II treatment began when improved performance on plosive behaviors transferred from Phase I treatment (Figure 5-5, Panel B, session 12). The resultant discrepancy between plosive and fricative skills was small (mean = 8.5%) and relatively unchanging before Phase II treatment began. Baseline accuracy for untrained word transfer items was 0 for DF behaviors and from 3% to 7% for IF behaviors (Figure 5-5, Panel C, Sessions 26-27).

**Acquisition.** While P1 achieved 94% accuracy on both behavior sets by the end of Phase II Treatment, his overall acquisition performance levels were better for the targets given IF (range 53%-94%) than those given DF (range 14%-94%), as predicted (Figure 5-5, Panel B, filled shapes).
Figure 5-5. P1 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 2. Filled shapes in panels A and B represent performance during Immediate and Delayed Feedback practice trials in treatment sessions. Unfilled shapes in panels A and B represent baseline and retention of trained behaviors in probe sessions. All probes after the withdrawal of Phase I treatment were elicited in the "modified" format (see text). Panel C represents transfer to related but untrained stimuli. (Note that different transfer stimuli were probed in each of the two phases.)

Retention and Transfer. Retention of the trained skills in the probe situation (Figure 5-5, Panel B, unfilled shapes) paralleled this acquisition pattern [i.e., there was stronger retention of skills given IF (range 57%-87%) than those given DF (range
33%-57%), contrary to the second prediction. Transfer of treatment effects to related items was highly variable in Phase II, and a consistent effect of feedback delay was not evident. Given that P1’s performance largely favored plosives in acquisition, transfer, and retention in both phases, differences may be attributable to a stimulus effect and may not reflect a response to feedback timing conditions.

5.2.2.2. Participant 2

Phase I

Baseline. P2 demonstrated stable zero-accuracy baseline performance on R-blend and fricative target behaviors (e.g., “BRAY-muh-nay” and “VAY-muh-nay”) before Phase I treatment began (Figure 5-6, Panel A, sessions 1-3). Baseline accuracy levels for untrained transfer items (Figure 5-6, Panels B-D, sessions 1-3) were at or near zero.

Acquisition. As predicted, skills given IF were performed better (range 19%-100%) during acquisition than those trained under conditions of DF (range 0-74%) in Phase I (Figure 5-6, Panel A, filled shapes).
Figure 5-6. P2 performance during baseline, treatment, and maintenance segments of the two phases of Experiment 2. Filled shapes in panels A and B represent performance during Immediate and Delayed Feedback practice trials in treatment sessions. Unfilled shapes in panels A and B represent baseline and retention of trained behaviors in probe sessions. Panels C and D represent transfer to related but untrained stimuli. (Note that different transfer stimuli were probed in each of the two phases.)
Retention and Transfer. Probe data suggest that retention (Figure 5-6, Panel A, unfilled shapes) and transfer (Figure 5-6, Panel D, Phase I) were enhanced by provision of DF. Although retention of treated skills in the IF condition initially led retention of treated skills in the DF condition (Figure 5-6, Panel A, sessions 22-25), one week into the maintenance phase (session 24), retention of IF targets began to decline (from 93% to 43%), whereas retention of DF targets remained stable, or showed slight improvement, throughout the maintenance phase (from 30% to 47%) and the second phase of the study (to 80%). Inspection of untreated word transfer items (Figure 5-6, Panel D, Phase I) in Phase I yields similar interpretation: transfer of IF treatment effects to untrained items appeared temporarily enhanced during the treatment phase (i.e., sessions 17 and 22), but some of this improvement fluctuated during the maintenance phase (range 13%-23%), whereas transfer to untrained words related to the DF condition continued to improve throughout the maintenance phase (range 10%-37%; Figure 5-6, Panel D, sessions 23-26). Transfer to untrained voiceless phonemes (Figure 5-6, Panel C, Phase I) is discussed below.

Phase II

Baseline. A stable or slightly declining baseline on syllable-order reversed R-blend and fricative behaviors (e.g., “muh-NAY-bray” and “muh-NAY-vay”) was demonstrated before Phase II treatment began (Figure 5-6, Panel B, sessions 27-28). Accuracy rates for training items averaged 12% for IF behaviors and 2% for DF behaviors. Baseline levels for untrained transfer items were at or near zero (Figure 5-6, Panels C-D, sessions 27-28).
**Acquisition.** In Phase II, targets subjected to IF were again performed better (range 3%-77%) during acquisition than those given DF (range 1%-63%), even though stimulus-condition pairings were reversed from Phase I (Figure 5-6, Panel B, filled shapes), supporting our first prediction.

**Retention and Transfer.** Enhanced learning under conditions of feedback delay was evident in most trained and untrained items. On initial inspection (data not shown), P2 did not evidence retention or transfer (no improvement above baseline) of the trained behaviors for either condition in the probe setting, wherein these six items were mixed with 18 other transfer items (elicited ten times each). A gradual change in response quality, however, was noticed by the experimenter, and thus a formal error analysis was subsequently conducted. Many of P2’s productions contained multiple errors, including incorrect stress, distorted phonemes, and reduced articulatory speed. As P2 acquired the new skills in treatment sessions, the number of different errors in each production decreased, and more productions, although still scored as “incorrect,” were characterized by a speed-only error. Therefore, to capture this improvement on retention and transfer items, the data were re-analyzed with a more lenient criterion; the speed-only errors were also considered “correct.” Using the more lenient criterion, Figure 5-6 illustrates better overall retention (Panel B, unfilled shapes) and transfer (Panels C-D, Phase II) in the DF condition, consistent with the interpretation suggested by the more stringent criterion for accuracy. Retention accuracy ranged from 17% to 70% in the DF condition, and from 13% to 57% in the IF condition (Panel B). Transfer to untrained words ranged from 0 to 83% in the DF condition, and
from 0 to 57% in the IF condition (Panel D). One exception to this pattern may be observed with inspection of the untrained voicing transfer items, which appear enhanced consistently with the fricative targets in both phases (Figure 5-6, Panel C). A possible cause for this is that these probes involved devoicing of a previously trained voiced consonant, and this skill may be more complex in the context of an R-blend than in the fricative (singleton) condition, thus obscuring potential effects of the feedback timing manipulation on this type of transfer probe.

### Table 5-4

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prediction 1: HFF enhances performance during acquisition</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction 2: LFF enhances learning (retention and transfer)</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prediction 1: IF enhances performance during acquisition</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prediction 2: DF enhances learning (retention and transfer)</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 5.3 General Discussion

Taken as a whole, the data reported here (see Table 5-4 for summary of results) provide some support for the a priori hypotheses, but also raise important challenges for a simple interpretation relative to feedback. The results of two of the four participants studied in Experiment 1 (P1 and P3) support the prediction that reducing
the frequency of feedback enhances learning (retention and transfer). Of these two participants, the first prediction (i.e., that HFF would be associated with temporary performance enhancement during initial speech skill acquisition), was only borne out in one participant and in only one phase of his treatment. More importantly however, despite these inconsistent acquisition patterns in P1 and P3, LFF was associated with enhanced long-term retention and/or transfer of trained speech skills in both participants. Furthermore, comparison of P1’s Phase I and Phase II target acquisition data suggests that LFF during training of a particular behavior (in this case, fricatives in CV syllables in Phase I) may also carry over to boost future acquisition of that behavior in a different context (e.g., VC syllables in Phase II). The differential effects of HFF and LFF in this participant, as well as his demonstration of stable baselines, indicates that these changes cannot be solely attributed to spontaneous recovery during the subacute stage. The effect of feedback frequency in the two remaining participants (P2 and P4) was masked by the probable influence of stimulus complexity. That is, enhanced performance across both phases of the study was associated with phonologically and motorically simpler targets (see Barlow & Gierut, 1999), namely singletons vs. L-blends for P2 and 2-element blends vs. 3-element clusters for P4.

The results of Experiment 2 offer qualified support for the prediction that IF enhances the initial acquisition of speech skills. P1’s performance in both phases of the study revealed a pattern of learning (i.e., retention and transfer) which roughly paralleled acquisition under each of the two feedback conditions. That is, enhanced learning was observed in whichever condition was also associated with enhanced
acquisition. Moreover, the enhanced performance was consistently associated with plosives relative to fricatives, suggesting that it may have been driven by a stimulus complexity effect that was not overridden by an effect of feedback delay. P2 demonstrated the predicted pattern of IF associated with enhanced acquisition and DF associated with enhanced learning (i.e., retention and transfer). This did not reflect a stimulus effect, as the same pattern was observed in both phases. That is, under all stimulus-condition pairings, IF and DF were consistently associated with enhanced acquisition and learning, respectively.

Motor learning theory asserts that the availability of outcome information is a crucial factor for learning and that the frequency and temporal locus of external feedback may determine the availability of the outcome information to the learner and its effect on long-term learning. Frequent and immediate provision of augmented feedback may be detrimental to long-term learning because both may interfere with information processing of internal outcome information. According to the “guidance hypothesis,” high-frequency feedback is thought to enhance performance during training by providing increased guidance and heightening energy and motivation (Lee et al., 1990; Salmoni et al., 1984). However, these performance effects are temporary, and are dependent on the continued delivery of the feedback. Presenting KR feedback with lower frequency is hypothesized to enhance long-term learning by facilitating the development of self-evaluation and error-detection skills that the learner can apply to situations in which external feedback is not available (Bruechert et al., 2003). Similarly, imposing a feedback delay should promote the development of these skills
by allowing the learner sufficient time to process and build relationships (schemata) based on the four kinds of information discussed above. Indeed, participants in the Anderson et al. (2005) novel aiming task study reported using a greater variety of intrinsic feedback sources under conditions of feedback delay than with immediate or no feedback. Alternatively, it has been suggested that reducing the frequency of augmented feedback may serve to relieve the performer of a state of internal attentional focus induced by constant feedback, and to allow the performer to assume an external attentional focus guided by automatic motor control processes (Wulf et al., 2002). This theory argues that an external attentional focus benefits both acquisition performance and learning. While the current studies were not aimed at testing this theory, it may provide an explanation that the guidance hypothesis cannot, for the lack of demonstrable acquisition benefits of IF or HFF for some participants in the present experiments.

For impaired learners, particularly those with cognitive-linguistic or attentional deficits, a hypothetical “optimal window” for temporal locus of feedback can be conceived in which the learner has sufficient time to process this information, but not so much time that the activation of the information cannot be sustained. Immediately and/or consistently filling the post-movement interval with external outcome information may initially facilitate acquisition by reducing the cognitive demands of activating and maintaining in working memory the internal outcome information, but instantaneous and frequent provision of external feedback may also block the learner
from learning how to activate and maintain this information when external outcome information is withdrawn.

The data from the current studies contribute new evidence to the growing body of literature supporting the application of principles of motor learning to the motor speech system, but also raise many important questions. One question pertains to why the feedback manipulations were strong factors affecting acquisition and learning for some individuals but not for others. Furthermore, when comparing across studies, different participants appear to have benefited from different feedback manipulations. Curiously, no single participant demonstrated a benefit from both reduced frequency and delayed feedback. The answer may involve individual variation in the hypothesized “optimal window” for receiving external feedback, and may be based on ability to meet the attentional demands for activation and maintenance of various information types (cf. Li & Wright, 2000). Another potential explanation might involve the particular manifestation of AOS in individuals. Clark and Robin (1998) have suggested that AOS may reflect disruption of generalized motor programs in some patients, but a disruption of parameterization (of intact motor programs) in others. If these categories indeed represent two different etiological subtypes of AOS, we may expect to find individual differences in the effectiveness of feedback manipulations, depending on the specific nature of the underlying impairment in an individual. However, our dependent measure of perceptual accuracy does not allow us to differentiate these potentially different subtypes. Finally, it is unclear which aspects of the stimuli used here are best-conceived of as governed by GMPs versus
parameters. Since it has been suggested that these feedback manipulations operate differently on GMPs and parameters (e.g., Wulf et al., 1993), potential modification of the predictions for learning of these skills should be addressed in future studies.

Another important issue in deciphering the effects of particular motor learning conditions involves the potentially additive or interactive nature of multiple conditions of practice. For example, Wulf and colleagues (2002) reported an interaction between relative feedback frequency and type of feedback. Specifically, they demonstrated that reduced feedback frequency was only influential in the learning of sport skills under internal-focus (as opposed to external-focus) feedback conditions. In the current studies, all feedback given in Experiment 1 (feedback frequency) was provided immediately after production, and in Experiment 2 (feedback delay) the feedback frequency was a constant 100%. In addition to the feedback frequency and delay manipulations, other principles of motor learning were inherent to the design of this study. For example, trials of all behaviors in a given set (e.g., three fricative targets) were practiced in random order, and random practice has been shown to enhance retention and transfer of motor skills relative to blocked practice, and may facilitate learning by increasing the difficulty of the learning environment and by approximating a natural context (Knock et al., 2000). In addition, as noted earlier, some participants demonstrated a learning pattern of consistent higher accuracy for less-complex skills (e.g., L-blends vs. S-clusters), regardless of the feedback condition with which they were paired, suggesting that stimulus complexity may override the effects of the feedback variables tested here. Finally, and perhaps most importantly, a very large
number of repetitions of each behavior was elicited over the course of the treatment phases in each of the studies. Perhaps the most influential of the training conditions, amount of practice, may have obscured the potentially more subtle effects of the feedback manipulations. These potential interactions between multiple conditions of practice and feedback would best be addressed in the future by a group treatment design.

A further concern is that most of the work that has been done on motor learning for limb movements has used kinematic measurements, not perceptual accuracy judgments, to assess learning of movement skills. Perceptual accuracy judgments were used in this study because of their clinical relevance and ecological validity. However, using perceptual accuracy measures to make inferences about processes occurring at the kinematic level may not be appropriate. It is likely that perceptual judgments, while valuable for functional measurement in clinical outcome studies, may not correlate with kinematic values, and that gross accuracy judgments do not capture more subtle changes that occur at the kinematic level as a result of training under different conditions (Ballard & Robin, 2002; Ballard et al., 2007).

Finally, it is clear that treatments for AOS (those reported here and those in the literature) are effective for some participants to varying degrees and are not effective for other participants, although it is important to note that we observed either benefits to learning of reduced or delayed feedback or no difference between conditions, whereas there were no instances in which 100% or immediate feedback unambiguously enhanced learning. Understanding why particular experimental
manipulations work or do not for a given individual will ultimately require examination of treatment outcomes for a given therapy approach with regard to pre-treatment variables that may help define for whom a given treatment is effective. Such pre-treatment variables that will be important to consider in future research include severity of concomitant cognitive, linguistic, and motoric symptoms, neurological factors such as lesion size, and demographic factors such as time post-onset and education level. For example, although participants in the present study were screened informally for their ability to read the orthographic stimuli, it is possible that the presence of concomitant aphasia interacted with the motor learning studied here.

In conclusion, although treatment protocols for AOS abound in the clinical practice of speech-language pathology; few studies have undertaken to investigate the effects of theory-driven practice factors in the remediation of this disorder in a systematic fashion. The data from the current studies, as part of a programmatic line of research, contribute new evidence to the growing body of literature supporting the application of principles of motor learning to the motor speech system, and support the larger notion that the speech motor learning and control system shares properties with limb motor learning and control. In addition, they endorse the principle well-documented in the motor learning literature, that acquisition performance does not necessarily predict true learning. Independent of feedback manipulations, these data also contribute to the growing body of literature supporting the efficacy of treatment for chronic AOS.
Acknowledgements

Chapter 5 is in press as a manuscript entitled “Effects of Feedback Frequency and Timing on Acquisition, Retention, and Transfer of Speech Skills in Acquired Apraxia of Speech,” by Shannon N. Austermann Hula, Donald A. Robin, Edwin Maas, Kirrie J. Ballard, and Richard A. Schmidt. Permission for its use in this dissertation was granted by the publisher. The dissertation author is the primary investigator and author of the manuscript.

Appendix A

Table A-1. Carrier phrases used in Experiment 1.

<table>
<thead>
<tr>
<th>Easy</th>
<th>Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>“He likes to ___ each day.”</td>
<td>“The thermoelectricity may ___ the town.”</td>
</tr>
<tr>
<td>“It’s a ____ a day.”</td>
<td>“The municipality may ___ the pen.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>“The hospitalization may ___ the week.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“The hypothyroidism may ___ the ray.”</td>
<td>“The misinterpretation may ___ the net.”</td>
</tr>
</tbody>
</table>


6.0. Learner-controlled Feedback in Normal Speech Motor Programming

Self-controlled learning conditions and strategies have been recognized as powerful variables for enhancing information processing and retention for a number of years (see Zimmerman, 1989, for review), although until recently, this area has received little attention in the motor learning domain. Explanations regarding the benefits of self-controlled learning in fields such as academic learning have emphasized the interrelatedness of self-regulation, motivation, and feelings of self-efficacy (e.g., Bandura, 1993; Boekaerts, 1996; Carver & Scheier, 1990), although a precise mechanism of these effects on learning has remained elusive (Boekaerts, 1996).

The motor learning literature, however, has recently developed a testable account that hypothesizes that giving learners control over an aspect of their learning environment, namely feedback schedule, enhances learning by allowing learners to tailor extrinsic feedback to meet their needs (Chiviacowsky & Wulf, 2002; 2005). For example, when two groups of subjects both have control over their feedback delivery schedules, but one group registers their requests for terminal (post-production) feedback before a trial begins and the other group registers their requests after the trial, the latter group shows a learning advantage (Chiviacowsky & Wulf, 2005). This learning advantage cannot be attributed to heightened motivation or feelings of self-efficacy associated with self-control of feedback per se, as both groups controlled their feedback delivery. It also cannot be attributed to the rewarding effects of receiving positive feedback (learners typically request feedback after “good” trials, and “good”
feedback enhances learning relative to “poor” feedback; Chiviacowsky & Wulf, 2007), because both groups received the same amount of “good” feedback (Chiviacowsky & Wulf, 2005). A complementary hypothesis is that the learning advantage that is evidenced when learners request feedback after (as opposed to before) a trial is related to their engagement in post-production self-evaluation and error estimation processes. That is, learners who can request feedback after a trial is completed have the opportunity to base their requests on their own evaluations of their performance. On the other hand, learners who can only request feedback on upcoming trials that have yet to be performed cannot use this information in making their requests, and therefore may not engage in post-movement evaluative processes to the same extent.

One major purpose of Experiment 3 was to extend the findings from the limb motor learning literature (e.g., Chiviacowsky & Wulf, 2002; 2005; Janelle et al., 1995; 1997), demonstrating the benefits of self-controlled feedback to the speech domain. A second major purpose was to test the hypothesis advanced by Chiviacowsky and Wulf (2005) that self-controlled feedback may be more advantageous to learning because it encourages learners to engage in error estimation, which reduces their reliance on extrinsic feedback sources and promotes learning. To test this hypothesis, and to extend this line of inquiry to the motor speech domain, Experiment 3 trained four groups of speakers on a speech motor timing task similar to Chiviacowsky and Wulf’s (2002, 2005) finger movement timing task. Two groups of participants were analogous to Chiviacowsky and Wulf’s (2005) “self-after” and “self-before” groups. That is, both
groups selected when to receive terminal feedback on a learner-determined 30% of practice trials, but one group (Learner-Controlled Before; LCB) made this determination before each trial was performed, whereas the other group (Learner-Controlled After; LCA) decided whether they wanted feedback upon the completion of each movement. A third group of subjects (Learner-Controlled Before with Self-Evaluation; LCB+E) submitted their feedback requests before trials like the LCB group, but were encouraged to engage in post-movement self-evaluation processes before feedback was provided. If Chiviacowsky and Wulf’s (2005) hypothesis about the learning advantage being attributable to more extensive post-movement self-evaluative processing is correct, then the LCB+E group would be expected to demonstrate a learning advantage over the LCB group, and to learn as well as the LCA group. The fourth group (Learner-Controlled After with Self-Evaluation; LCA+E) was added as a check on the hypothesis that the LCA group would indeed engage in spontaneous error estimation, and that their learning advantage could be attributable to this. If this is the case, then addition of a post-movement self-evaluation task should impart no added benefit for the LCA+E group on learners. If, however, the LCA group’s predicted learning advantage is due to something other than spontaneous self-evaluation, then even further learning advantages might be expected with the addition of the self-evaluation task.

Following Chiviacowsky and Wulf’s (2005) findings, it was expected that all learners would execute responses more accurately on feedback trials, compared to no-feedback trials. Participant responses to a questionnaire similar to that used by
Chiviacowsky and Wulf (2002) were expected to provide a qualitative account of learners’ introspection about their feedback-requesting behavior.

An abundance of converging evidence from recent limb and speech motor learning studies has demonstrated that multiple measures of performance are required to capture the full effect of learning condition manipulations, as these manipulations often affect acquisition, retention, and transfer measures differently (see Maas, Robin, Austermann-Hula et al., in press, for review). For example, Chiviacowsky and Wulf (2002, 2005) found significant effects of self-controlled or “self-after” feedback on learners’ abilities to transfer trained skills to an untrained but related task, but not on their performance during acquisition or retention testing (cf. Chiviacowsky & Wulf, 2002, practice block 6). Developing training techniques that promote transfer of newly-acquired skills to related tasks and situations is of great relevance in applied domains. Therefore, it was important that for the current study, the effects of learner-controlled feedback and self-evaluation be examined not only with respect to acquisition and retention, but also for multiple types of transfer.

The motivation for the different types of transfer tasks administered in Experiment 3 is derived from Schema Theory (Schmidt, 1975). It is also based on the assumption that when feedback is provided after self-evaluation or after post-movement learner request, learners are more likely to engage in comparative processing of the extrinsic feedback and their own intrinsic feedback, thereby “tuning” their self-evaluation skills. It is reasonable to predict that this would lead to more elaborate processing and better storage of outcome information, sensory
consequences, initial conditions, and parameters assigned to the movement structure. This in turn would lead to development of stronger recall and recognition schemata, which would facilitate the programming of the practiced movement or a related one on a future trial. In contrast, learners who are not encouraged to engage in extensive post-movement processing (e.g., learners who must make feedback requests prior to movement execution and who are not required to self-evaluate upon movement completion) may exhibit weaker learning effects.

In the current study, participants practiced four different temporally-specified speech targets in homogenous blocks during the first session, and were given feedback about their absolute timing performance as indicated by their group assignment. In the second session, they were asked to execute these movements again in homogenous blocks, this time without feedback, for assessment of retention. Learners were also asked during the second session to execute the movements that they practiced in four novel but related transfer situations or tasks. The first of these was the random transfer task in which the exact four movements that had been practiced one day earlier were presented in random order within a single block. As demonstrated by Immink and Wright (1998, 2001), randomly ordering different types of trials creates contextual interference, requiring the participant to reconstruct, or re-program, the appropriate movement before each trial, instead of holding a single movement in working memory as would be possible for a block of homogenous trials. The remaining three transfer tasks, in which trials were also presented in random order, asked participants to change the parameterizations of the trained movements, or to build new movement
structures altogether. For the phoneme transfer task, learners were required to produce the targets that they practiced using the syllable “ba” with the syllable “chee” instead. If the relative timing patterns that speakers acquired during practice can be thought of as GMPs, then executing the same relative patterns with different sounds may be considered different parameterizations of those patterns. In the absolute difference transfer task, learners were asked to produce the practiced responses with longer absolute durations, analogous to the transfer tasks used by Chiviacowsky and Wulf (2002; 2005). Since the same relative timing structures were used, the application of different absolute durations constituted different parameterizations of GMPs. Finally, for the relative difference transfer probe, participants were tested on their ability to transfer the acquired absolute timings to new relative timing patterns (e.g., Magnuson & Wright, 2004). That is, the durations of the syllables and overall pattern durations remained unchanged, but the relationships between successive syllables in 4-syllable patterns were altered. Presumably, in order to execute these patterns efficiently, the practiced parameterizations (absolute timings) would have to be applied to new GMPs.

Recent literature has demonstrated that manipulating certain conditions of practice such as relative frequency of feedback can produce differential effects on GMP learning versus parameterization learning (Wulf et al., 1993). The purpose of the

---

6 Conversely, it has been suggested that GMPs may govern sound classes such as manner of articulation (Ballard, et al., 2007), although there is a great deal of evidence for generalization across this and other distinctive feature classes (e.g., McNeil et al., 2007), suggesting that it is possible for GMPs to be larger than phoneme class. At this juncture, with debate about the fundamental units of speech production ongoing, it seems reasonable to conclude that the basic units of speech are flexible, and vary with conditions such as speaker ability and frequency of use (e.g., Varley et al., 2006).
four different kinds of transfer tests in the current experiment was to explore whether the current manipulations in learner-controlled feedback also differentially affected these dimensions of learning. Another way to approach this question is to examine different aspects of response execution within the same response. For example, measuring the absolute timing error (e.g., absolute error, “AE” or root mean square error, “E”) can be used as an index of parameterization accuracy, whereas measuring relative timing error (i.e., relative error, “RE” or “AE_{prop}”) can serve as an index of structural integrity of the GMP (Schmidt, 1975). Following other literature in this area (e.g., Chiviacowsky & Wulf, 2002; 2005; Maas, 2006; Magnuson & Wright, 2004; Wright & Shea, 2001), both absolute and relative timing errors were examined as dependent variables in the current study to determine if the present manipulations affected GMP learning versus parameterization learning differently. It was predicted that all groups except LCB would presumably estimate intrinsic errors frequently, likely focusing on absolute temporal error mirroring the form of the extrinsic feedback, and that they would use these error signals to correct their parameterizations on subsequent attempts and thus would become proficient on tasks and measures of parameterization (i.e., absolute error, phoneme transfer, absolute difference transfer). On the other hand, the LCB group, presumably having less motivation to engage in extensive movement processing involving absolute temporal outcomes, might exhibit a disadvantage in parameterization skills. In other words, an interaction between Request Time (After vs. Before) and Self-Evaluation (Evaluation vs. No Evaluation) was predicted for absolute timing error, such that evaluation would reduce AE for the
“before” condition, but would not affect the “after” condition, and that LCB would have the greatest AE. Since parameterization reliability presumably supports the development of stable GMPs, this group interaction was also predicted for the measure of RE. Following the results of Chiviacowsky and Wulf (2002, 2005), it was predicted that error measures would only be significantly different between groups on transfer tasks.

A secondary purpose of Experiment 3 was to attempt to further understand the learner-controlled feedback phenomenon by examining its effects on pre-movement processing, as measured during different stages of motor programming. The self-selection paradigm (Immink & Wright, 1998, 2001) has been productively applied to investigating the operation of principles of motor learning such as random versus blocked practice on motor programming (e.g., Immink & Wright, 1998; 2001; Wright et al., 2004), and to demonstrating group differences between patients with and without putative selective deficits involving a single stage of motor programming (Maas, Robin, Wright et al., in press). Experiment 3 employed stimuli with the same temporal requirements that have been used to measure the INT and SEQ processes in previous work in finger movements and speech (e.g., Maas, Robin, Wright et al., in press; Immink & Wright, 2001, Klapp, 1995). Study time (ST) was used as a measure of the INT process of pre-programming the internal structure of individual units of movement, and reaction time (RT) served as an index of the SEQ process of retrieving and serially ordering multiple units of action.
Following the literature in this area, it was predicted that ST would be affected by the duration of single-syllable responses (i.e., longer ST for longer syllables), and that RT would be sensitive to the number of syllables to be executed (i.e., longer RT for greater number of syllables). Also, as demonstrated in other learning studies (e.g., Immink & Wright, 2001, Wright et al., 2004; Maas, Robin, Wright et al., in press), ST and RT were both expected to decrease with practice, as motor programming processes became more efficient, and to increase in the retention test 24 hours after practice was withdrawn. It was also predicted that, since groups LCB+E, LCA, and LCA+E were hypothesized to engage in more extensive movement processing and presumably more frequent response re-parameterization (i.e., to correct their absolute duration errors) based on their movement duration evaluations, they would demonstrate longer STs during acquisition. On the other hand, since practice was conducted in blocked format, learners could simply maintain one programmed response in the motor buffer for the duration of the homogenous block, bypassing INT processing if they chose not to re-parameterize the movement structure. Since learners in the LCB group were predicted to not evaluate their performance as often or as extensively, they might not have motivation to engage in INT processing as frequently, and would thus demonstrate shorter average STs. This would be substantiated by an interaction between the two group factors, Request Time (After vs. Before), and Self-Evaluation (Evaluation vs. No Evaluation) indicating that the self-evaluation task lengthened STs for the “before” condition, while it did not affect the “after” condition, and that LCB had shorter STs than the other three groups. This
pattern was expected to reverse in transfer conditions in which new parameterizations must be applied (i.e., phoneme transfer, absolute difference transfer), as the LCB group, which will have presumably had less experience with re-parameterization during acquisition, will therefore demonstrate inefficient INT processing. Group was not expected to have an effect on SEQ processing, as measured by RT.

6.1. Methods

6.1.1. Participants

Eighty undergraduate and graduate students from the University of Pittsburgh (49 female, 31 male) with no previous experience with this paradigm met the eligibility requirements and participated in both sessions of the experiment. Participants were right-handed native English speakers over the age of 18 ($M = 19.7$, range $= 18-25$) with normal (aided or unaided) hearing and vision acuity and no self-reported history of communication disorder, learning disability, neurological illness, psychiatric illness, or head injury. Monetary compensation was not provided; however, some students fulfilled a course requirement by participating in this research experiment. All participants provided informed consent as regulated by the University of Pittsburgh, VA Pittsburgh Healthcare System, and San Diego State University Institutional Review Boards.

6.1.2. Materials and Equipment
The experimental protocol was administered with a Dell laptop computer (Inspiron 600m) using E-Prime software, version 1.2 (Psychology Software Tools, Inc.). Vocal responses were recorded on a second Dell laptop computer (Latitude X1) using an omnidirectional stand microphone placed on the table in front of the participant. Vocal reaction times were collected using a second microphone and the voice key component of the Serial Response Box (Psychology Software Tools, Inc.), which also collected participants’ manual button-press responses. In order to produce optimal vocal recordings and prevent mis-triggering of the voice key, the distance between the microphones and the participant was determined by each participant’s habitual loudness while performing the task. Finally, a second response box (X-Keys SE) was connected to the experimental computer and used by the examiner to code response correctness and input values for feedback.

Target responses consisting of single syllables or four-syllable patterns of “ba” or “chee” were paired with orthographic symbols presented on the computer screen (see Table 6-1) and modeled for participants prior to each experimental task. The auditory models were generated by digitally recording (at 20 kHz with 10 kHz low-pass filtering) two male native English speakers (each set of stimuli consisted of only one speaker’s voice) producing “ba” or “chee” single syllables for slightly longer than the desired durations. These waveforms were then edited using Adobe Audition version 1.5 (Adobe Systems, Inc.) by excising a portion of the end of the vowel so that the exact desired durations were achieved (see Table 6-1). These single syllable

---

7 Due to observed imprecision of the voice key, all RT data presented here were measured directly using the recorded waveforms from the offset of the tone accompanying the imperative signal.
recordings of different durations were strung together with 100-millisecond silences intervening to construct the four-syllable non-isochronous patterns.

6.1.3. Tasks and Procedures

The self-selection paradigm (see Figure 6-1) was used for all experimental trials during practice, retention, and transfer. For each trial, the participant first saw a signal (“***”) for one second on the center of the computer screen indicating that the trial was going to begin. This signal was replaced by the pre-cue stimulus displaying the cue for the desired response as well as its total target duration (e.g., “Prepare 4S (1500). Press READY when you are Ready.”). The pre-cue stimulus remained on the screen until the participant used his or her right index finger to press a key on the Serial Response Box labeled “READY.” Participants were instructed to prepare the required response as much as possible before pressing the “READY” button. This interval was measured as study time (ST). Next, after a variable delay of 800 to 1200 ms (used to prevent anticipatory responding), the participant was given the imperative signal “GO!” (displayed on the center of the computer screen, accompanied by a 75 ms 2000 Hz tone) to execute the appropriate response (e.g., “ba-baaa-baa-ba”). The “Go!” signal remained in the center of the computer screen until the voice key detected a response. Participants were instructed to respond to the imperative signal as quickly and accurately as possible. Reaction time (RT), the latency between the “Go” signal and the onset of vocal response was measured by the software, which presented an error message if it was less than 100 ms (“Too early. Wait for the “Go” signal.”) or
greater than 1000 ms ("Too slow."). Following the participant’s response, the examiner used the second response box to code the response as correct or incorrect. If the participant did not produce the correct pattern or the correct phonemes, the error message “INCORRECT” was displayed on the computer screen for 2 seconds. The next trial was begun 2 seconds later.

Patterns were considered incorrect if the relative syllable durations were not achieved (i.e., both long syllables had to be longer than both short syllables). Incorrect trials, or trials in which RT did not meet criteria for inclusion, were re-run at the end of each block until the required number of correct trials was obtained. The maximum allowable number of trial attempts was equal to two times the number of correct trials required for that block. If this limit was reached before the participant correctly completed the required number of trials, the block ended and the participant proceeded to the next block or task.

Participants, who were blind to the purpose of the experiment, were randomly assigned to one of four experimental groups. The four groups, “Learner-Controlled Feedback Before” (LCB), “Learner-Controlled Feedback After” (LCA), “Learner-Controlled Feedback Before with Self-Evaluation” (LCB+E), and “Learner-Controlled Feedback After with Self-Evaluation” (LCA+E) had 20 participants each. Figure 6-1 illustrates the differences in trial events for each group. All groups received feedback on 30% of acceptable trials during the acquisition phase. Further, participants in all groups selected which trials they received feedback on, with the restriction that for every block of ten correct practice trials, feedback was requested three times.
Participants in the “Before” groups registered their feedback requests before responding in each trial, whereas participants in the “After” groups requested feedback after their responses. Both of the “Before” groups (i.e., LCB and LCB+E) were shown a display on the experimental computer at the beginning of each trial that read, “Do you want feedback on the next trial? Press Yes or No.” Participants were able to monitor their progress through practice blocks and keep track of feedback requests by referring to a progress bar shown on this screen, which filled as the participant progressed through 10 correct trials per block, as well as a schematic which indicated how many times they had received feedback in the current block, and how many feedback chances remained. When the participant had used all three feedback chances in a given block, this display changed to “You are out of feedback chances. Press No to continue” for the remaining trials. Immediately after the participant pressed a button labeled “Yes” or “No” on the Serial Response Box to indicate his feedback request, the trial was begun (i.e., starting with warning signal). If the trial was acceptable (RT was between 100 and 1000 ms, and the examiner perceptually judged the response as “correct”), and the participant requested feedback, the examiner immediately measured the total duration of the response using the recorded waveform and Adobe Audition, and transmitted this value to the experimental computer via the examiner response box. The actual response duration in milliseconds was then displayed, along with the target duration, to the participant on the experimental computer. This information remained on the screen until the participant pressed the “Ready” button to go on to the next trial. If the participant had requested feedback on
a trial that turned out to be unacceptable, this feedback phase was bypassed (without being counted as one of the three feedback requests), and the next trial was begun after the “Incorrect” message was displayed.

Groups LCA and LCA+E followed the same procedures for requesting feedback, except that they were asked after each correct trial whether or not they wanted feedback on the trial that they had just completed. Immediately after the examiner coded an LCA or LCA+E participant’s response as correct, the display described above was shown, prompting the participant to indicate whether or not feedback was desired. Actual duration was measured and displayed after the participant requested feedback. Participants in the “After” groups were not given the option to request feedback after unacceptable trials.

For the Evaluation groups (i.e., LCB+E and LCA+E), an additional task was completed just before feedback delivery, and only on trials for which feedback was requested. The self-evaluation task presented a display that asked the participant to estimate his or her error on the just-completed trial by using a 5-point equal-appearing Likert-type interval scale with labels 1 = “Too short,” 3 = “Just right,” and 5 = “Too long,” similar to the one Liu and Wrisberg (1997) used to have their participants provide subjective ratings of movement force. Participants were instructed to consider their performance on the just-completed trial and press a button (1-5) to indicate their self-assessment of its duration relative to the target duration. Participants were then given feedback in terms of absolute response duration, as described above.
<table>
<thead>
<tr>
<th>Request feedback</th>
<th>***</th>
<th>precue</th>
<th>press READY</th>
<th>“GO!”</th>
<th>Vocal onset</th>
<th>Vocal offset</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Time (ST)</td>
<td>INT</td>
<td>Variable Delay</td>
<td>Reaction Time (RT)</td>
<td>SEQ</td>
<td>Response Duration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Group LCB**

<table>
<thead>
<tr>
<th>Request feedback</th>
<th>***</th>
<th>precue</th>
<th>press READY</th>
<th>“GO!”</th>
<th>Vocal onset</th>
<th>Vocal offset</th>
<th>Self-Evaluate</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Time (ST)</td>
<td>INT</td>
<td>Variable Delay</td>
<td>Reaction Time (RT)</td>
<td>SEQ</td>
<td>Response Duration</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Group LCB+E**

<table>
<thead>
<tr>
<th>***</th>
<th>precue</th>
<th>press READY</th>
<th>“GO!”</th>
<th>Vocal onset</th>
<th>Vocal offset</th>
<th>Request feedback</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Time (ST)</td>
<td>INT</td>
<td>Variable Delay</td>
<td>Reaction Time (RT)</td>
<td>SEQ</td>
<td>Response Duration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Group LCA**

<table>
<thead>
<tr>
<th>***</th>
<th>precue</th>
<th>press READY</th>
<th>“GO!”</th>
<th>Vocal onset</th>
<th>Vocal offset</th>
<th>Request feedback</th>
<th>Self-Evaluate</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Time (ST)</td>
<td>INT</td>
<td>Variable Delay</td>
<td>Reaction Time (RT)</td>
<td>SEQ</td>
<td>Response Duration</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Group LCA+E**

---

**Figure 6-1.** Events on each practice trial in the self-selection paradigm for each group in Experiment 3. Note that Feedback and Self-evaluation events were only components of trials in which feedback was requested. LCB = Learner-Controlled Before; LCB+E = Learner-Controlled Before with Evaluation, LCA = Learner-Controlled After; LCA+E = Learner-Controlled After with Evaluation.
Acquisition for all groups consisted of 24 blocks of ten correct trials each, and took approximately 60 to 90 minutes to complete. Each block consisted of a single stimulus type. The order of blocks was organized such that each set of four consecutive blocks consisted of one block of each trial type (e.g., 1S, 1L, 4S, 4L) presented in random order. The first four blocks, being the first practice block for each trial type, were referred to in general as “practice block 1,” the second four blocks comprised “practice block 2,” and so on until the last four blocks, or “practice block 6.” Prior to the beginning of acquisition, participants were given written and verbal instructions, and had an opportunity to ask questions about the experimental protocol. Instructions included verbal models as well as written schematic representations of each of the four response types. In addition, at the beginning of each practice block, an auditory model of the appropriate response for that block was played by the computer at a comfortable loudness.
Table 6-1. Order of experimental tasks and orthographic target cues with associated temporal patterns for Experiment 3. S = short; L = long. Syllable durations given in milliseconds.

<table>
<thead>
<tr>
<th>Experimental Task Order</th>
<th>Target Cues</th>
<th>Syllable Durations (&amp; Totals) (ms)</th>
<th>Order of Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Practice (Session 1)</td>
<td>1S 1L 4S 4L</td>
<td>150-450-450-150 (1500)</td>
<td>Blocked</td>
</tr>
<tr>
<td></td>
<td>1S 1L 4S 4L</td>
<td>450-150-150-450 (1500)</td>
<td></td>
</tr>
<tr>
<td>2. Retention (Session 2)</td>
<td>1S 1L 4S 4L</td>
<td>150-450-450-150 (1500)</td>
<td>Blocked</td>
</tr>
<tr>
<td></td>
<td>1S 1L 4S 4L</td>
<td>450-150-150-450 (1500)</td>
<td></td>
</tr>
<tr>
<td>3. Random Transfer (Session 2)</td>
<td>1S 1L 4S 4L</td>
<td>150-450-450-150 (1500)</td>
<td>Random</td>
</tr>
<tr>
<td></td>
<td>1S 1L 4S 4L</td>
<td>450-150-150-450 (1500)</td>
<td></td>
</tr>
<tr>
<td>4. Phoneme Transfer (Session 2)</td>
<td>1S 1L 4S 4L</td>
<td>150-450-450-150 (1500)</td>
<td>Random</td>
</tr>
<tr>
<td></td>
<td>1S 1L 4S 4L</td>
<td>450-150-150-450 (1500)</td>
<td></td>
</tr>
<tr>
<td>5. Relative Difference Transfer (Session 2)</td>
<td>LLSS SSLL LSL LSL</td>
<td>450-450-150-150 (1500)</td>
<td>Random</td>
</tr>
<tr>
<td></td>
<td>1S 1L 4S 4L</td>
<td>150-150-450-450 (1500)</td>
<td></td>
</tr>
<tr>
<td>6. Absolute Difference Transfer (Session 2)</td>
<td>1S+ 1L+ 4S+ 4L+</td>
<td>225-675-675-225 (2100)</td>
<td>Random</td>
</tr>
<tr>
<td></td>
<td>1S+ 1L+ 4S+ 4L+</td>
<td>675-225-225-675 (2100)</td>
<td></td>
</tr>
</tbody>
</table>

Approximately 24 hours after the acquisition session, participants returned for a second session to complete retention and transfer testing. One retention probe and four transfer probes were administered in the same order for every participant (see...
Table 6-1). No feedback was provided except for the messages regarding RT. Responses were still coded as incorrect or correct by the examiner, and unacceptable trials were re-run at the end of each block to achieve the desired number of acceptable responses. The random transfer probe consisted of one block in which the four trained responses were elicited in random order, with the restriction that a given stimulus type not be presented more than twice in any group of four consecutive trials, until four correct productions of each type were collected. This was followed by the blocked retention probe, which consisted of four uniform blocks of five correct trials each. A block of 1S trials was completed first, followed by 1L, then 4S, and finally, 4L. Next, participants completed a phoneme transfer probe that was identical to the random transfer probe except that the trained syllable “ba” was replaced with “chee.” For the relative difference transfer probe, new four-syllable patterns were constructed that used the same absolute durations of “ba” as were practiced during acquisition, but that differed in their relative timing aspects (i.e., the short and long duration syllables were in different orders than the trained 4S and 4L patterns). The four 4-syllable patterns in this probe were randomly presented within a single block until four correct productions of each were collected. Finally, participants completed the absolute difference transfer probe, which was identical to the random transfer probe, except that the absolute durations of the “ba” syllables were lengthened by 33%. The relative patterns remained unchanged. Because this probe required participants to produce new

---

8 Although direct feedback about accuracy was not provided, the appearance of extra trials at the end of a block might have provided participants with indirect feedback about accuracy, particularly if only one or two response types were re-run a number of times. However, this would not have provided participants with any information about the temporal features of their responses, which were of primary relevance to the experimental manipulations about feedback.
durations that they were unfamiliar with, they first completed a warm-up exercise in which a model of each response was provided, followed by four practice trials. No feedback was given during this warm-up task except the RT messages described above.

For the blocked-presentation retention probe, participants were given one auditory model of each response type prior to the corresponding block of trials. For the remaining transfer probes, in which trials were presented in random fashion, auditory models of all four response types (one each) for that block were presented at the beginning of the probe. In addition, for retention and transfer probes, whenever a participant produced three consecutive incorrect responses to a particular stimulus type, an auditory model was re-played before the next trial.

Finally, after participants completed all experimental speech tasks in the second session, they were given a brief multiple-choice questionnaire to describe their thoughts about when and why they chose feedback, how much they thought the feedback helped them improve their durations, and whether or not they would have preferred to receive feedback on trials other than the ones they selected (adapted from Chiviacowsky & Wulf, 2002). These questions are listed in Table 6-2.
Table 6-2. Questionnaire questions and number of responses for each group in Experiment 3 (adapted from Chiviacowsky & Wulf, 2002). LCB = Learner-Controlled Before, LCB+E = Learner-Controlled Before with Evaluation, LCA = Learner-Controlled After, LCA+E = Learner-Controlled After with Evaluation.

<table>
<thead>
<tr>
<th>Q: When/why did you ask for feedback?</th>
<th>LCB</th>
<th>LCB+E</th>
<th>LCA</th>
<th>LCA+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) mostly after (before) good trials</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>(2) mostly after (before) bad trials</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(3) after (before) good and bad trials equally</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>(4) randomly</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>(5) none of the above</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q: When did you NOT ask for feedback?</th>
<th>LCB</th>
<th>LCB+E</th>
<th>LCA</th>
<th>LCA+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) after (before) good trials</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(2) after (before) bad trials</td>
<td>3</td>
<td>6</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>(3) none of the above</td>
<td>15</td>
<td>9</td>
<td>6</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q: How much do you think the feedback helped you improve your times?</th>
<th>LCB</th>
<th>LCB+E</th>
<th>LCA</th>
<th>LCA+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) not at all</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(2) a little</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>(3) somewhat</td>
<td>11</td>
<td>8</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>(4) a lot</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q: Do you think you received feedback on the right trials?</th>
<th>LCB</th>
<th>LCB+E</th>
<th>LCA</th>
<th>LCA+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) yes</td>
<td>18</td>
<td>14</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>(2) no</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q: (If no) When would you have preferred to receive feedback?</th>
<th>LCB</th>
<th>LCB+E</th>
<th>LCA</th>
<th>LCA+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) after (before) good trials</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(2) after (before) bad trials</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(3) doesn’t matter</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(4) none of the above</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

6.1.4. Design and Analysis

The four primary dependent variables of interest were study time (ST), reaction time (RT), absolute timing error (AE), and relative timing error (RE). ST and
RT were included as indices of the two motor programming stages, INT and SEQ, respectively. AE and RE were measures of the response execution stage. In addition to examining questionnaire responses about feedback choices, AE was also examined for feedback versus no-feedback practice trials to investigate whether or not actual error in movement durations was related to participants’ feedback-requesting behavior. Accuracy (in terms of proportion of acceptable trials) was also examined as a gross measure of group differences.

Analyses were conducted on all acceptable trials in retention and transfer blocks, and for the first (block 1), intermediate (block 3), and last (block 6) practice blocks. Practice blocks 2, 4, and 5 were not analyzed. Unacceptable trials that were identified on-line as described above were excluded from the analyses. Next, trials that were determined to be unacceptable off-line, via acoustic measurement of recorded waveforms, were removed. These included trials which violated the RT criteria but, because of voice key imprecision, were not aborted during the experiment, and trials of 4-syllable sequences that did not conform to the appropriate relative patterns (both long syllables must have been longer than both short syllables), yet were not perceptually judged as incorrect by the examiner during the experiment. To assess study time (ST) performance, trials that were deemed acceptable on-line, with STs between 100 ms and 10 seconds were included. RT analyses were conducted on correct trials that met the criteria of being greater than 100 ms and less than 1000 ms. Of the 10,392 practice trials administered, 4.8% were excluded for RT violations, 0.7% for ST violations, and 3.3% for incorrect patterns or phonemes. Of the 7,838
probe trials that were administered, 3.7% were excluded for RT violations, 1.6% for ST violations, and 12.0% for incorrect patterns or phonemes.

Absolute and relative error values were calculated from acoustic measurements made on recorded waveforms for all acceptable trials in retention and transfer blocks, and in practice blocks 1, 3, and 6. Absolute error (AE), a measure of overall accuracy in performance (Schmidt & Lee, 2005), was chosen as a key dependent measure of the response execution stage because it resembled (non-directionally) the form of feedback that participants received during acquisition (i.e., total target duration and total actual response duration). AE was defined as the average absolute deviation between the total target duration and the total response duration, in milliseconds. Duration for single-syllable responses was measured from the onset of the consonant (release of plosive burst) to the offset of the vowel (end of periodic oscillations in waveform). For the 4-syllable responses, total duration was measured from the onset of the first syllable to the offset of the final syllable.

Relative error (RE) was also calculated from the acoustic measurements. Relative timing performance provides an index of coherence of the putative GMP for multi-unit patterns, regardless of absolute temporal structure (e.g., Shea, Wulf, Park, & Gaunt, 2001). To compute RE, proportions were calculated for each syllable and inter-syllable interval relative to the total duration of the response, and compared to the target proportions. RE, expressed as a proportion, was calculated by summing the absolute deviations between target and actual proportions for each segment within a response. Accuracy was defined for each task as the proportion of correct trials to the
total number of trials administered. The total number of trials depended on how many trials were re-run at the end of the block in order to satisfy the requirement for minimum number of correct trials. Because many different factors influenced whether or not a trial had to be re-run (i.e., RT, pattern correctness, and phoneme correctness), accuracy was only used as a secondary measure to provide a gross indication of overall performance.

Acoustic measurements were performed by two raters using Adobe Audition software. The primary rater made measurements of RT and syllable and inter-syllable interval durations for all trials, and the secondary rater independently measured a randomly selected 10% of trials for the blocked retention and random transfer probes. Further, the primary rater re-rated 10% of all trials (in each experimental task) and intra-rater reliability was computed. Reliability was satisfactory for practice block 1 (Intra-rater r = 0.993, mean absolute difference = 12 ms, SD = 14 ms), practice block 3 (Intra-rater r = 0.990, mean absolute difference = 10 ms, SD = 21 ms), practice block 6 (Intra-rater r = 0.989, mean absolute difference = 14 ms, SD = 16), the blocked retention probe (Inter-rater r = 0.993, mean absolute difference = 13 ms, SD = 16 ms; Intra-rater r = 0.993, mean absolute difference = 13 ms, SD = 15 ms), the random retention probe (Inter-rater r = 0.990, mean absolute difference = 16 ms, SD = 17 ms; Intra-rater r = 0.965, mean absolute difference = 21 ms, SD = 37 ms), the phoneme transfer probe (Intra-rater r = 0.997, mean absolute difference = 8 ms, SD = 9 ms), the relative difference transfer probe (Intra-rater r = 0.995, mean absolute difference = 10
ms, SD = 15 ms), and the absolute difference transfer probe (Intra-rater r = 0.997, mean absolute difference = 12 ms, SD = 13 ms).

The four groups in this experiment were compared in a series of analyses involving two crossed between-subjects factors, Request Time (After groups vs. Before groups) and Evaluation (Self-evaluation groups and No Evaluation groups). As is the standard in the literature from which this study draws (e.g., Deger & Ziegler, 2002; Immink & Wright, 2001; Klapp, 1995; 2003; Maas, Robin, Wright et al., in press; Wright et al., submitted), separate analyses were conducted to focus on the most critical aspects of the data for examining the different hypotheses (e.g., duration effect, sequence length effect). For ST and RT during acquisition, the analyses involving the duration effect (i.e., comparing 1S and 1L responses) were conducted on one-syllable responses using 2 (Request Time) x 2 (Evaluation) x 2 (Duration) x 3 (Block) ANOVAs with repeated measures on Duration and Block. Analyses involving the Sequence Length effect were conducted using 2 (Request Time) x 2 (Evaluation) x 2 (Sequence Length; 1-syllable vs. 4-syllable) x 3 (Block) ANOVAs with repeated measures on the last two factors. Timing error (i.e., AE and RE) analyses for acquisition were conducted separately for 1-syllable and 4-syllable responses using 2 (Request Time) x 2 (Evaluation) x 2 (Duration; 1S vs. 1L or 4S vs. 4L) x 3 (Block) ANOVAs with repeated measures on the last two factors. Similar analyses for all of these variables were conducted for relative retention by changing the repeated factor of Block to the repeated factor of Phase (practice block 6 vs. retention). Transfer tasks were analyzed separately, also involving repeated measures on Duration and Sequence
Length (with no Block or Phase factor), except for the relative difference transfer probe, in which all four stimuli were 4-syllable patterns, and were included in the model as a factor called “Pattern” with four levels (see Table 6-1). Analysis of accuracy was conducted for each task type as a whole, without separating data by Duration, Sequence Length, Block, or Phase. Significant ($p < .05$) and marginally-significant ($0.05 \leq p < .10$) ANOVA results, using Huynh-Feldt-adjusted degrees of freedom when the assumption of sphericity was not met, are reported. Follow-up pairwise comparisons with Tukey-adjusted alpha levels of less than .05 are also reported when meaningful.

6.2. Results

A general overview of predictions and results for each dependent variable can be found in Table 6-3.

6.2.1. Acquisition

Feedback. Participants from all groups requested feedback three times per block of practice trials. Average AE for feedback trials and the first three no-feedback trials per block was computed and compared within-subject. A main effect of Feedback, $F(1, 76) = 14.19, p < .001, \eta^2 = .157$, revealed that overall, participants had greater absolute timing error on trials in which they requested feedback ($M = 79$ ms, $SD = 31$ ms) than trials in which they did not request feedback ($M = 74$ ms, $SD = 32$ ms).
ms). There were no differences between groups in terms of AE for feedback versus no-feedback trials.

The responses to the Questionnaire (see Table 6-2) revealed that more learners in the LCA group (11) thought that they had requested feedback on good trials compared to other groups (LCB = 3; LCB+E = 2; LCA+E = 3), and that the only learners who thought that they had asked for feedback mostly on bad trials were in the LCB (1) and LCB+E (2) groups. Substantially more learners in the LCA+E group (11) reported requesting feedback on good and bad trials equally, relative to other groups (LCB = 6; LCB+E = 6; LCA = 3). Half of learners in “before” groups (i.e., LCB, LCB+E) reported requesting feedback “randomly” or “none of the above,” while 30% of learners in the “after” groups (i.e., LCA, LCA+E) responded this way. As for when learners reported that they did not request feedback, the LCA group was the only group to respond predominantly with “after bad trials,” whereas the other three groups did not strongly endorse either good or bad trials. Learners’ perceptions about the helpfulness of the feedback also varied between groups. The only participants who reported that the feedback helped them “not at all” were in the “before” conditions (i.e., LCB = 1; LCB+E = 2). More participants in the LCA+E group (11) thought that the feedback helped them “a lot,” compared to the other groups (LCB = 6; LCB+E = 7; LCA = 6). Additionally, although the majority of participants in each group thought that they had received feedback on the “right” trials, at least twice as many LCB+E learners (6) did not think that they received feedback on the right trials, relative to
each of the other groups (LCB = 2; LCA = 3; LCA+E = 1). This subset of learners was ambiguous about when they would have preferred to receive feedback.

**Accuracy.** Inaccurate trials and trials in which RT was less than 100 ms or greater than 1000 ms accounted for 8.2% of practice data. A 2 (Request Time) x 2 (Evaluation) x 3 (Block) ANOVA on accuracy revealed a main effect of Block, $F(1.66, 125.96) = 25.91, p < .001, \eta^2 = .254$, indicating that the percentage of accurate trials increased from Block 1 to 3, and from 1 to 6, but not from Block 3 to 6 (Figure 6-2). There were no effects of group on accuracy during practice.

**Figure 6-2.** Mean accuracy (as proportion of acceptable trials) for each group during practice, retention, and transfer blocks. LCB = Learner-Controlled Before, LCB+E = Learner-Controlled Before with Evaluation, LCA = Learner-Controlled After, LCA+E = Learner-Controlled After with Evaluation.

**Study Time (INT process).** The ANOVA testing the effects of groups on study time (ST) for short versus long duration single syllables throughout practice (Figure 6-3) revealed main effects of Block, $F(1.6, 121.55) = 187.96, p < .001, \eta^2 = .712$, and
Request Time, $F(1, 76) = 13.24, p < .001, \eta^2 = .148$, with marginally-significant main effects of Evaluation, $F(1, 76) = 2.97, p = .089, \eta^2 = .038$, and Duration, $F(1, 76) = 2.91, p = .092, \eta^2 = .037$. Interactions between Duration and Evaluation, $F(1, 76) = 4.05, p < .05, \eta^2 = .051$, Duration, Evaluation, and Request Time, $F(1, 76) = 6.28, p < .05, \eta^2 = .076$, and Duration, Request Time, and Block, $F(1.28, 97.34) = 3.64, p < .05, \eta^2 = .046$, were all significant. The interpretations of these results were superseded by a significant 4-way interaction between Block, Duration, Request Time, and Evaluation, $F(1.28, 97.34) = 3.880, p < .05, \eta^2 = .049$, indicating that (a) ST decreased between blocks 1 and 3 (and between 1 and 6), but not between blocks 3 and 6, for all levels of the other factors, (b) the marginal effect of duration was driven by nonsignificant differences in the direction of Short > Long for the After groups in the first block, but no other pair-wise comparisons involving the Duration effect were significant, and (c) at the level of pair-wise comparisons, LCB and LCB+E had significantly shorter STs than LCA+E for the 1S response in all blocks (for Block 1, LCA+E also had longer ST than LCA), and for the 1L response in Block 1, LCB had significantly shorter ST than both LCA and LCA+E, and LCB+E had shorter ST than LCA.
Figure 6-3. Study Time (ST, in milliseconds) for groups across practice, retention, and transfer phases of Experiment 3 (top panel). Bottom panel shows ST for each group and each response type in Experiment 3. Abbreviations as before.

The ANOVA testing the effects of groups on ST for 1-syllable versus 4-syllable sequences revealed main effects of Block, $F(1.81, 137.55) = 252.51, p < .001$, $\eta^2 = .769$, and Sequence Length, $F(1,76) = 29.14, p < .001, \eta^2 = .277$. The
interpretation of these main effects, however, was superseded by a Block by Sequence Length interaction, $F(1.44, 109.47) = 5.99$, $p < .01$, $\eta^2 = .073$, indicating that the sequence length effect (4 syllables > 1 syllable) was significant only for practice blocks 1 and 3, and that the block effect (1 > 3 > 6) was not significant for the 1-syllable responses between the last two practice blocks. Further, there were effects of both between-subjects factors: “After” groups ($M = 1324$ ms, $SD = 860$ ms) had longer STs than “Before” groups ($M = 1021$ ms, $SD = 726$ ms), $F(1, 76) = 8.64$, $p < .005$, $\eta^2 = .102$, and Evaluation groups ($M = 1266$ ms, $SD = 888$ ms) had marginally longer STs than No Evaluation groups ($M = 1079$ ms, $SD = 711$ ms), $F(1, 76) = 3.29$, $p = .074$, $\eta^2 = .042$. The two group factors did not interact with other factors or with each other.

**Reaction Time (SEQ process).** The RT analysis for single-syllable targets during practice (Figure 6-4) revealed significant main effects of Block, $F(1.62, 123.28) = 90.07$, $p < .001$, $\eta^2 = .542$, and Duration, $F(1, 76) = 201$, $p < .001$, $\eta^2 = .726$. The interpretation of these results was superseded by a significant Block by Duration interaction, $F(1.5, 113.98) = 5.39$, $p < .05$, $\eta^2 = .066$, indicating that RT for the 1L response decreased between all blocks, whereas RT for 1S was only different between the first and final blocks. The Duration effect (1L > 1S) was significant in each block. There was also a marginally-significant interaction between Request Time and Evaluation, $F(1, 76) = 3.79$, $p = .055$, $\eta^2 = .048$, suggesting a trend such that, although no pair-wise comparisons were significant, the addition of the self-evaluation task affected the “After” groups (LCA $M = 289$ ms, $SD = 101$ ms; LCA+E $M = 328$ ms, $SD$
= 125 ms) differently than the “Before” groups (LCB $M = 330$ ms, $SD = 103$ ms; LCB+E $M = 290$ ms, $SD = 115$ ms).

The analysis of 1- versus 4-syllable responses⁹ yielded significant main effects of Block, $F(1.71, 129.8) = 115.91, p < .001, \eta^2 = .604$, Sequence Length, $F(1, 76) = 195.99, p < .001, \eta^2 = .721$, and Duration, $F(1,76) = 121.55, p < .001, \eta^2 = .615$. The interpretations of these effects were superseded by significant interactions between each pair of these factors, as well as an interaction between all three $F(1.6, 122.1) = 7.26, p < .01, \eta^2 = .087$. These interactions indicated that (a) RTs shortened significantly between all blocks for all targets, except RT for the 1S target did not change between Blocks 3 and 6, (b) there was a significant sequence length effect for the 1S-4S pair, but not for the 1L-4L pair, and (c) there was a significant duration effect (Long > Short) for 1-syllable responses only. Additionally, two interactions involving groups were revealed. First, a significant Request Time by Sequence Length interaction, $F(1, 76) = 4.11, p < .05, \eta^2 = .051$, indicated that, although there were no significant mean differences in overall RT between “After” and “Before” groups, the “Before” group exhibited a significantly greater sequence length effect (1-syllable $M = 310$ ms, $SD = 111$ ms, 4-syllable $M = 383$ ms, $SD = 124$ ms) than the “After” group (1-syllable $M = 313$ ms, $SD = 114$ ms, 4-syllable $M = 368$ ms, $SD = 122$ ms). Second, a marginal interaction between Request Time and Evaluation, $F(1, 76) = 3.7, p = .058, \eta^2 = .046$, suggested a trend such that, although no pair-wise comparisons were significant, the addition of the self-evaluation task affected the “After” groups (LCA

---

⁹ The duration factor is included in the ANOVA model testing the sequence length effect on RT during practice because it was found to interact with the Sequence Length factor such that a sequence length effect was present only for the 4S vs. 1S comparison, and not the 4L vs. 1L comparison.
\[ M = 329 \text{ ms}, \ SD = 111 \text{ ms}, \ \text{LCA+E } M = 352 \text{ ms}, \ SD = 130 \text{ ms} \] differently than the “Before” groups (LCB \[ M = 370 \text{ ms}, \ SD = 118 \text{ ms}, \ \text{LCB+E } M = 322 \text{ ms}, \ SD = 124 \text{ ms} \]), and group LCB had numerically longer RTs than all other groups in eleven of twelve conditions (4 responses x 3 practice blocks).

**Figure 6-4.** Reaction Time (RT, in milliseconds) for groups across acquisition, retention, and transfer phases of Experiment 3 (top panel). Bottom panel shows RT for each group and each response type in Experiment 3. Abbreviations as before.
**Absolute Error.** For absolute timing error of single syllables (Figure 6-5), analysis of acquisition data for the four groups revealed main effects of Block, $F(1.67, 126.99) = 31.84, p < .001, \eta^2 = .295$, Duration, $F(1, 76) = 10.76, p < .01, \eta^2 = .124$, and Request Time, $F(1, 76) = 8.51, p < .01, \eta^2 = .101$. The interpretations of Block and Duration effects were superseded by a significant interaction between these two factors, $F(1.85, 140.24) = 7.82, p < .005, \eta^2 = .093$, indicating that the duration difference (Short > Long) was only reliable in blocks 1 and 3, and that absolute timing of long syllables did not improve between blocks 3 and 6, whereas timing of short syllables improved in each block. In addition, a significant Duration by Request Time interaction, $F(1, 76) = 7.61, p < .01, \eta^2 = .091$, indicated that AE depended on duration for the “Before” groups (1S $M = 79$ ms, $SD = 48$ ms, 1L $M = 56$ ms, $SD = 41$ ms) but not the “After” groups (1S $M = 54$ ms, $SD = 33$ ms, 1L $M = 52$ ms, $SD = 25$ ms), and that for the 1S response, the “Before” groups had greater AE than the “After” groups.\(^{10}\)

Analysis of acquisition data for 4-syllable sequences revealed significant main effects of Block, $F(1.38, 104.76) = 23.51, p < .001, \eta^2 = .236$, and Duration, $F(1, 76) = 6.46, p < .05, \eta^2 = .078$. These effects were qualified by a significant Block by

---

\(^{10}\) Inspection of single syllable Absolute Error data for Practice uncovered two extreme subject outliers (individual trials for these subjects were uniform; that is, there were no trials outliers for these subjects). Subject 48’s mean AE for 1L in Block 3 was greater than 4 standard deviations above the mean, and Subject 93’s mean for 1S in Block 1 was greater than 3.5 standard deviations above the mean. Both subjects were in the LCB+E group. When the data were re-analyzed without these subjects, all of the original effects remained, but two additional effects emerged. First, at Block 1, the Evaluation groups ($M = 62, SD = 33$) had less error than the No Evaluation groups ($M = 82, SD = 51$). Second, a Block by Request Time interaction emerged, indicating that at Block 1, the After groups ($M = 63, SD = 32$) had less error than the Before groups ($M = 81, SD = 52$), and that the After groups improved significantly only between the last two blocks, whereas the Before groups improved significantly only between the first two blocks.
Duration interaction, \( F(1.29, 98.2) = 9.77, p < .01, \eta^2 = .114 \), indicating that AE was greater for 4S than 4L only in the first block, and that the only significant difference between blocks was for the 4S pattern from Block 1 to Block 3. There were no group effects for AE during practice.\(^{11}\)

\(^{11}\) When the AE data for 4-syllable patterns during practice were examined, nine extreme subject outliers were found (LCB = 2, LCB+E = 3, LCA = 3, LCA+E = 1) to have an AE mean at least 3 standard deviations above their group means. When the analysis was re-run excluding these extreme outliers, a similar pattern of results was obtained. Although the Duration effect was no longer significant, the Duration by Block interaction persisted. In addition, the analysis of this constrained set of data revealed that the decrease in AE from Blocks 1 and 3 also pertained to the 4L pattern. A significant Block by Request Time interaction also emerged, \( F(1.64, 109.81) = 4.91, p < .05, \eta^2 = .068 \), indicating that, in the first block, the After groups (\( M = 114 \) ms, \( SD = 84 \) ms) had greater absolute timing error than the Before groups (\( M = 90 \) ms, \( SD = 42 \) ms).
Relative Error. The analysis of relative timing error for the four-syllable patterns during practice for the four groups (Figure 6-6) revealed main effects of Block, $F(1.7, 129.4) = 51.91, p < .001$, $\eta^2 = .406$, and Duration, $F(1,76) = 42.3, p <$
.001, $\eta^2 = .358$. The interpretation of these results was superseded by a significant interaction between Block and Duration, $F(1.73, 131.32) = 10.47, p < .001$, $\eta^2 = .121$, indicating that there was more error for 4S than 4L across practice, and that the amount of error for the 4S response decreased with each practice block, whereas the only difference in error for the 4L pattern was between the first and last blocks. There were no significant group differences in RE during practice.
Figure 6-6. Relative Error (RE, as a proportion) for groups across acquisition, retention, and transfer phases of Experiment 3 (top panel). Bottom panel shows RE for each response type for each group during practice blocks. Abbreviations as before.

6.2.2. Retention

In order to control for any potential group differences that emerged during acquisition and observe the degree of performance decrement after the withdrawal of
training, the retention probe was analyzed in relation to the last block of practice (block 6) by including Phase (practice block 6 vs. retention block) as a repeated factor in the following analyses.

Accuracy. Inaccurate trials and trials violating RT criteria accounted for 7.6% of retention data. The 2 (Request Time) x 2 (Evaluation) x 2 (Phase) ANOVA on accuracy (Figure 6-2) revealed no main effects. There was, however, a marginally significant Request Time by Evaluation interaction, $F(1, 76) = 3.21, p = .077, \eta^2 = .041$, although no pair-wise comparisons involved in this interaction were significant. Group LCA ($M = 98.9\%, SD = 2.1\%$) was marginally more accurate than LCA+E ($M = 95.1\%, SD = 8.5\%$).

Study Time (INT process). The analysis of the effects of Group and Duration on single-syllable STs across Phase (Practice 6 vs. Retention) (Figure 6-7) revealed significant main effects of Phase, $F(1, 76) = 138.57, p < .001, \eta^2 = .646$, and Duration, $F(1, 76) = 37.29, p < .001, \eta^2 = .329$, with a marginally-significant effect of Request Time, $F(1, 76) = 3.6, p = .062, \eta^2 = .045$. A significant Phase by Duration interaction, $F(1, 76) = 15.18, p < .001, \eta^2 = .167$, qualified the duration effect such that ST for 1S was longer than for 1L at the retention block, but not at the last practice block. The interpretation of the Request Time effect was superseded by a significant Phase by Request Time interaction, $F(1, 76) = 7, p < .05, \eta^2 = .084$, and a marginal Phase by Request Time by Evaluation interaction, $F(1, 76) = 2.96, p = .089, \eta^2 = .037$, indicating that all groups had longer STs in the retention block than in practice.
block 6, and that group LCA+E had longer ST than groups LCB and LCB+E in practice block 6, but that no groups were significantly different in the retention block.

The analysis of the effects of Group and Sequence Length on 1- and 4-syllable target STs across Phase yielded a significant main effect of Request Time, $F(1, 75) = 4.39, p < .05, \eta^2 = .055$, indicating that the “After” groups ($M = 1159$ ms, $SD = 699$ ms) had significantly longer STs than the “Before” groups ($M = 946$ ms, $SD = 695$ ms). Also present were significant main effects of Phase, $F(1, 75) = 103.82, p < .001, \eta^2 = .581$, and Sequence Length, $F(1, 75) = 21.78, p < .001, \eta^2 = .225$, as well as a significant interaction between Phase and Sequence Length, $F(1, 75) = 6.12, p < .05, \eta^2 = .075$, indicating that ST increased significantly from practice block 6 to the retention test, and that an effect of sequence length (4 syllables > 1 syllable) was present only at the retention test.\(^{12}\)

\(^{12}\) A potential concern about this analysis was that the group effect (After > Before) and/or the sequence length effect in the retention block (4 > 1) might be driven only by the 4S stimulus. Analyses were conducted for the 1S versus 4S and the 1L versus 4L comparisons separately, and it was confirmed that the sequence length effect was demonstrated at both levels of duration. Further, although there was a significant Duration by Request Time interaction when Duration was included in the model, none of the relevant pair-wise tests comparing After and Before groups was significant, so the Request Time effect was not localized to a particular stimulus type.
Figure 6-7. Study Time (ST, in milliseconds) for groups and response types across practice and retention phases of Experiment 3. Abbreviations as before.

Reaction Time (SEQ process). The RT analysis for single syllables across practice and retention phases (Figure 6-8) yielded a significant main effect of Duration, $F(1,76) = 109.44$, $p < .001$, $\eta^2 = .590$, and a Duration by Phase interaction, $F(1,76) = 14.04$, $p < .001$, $\eta^2 = .156$. These results indicated a duration effect (1L > 1S) in each phase and an increase in RT for 1S (but not 1L) responses between practice and retention. Although there were no main effects of groups, marginally-significant interactions between Request Time and Evaluation, $F(1,76) = 2.98$, $p = .088$, $\eta^2 = .038$, and Request Time, Evaluation, and Duration, $F(1,76) = 3.137$, $p = .081$, $\eta^2 = .04$, were noted, suggesting the Evaluation factor tended to have different
effects on the “After” versus the “Before” groups, and the strength of these effects differed across Duration.

The analysis of RT across sequence length revealed main effects of Phase, $F(1,75) = 8.27, p < .01, \eta^2 = .099$, and Sequence Length, $F(1,75) = 142, p < .001, \eta^2 = .654$. Superseding the interpretations of these effects, however, were significant interactions between Phase and Sequence Length, $F(1,75) = 7.87, p < .01, \eta^2 = .095$, Request Time and Evaluation, $F(1,75) = 5.1, p < .05, \eta^2 = .064$, and Sequence Length, Request Time, and Evaluation, $F(1,75) = 6.21, p < .05, \eta^2 = .076$. In addition, the four-way interaction between Phase, Sequence Length, Request Time, and Evaluation was marginally-significant, $F(1,75) = 3.11, p = .082, \eta^2 = .04$. Together, these results indicated that (a) RT significantly increased from practice to retention only for group LCA+E, (b) all groups demonstrated sequence length effects ($4 > 1$) in both practice block 6 and retention, and (c) the addition of the self-evaluation task affected the “After” groups (LCA $M = 279$ ms, $SD = 88$ ms, LCA+E $M = 315$ ms, $SD = 119$ ms) differently than the “Before” groups (LCB $M = 328$ ms, $SD = 97$ ms, LCB+E $M = 281$ ms, $SD = 102$ ms), with group LCB having the numerically longest RTs in six of eight conditions, although no pair-wise comparisons were significant.
Figure 6-8. Reaction Time (RT, in milliseconds) for groups and response types across acquisition and retention phases of Experiment 3. Abbreviations as before.

Absolute Error. When the single syllable absolute timing error data for practice block 6 and retention (Figure 6-9) were submitted to analysis, main effects of Phase, $F(1,76) = 61, p < .001, \eta^2 = .445$, and Request Time, $F(1,76) = 7.236, p < .01, \eta^2 = .087$, were revealed, indicating that there was a significant absolute timing performance decrement between the last block of practice and retention, and that overall, the “After” groups ($M = 55 \text{ ms}, SD = 42 \text{ ms}$) had significantly less absolute timing error than the “Before” groups ($M = 71 \text{ ms}, SD = 51 \text{ ms}$).\footnote{In the single-syllable retention analyses, five subjects (LCA = 1, LCA+E = 2, LCB = 0, LCB+E = 1) displayed mean performance that was more than 3 standard deviations above their group means for a particular phase and duration. When the analysis was constrained by excluding these subjects, the main effects of Phase and Request Time persisted, but a Phase by Duration by Request Time interaction emerged which indicated that there was significantly greater error in Retention than in Practice 6 for all
For the 4-syllable patterns, there were no significant differences between groups in retention. However, there were significant main effects of Phase, $F(1,75) = 42.21, p < .001, \eta^2 = .36$, and Duration, $F(1,76) = 14.63, p < .001, \eta^2 = .163$, as well as an interaction between Phase and Duration, $F(1,75) = 9.66, p < .005, \eta^2 = .114$, indicating that there was a significant decrement in performance between practice block 6 and retention, and that, while 4L and 4S responses had similar error in practice 6, there was greater error associated with 4L than 4S in the retention phase.14

14 In the retention analysis for sequences, five subjects (LCA = 2, LCA+E = 0, LCB = 73, LCB+E = 2) had mean AEs that were greater than 3 standard deviations above those of their groups for particular levels of Phase and Duration. When the analysis was re-run on the data set constrained by the exclusion of these subjects, the significant effects persisted, and a marginal interaction emerged between Request Time, Evaluation, Phase, and Duration, $F(1,70) = 3.94, p = .051, \eta^2 = .053$, suggesting that the LCB+E group did not change across Phase.
Relative Error. The analysis of relative timing error (RE) for 4-syllable patterns across practice block 6 and retention (Figure 6-10) revealed main effects of Phase, $F(1, 75) = 18.95, p < .001, \eta^2 = .202$, and Duration, $F(1, 75) = 17.84, p < .001, \eta^2 = .271$. The interpretation of these results was qualified by significant interactions of Phase by Request Time by Evaluation, $F(1, 75) = 4.98, p < .05, \eta^2 = .062$, and Phase by Request Time by Duration, $F(1, 75) = 4.024, p < .05, \eta^2 = .051$, indicating that (a) the “After” groups had greater RE for the 4S pattern than the 4L pattern, (b) from practice block 6 to retention, error on the 4S pattern increased for the “After” groups, while error on the 4L pattern increased for the “Before” groups, and (c) groups
LCA+E and LCB demonstrated deterioration in performance at retention, while groups LCA and LCB+E maintained stable levels of relative error from the last practice block to the retention test.

![Figure 6-10](image-url)

**Figure 6-10.** Relative Error (RE, as a proportion) for groups and response types across acquisition and retention phases in Experiment 3. Abbreviations as before.

6.2.3. Transfer

**Accuracy.** Trials that were incorrect or violated RT criteria accounted for 18.0% of random transfer trials, 15.2% of phoneme transfer trials, 25.4% of relative difference transfer trials, and 11.9% of absolute difference transfer trials. Group had an effect on accuracy for the random transfer and phoneme transfer probes (Figure 6-2). The 2 (Request Time) x 2 (Evaluation) ANOVA on accuracy for random transfer
revealed that the “After” groups (\(M = 91.0\%, SD = 9.5\%\)) were more accurate than the “Before” groups (\(M = 84.6\%, SD = 12.1\%\)), \(F(1,76) = 7.54, p < .01, \eta^2 = .09\), and that the No-Evaluation groups (\(M = 91.0\%, SD = 10.2\%\)) were more accurate than the Evaluation groups (\(M = 84.6\%, SD = 11.5\%\)), \(F(1,76) = 7.57, p < .01, \eta^2 = .091\). The phoneme transfer analysis revealed no main effects, but an interaction between Request Time and Evaluation, \(F(1,76) = 5.01, p < .05, \eta^2 = .062\), indicated that Group LCA (\(M = 94.7\%, SD = 5.5\%\)) was more accurate than LCA+E (\(M = 84.5\%, SD = 13.7\%\)).

**Study Time (INT process).** ST data for all Transfer blocks are shown in Figure 6-11. For the random transfer probe, which involved producing the four trained responses in random order within a single block, the analysis testing the effects of Groups and Duration for single syllable response STs revealed a marginal main effect of Evaluation, \(F(1, 76) = 2.83, p = .097, \eta^2 = .036\), suggesting that ST was longer for the Evaluation groups (\(M = 2117\ ms, SD = 864\ ms\)) than the No Evaluation groups (\(M = 1845\ ms, SD = 697\ ms\)). There was not a significant effect of Duration for 1-syllable responses in random transfer. The analysis of 4- versus 1- syllable responses for random transfer revealed a main effect of Sequence Length, \(F(1, 76) = 90.31, p < .001, \eta^2 = .543\), indicating that ST was longer for 4-syllable than 1-syllable responses, with no interactions or effects of groups.

For the phoneme transfer probe, which involved producing the four trained responses in random order using different phonemes, analysis revealed a marginal effect of Request Time for single syllable responses, \(F(1, 76) = 2.85, p = .096, \eta^2 =\)
suggesting that “After” groups ($M = 1625 \text{ ms}, SD = 687 \text{ ms}$) had longer STs than “Before” groups ($M = 1414 \text{ ms}, SD = 581 \text{ ms}$). A marginal effect of Request Time, $F(1, 76) = 3.34, p = .071, \eta^2 = .042$ (After $M = 2131 \text{ ms}, SD = 1202 \text{ ms}$; Before $M = 1790 \text{ ms}, SD = 1067 \text{ ms}$) was also present for the analysis comparing 4- and 1-syllable responses in the phoneme transfer probe, as was a significant main effect of Sequence Length (4 syllables $> 1$ syllable), $F(1, 76) = 77.89, p < .001, \eta^2 = .506$.

No group effects were found in the analysis of the relative difference transfer probes, which required participants to produce the trained syllable durations in novel relative patterns. However, a main effect of Pattern type, $F(2.98, 226.7) = 4.18, p < .01, \eta^2 = .052$, indicated that ST was longer for the LLSS response pattern than patterns SLSL and LSLS.

The analyses of single syllables in the absolute difference transfer probe, wherein participants produced the trained responses with 33% longer syllable durations, revealed a main effect of Duration, $F(1, 76) = 4.74, p < .05, \eta^2 = .056$, which was qualified by a Duration by Request Time interaction, $F(1, 76) = 12.87, p < .005, \eta^2 = .145$, localizing the observed duration effect ($1S+ > 1L+$) to the After group alone. The analysis of 1- versus 4-syllable responses revealed a main effect of Sequence Length, $F(1, 76) = 105.51, p < .001, \eta^2 = .581$. A marginally-significant interaction between Sequence Length and Request Time, $F(1, 76) = 2.81, p = .098, \eta^2 = .036$, and a significant interaction between Sequence Length, Request Time, and Evaluation, $F(1, 76) = 8.10, p < .01, \eta^2 = .096$, indicated that, while the sequence length effect held for all groups, LCA+E ($M = 3065 \text{ ms}, SD = 1481 \text{ ms}$) had
significantly longer STs for 4-syllable responses than LCB+E ($M = 2099$ ms, $SD = 1139$ ms).
Figure 6-11. Study Time (ST, in milliseconds) for groups and response types in the four transfer probes in Experiment 3. Abbreviations as before.
**Reaction Time (SEQ process).** All RT data pertaining to transfer probes are shown in Figure 6-12. The analysis of single syllable RTs for the random transfer probe revealed a significant main effect of Duration, $F(1,76) = 66.42, p < .001, \eta^2 = .466$. The interpretation of this result was superseded by a significant Duration by Request Time interaction, $F(1,76) = 4.39, p < .05, \eta^2 = .055$, indicating that the duration effect (Long > Short), although significant for both groups, was greater in the “Before” groups (Long $M = 330$ ms, $SD = 82$ ms, Short $M = 272$ ms, $SD = 72$ ms) than the “After” groups (Long $M = 320$ ms, $SD = 95$ ms, Short $M = 286$ ms, $SD = 101$ ms). An interaction between Request Time and Evaluation was also found, $F(1,76) = 4.22, p < .05, \eta^2 = .053$, indicating that the addition of the self-evaluation task affected the “After” groups (LCA $M = 284$ ms, $SD = 89$ ms; LCA+E $M = 323$ ms, $SD = 105$ ms) and the “Before” groups (LCB $M = 319$ ms, $SD = 75$ ms; LCB+E $M = 282$ ms, $SD = 86$ ms) differently, although no pair-wise differences were significant. For the 1-syllable versus 4-syllable comparison, there was also a significant Request Time by Evaluation interaction, $F(1,76) = 4.5, p < .05, \eta^2 = .056$, again illustrating the different effect of the self-evaluation task on the “After” (LCA $M = 304$ ms, $SD = 90$ ms; LCA+E $M = 347$ ms, $SD = 117$ ms) and “Before” (LCB $M = 355$ ms, $SD = 92$ ms; LCB+E $M = 313$ ms, $SD = 109$ ms) groups. The main effect of Sequence Length was also significant, $F(1,76) = 72.38, p < .001, \eta^2 = .488$, demonstrating longer RTs for patterns than single syllables.

For the phoneme transfer probe on single syllables, a main effect of Duration, $F(1,76) = 73.54, p < .001, \eta^2 = .492$, indicated that RT was longer for 1L than 1S
responses. An interaction between Request Time and Evaluation, $F(1,76) = 7.84, p < .01, \eta^2 = .094$, indicated that the self-evaluation task affected the “After” groups ($LCA M = 237$ ms, $SD = 85$ ms; $LCA+E M = 292$ ms, $SD = 116$ ms) differently than the “Before” groups ($LCB M = 285$ ms, $SD = 67$ ms; $LCB+E M = 234$ ms, $SD = 79$ ms), although no pair-wise differences were significant. The Request Time by Evaluation interaction was also observed for the 1-syllable versus 4-syllable responses, $F(1,76) = 6.46, p < .05, \eta^2 = .078$ with “After” groups ($LCA M = 278$ ms, $SD = 84$ ms; $LCA+E M = 317$ ms, $SD = 115$ ms) responding differently to the Evaluation factor than “Before” groups ($LCB M = 340$ ms, $SD = 89$ ms; $LCB+E M = 288$ ms, $SD = 101$ ms).

Further, analysis of phoneme transfer data revealed a main effect of Sequence Length, $F(1,76) = 72.88, p < .001, \eta^2 = .489$, which was qualified by a significant Sequence Length by Request Time interaction, $F(1,76) = 4.42, p < .05, \eta^2 = .055$, indicating that the sequence length effect (4 syllables > 1 syllable) was stronger for the “Before” groups (1-syllable $M = 259$ ms, $SD = 77$ ms; 4-syllable $M = 314$ ms, $SD = 98$ ms) than the “After” groups (1-syllable $M = 265$ ms, $SD = 105$ ms; 4-syllable $M = 297$ ms, $SD = 102$ ms).

The 2 (Request Time) x 2 (Evaluation) x 4 (Stimulus Type) ANOVA for the relative difference transfer probe revealed a main effect of Pattern, $F(2.7, 205.8) = 3.485, p < .05, \eta^2 = .044$, indicating that RTs to the SSLL target were significantly longer than both the LLSS and the LSLS targets. In addition, a marginally-significant interaction between Request Time and Evaluation, $F(1,76) = 3.7, p = .058, \eta^2 = .046$, suggested that the Evaluation factor tended to affect the “After” groups ($LCA M = 313$
ms, $SD = 104$ ms; LCA+E $M = 357$ ms, $SD = 109$ ms) and the “Before” groups (LCB $M = 387$ ms, $SD = 115$ ms; LCB+E $M = 339$ ms, $SD = 127$ ms) differently.

For the single-syllable responses in the absolute difference transfer probe, a main effect of Duration, $F(1,76) = 30.78$, $p < .001$, $\eta^2 = .288$, indicated that RTs were longer for 1L+ responses than to 1S+ responses. In addition, a significant Request Time by Evaluation interaction, $F(1,76) = 4.04$, $p < .05$, $\eta^2 = .05$, demonstrated that the “After” groups (LCA $M = 315$ ms, $SD = 113$ ms; LCA+E $M = 355$ ms, $SD = 107$ ms) and the “Before” groups (LCB $M = 366$ ms, $SD = 99$ ms; LCB+E $M = 313$ ms, $SD = 100$ ms) responded differently to the self-evaluation task. The Request Time by Evaluation interaction was also marginally-significant for the 1-syllable versus 4-syllable responses in the absolute difference transfer probe, $F(1,76) = 3.12$, $p < .081$, $\eta^2 = .039$ (LCA $M = 355$ ms, $SD = 125$ ms; LCA+E $M = 380$ ms, $SD = 102$ ms; LCB $M = 417$ ms, $SD = 107$ ms; LCB+E $M = 373$ ms, $SD = 120$ ms). The main effect of Sequence Length was also significant, $F(1,76) = 49.33$, $p < .001$, $\eta^2 = .394$, indicating longer RT for sequences than for single syllables. A marginally-significant interaction between Sequence Length and Request Time, $F(1, 76) = 3.21$, $p = .077$, $\eta^2 = .041$, suggested that “Before” groups (1-syllable $M = 340$ ms, $SD = 102$ ms; 4-syllable $M = 395$ ms, $SD = 115$ ms) tended to have a stronger sequence length effect than the “After” groups (1-syllable $M = 335$ ms, $SD = 112$ ms; 4-syllable $M = 368$ ms, $SD = 114$ ms).
Figure 6-12. Reaction Time (RT) for groups and response types across transfer probes in Experiment 3. Abbreviations as before.
**Absolute Error.** AE data for Transfer probes are shown in Figure 6-13. The analyses of absolute timing error (AE) in the random transfer probe revealed for single-syllable responses an interaction between Duration, Request Time, and Evaluation, $F(1, 76) = 4.5, p < .05, \eta^2 = .056$, indicating that the self-evaluation task had different effects on the “After” groups than the “Before” groups, and that this interaction depended on the duration of the response, although no pair-wise comparisons in this interaction were significant.\(^{15}\) There were no group effects for the 4-syllable patterns in the random transfer probe, but a main effect of Duration, $F(1, 76) = 4.47, p < .05, \eta^2 = .056$, revealed that error was greater for the 4L pattern than the 4S pattern.\(^{16}\)

Analysis of single-syllable responses in the phoneme transfer probe revealed a main effect of Duration, $F(1, 76) = 11.11, p < .005, \eta^2 = .128$, indicating more error for the 1S than the 1L target. A marginal effect of Request Time for the 1-syllable stimuli, $F(1, 76) = 3.67, p = .059, \eta^2 = .046$, suggested that the “After” groups ($M = 83, SD = 45$) had less error than the “Before” groups ($M = 96, SD = 52$).\(^{17}\) There were no significant effects involving the 4-syllable patterns in the phoneme transfer probe.\(^{18}\)

---

\(^{15}\) When one LCA subject outlier, whose mean for 1L was greater than 3 standard deviations above the group mean, was excluded from the analysis of single syllables, the 3-way interaction result remained, and the main effect of Request Time became marginally significant.

\(^{16}\) This result was unchanged when two outliers (LCA = 1, LCB+E = 1) with means for 4S greater than 3 standard deviations above the group means were excluded from the analysis.

\(^{17}\) When one LCA subject outlier, whose mean for 1S was greater than 3 standard deviations above the group mean, was excluded, the marginal effect of Request Time (After < Before) became significant, and a marginal Duration by Evaluation interaction emerged, suggesting that the Evaluation group, but not the No Evaluation group, had greater error for 1S than 1L.

\(^{18}\) Excluding four subject outliers (LCA = 1, LCA+E = 1, LCB+E = 2) from this analysis revealed a marginal Duration effect (4L > 4S) and a significant Request Time by Evaluation interaction indicating that the Before group benefited from the self-evaluation task, whereas the After group did not.
There were also no significant effects of groups or stimulus type in the relative difference transfer probe.\textsuperscript{19}

For the absolute difference transfer probe, a significant effect of Duration, $F(1, 76) = 49.74, p < .001, \eta^2 = .396$, was found for single-syllable responses, indicating that there was greater error associated with the 1S+ target than the 1L+ target. Main effects of Evaluation were found for both the 1-syllable responses, $F(1, 76) = 5.35, p < .05, \eta^2 = .066$, and the 4-syllable responses, $F(1, 76) = 8.55, p < .01, \eta^2 = .101$, indicating that the Evaluation groups (1-syllable $M = 128$ ms, $SD = 80$ ms; 4-syllable $M = 156$ ms, $SD = 100$ ms) had less error than the No Evaluation groups (1-syllable $M = 156$ ms, $SD = 83$ ms; 4-syllable $M = 250$ ms, $SD = 190$ ms) in the absolute difference transfer probe.\textsuperscript{20}

\textsuperscript{19} Excluding one LCA+E subject outlier whose SLSL response mean was greater than 3 standard deviations above the group mean yielded a marginally-significant main effect of Evaluation such that error was greater for the No Evaluation groups than the Evaluation groups.

\textsuperscript{20} These results persisted when one LCA subject outlier was excluded for the 1-syllable analysis and one LCA subject outlier was excluded for the 4-syllable analysis.
Figure 6-13. Absolute Error (AE, in milliseconds) for groups and response types across transfer probes in Experiment 3. Abbreviations as before.
**Relative Error.** All RE data for transfer probes are shown in Figure 6-14. The analysis of RE for 4-syllable patterns in the random transfer probe revealed a main effect of Duration (4S > 4L), \( F(1, 76) = 12.92, p < .005, \eta^2 = .145, \) and a significant Request Time by Evaluation interaction, \( F(1, 76) = 4.26, p < .05, \eta^2 = .053. \) This interaction indicated that the addition of the self-evaluation task affected the “After” groups (LCA \( M = .203, SD = .089, \) LCA+E \( M = .241, SD = .11 \)) differently than the “Before” groups (LCB \( M = .255, SD = .116, \) LCB+E \( M = .213, SD = .077 \)), although no pair-wise comparisons were significant. The analysis of the phoneme transfer probe also demonstrated a main effect of Duration, \( F(1, 76) = 14.66, p < .001, \eta^2 = .162, \) such that RE was greater for the 4S response than the 4L response. In addition, there was a significant main effect of Request Time, \( F(1, 76) = 4.03, p < .05, \eta^2 = .05, \) indicating that the “Before” groups \( (M = .247, SD = .081) \) had greater RE than the “After” groups \( (M = .219, SD = .063) \) in the phoneme transfer probe. For the relative difference transfer probe, significant main effects of Pattern, \( F(3, 228) = 6.48, p < .001, \eta^2 = .079, \) and Request Time, \( F(1, 76) = 5.27, p < .05, \eta^2 = .065, \) were found. The interpretation of these effects was superseded by a marginal Request Time by Evaluation interaction, \( F(1, 76) = 3.89, p = .052, \eta^2 = .049, \) and a significant Request Time by Evaluation by Pattern interaction, \( F(3, 228) = 2.89, p < .05, \eta^2 = .037, \) indicating the LCA group \( (M = .228, SD = .087) \) had less error than the LCB group \( (M = .265, SD = .092) \), but that this pattern was not consistent across stimulus patterns. Finally, the analysis of the absolute difference transfer probe revealed a main effect of Duration (4S > 4L), \( F(1, 76) = 16.7, p < .001, \eta^2 = .18, \) and no significant effects of
Figure 6-14. Relative Error (RE, as a proportion) for groups and response types across transfer probes in Experiment 3. Abbreviations as before.
6.3. Discussion

Although the benefits of self-controlled learning conditions have been recognized for a number of years (e.g., Janelle, et al., 1995; 1997; Zimmerman, 1989), only recently have the possible mechanisms underlying this phenomenon been investigated in a systematic fashion (e.g., Chiviacowsky & Wulf, 2002, 2005, 2007). The limb motor learning literature has provided evidence suggesting that allowing learners to control their feedback schedules, as well as other aspects of their learning environment, enhances learning because it allows learners to receive information (or assistance) when they need it (Chiviacowsky & Wulf, 2002; 2005; 2007; Wulf et al., 2001; 2005; Wulf & Toole, 1999). More specifically, it was recently hypothesized that learners who are able to make decisions about their feedback schedules after completing movements may attend more to their intrinsic sources of feedback, thus reducing their dependence on extrinsic feedback and contributing to their advantage over learners who must make the feedback decision before the movement trial (Chiviacowsky & Wulf, 2005). The purpose of Experiment 3 was to test this hypothesis in the domain of speech motor learning by comparing these groups of learners with learners who were encouraged to attend to their intrinsic feedback via a self-evaluation task. Predictions were made based on Schema Theory (Schmidt, 1975) and in the context of the INT/SEQ two-process model of motor programming (Klapp, 1995; 2003). It was expected that learners who made feedback choices before movement trials would benefit from the addition of the self-evaluation task, whereas learners who made feedback choices upon movement completion would not. First, this
hypothesis will be examined with respect to the results concerning motor programming (ST and RT), and then response execution (AE and RE) will be considered. Table 6-3 presents a general overview of Experiment 3 predictions and results.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Prediction</th>
<th>Result</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decrease with practice</td>
<td>Decrease with practice except single syllables in blocks 3 vs. 6.</td>
<td>INT more efficient with practice.</td>
</tr>
<tr>
<td></td>
<td>Increase at retention</td>
<td>Increase at retention.</td>
<td>INT less efficient after withdrawal of practice.</td>
</tr>
<tr>
<td></td>
<td>Syllable duration effect (1L &gt; 1S)</td>
<td>No duration effect except 1S &gt; 1L at retention and absolute difference transfer.</td>
<td>INT not sensitive to syllable duration or INT not confined to ST interval.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitive to sequence length.</td>
<td>All units or all unique units programmed during INT</td>
</tr>
<tr>
<td>ST</td>
<td>Request Time x Evaluation interaction such that LCB &lt; other groups during practice, retention, random transfer</td>
<td>Request Time x Evaluation interaction: 1-syllable in practice, LCB numerically shortest in practice, retention, random transfer. After &gt; Before for 1 vs. 4 in practice, retention. Evaluation &gt; No evaluation for 1 vs. 4 in practice, random transfer.</td>
<td>LCB engaged in INT less extensively. “After” groups engaged in INT more extensively. Self-evaluation promoted more extensive INT for both “before” and “after” conditions.</td>
</tr>
<tr>
<td></td>
<td>Request Time x Evaluation interaction such that LCB &gt; other groups in transfer</td>
<td>After &gt; Before - phoneme transfer. LCA+E &gt; LCB+E - absolute difference transfer.</td>
<td>More extensive INT processing in “after” feedback - transfer of parameterization to new phoneme. More extensive INT processing in LCA+E than LCB+E for transfer to new temporal parameterizations of trained GMPs</td>
</tr>
<tr>
<td>RT</td>
<td>Decrease with practice</td>
<td>Decrease with practice except 1S in blocks 3-6</td>
<td>SEQ more efficient with practice</td>
</tr>
<tr>
<td></td>
<td>Increase at retention</td>
<td>Increase at retention for 1S, and for sequences for LCA+E.</td>
<td>SEQ less efficient with withdrawal of practice for LCA+E</td>
</tr>
<tr>
<td></td>
<td>Sequence length effect (4&gt;1 syllable)</td>
<td>Sensitive to sequence length (4&gt;1) for all but 1L-4L in practice.</td>
<td>Sequences were programmed as multiple units, except 4L, which was “chunked” during practice.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitive to duration</td>
<td>INT during RT interval, or kinematic confound.</td>
</tr>
<tr>
<td></td>
<td>No group effects</td>
<td>Before &gt; After for sequence length effect. Ubiquitous Request Time x Evaluation interaction.</td>
<td>“After” groups may have begun to chunk sequences. Evaluation improved SEQ efficiency for “Before,” but hindered it for “After.”</td>
</tr>
<tr>
<td>Variable</td>
<td>Prediction</td>
<td>Result</td>
<td>Interpretation</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>AE</td>
<td>( AE_{\text{feedback trials}} &lt; AE_{\text{no-feedback trials}} )</td>
<td>( AE_{\text{feedback trials}} &gt; AE_{\text{no-feedback trials}} )</td>
<td>Might reflect error correction on no-feedback trials following feedback trials</td>
</tr>
<tr>
<td>No group differences in acquisition</td>
<td>No group differences in acquisition for sequences. After &lt; Before - single syllables.</td>
<td>“After” feedback enhances parameterization acquisition for single syllables.</td>
<td></td>
</tr>
<tr>
<td>No group differences in retention</td>
<td>No group differences in retention for sequences. After &lt; Before - single syllables.</td>
<td>“After” feedback enhances parameterization retention for single syllables.</td>
<td></td>
</tr>
<tr>
<td>Request Time x Evaluation interaction such that LCB &gt; other groups in transfer</td>
<td>After &lt; Before - phoneme transfer. Evaluation &lt; No evaluation – absolute difference transfer.</td>
<td>“After” feedback enhances parameterization transfer to different phonemes. Self-evaluation enhances new parameterizations of trained GMPs</td>
<td></td>
</tr>
<tr>
<td>No group differences in acquisition</td>
<td>No group differences in acquisition.</td>
<td>Feedback request time and self-evaluation do not affect GMP coherence during acquisition.</td>
<td></td>
</tr>
<tr>
<td>No group differences in retention</td>
<td>LCB, LCA+E worsened in retention; LCB+E, LCA stable.</td>
<td>“After” feedback or self-evaluation enhance retention GMP coherence of trained responses</td>
<td></td>
</tr>
<tr>
<td>Request Time x Evaluation interaction such that LCB &gt; other groups in transfer</td>
<td>Request Time x Evaluation interaction – random transfer. After &lt; Before - phoneme transfer. LCA &lt; LCB - relative difference transfer.</td>
<td>“After” feedback or self-evaluation enhance GMP coherence of trained responses in new context. “After” feedback promotes GMP coherence with new parameters “After” feedback promotes GMP coherence of new related patterns for no-feedback groups</td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>4S &gt; 4L</td>
<td>4L may have been programmed as a single unit</td>
<td></td>
</tr>
</tbody>
</table>
6.3.1. Study Time (INT process)

Study Time represents the INT stage of motor programming in which a response is selected and its internal structure is specified. In terms of Schema Theory, a GMP is activated and parameterized during INT. The fundamental predictions of block, phase, and single-syllable duration effects on ST will be examined before group differences in ST are considered. The discussion pertaining to the effect of sequence length on ST will be postponed until RT findings are discussed.

Practice influences on ST: The acquisition results confirmed that ST decreased with practice, although the decrease between the middle and last practice blocks was not significant for one-syllable responses. ST also increased significantly following the withdrawal of practice, as indicated on the retention probe. These results suggest that INT processing became more efficient with practice, but that not all of the gains made during practice were retained. Given that INT processing is thought to be impoverished under blocked practice, such as was used here, as opposed to random practice (Imminck & Wright, 1998, 2001, Wright et al., 2004), the decrement between practice and retention was not unexpected. It is interesting to note, however, that in the present experiment, the ST increase from practice to blocked retention 24 hours later was not as dramatic as that demonstrated by other recent experiments (e.g., Imminck & Wright, 2001, Experiment 1; Wright et al., 2004), in which blocked practice participants’ 24-hour blocked retention STs exceeded those of the initial practice block. This might be partially explained by the fact that participants in the present study received twice as many practice trials as those in the Imminck and Wright (2001)
study, suggesting that the number of practice trials may have an influence on INT retention 24 hours later. Also, participants in the present study practiced four different kinds of targets, whereas Immink and Wright’s participants only practiced the single-syllable responses. Practicing a greater variety of responses, albeit in blocked fashion, may have imparted a benefit to overall INT efficiency. However, blocked practice participants in Wright and colleagues’ (2004) experiment did practice the four different patterns that were used in the present study, and also performed twice as many practice trials of each type, yet they still exhibited longer STs in retention than they did in the first practice block. A possible explanation for this is that although the retention test was administered 24 hours after practice was completed, because practice of different targets in Wright et al. (2004) was blocked across days, for half of the participants, it had been at least 72 hours since they had practiced a given target. Of course, response modalities differed across these two studies--vocal in the present study, and manual in the other studies (Immink & Wright, 2001; Wright et al 2004)--which could also be a relevant factor.

**Duration effect on ST:** Since INT is the stage of motor programming in which the internal structure of a unit of movement is assembled, ST is sensitive to the complexity of individual units. For syllables and single button-presses, complexity has been defined in terms of the duration of the movement (Klapp, 1995; Maas, Robin, Wright et al., in press; cf. Wright, et al., submitted). Thus, it was predicted that ST would be longer for the longer-duration syllable (1L) than the shorter one (1S). No reliable effect of duration was found during acquisition, and a significant difference
emerged at retention that was contrary to the prediction. For the blocked retention
task, ST for short (1S) syllables was longer than ST for long (1L) syllables. The same
result was also found in the absolute difference transfer probe, although it was only
significant for the “After” groups.

Although the duration effect on ST does not appear to be a particularly robust
finding for speech motor learning (Maas, Robin, Wright et al., in press) or blocked
practice (Immink & Wright, 2001; Wright et al., 2004), it was predicted for the present
experiment because of the detailed duration feedback that participants received, as
well as the three group conditions that were expected to encourage more frequent re-
parameterization than might typically be expected in blocked practice. One possible
explanation for the lack of a consistent duration effect on single syllable responses
could be that participants simply did not differentiate the short and long responses.
However, this does not seem plausible in the present study, as average response
durations indicated that participants were successful at distinguishing responses with
different total durations (Figure 6-15). Another explanation that has been suggested
before (Maas, Robin, Wright et al., in press) is that speech is such an over-learned
motor skill that simple differences in syllable duration such as between 1S and 1L do
not sufficiently tax the INT processor to register an effect. Indeed, recent studies
employing the same speech stimuli used here have failed to find a duration effect on
ST (Maas Robin, Wright et al., in press; Wright, et al., submitted), although some have
argued for an effect of syllable complexity in terms of phonetic gestures (e.g., /ta/ vs.
/stra/; Wright, et al., submitted). If it is the case for speech that syllabic duration
assignment is highly automatic and does not affect INT, it is curious that a significant
effect of duration was found in the opposite direction as was predicted (i.e., 1S > 1L)
in two experimental tasks (i.e., retention and absolute difference transfer). It seems
possible that this finding has to do with response difficulty in terms of the learner’s
ability to minimize the absolute error of the response. Overall, participants had greater
error on 1S responses than 1L, although both improved with practice. This might have
had to do with habitual speaking rate for single-syllable words in isolation being closer
to the 450 ms duration than the 150 ms duration. Recognizing that their errors for 1S
were greater, participants might have applied added effort to the INT processing for
the 1S responses, in an attempt to parameterize more appropriately.
Figure 6-15. Stimuli target durations (solid lines) and participant total response durations (broken lines) across three blocks of practice for each group. Stimulus abbreviations defined in Table 6-1. LCB = Learner-Controlled Before, LCB+E = Learner-Controlled Before plus Evaluation, LCA = Learner-Controlled After, LCA+E = Learner-Controlled After with Evaluation.

**Group differences in ST:** Effects of feedback request time and the self-evaluation task were expected to be evident in ST, as groups LCB+E, LCA, and LCA+E were predicted to engage in more extensive INT processing than group LCB.

This pattern was predicted for acquisition, as well as for the blocked retention and random transfer tasks in which the practiced responses were also elicited. Indeed, for half of the stimuli in practice block 1 and for all of the stimuli in blocks 3 and 6, the
LCB group displayed the shortest mean STs (Figure 6-3). Further, group LCB had the shortest mean STs for all but the 4L response in blocked retention (Figure 6-7), and for all but the 4S response in random transfer (Figure 6-11). These findings, however, only provide weak support for the prediction because the statistical interaction between the two group factors was not consistent across all levels of block, duration, and sequence length. For the single-syllable responses, LCB had significantly shorter STs than LCA+E for the short syllable across blocks, and for the long syllable only in block 1. The pattern of significant differences between other groups during practice and retention does not point to a straightforward interpretation. For single-syllable responses in the random transfer probe, the groups that performed the self-evaluation task had marginally-longer STs than those who were not encouraged to self-evaluate (difference = 272 ms). Contrary to expectation, this effect was not dependent on request time, suggesting that the self-evaluation task alone influenced the length of INT processing for single-unit responses when tested in random order.

For the comparison of sequences and single syllables, the two group factors did not interact, although “After” groups had significantly longer STs than “Before” groups across acquisition (difference = 303 ms) and retention (difference = 213 ms), and Evaluation groups had marginally longer STs than No-evaluation groups during acquisition (difference = 187 ms). It is not surprising that requesting feedback after movements would be associated with longer STs than requesting feedback before movements. In the “After” condition, learners presumably engaged in post-movement error estimation to inform their feedback requests, and could use these estimates to re-
parameterize the program for the next response, albeit at an additional cost to INT. Because acquisition trials were delivered in a blocked practice format, bypassing the INT process and simply maintaining a single programmed response in the motor buffer for the duration of the block was a less-costly option (i.e., shorter ST) for learners who did not attempt to re-parameterize as often. The marginal main effect of the self-evaluation task during acquisition and the absence of a Request Time by Evaluation interaction, however, calls this interpretation into question. The self-evaluation task was expected to increase INT processing, and thus ST, in the “Before” condition, as it encouraged learners to estimate their duration errors when otherwise they might not have. However, for learners who chose feedback after movement execution, presumably using some evaluation of their performance to do so, why would the addition of a self-evaluation task further increase ST? One possibility is that the self-evaluation task prompted even greater attention to intrinsic feedback, leading to more extensive INT processing. Indeed, it has been suggested that under normal conditions learners spontaneously and unavoidably engage in some degree of intrinsic feedback processing (Swinnen, 1990; Swinnen et al., 1990). Perhaps the factors manipulated between-subjects in the present study simply increased that processing further, making independent and additive contributions to ST. Another possibility is that groups may have engaged in similar degrees of error estimation, but that these factors (i.e., after-movement feedback request and self-evaluation task) might have each increased the probability that the error estimate was used for updating the specifications of the next trial.
An alternate, and perhaps complementary approach to explaining the finding that the “Before” groups had shorter STs than the “After” groups involves viewing the difference as a shortening of ST for the “Before” groups, rather than a lengthening for the “After” groups. On this view, during acquisition, instead of the post-movement decision about feedback increasing INT processing for the “After” groups, the pre-movement decision about feedback might have had a facilitating effect on INT for the “Before” groups. Perhaps these learners even began INT processing during the feedback decision stage instead of waiting for the beginning of the formal ST interval. Since trials were presented in blocked fashion, there was no reason that learners would have to wait until the presentation of the pre-cue to begin INT processing. However, this option was not limited to the “Before” groups. The “After” groups had the same inter-trial interval duration and events, although in a different order (see Figure 6-1), and thus had similar opportunity to begin INT processing prior to the formal ST interval. However, the “After” groups still required more time to prepare their responses once the pre-cue was given. This was the case for acquisition as well as blocked retention, in which feedback decisions and the self-evaluation task were withdrawn. Interestingly, however, when INT processing was confined to the ST interval, as in the random transfer probe in which the pre-cue that initiated ST was the earliest indication of which response to prepare, group differences vanished (except for a marginal effect of evaluation for single-syllable responses), lending tentative support to this hypothesis.21

---

21 As will be discussed below, the presence of the duration effect (1L > 1S) in the RT data might be taken as evidence that some INT processing was carried out during the RT interval. Interestingly, the
Although this discussion has heretofore assumed, following the extant literature (e.g., Immink & Wright, 2001; Wright et al., 2004), that a lengthening of ST reflects more complete response preparation, it could conversely be interpreted as an indication of less efficient INT processing (e.g., Maas, Robin, Wright et al., in press). Perhaps requiring learners to make post-motion feedback decisions or perform a self-evaluation task disrupts the normal production process and slows INT processing. This interpretation would be consistent with the ST results found here. However, as will be discussed later, the analysis of movement execution indicated that the “After” groups had significantly smaller absolute timing error than “Before” groups for single syllables during acquisition and retention, which is more compatible with the view that longer ST reflects more complete or “better” INT processing.

The ST pattern across groups was expected to reverse under transfer conditions requiring new parameterizations (i.e., phoneme transfer, absolute difference transfer), as the LCB group, which presumably had less experience with INT processing during acquisition, would demonstrate inefficient response planning for new parameterizations of the practiced programs relative to the other groups. Although the LCB group no longer exhibited the shortest mean STs for the majority of stimuli in phoneme and absolute transfer probes (as they did during practice, retention, and random transfer), the pattern did not reverse. For transfer that involved the same relative and absolute timing patterns instantiated with an unpracticed syllable, the duration effect on RT was stronger for the “Before” groups than the “After” groups in the random retention probe, possibly suggesting that when “Before” learners could not begin INT processing early, it was carried over into the RT interval (more than it was for “After” learners), suggesting that they were less able to efficiently and/or effectively complete INT during the ST interval.
“After” groups continued to demonstrate longer overall STs than “Before” groups (difference = 341 ms). For the absolute difference transfer probe, in which the practiced GMPs and syllables were re-parameterized with longer absolute durations, the significant difference between “After” and “Before” groups was restricted to only the 4-syllable targets, and only for groups that performed the self-evaluation task during practice (difference = 966 ms).

The tendency of the “After” groups to engage in lengthier response preparation than “Before” groups is compatible with the hypothesis that allowing learners to select feedback after movement execution promotes their attention to intrinsic feedback sources (Chiviacowsky & Wulf, 2005) and/or their likelihood of using feedback information to update the movement parameters for the next response. Although this result was not expected for transfer tasks involving novel parameterizations, it suggests that once trained, the tendency to engage in more extensive response preparation persists after the factors that promoted it are withdrawn and transfers to similar but untrained tasks. It also accords well with the present response execution data as well as previous findings that learners who self-selected feedback after movement completion had less error in absolute duration transfer than participants who did not self-select feedback (Chiviacowsky & Wulf, 2002) or learners who self-selected feedback before the movement (Chiviacowsky & Wulf, 2005). Interestingly, no group differences in ST were found in the relative difference transfer probe, suggesting that either the group patterns of INT efficiency were dependent on the
relative timing patterns on which they were developed, or that they were obscured by a much larger impact of organizing completely new structures.

### 6.3.2. Reaction Time (SEQ process)

Turning to the RT results, recall that the fundamental predictions were that (a) RT would decrease during practice as the SEQ process became more efficient, and increase again following the withdrawal of practice, and (b) RT would be influenced by sequence length, increasing as a function of the number of units to be serially ordered during SEQ processing. Group differences were not predicted for RT, as feedback and self-evaluation concerning absolute timing were not expected to influence the sequencing of pre-programmed units.

**Practice influences on RT:** The first of these predictions for RT was supported by the present results. That is, RT significantly decreased from practice block 1 to block 3, and from block 3 to block 6 for all but the 1S response, which did not decrease significantly between block 3 and block 6. Further, when practice conditions (i.e., feedback, self-evaluation task) were withdrawn, overall RTs lengthened slightly, as measured on the retention test 24 hours later, although not consistently across stimuli and groups. For the single-syllable responses, RT significantly increased between practice and retention for 1S, but not for 1L. For the 4-syllable patterns, the RT decrement at retention was only significant for one group, LCA+E.

**Sequence Length effect on RT:** The second fundamental prediction regarding RT involved the demonstration of the sequence length effect. The time it takes to
retrieve units from the motor buffer in the correct serial order and initiate movement is dependent on the number of elements in the response (Klapp, 1995, 2003). The acquisition RT data from the present study revealed a sequence length effect for the 1S-4S pair, but not for the 1L-4L pair. For the retention block, as well as for all 1-syllable versus 4-syllable transfer blocks, the sequence length effect was significant for both pairs of responses.

The finding of the fairly ubiquitous sequence length effect in the present RT data replicates previous findings using simple reaction time (e.g., Deger & Ziegler, 2002; Klapp, 1995, 2003; Sternberg et al., 1978; Verwey, 1999) or the self-selection paradigm (e.g., Immink & Wright, 2001; Maas, 2006 Experiment 2; Magnuson et al., in press). However, the sequence length effect is influenced by a number of factors, including amount of practice (Klapp, 1995), practice schedule (Immink & Wright, 2001; Wright et al., 2004), homogeneity of syllable type (Deger & Ziegler, 2002; Klapp, 2003), speech rate (Klapp, 2003), and prosodic patterns of speech sequences (Maas, 2006). The general interpretation of these findings is that these factors affect the tendency for sequences of units to be integrated or concatenated into single larger units or “chunks.” When a series of units of action is reorganized into a single longer unit, the movement can be initiated more quickly because only one element must be retrieved from the motor buffer. For example, Klapp (2003, Experiment 4) reported that the number of repeated syllables in a sequence affected the time required for SEQ processing, but not INT processing, because those syllables were programmed as separate units. When alternating syllables were used, however, the opposite pattern
obtained, suggesting that alternating syllables were chunked into a single unit (Klapp, 2003, Experiment 1). Maas (2006, Experiment 1) found the same pattern, that is, an effect of number of repeated syllables on INT (ST), but not SEQ (RT), and hypothesized that the prosodic pattern applied to the syllable repetitions in that experiment encouraged speakers to program sequences as single units.

The present results permit a direct comparison to Maas (2006, Experiment 1), as the self-selection paradigm was used in both experiments to target identical responses. Interestingly, the sequence length effect on RT was fairly robust in the present study, in contrast to Maas (2006). Why would participants in only one study (i.e., Maas) consolidate four-syllable patterns? One suggestion having to do with a methodological difference between these two studies comes from recent work by Wright and colleagues (2004), who found differences in sequence consolidation for learners practicing targets in random versus blocked practice order. Specifically, blocked practice (as in the present study) learners only temporarily integrated sequences into single units, whereas more-permanent movement restructuring was facilitated by the random practice condition (as in Maas, 2006). In the present blocked practice experiment, although the RT sequence length effect was reliable for the 4S-1S pair throughout acquisition, the size of the sequence length effect diminished (nonsignificantly) across practice blocks (block 1 difference = 130 ms, block 2 difference = 96 ms, block 3 difference = 70 ms), suggesting that the learners might have been attempting to “repackage” the 4S pattern as a single unit. The 4L-1L pair did not evidence a significant sequence length effect across blocks (block 1 difference
= 29 ms, block 2 difference = 31 ms, block 3 difference = 27 ms), but did in blocked retention (difference = 55 ms), and random transfer (difference = 37 ms) suggesting that the 4L sequence may have been integrated early in practice, but the integration was only temporary, consistent with Wright and colleagues’ (2004) findings for blocked practice. Learners in the present study exhibited increased absolute and relative timing error for the 4S response (which was not successfully concatenated during acquisition), suggesting that inability to program a sequence as a single unit is related to performance accuracy, in that programming a sequence as multiple units is associated with larger absolute and relative timing error.

Several researchers have pointed out that when a sequence of movements becomes “chunked” into a single unit, it might be expected that the cost of preparing the larger unit would be redistributed from SEQ to the INT process, resulting in an inverse relationship between the sequence length effect on RT (or simple RT) and on ST (or choice RT) (e.g., Klapp, 1995; Magnuson et al., in press; Wright et al., 2004). However, two factors make this shift in processing cost difficult to observe. First, in learning studies, response preparation (INT) efficiency is concurrently increasing while movement restructuring would be taking place (Magnuson et al., 2004). This could explain why, in the present experiment, a redistribution of processing cost to INT was not observed in ST when the RT sequence length effect for the 4S-1S pair diminished with practice. Second, it is possible that the number of elements in a response affects ST, regardless of whether they are integrated into a single unit or programmed as separate units. The prevalence of sequence length effects on ST
throughout the present experiment (in the presence of sequence length effects on RT) lends support to this notion. For example, there was no difference in ST between 4L (which was assumed to be integrated, given the RT results) and 4S (which was assumed not to be integrated) during acquisition. This suggests that each element in the sequences (i.e., two short “ba” syllables and two long “baaa” syllables) (Magnuson et al., in press) or each unique element (i.e., short “ba” and long “baaa”) (Immink & Wright, 2001; Wright et al., submitted) may be prepared during INT, resulting in longer STs than when the single-syllable responses (i.e., 1S and 1L) were prepared. The fact that other studies have demonstrated an absence of sequence length influences on INT (e.g., Klapp, 2003, Experiments 3 and 4) is likely a function of different experimental paradigms. That is, when INT demands are measured by choice RT, the number of units in a response may not be influential because the movement may be initiated once the first unit is prepared. When INT demands are measured by ST, as discussed above, a number of units may be prepared before the participant indicates that he or she is “ready.”

Duration effects on RT: Before moving to the discussion of the group findings, an unexpected RT result requires comment. Recall that unit complexity, typically defined in terms of duration (cf. Wright et al., submitted), is generally expected to affect the length of INT processing (ST), and not the SEQ processing stage (RT), which is thought to be influenced by the number of units occupying the motor buffer, not their internal complexities. The absence of a complexity (duration) effect on ST observed in the present experiment is not without precedent, and hypotheses regarding
it have been discussed (e.g., Maas 2006; Wright et al., submitted). Less often, however, has an effect of single unit duration on RT been reported in the INT/SEQ literature, and when it has, its presence has typically not been addressed (e.g., Immink & Wright, 2001; Experiment 1 Block 4; Wright et al., submitted, Experiment 1). The robust effect of single-syllable duration on RT (long “baaa” > short “ba”) in the present experiment, however, requires attention, particularly in light of the absence of the predicted duration effect on ST. Two possible explanations could account for this finding. First, a methodological difference in the acceptability of RTs might have resulted in learners shifting some INT processing to the relatively long RT interval permitted in this experiment. Specifically, other studies that have successfully demonstrated a unit complexity effect during the INT stage have constrained the maximum allowable RT by either screening out subjects whose simple RTs on a pre-test exceeded 350 ms (Klapp, 2003), giving “slow start” feedback for RTs greater than 750 ms (Klapp, 1995), or counting trials with RTs greater than 400 ms (Immink & Wright, 2001; Wright et al., 2004) or 800 ms (Magnuson et al., in press) as errors. In the present experiment, RTs between 100 and 1000 ms were admissible. Perhaps this longer RT window, as Maas (2006) conjectured, may have caused learners to adopt different programming strategies or devalue speedy responding. Indeed, Maas noted that RTs for normal learners in his speech programming task (Maas 2006; Experiment 1) were substantially longer than those reported in Klapp (2003) and Deger and Ziegler (2002). However, average RTs during acquisition in the present experiment were not unlike those of earlier speech motor programming studies (e.g., Klapp, 2003;
Deger & Ziegler, 2002), suggesting that it was unlikely, although theoretically possible (Klapp, 1995) for INT and SEQ to occur in parallel during the RT interval.

Another possible explanation for the duration effect observed on RT is that perhaps something about the kinematic dynamics of the two different responses (i.e., 150 ms vs. 450 ms “ba”) affected RT. In fact, it has been suggested that movement initiation time is better understood as a function of average velocity (Falkenberg & Newell, 1980) or rate of force production (Carlton, Carlton, & Newell, 1987) than of movement duration. Indeed, consistent with the present results, this literature has shown that faster movement velocities, as would be expected with shorter syllables, are usually associated with faster RTs (Falkenberg & Newell, 1980). These factors may be confounded in experiments like the present one, although they have historically been considered to be effectively controlled since effects of single-unit kinematics are not usually observed on RT in the INT/SEQ literature (see Klapp, 1995). Relatedly, speech movements present another potential confound for measuring RT, given that producing longer syllables requires a greater lung volume than shorter syllables, and thus longer inhalation time might contribute to lengthened RTs for longer syllables. Klapp (2003) attempted to minimize the influence of inhalation time by reminding participants before every trial to inhale before the imperative signal; however, Maas, Robin, Wright, and colleagues (in press), who did not remind participants before every trial to inhale, did not observe a duration effect on RT. The present data suggest that these factors may be worth further consideration in future studies.
Group effects on RT: Returning to the discussion of the sequence length effect, an examination of the group effects on RT in the present experiment could shed additional light on the differences between this study and the Maas (2006, Experiment 1) study. That is, perhaps the learner-control of absolute timing feedback and the implementation of the self-evaluation task had an influence on learners’ abilities to chunk sequences. Although the sequence length effect was observed on RT in acquisition, retention, and transfer, suggesting that, in general, the 4-syllable patterns were not treated as single-unit chunks, an interesting group difference in the magnitude of the sequence length effect raises the possibility that some learners had integrated, or were in the process of integrating the multisyllabic responses. Specifically, learners who requested feedback before performing a trial (regardless of whether or not they were asked to self-evaluate) demonstrated a larger effect of sequence length on RT than learners in the “After” condition during acquisition, phoneme transfer, and absolute difference transfer. This could be interpreted as an indication that learners in the “After” groups had achieved some initial degree of integration of the 4-syllable sequences that the “Before” groups had not. Perhaps learners who chose feedback after movements used the total movement duration feedback in a way that “Before” group learners did not, which facilitated the beginnings of a transformation of the movement from four units to one. Further, the LCA+E group was the only group to exhibit an increase in RT for sequences at retention, suggesting that the addition of the self-evaluation task to the “After” condition may have made these early steps toward integration less stable.
Along these lines, a strikingly reliable pattern of interactions was revealed in the analysis of group differences in RT during acquisition, retention, and transfer. For every response type, in every block and phase and transfer test of the experiment, there was a significant or marginally-significant interaction between request time and evaluation such that adding the self-evaluation task to the “Before” condition was associated with faster RTs, whereas adding the self-evaluation task to the “After” condition was associated with slower RTs. Moreover, improvements (decreases) in RT with acquisition were more transient when the self-evaluation task was added to the “After” condition, as demonstrated by the LCA+E group. Given that group was not predicted to have any effect on RT, and that none of the pair-wise comparisons between groups in these interactions was significant, it is difficult to interpret this interaction, other than to say that there was a trend such that learners who were given the self-evaluation task were able to complete the SEQ process more quickly if they selected feedback before trials, and learners who were not given the self-evaluation task completed SEQ more quickly if they selected feedback after trials (or, equivalently, learners who chose feedback before trials completed SEQ processing more quickly if they were also given the self-evaluation task, and learners who chose feedback after trials completed SEQ more quickly if they were not given the self-evaluation task).

A possible interpretation of this unexpected finding is that for learners who chose feedback before trials, perhaps knowing that they would have to perform the self-evaluation task (i.e., LCB+E) heightened their awareness of not only accuracy,
but also speed of response initiation. Learners in the LCA+E group, who presumably did not know before each movement whether or not they would seek feedback and consequently be asked to perform the self-evaluation task, perhaps spent time during the RT interval (and maybe also the ST interval) setting up or maintaining evaluation processes (e.g., activation of the expected response outcome for comparative purposes) to use in the event of having to complete the self-evaluation task. That is, perhaps post-movement processing for the LCA+E group involved comparing expected and actual sensory consequences for the purposes of the self-evaluation task, and then comparing this information to the extrinsic feedback. For the LCA group, it may not have been necessary to activate (or maintain) an expected outcome representation prior to movement as post-movement processing could have just involved comparison of the sensory consequences to the extrinsic feedback. As will be discussed in the next section, if use of a different post-movement processing strategy by the LCA+E group is what underlies these RT findings, the strategy did not appear to be effective for promoting retention and random transfer of relative timing skills.

6.3.3. AE and RE (response execution)

Feedback selection: Turning now to the analyses of the execution of these responses, contrary to prediction, during acquisition, learners from all groups had greater AE on feedback trials than on non-feedback trials. This stands in direct contrast to Chiviacowsky and Wulf’s (2002; 2005) findings that learners perform better on feedback trials (cf. Chiviacowsky & Wulf, 2002 “yoked” group). A possible
reason for the discrepancy is that Chiviacowsky and Wulf compared feedback trials to all non-feedback trials, whereas in the present experiment, the three feedback trials per block were compared to the first three non-feedback trials per block. This was done to minimize the influence of non-feedback trials in which the learners did not actually have a choice about whether or not to receive feedback (i.e., when they had already exhausted their feedback chances for the block). Because learners tended to request feedback early in practice blocks, and did not typically request feedback on three consecutive trials, the present result suggests that after receiving feedback on a given trial, learners corrected their errors on the next trial in which they did not receive feedback, leading to substantially less error on early non-feedback trials.

It is not likely that the finding of greater error on feedback trials reflected an intentional strategy of requesting feedback about poor performance, because half of the learners (i.e., those in the “Before” condition) would not have known the outcome of the trial when they requested feedback, and only three of 80 participants reported that they primarily selected feedback on poor trials. Participants’ introspection about their feedback choices suggested that the LCA group may have attempted to use the strategy of requesting feedback on relatively good trials but avoiding feedback on relatively bad trials. However, the three participants in the “Before” condition, in reporting the use of the maladaptive strategy of requesting feedback on relatively bad trials, were actually the only participants whose subjective report matched the result for AE. Overall, these results suggest that learners were not accurate judges and/or reporters of their feedback-requesting behavior. One methodological difference with
Chiviacowsky and Wulf (2002) that may have contributed to this was that learners in that study completed the questionnaire on the same day as the practice session, whereas in the present experiment, the questionnaire was administered 24 hours later at the end of the entire experiment, in an effort to avoid suggesting anything to the participants about the purpose of the experiment that might bias their retention and transfer performance. These results suggest that learners might be more reliable judges and/or reporters about their use of feedback when asked immediately after practice.

**General patterns in timing error:** Even though learners were only given feedback about the total durations of their responses, both absolute (AE) and relative (RE) timing error decreased with practice. RE was significantly greater for the 4S response pattern than the 4L pattern across all phases of the experiment, supporting the earlier claim (see discussion related to RT) that learners were better able to integrate the 4L sequence into a unitary response pattern.

Learners had more difficulty accurately assigning absolute temporal parameters to short syllables than long syllables during acquisition, phoneme transfer, and absolute difference transfer, which might be related to the pattern of INT demands as described for ST above. At first glance, one might expect that the shorter duration (150 ms) would be easier to accurately execute than the longer duration (450 ms). Indeed, this is what has been found in recent finger movement (Immink & Wright, 2001; Maas, 2006, Experiment 2), and speech (Maas, 2006, Experiment 1) experiments for root mean square error, although the result was only marginally significant for speech. However, in the present experiment, speakers demonstrated
obvious difficulty making the short “ba” short enough (see Figure 6-15). A methodological difference between this experiment and those showing the opposite duration effect (Immink & Wright, 2001; Maas, 2006) possibly contributed to this discrepancy. In the other studies, auditory models of the correct response patterns were provided during or after every trial, whereas in the present experiment, auditory models were generally only provided at the beginning of each block of trials, in an attempt to encourage learners to rely more on the augmented feedback, which was a critical factor for testing group hypotheses. During acquisition, mean response durations for 1L more closely approximated the target duration in the present experiment than in Maas (2006, Experiment 1); however, the opposite was true for the 1S target. This suggests that auditory models, which have been shown to enhance learning (e.g., Lai, Shea, Bruechert, & Little, 2002; Shea et al., 2001), when provided frequently during acquisition, may be more effective for parameterization learning of very short speech responses than infrequent duration feedback.

**Group patterns in timing error:** Following the acquisition phase, learners experienced a decrement in absolute timing performance (AE) for both single syllables and four-syllable patterns. No group differences were found in AE for four-syllable patterns across practice or retention, consistent with Chiviacowsky and Wulf (2002, 2005). However, learners in the present experiment who had requested feedback after trials demonstrated significantly less absolute timing error for single-syllable responses in acquisition, retention, and phoneme transfer than learners who had requested feedback before responses, supporting the hypothesis that learners who
could make feedback requests based on movement performance would better learn to parameterize movements when extrinsic feedback was unavailable.

Interestingly, request time interacted with the self-evaluation factor in the assessment of relative error (RE) for sequences across acquisition and retention phases. Although there were no group differences in RE during acquisition, a key interaction emerged when retention was assessed. Learners in the LCB and LCA+E groups showed significant RE decrements between the last practice block and the retention test, while learners in the LCA and LCB+E groups maintained a level of RE similar to the last practice block on the retention test 24 hours later. This is particularly interesting given that feedback about relative timing error was not provided to learners during acquisition. Learners who made decisions about feedback after trials or who were asked to self-evaluate their performance (but not both) were significantly better at remembering the relative timing patterns of four-syllable sequences on a retention test 24 hours later. This finding suggests that forced self-evaluation benefited movement structure coherence for learners who chose feedback before trials, and therefore might not have spontaneously self-evaluated as often, but that forced self-evaluation was detrimental to learners who chose feedback after trials, as it might have added another dimension to the post-movement processing phase (see discussion of RT results) that distracted learners from building relative timing patterns for response sequences. This pattern was also corroborated by RE during random transfer as well as the overall accuracy (proportion of acceptable trials) in retention and phoneme transfer.
For transfer tasks in which learners were required to apply new parameterizations to learned relative patterns (phoneme transfer and absolute difference transfer probes), a mixed pattern of results was obtained. When executing the learned patterns with novel phonemes, learners in the “After” condition had smaller absolute timing error for single syllables, and smaller relative timing error for sequences than learners in the “Before” condition. Chiviacowsky and Wulf (2002, 2005) also demonstrated this general pattern of “self-after” learners exhibiting smaller error than “self-before” learners or learners who did not have self-control of feedback in transfer. The transfer task used by Chiviacowsky and Wulf, however, was more similar to the present absolute difference transfer probe, in which no group differences were found for RE, and an advantage of “no evaluation” (as compared to self-evaluation) was found for AE. Finally, for the relative difference transfer task in which learners were asked to apply learned parameterizations to new relative timing patterns, the only group difference that emerged was an advantage of LCA over LCB for relative error, although, numerically, the LCB group did demonstrate the greatest AE and RE for all relative difference transfer patterns. This suggests that for LCB-condition learners, the practiced parameterizations may not have been stable enough to be successfully applied in new relative patterns.

6.3.4. Summary and Conclusions

The literature on limb motor learning has shown that allowing subjects to request feedback after performance is beneficial to learning. The present experiment is
the first to demonstrate this effect in speech motor learning. Differences between “After” and “Before” feedback conditions generally indicated an advantage of the “After” condition in terms of greater accuracy in random transfer, less absolute timing error for monosyllabic responses, less relative timing error for phoneme transfer, longer study times suggestive of more extensive INT processing, and RT evidence for reorganization of sequences into unitary representations. Although participants did not tend to request feedback on good performance as was predicted, learners in the LCA group reported that this had been their intention. Learners in the “After” groups also tended to have generally more positive attitudes about the augmented feedback than learners in the “Before” groups, as indicated by their responses to the questionnaire.

The present results also provide qualified support for an interaction between feedback request time and forced self-evaluation, although not exactly in the way predicted by Chiviacowsky and Wulf’s (2005) hypothesis that the learning advantage demonstrated when feedback is chosen after a trial is related to a greater propensity to attend to intrinsic feedback sources and spontaneously evaluate performance. If spontaneous performance evaluation is responsible for this learning advantage, then an interaction would be expected such that forced self-evaluation would help learners who make feedback choices before trials, but would not help (or help less) learners who choose feedback after trials. Support for the first half of this prediction was found in several analyses, that is, that forced self-evaluation helped learners who had to make feedback choices before movements. However, evidence was also uncovered that suggested that forced self-evaluation actually had a deleterious effect on learners
who chose feedback after trials. The most striking example of this was the finding that the LCB+E and LCA groups maintained stable relative timing error between practice and the delayed retention test, whereas groups LCB and LCA+E significantly worsened. The advantage of the LCB+E and LCA groups persisted when they were required to perform the practiced movements in random order. Overall accuracy yielded converging evidence for this pattern in the delayed retention test as well as the phoneme transfer probe. Both tasks demonstrated an interaction between the two group factors such that forced self-evaluation was associated with greater accuracy for “Before” learners, and with reduced accuracy for “After” learners. The RT data demonstrated this interaction most consistently, indicating that forced self-evaluation was associated with faster initiation of all responses in the “Before” condition, but slower initiation in the “After” condition.

Why would an additional opportunity to self-evaluate have a detrimental effect on learners who chose feedback after responses? Perhaps the insertion of an additional task into the KR-delay interval disrupted learners’ movement processing enough to have a detrimental effect on initiation time and execution even though performing the self-evaluation task should be compatible with, not disruptive to, the normal processing occurring during the KR-delay interval (e.g., Hogan & Yanowitz, 1978; Swinnen, 1990). The pattern of ST results, viewed in relation to the other findings, suggests that too-short or too-long response preparation time may be associated with reduced performance in terms of movement initiation (RT) and execution (AE, RE, Accuracy). Although the pattern of statistically-significant results was not
unequivocal, there was an overall trend, particularly in acquisition and retention, for group LCB to have the shortest STs and for group LCA+E to have the longest STs. As described above, these groups did not perform as well as the LCB+E and LCA groups on other aspects of performance. The less-extreme (i.e., neither long nor short) STs evidenced by the latter groups appear to be associated with better performance.

Put in the context of Schema Theory, the present results indicate that the group factors manipulated here influenced both GMP and parameterization learning. Despite the fact that augmented feedback was given only in terms of total duration, learners appear to have paid attention to both relative and absolute timing aspects of the targets, as both RE and AE decreased significantly with practice. Further, group differences, although not consistent, were found for both measures, indicating that learners who were theoretically able to make feedback choices based on the outcomes of their responses generally demonstrated better ability to parameterize trained responses and transfer trained duration parameterizations to different speech sounds. Further, learners who were forced to self-evaluate their movements prior to receiving feedback were more successful at re-parameterization of trained sequences than learners who were not forced to self-evaluate. Interestingly, when coherence of the trained movement patterns was assessed with the relative error measure, interactions between the two group factors emerged, such that in general, forced self-evaluation was associated with improved GMP coherence for learners who were required to request feedback before movements, but with reduced GMP coherence for learners who chose feedback after trials. These results suggest that requesting feedback after
response completion or self-evaluating movement execution is beneficial to GMP learning. Being asked to both make a feedback choice and complete a self-evaluation task during the KR delay appears to disrupt the optimal processing circumstances that either of these conditions in isolation promotes.

Finally, the present results raise interesting issues concerning the two-process INT/SEQ model of motor programming (e.g., Klapp, 1995) and its demonstration in speech with the self-selection paradigm (e.g., Immink & Wright, 1998; 2001). First, duration of single-syllable responses did not affect ST, as was predicted, but rather impacted RT, raising questions about when INT processing took place, and whether other kinematic-related variables were more influential in this experiment than in prior studies. Further, Experiment 3 was the third recent study to fail at finding a duration effect on ST, highlighting the relevance of the debate about whether single-syllable duration is an appropriate way to define motoric complexity for speech units (see Maas, Robin, Wright et al., in press; Wright et al., submitted) or whether methodological confounds have obscured this subtle effect. Second, while the predicted effect of sequence length on the SEQ process was demonstrated in RT, it was also robust in the ST interval, indicating that all units or all unique units in sequences were pre-programmed. While other recent work using the same stimuli (Maas, Robin, Wright et al., in press) has concluded that non-isochronous syllable repetitions are organized as single units, the present RT results provided evidence that suggested that only one of the patterns (i.e., 4L) was unitized by the participants in this experiment.
Finally, Experiment 3 demonstrates that the self-selection paradigm can be productively applied to the speech motor domain for the purpose of developing an understanding about the mechanism of feedback manipulations on two motor programming processes, INT and SEQ. Future work might take advantage of this paradigm to investigate the effects of other principles of motor learning on normal and disordered speech motor programming.
7.0. Learner-Controlled Feedback in Speech Motor Programming in AOS

The primary purpose of Experiment 4 was to extend the principle of learner-controlled feedback to speech motor learning in apraxia of speech (AOS). Although a growing body of literature has begun to examine the effects of feedback manipulations on treatment for AOS (e.g., Austermann Hula et al., in press, see Experiments 1 & 2; Ballard et al., 2007; Fossett et al., 2008; McNeil et al., 2007), recent evidence has also suggested that individuals with AOS have an impaired ability to use feedback to improve temporal accuracy of movements (Ballard & Robin, 2007). Based on the hypothesis that allowing learners to choose when to receive feedback enhances learning by giving them information when they need it (Chiviacowsky & Wulf, 2002; 2005), it was predicted that allowing individuals with AOS to control their own feedback schedules would improve their learning of a novel speech timing task, as measured by absolute and relative timing error of their responses. Following the results of Chiviacowsky and Wulf (2002), it was predicted that this difference would emerge not during acquisition, but in transfer probes. These predictions were tested in a multiple baselines across behaviors single subject experimental design with four participants. Learner-controlled (LCF) and non-learner-controlled (yoked; YF) feedback conditions were compared across two phases of the treatment design for each participant.

Like Experiments 1 and 2 described in Chapter 5, the present experiment was designed to test the application of a principle of motor learning (i.e., learner-controlled feedback) to speech motor skill learning, and not to test the efficacy of a particular
kind of treatment. However, unlike Experiments 1 and 2, which employed a commonly-used AOS treatment method, the present experiment trained more-abstract speech behaviors (i.e., temporal patterns of syllables) in the context of the self-selection paradigm (Immink & Wright, 1998; 2001), as in Experiment 3. By conducting the present learning study with the self-selection paradigm within the framework of the INT/SEQ model of motor programming (Klapp, 1995; 2003), a secondary objective was achieved. Namely, the effects of self-controlled feedback were examined for INT and SEQ stages of response preparation and initiation. The key model predictions of a syllable duration effect on ST and a sequence length effect on RT were predicted. It was also predicted that INT processing would become more efficient, as measured by decreases in ST, with practice, and that this would be better promoted by LCF as opposed to YF.

Although AOS is widely understood to be a disorder of motor programming (e.g., Aichert & Ziegler, 2004; Clark & Robin, 1998; Deger & Ziegler, 2002; Hageman et al., 1994; McNeil et al., 1990; 1997; 2000; Wambaugh et al., 2006), the precise nature of the deficit is not yet understood. Recent evidence using the self-selection paradigm has provided evidence for a general motor programming disruption in AOS isolated to the INT processing stage across speech and finger-movement modalities (Maas, Robin, Wright et al., in press). Although the present study does not permit direct comparison between impaired and normal processing, and thus cannot confirm Maas’ findings, it affords an opportunity to observe learning-induced changes
in the two processing stages as well as in motor execution, and allows qualitative comparison of patterns of results with normal performance.

7.1. Methods

7.1.1. Participants

Four participants (3 men, 1 woman) with AOS participated in the study. Participants provided informed consent and were paid $10 per session for their services. None had previous experience with the paradigm, and all were blind to the specific purpose of the experiment and the nature of the independent variable manipulations. Participants were all premorbidly right-handed native English speakers over the age of 18 ($M = 57$, range = 46-63) with AOS and aphasia secondary to stroke (months post-onset $M = 31$, range = 6-57), normal (aided or unaided) hearing and vision acuity and negative self-reported premorbid history of communication disorder, learning disability, neurological illness, psychiatric illness, and head injury. All participants were diagnosed with AOS by at least three certified speech-language pathologists with expertise or experience in motor speech disorders, and demonstrated all of the cardinal perceptual features consistent with the disorder as proposed by McNeil, Robin, and Schmidt (1997) and adopted by the Academy of Neurological Communication Disorders and Sciences (Wambaugh et al., 2006a; 2006b), including overall slow rate due to lengthened segment and inter-segment durations, phonemic distortions, and prosodic abnormalities such as stress neutralization and sound, syllable, and word segregation.
Approximately one week before beginning the experimental protocol, participants were tested with the Western Aphasia Battery (WAB; Kertesz, 1982) to characterize their aphasia, the Apraxia Battery for Adults-2 (ABA-2; Dabul, 2000) to characterize their AOS, the Arizona Battery for Communication Disorders in Dementia Story-Retelling subtest (ABCD; Bayles & Tamoeda, 1993) to characterize their short-term verbal memory, and an oral-mechanism examination to characterize any oral-motor impairments or dysarthria. Results of these tests are shown in Table 7-1.
Table 7-1. Participant information for Experiment 4. AOS = apraxia of speech; WAB AQ = Western Aphasia Battery Aphasia Quotient; WAB CQ = Western Aphasia Battery Cognitive Quotient; ABCD = Arizona Battery for Communication Disorders of Dementia.

<table>
<thead>
<tr>
<th></th>
<th>AOS1</th>
<th>AOS2</th>
<th>AOS3</th>
<th>AOS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Age</td>
<td>62;11</td>
<td>69;8</td>
<td>45;11</td>
<td>49;6</td>
</tr>
<tr>
<td>Education (years)</td>
<td>23</td>
<td>17</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Hand used</td>
<td>R</td>
<td>L</td>
<td>R</td>
<td>L</td>
</tr>
<tr>
<td>Time post-onset</td>
<td>39 months</td>
<td>57 months</td>
<td>23 months</td>
<td>6 months</td>
</tr>
<tr>
<td>Type of stroke</td>
<td>hemorrhagic</td>
<td>thromboembolic with hemorrhagic transformation</td>
<td>thromboembolic</td>
<td>thromboembolic</td>
</tr>
<tr>
<td>Lesion location</td>
<td>Left MCA territory (frontal, temporal, parietal)</td>
<td>Left MCA involving frontoparietal region and insula</td>
<td>Left MCA involving sylvian region (frontal, parietal, and anterior temporal lobes), including lentiform nucleus and portion of caudate nucleus.</td>
<td>Left MCA involving frontoparietal region</td>
</tr>
<tr>
<td>WAB AQ</td>
<td>94.8</td>
<td>16.9</td>
<td>62.9</td>
<td>92.7</td>
</tr>
<tr>
<td>WAB CQ</td>
<td>96.3</td>
<td>35.5</td>
<td>71.4</td>
<td>89.8</td>
</tr>
<tr>
<td>Aphasia Severity</td>
<td>Mild</td>
<td>Severe</td>
<td>Moderate</td>
<td>Mild</td>
</tr>
<tr>
<td>AOS Severity</td>
<td>Mild-moderate</td>
<td>Severe</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Oral/Limb Apraxia</td>
<td>None</td>
<td>Moderate Oral and Limb</td>
<td>Mild Oral and Limb</td>
<td>None</td>
</tr>
<tr>
<td>Oral Motor</td>
<td>Normal</td>
<td>Normal</td>
<td>Mild unilateral weakness</td>
<td>Mild unilateral weakness</td>
</tr>
<tr>
<td>ABCD Story- Retell Immediate-Delayed Proportion</td>
<td>100</td>
<td>Unable to complete due to communication severity</td>
<td>100</td>
<td>85.7</td>
</tr>
</tbody>
</table>

7.1.2. Materials and Equipment
Materials and equipment were the same as those used in Experiment 3, except that speech sound stimuli (i.e., syllable type, syllable durations, and durational patterns) varied within subjects across the two treatment phases of the experiment (see Table 7-3). All participants learned monosyllables and multisyllabic (two or four syllable) patterns using the syllable “pa” in the first phase and “chee” in the second phase, except where individual capabilities for producing these sounds varied, and modifications were made for individual participants (see Table 7-2). In one phase of the experiment, targets were constructed using short (150 ms) and long (450 ms) versions of one of these syllables, and in the other phase, medium (275 ms) and extra-long (575 ms) durations of the other syllable. The order of these duration targets was counterbalanced across subjects, and across levels of the independent variable (learner-controlled versus yoked feedback schedule), as shown in Table 7-2.
Table 7-2. Treatment stimulus and independent variable pairings for each participant in Experiment 4.

<table>
<thead>
<tr>
<th></th>
<th>Phase 1</th>
<th>Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Syllable</td>
<td>Durations</td>
</tr>
<tr>
<td>AOS1</td>
<td>pa</td>
<td>short-long</td>
</tr>
<tr>
<td>AOS2</td>
<td>pa</td>
<td>medium-extra long</td>
</tr>
<tr>
<td>AOS3</td>
<td>la</td>
<td>short-long</td>
</tr>
<tr>
<td>AOS4</td>
<td>pa</td>
<td>medium-extra long</td>
</tr>
</tbody>
</table>

Another difference between the material used for this experiment and Experiment 3 had to do with the nature of the stimuli used for the transfer probe tasks. Treatment targets and transfer probe stimuli are shown in Table 7-3 (Participant AOS1’s order of presentation is shown as an example). All participants were administered blocked retention and random transfer probes involving the treated responses. Regardless of phase, the random-order transfer probes for both conditions were administered at each baseline or retention/transfer session. The blocked retention probe for each syllable type was administered only in the relevant phase (i.e., blocked retention probes for the “pa” stimuli were only conducted in Phase I, and for the “chee” stimuli were only conducted in Phase II), except for AOS2, who received both the blocked and random retention probes for each set of stimuli across the entire experiment.
Table 7-3. Example target cues and associated patterns for Experiment 4 (AOS1’s stimuli-condition pairings shown). S = short; L = long; M = medium; X = extra-long.

<table>
<thead>
<tr>
<th>Block Name</th>
<th>Target Cues</th>
<th>Syllable Durations (Total) (ms)</th>
<th>Order of Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice, Blocked Retention</td>
<td>1S</td>
<td>150</td>
<td>Blocked</td>
</tr>
<tr>
<td></td>
<td>1L</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2S</td>
<td>150-450 (700)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4L</td>
<td>450-150-150-150-450 (1500)</td>
<td></td>
</tr>
<tr>
<td>Phase I</td>
<td>1S</td>
<td>150</td>
<td>Random</td>
</tr>
<tr>
<td>/pa/</td>
<td>1L</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>Random Transfer</td>
<td>2S</td>
<td>150-450 (700)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4L</td>
<td>450-150-150-150-450 (1500)</td>
<td></td>
</tr>
<tr>
<td>Practice, Blocked Retention</td>
<td>1M</td>
<td>275</td>
<td>Blocked</td>
</tr>
<tr>
<td></td>
<td>1X</td>
<td>575</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2X</td>
<td>575-275 (950)</td>
<td></td>
</tr>
<tr>
<td>Phase II</td>
<td>1M</td>
<td>275</td>
<td>Random</td>
</tr>
<tr>
<td>/chi/</td>
<td>1X</td>
<td>575</td>
<td></td>
</tr>
<tr>
<td>Random Transfer</td>
<td>2X</td>
<td>575-275 (950)</td>
<td></td>
</tr>
</tbody>
</table>

7.1.3. Tasks and Procedures

The self-selection paradigm, as described in Experiment 3, was used for elicitation of all treatment, retention, and transfer responses in Experiment 4. A few exceptions to the tasks as described in Experiment 3 were made. First AOS2 (severe apraxia) was allowed a longer RT window (100 ms to 3000 ms) because he could not otherwise achieve the minimum number of acceptable trials in the first baseline phase;
however after the first baseline, very few of his trials had RTs greater than 1000 ms. The other participants were constrained to an RT window of 100 to 1000 ms, as was used in Experiment 3. Also, AOS3 was shown only “!” as the imperative signal (instead of “Go!”) because of her difficulty inhibiting the syllable “go” in place of the target syllable production. In addition, participants provided manual button-press responses with whichever hand they considered to be their most functional hand, which meant that for two participants who were affected by significant hemiparesis, the left hand was used. Finally, study times (STs) were not screened with a hard maximum as they were in Experiment 3. Rather, Experiment 4 STs were excluded from analysis only when the participant was demonstrably off-task (e.g., talking, coughing). These exclusions accounted for 0.04 % of trials for AOS1, 0.35% for AOS2, 1.2% for AOS3, and 0.99% for AOS4.

The independent variable of feedback control was manipulated across the two treatment phases of the experiment for each participant. As in Experiment 3, learners in Experiment 4 received feedback on 30% of correct trials, or 3 trials for every block of 10 acceptable responses. Trial schematics for the two feedback conditions are shown in Figure 7-1. It should be noted the KR-delay was longer in the LCF condition than the YF condition, because this was when participants chose whether or not to receive feedback. Inter-trial intervals, which are considered more influential to motor skill learning (Salmoni, et al., 1984), were constant across both feedback conditions. In the learner-controlled feedback (LCF) phase, participants registered their feedback requests immediately after each response, and total duration feedback (in ms) was
subsequently provided as described in Experiment 3 for the LCA group. In the yoked feedback (YF) condition, after each response, participants were shown a display that instructed them to press a button to go on (in place of the feedback request display that was used for LCF), and then feedback was provided (or not), depending on the yoked feedback schedule. Each participant demonstrated satisfactory ability to understand numbers and numeric relationships of the kind used for feedback by completing an informal screening before the experiment began. In the screening, participants were required to correctly place numbers (similar to the values that would be encountered during the experiment) on a number line and classify the magnitudes of differences between pairs of values (e.g., much larger, a little bit larger).
Figure 7-1. Events during each practice trial in the self-selection paradigm for both conditions of Experiment 4. Note that feedback was only provided on 30% of correct trials. LCF = learner-controlled feedback; YF = yoked feedback.

The first participant, AOS1, received LCF for the first phase of the experiment. For the second phase of the experiment, “yoked” feedback (YF) was provided on the same Phase II trials that AOS1 had requested feedback for in Phase I. Thus, AOS1’s self-determined Phase I feedback schedule was used as the schedule of Phase II feedback, with the only difference being that in Phase II, he did not have control over feedback delivery and could not predict its arrival. (He later reported that he thought that it had been randomly presented.) Participant AOS2 began the experiment under
the YF condition, which was again delivered by yoking his feedback schedule to that of AOS1. In this manner, each participant received feedback on the same 30% of practice trials during YF treatment.

Acquisition consisted of four practice sessions in each phase, typically occurring every other day. Due to participants’ availability limitations and resource restrictions, it was decided \textit{a priori} that four practice sessions per phase would be conducted, instead of using a performance-based criterion for practice termination. In each practice session, 16 blocks (four for each of four stimulus types) of ten correct trials each were administered, taking approximately 60 minutes. As in Experiment 3, each block consisted of a single stimulus type, and the order of blocks was organized such that each set of four consecutive blocks contained one block of each trial type (e.g., 1S, 1L, 2S, 4L) presented in random order. In the first treatment session, the first four blocks, being the first practice block for each trial type, were collectively referred to as “treatment 1.1,” the second four blocks comprised “treatment 1.2,” and so on until the last four blocks, or “treatment 1.4.” The second treatment session began with “treatment 2.1,” and so on. Verbal and written instructions and auditory models were provided as in Experiment 3.

Learners also participated in probe sessions to evaluate their baseline, retention, and transfer abilities in the absence of augmented feedback. At least four baseline sessions (participant AOS1 received five baselines), conducted approximately every other day, preceded the commencement of the first treatment phase. Thereafter, retention and transfer probes were administered on the same days as practice sessions,
before practice began, and took approximately 30-45 minutes to complete. During probe periods, a blocked retention test of the currently-treated responses was administered first, followed by the random transfer probe. Finally, productions of the treatment targets for the other (upcoming or already completed) treatment phase were elicited in random order at the end of each retention and transfer probe session. Each of these three probe sets consisted of 12 acceptable trials: three for each of the four responses. Unacceptable trials (i.e., incorrect responses or RT window violations) were re-run at the end of each block until 12 acceptable trials were completed. A final retention and transfer probe session was completed four days after the withdrawal of treatment in each phase. Following initial testing, which took two to three sessions, learners participated in 12 to 13 experimental sessions, spanning approximately five weeks.

7.1.4. Design and Analysis

A single-subject multiple baselines across behaviors experimental design was used to assess treatment-related changes in speech motor programming and execution in four participants with AOS. As in Experiment 3, the execution of speech movements was measured by analyzing absolute and relative timing error (AE and RE, respectively) of speech productions. Computation of AE and RE was the same as in Experiment 3. Inter- and intra-rater reliability was examined on 16.4% of probe data and 15.4% of treatment data for each participant. Reliability was satisfactory for AOS1 (Inter-rater r = 0.983, Inter-rater mean absolute difference = 24 ms, SD = 27
ms; Intra-rater \( r = 0.984 \), Intra-rater mean absolute difference = 24 ms, SD = 26 ms),
AOS2 (Inter-rater \( r = 0.986 \), Inter-rater mean absolute difference = 26 ms, SD = 24
ms; Intra-rater \( r = 0.992 \), Intra-rater mean absolute difference = 21 ms, SD = 20 ms),
AOS3 (Inter-rater \( r = 0.995 \), Inter-rater mean absolute difference = 12 ms, SD = 16
ms; Intra-rater \( r = 0.988 \), Intra-rater mean absolute difference = 20 ms, SD = 24 ms),
and AOS4 (Inter-rater \( r = 0.995 \), Inter-rater mean absolute difference = 14 ms, SD =
14 ms; Intra-rater \( r = 0.987 \), Intra-rater mean absolute difference = 30 ms, SD = 25
ms).

To compare treatment effects across the two different feedback conditions,
effect sizes were calculated on AE and RE for the random transfer probes across all
phases of the experiment. Busk and Serlin’s (1992) variation of Cohen’s \( d_1 \) statistic
was used, as advocated by Beeson and Robey (2006):

\[
   d_1 = \frac{\bar{x}_{A_2} - \bar{x}_{A_1}}{S_{A_1}}
\]

where \( \bar{x}_{A_1} \) and \( \bar{x}_{A_2} \) represent means of baseline and post-treatment data, respectively,
and the denominator is the standard deviation of the baseline data. For the present
data, improvements (i.e., decreases) in AE and RE are referred to with positive \( d_1 \)
values, and decrements (i.e., increases) in AE and RE are referred to with negative \( d_1 \)
values. Effect size interpretations have not been established for AOS single-subject
designs, and therefore the following benchmarks derived from the aphasia treatment
literature by Robey, Schultz, Crawford, and Sinner (1999) were used for \( d_1 \): 2.6 =
small effect, 3.9 = medium effect, 5.8 = large effect.
The calculation of the $d_l$ statistic for Phase II was complicated by a substantial influence of Phase I treatment (for “pa”) on the Phase II stimuli (“chee”) being held in baseline. The magnitude of this transfer effect was measured by comparing performance on the random transfer probe for “chee” during the true baseline (i.e., prior to any treatment) to performance on the same probe during the first treatment phase (in which “pa” was treated) (see $d_{lP}$). In contrast, the effect sizes pertaining to Phase II treatment were computed by comparing treatment-phase probes to the most recent four (or five, in the case of AOS1) pre-treatment probes (see $d_{lT}$). In this way, the same number of “baseline” data points was used to calculate treatment effects in both phases. Table 7-4 describes the different effect sizes that were calculated for Experiment 4.

Table 7-4. Four different effect size ($d_l$) calculations used in Experiment 4, and the random transfer probe data used to compute them. $A_1 = $ pre-treatment period, $A_2 = $ post-treatment period, $T = $ treatment, $W = $ withdrawal, $P = $ pre-treatment, $M = $ maintenance.

<table>
<thead>
<tr>
<th>Effect size label</th>
<th>$A_1$ (pre-treatment data)</th>
<th>$A_2$ (post-treatment data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{lT}$</td>
<td>Four probes (five for AOS1) administered immediately prior to onset of treatment; Phase I and Phase II</td>
<td>Four probes administered during treatment phase, including 4-day delayed retention probe; Phase I &amp; II</td>
</tr>
<tr>
<td>$d_{lW}$</td>
<td>Four probes (five for AOS1) administered immediately prior to onset of treatment; Phase I and Phase II</td>
<td>Single probe administered four days after the withdrawal of treatment; Phase I and II</td>
</tr>
<tr>
<td>$d_{lP}$</td>
<td>Four “chee” probes (five for AOS1) administered during “true” baseline (prior to any treatment)</td>
<td>Four “chee” probes (five for AOS1) administered pre-Phase II treatment (i.e., during Phase I treatment)</td>
</tr>
<tr>
<td>$d_{lM}$</td>
<td>Four “pa” probes (five for AOS1) administered immediately prior to onset of Phase I treatment</td>
<td>All “pa” probes administered after the end of Phase I treatment, during its maintenance phase</td>
</tr>
</tbody>
</table>
It should also be noted that since retention probes were administered throughout the treatment phases in the present experiment, the measurement reported as the “treatment effect” is not calculated from “post-all-treatment” data, but rather from retention data that were collected after each treatment session. A truly post-treatment effect was calculated by using the one retention probe administered four days following the withdrawal of treatment in each phase (see $d_{1W}$). Probes were also administered on Phase I stimuli throughout Phase II in what might be seen as a maintenance period for Phase I targets (see $d_{1M}$), but there was not a corresponding post-treatment period in Phase II with which these data could be compared.

The analysis of treatment-related changes in response execution focused primarily on the random transfer probe. However, the blocked retention data were also inspected for treatment-related differences in absolute and relative timing error. First, the acquisition (i.e., practice session) data were examined by analyzing the rate of change (i.e., the slope of the best fit linear function for the acquisition data plotted against session number) of AE and RE during treatment sessions and comparing these across treatment phases. Negative slope values were used to refer to decreasing AE or RE. Second, blocked retention was examined by comparing the overall mean performance on blocked retention probes (after treatment initiation) to overall mean performance in treatment blocks. Although this method does not permit a comparison of the absolute levels of performance, it does allow the blocked retention probe to be inspected across phases relative to the degree of acquisition.
Finally, to identify the presence of any patterns in participants’ feedback-requesting behavior, two methods were used. First, for each practice block in each phase, ST, RT, AE, and RE were computed separately for the three feedback trials and the first three no-feedback trials. Means and standard deviations for these dependent variables for feedback versus no-feedback trials were computed for each phase and compared descriptively as an estimate of any differences. Second, after each treatment condition was completed, participants were asked questions similar to those in the questionnaire given in Experiment 3 (see Table 6-2) about when they did and did not request (in the case of the LCF condition) or think that they received (in the case of the YF condition) feedback.

7.2. Results

7.2.1 AOS1

Response execution: AE and RE

Participant AOS1 was able to reliably differentiate the four different response durations in his productions during acquisition and probes, as shown in Figures 7-2 and 7-3, respectively. Absolute and relative error data relating to these productions are displayed in Figures 7-4 and 7-5, respectively. Each data point in Figures 7-4 and 7-5 represents average performance for all response types (e.g., 1S, 1L, 2S, 4L) administered in each particular task.

Baseline: During baseline, AOS1 did not demonstrate very stable performance on either set (“pa” or “chee”) of target behaviors (see Figure 7-4 and 7-5, first five
sessions). However, particularly for the “pa” responses, a reliable trend in error reduction was not observed either. Phase I treatment commenced after five baseline sessions. Despite the initial instability, treatment effects in the random transfer probe were observed and measured with the effect size statistic, which took into account baseline variability.

**Feedback during acquisition:** AOS1 received LCF treatment on short- and long- duration “pa” syllables and patterns in Phase I and YF treatment on medium- and extra-long-duration “chee” syllables and patterns in Phase II (see Table 7-3). During Phase I LCF treatment, there was no substantial difference between feedback and no-feedback trials in AE (feedback: $M = 59$ ms, $SD = 46$ ms; no-feedback: $M = 55$ ms, $SD = 43$ ms) or RE (feedback $M = 0.139$, $SD = 0.062$; no-feedback: $M = 0.149$; $SD = 0.056$). Likewise, during Phase II YF treatment there was no difference between feedback and no-feedback trials in AE (feedback $M = 69$ ms, $SD = 60$ ms; no-feedback: $M = 71$ ms; $SD = 56$ ms) or RE (feedback: $M = 0.104$, $SD = 0.049$; no-feedback: $M = 0.109$, $SD = 0.054$). After Phase I LCF treatment, AOS1 reported that he had requested feedback randomly, that he had chosen not to request feedback on “neither good nor bad” trials, and that the feedback helped him improve his response durations “somewhat.” After Phase II YF treatment, AOS1 reported that he thought he had received feedback on predominately good trials, that he had not received feedback on “neither good nor bad” trials, and that the feedback helped him improve his responses “a lot.”
**Acquisition:** Performance during acquisition was assessed by comparing the slopes of Phase I and Phase II treatment session absolute and relative error data (see open symbols in Figure 7-4 and 7-5, top and bottom panels, respectively). Neither treatment set (i.e., “pa” or “chee”) changed substantially during practice blocks, although a very slight improvement was noted with LCF treatment (Phase I LCF slope = -0.28, \( r^2 = 0.27 \); Phase II YF slope = -0.07, \( r^2 = 0.03 \)). Conversely, Phase II YF treatment was associated with a small reduction in relative error (slope = -1.14, \( r^2 = 0.44 \)), whereas Phase I LCF treatment was not (slope = -0.06, \( r^2 = .001 \)).
Retention: Retention of responses trained under blocked practice conditions was examined by comparing average performance on the blocked retention probe (see filled gray symbols in Figures 7-4 and 7-5) to overall practice performance in each phase. Relative to practice performance, AOS1 demonstrated greater (18%) AE in blocked retention during the Phase I LCF treatment period, and reduced (-25%) AE in blocked retention during Phase II YF treatment. There was less relative timing error (RE) in blocked retention probes than in practice performance, with an apparent

Figure 7-2. Stimuli target durations (in ms, solid lines) and actual response durations (in ms, broken lines) for AOS1 for each practice block during acquisition (Phase I on left; Phase II on right). See Table 7-3 for stimulus cue definitions. Session labels X.1-X.4 represent the four practice blocks within the X practice session. LCF = Learner-Controlled Feedback; YF = Yoked Feedback.
advantage to the Phase I LCF condition (-23%) compared to the Phase II YF condition (-8%). However, these data must be interpreted with caution given the differences in baseline performance for these items immediately prior to each treatment, which can explain these differences.

Transfer: Transfer of the target responses trained in blocked order to a random-order probe task was assessed across all phases of the experiment for both stimulus sets (see filled black symbols in Figures 7-4 and 7-5). During the first treatment phase (“pa”), there were small effects of LCF treatment on AE ($d_{1T} = 3.43$) and RE ($d_{1T} = 2.64$) for the random transfer probes in that condition. However, there were also very small effects of Phase I LCF treatment (“pa”) on AE ($d_{1P} = 1.69$) and RE ($d_{1P} = 1.47$) for the untrained “chee” stimuli that were to be held in baseline until Phase II treatment began (see Figures 7-4 and 7-5, bottom panels, sessions P1-P4), demonstrating that training short- and long- duration “pa” syllables and patterns with LCF treatment may have influenced performance on putatively unrelated medium- and extra long-duration “chee” syllables and patterns that had not yet been treated. Another possibility is that the error reduction observed in “chee” responses during Phase I treatment for “pa” was simply an extension of the unstable and slightly decreasing baseline for “chee” responses. Despite this limitation, slightly greater decreases in error (AE: $d_{1T} = 2.74$; RE: $d_{1T} = 1.95$) were observed for “chee” responses during Phase II YF treatment than during the pre-treatment period. Although these effects were smaller than the treatment effects described above for “pa” responses
during Phase I LCF treatment, they were suggestive of a small effect of Phase II YF treatment on “chee” responses.

To measure the persistence of these treatment effects shortly after practice conditions (i.e., feedback) were withdrawn, the 4-day delayed probe (see Figures 7-4 and 7-5; sessions P4 and P8) was compared to baselines for the random transfer probe for each behavior set. This comparison indicated a medium-sized effect of Phase I LCF treatment on AE for random-order probes \(d_{lw} = 4.27\) and no effect of Phase II YF treatment \(d_{lw} = 0.02\). For RE, the advantage of LCF treatment persisted, albeit less dramatically, with small effect sizes for both Phase I (LCF) \(d_{lw} = 2.32\) and Phase II (YF) \(d_{lw} = 1.80\). Despite this apparent short-term learning advantage for the LCF condition, the continued collection of “pa” probes throughout Phase II YF treatment indicated that these gains were not well-maintained once treatment for the “chee” responses began, as a small performance decrement was observed for AE \(d_{im} = -0.93\) and RE \(d_{im} = -0.087\), suggesting that “chee” treatment may have interfered with retention of “pa” responses, or that the improvements with “pa” were dependent on the continued provision of augmented feedback during treatment sessions.
Figure 7-3. Stimuli target durations (in ms, solid lines) and actual response durations (in ms, broken lines) for AOS1 for random transfer probes during each probe session (Top panel: “pa” responses; Bottom panel: “chee” responses). See Table 7-3 for stimulus cue definitions. LCF = Learner-Controlled Feedback; YF = Yoked Feedback. Tx = treatment.
Figure 7-4. Participant AOS1 average absolute error (AE) for responses during treatment sessions and retention and transfer probes. Top panel = Phase I targets; Bottom panel = Phase II targets. B = Baseline, T = Treatment, P = probe, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Figure 7-5. Participant AOS1 average relative error (RE) for responses during treatment sessions and retention and transfer probes. Top panel = Phase I targets; Bottom panel = Phase II targets. B = Baseline, T = Treatment, P = probe, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Motor Programming

To assess the effects of the different feedback conditions on pre-movement motor programming in AOS1, study times (STs) and reaction times (RTs) for one-, two-, and four-syllable responses were examined within the context of random-order probes across all phases of the experiment (see Figure 7-6). For the most part, ST (left panels) and RT (right panels) decreased for “pa” and “chee” responses across all phases of the experiment, although fluctuations were evident at the level of different response types. One exception to this pattern was the case of STs for the “pa” responses, which did not continue to decrease after Phase I LCF treatment (Figure 7-6, top left panel). This pattern of ST results corresponded to a similar pattern in response execution results discussed above. That is, AE and RE decreased across the experiment, except for the “pa” responses after Phase I LCF treatment.

ST and RT did not appear to be related to whether or not AOS1 requested feedback on a given trial. During LCF treatment, there was little difference between feedback and no-feedback trials in ST (feedback: $M = 1209$ ms, $SD = 956$ ms; no-feedback: $M = 1076$ ms, $SD = 636$ ms) and RT (feedback $M = 484$ ms, $SD = 176$ ms; no-feedback: $M = 476$ ms; $SD = 173$ ms). During YF treatment (in which AOS1 did not request feedback at all), there was also little difference between feedback and no-feedback trials in ST (feedback: $M = 971$ ms, $SD = 524$ ms; no-feedback: $M = 933$ ms, $SD = 576$ ms) and RT (feedback $M = 458$ ms, $SD = 184$ ms; no-feedback: $M = 442$ ms; $SD = 178$ ms).
Study Time (INT process): Based on the INT/SEQ model, it was predicted that ST would be influenced by the duration of the single-syllable responses. As shown in Figure 7-6 (left panels), STs were, on average, 760 ms longer for the long-duration ($M = 2063$ ms, $SD = 444$ ms) than the short-duration ($M = 1303$ ms, $SD = 264$ ms) “pa” (LCF-treated) response (top left panel), and 663 ms longer for the extra-long-duration ($M = 2291$ ms, $SD = 661$ ms) than the medium-duration ($M = 1629$ ms, $SD = 425$ ms) “chee” (YF-treated) response (bottom left panel). This pattern of average ST was reliable across every stage of the experiment (i.e., Baseline, Phase I, Phase II). It is also of interest to note that in baseline, ST was related to target duration such that AOS1 displayed the shortest average STs for target 1S ($M = 1513$ ms, $SD = 274$ ms), followed by 1M ($M = 1838$ ms, $SD = 267$ ms), then 1L ($M = 2333$ ms, $SD = 491$ ms), and finally 1X ($M = 2737$ ms, $SD = 732$ ms), although this pattern fluctuated as different treatments were applied.

A further expectation for ST, given the findings of Experiment 3 and those of others using the self-selection paradigm (e.g., Immink & Wright, 2001; Mass, Robin, Wright et al., in press; Wright et al., 2004; Magnuson et al., in press), was that multi-element or multi-unit responses would incur higher INT processing costs than single syllables. On average, this was not the case, as numeric differences between mean STs indicated a quite different pattern. For the short and long “pa” responses (Figure 7-6, top left panel), the 1L target response had longer average ST ($M = 2061$ ms, $SD = 444$ ms) than either of the multi-syllable responses (2S $M = 2025$ ms, $SD = 679$ ms; 4L $M = 1658$ ms, $SD = 461$ ms) in each stage of the experiment, except for baseline, when
average ST was longer for the 2S response. For the medium and extra-long “chee” responses (Figure 7-6, bottom left panel), the 1X target response had slightly longer average ST \((M = 2291 \text{ ms}, SD = 661 \text{ ms})\) than the 4M response \((M = 2171 \text{ ms}, SD = 1092 \text{ ms})\), and the 2X response had slightly longer average ST than all of the others \((M = 2466 \text{ ms}, SD = 1087 \text{ ms})\), although there was quite a bit of fluctuation in these patterns across individual sessions. STs for 2S “pa” and 4M “chee” responses appear to have decreased during their respective treatment phases, however, a systematic change in ST across sessions was not discernible for the other multi-element responses.

**Reaction Time (SEQ process):** For RT, it was predicted, based on the INT/SEQ model, that increasing the number of elements (here, syllables) in a response would increase the time required to initiate the movement unless those elements were integrated into a single unit. There was a difference in average RT between one- and two-syllable responses (i.e., 2 > 1), but it was greater (and more reliable across sessions) for “pa” (LCF) \((173 \text{ ms}; \text{Figure 7-6, top right panel})\) than for “chee” (YF) \((11 \text{ ms}; \text{bottom right panel})\) responses. Average RT for two-syllable responses was also longer than for four-syllable responses in the “pa” (LCF) condition \((-76 \text{ ms sequence length effect})\), a pattern which emerged after the fourth baseline. In the “chee” (YF) condition (bottom right panel), RT was on average 93 ms longer for four-syllable responses than two-syllable responses, and this pattern was observed at each stage of the experiment (i.e., Baseline, Phase I, Phase II). These results were likely attributable to overall longer RTs for two-syllable “pa” responses \((M = 628 \text{ ms}, SD =\)
than the two-syllable “chee” responses ($M = 457 \text{ ms}, SD = 66 \text{ ms}$). Interestingly, the steep increase in RT for the 2S “pa” response following the withdrawal of Phase I treatment coincided only with slightly decreasing ST, AE, and RE. For the Phase I LCF condition, the average sequence length effect for one- versus four-syllable responses decreased with the application of LCF treatment (from 145 ms to 88 ms) and continued to decrease after the withdrawal of treatment (from 88 ms to 54 ms). For the Phase II YF “chee” condition, on the other hand, this effect increased with the application of LCF treatment to “pa” (from 37 ms to 129 ms), and increased again slightly with the application of YF treatment to “chee” (from 129 ms to 150 ms).
Figure 7-6. Participant AOS1 study time (ST, left panels) and reaction time (RT, right panels) for treated responses elicited as random-order transfer probes during baseline, treatment, and retention phases. Top panels = Phase I targets (top legend); Bottom panels = Phase II targets (bottom legend). See Table 7-3 for stimulus cue definitions. B = Baseline, Tx = Treatment, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Summary: AOS1

To summarize, there was no difference in any dependent measure between feedback and no-feedback trials for AOS1, although this participant thought that he received feedback on “good” trials during YF treatment, and that YF treatment helped him “a lot” whereas LCF treatment helped him “somewhat.” During acquisition, LCF treatment was associated with slight improvement in AE, whereas YF treatment was linked to improvement in RE. The random transfer probes indexed greater transfer related to LCF treatment for “pa,” however, the “chee” baseline was influenced either by “pa” treatment or an independent improvement trend. Differences in blocked retention probes between the two feedback conditions were likely related to baseline inequalities. In programming responses during random transfer probes, AOS1 demonstrated a complexity (duration) effect of single syllable responses on average ST during each phase and no reliable effect of sequence length on ST, although ST for the 2S and 4M responses did appear to shorten during their respective treatment phases. The sequence length effect on RT for four-syllable versus one-syllable responses decreased over time for LCF-treated “pa” responses, but increased over time for YF-treated “chee” responses.

7.2.2 AOS2

Response execution: AE and RE

Participant AOS2 was able to reliably differentiate the four different response durations in his productions during acquisition and probes, as shown in Figures 7-7.
and 7-8, respectively. Absolute and relative error data relating to these productions are displayed in Figures 7-9 and 7-10, respectively. Each data point in Figures 7-9 and 7-10 represents average performance for all response types (e.g., 1S, 1L, 2S, 4L) administered in each particular task. Because AOS2 had extreme difficulty producing the “chee” syllable, the syllable “tee” was used instead.

**Baseline:** Like AOS1, AOS2 also did not demonstrate very stable baseline error rates for either set (“pa” or “tee”) of target behaviors (see Figures 7-9 and 7-10, first four sessions); however a clear unidirectional trend was also not apparent. Phase I treatment commenced after AOS2 had completed four baseline sessions. Despite the initial instability, treatment effects in the random transfer probe and blocked retention probe were observed and measured with the effect size statistic, accounting for baseline variability.

**Feedback during acquisition:** Participant AOS2 received treatment conditions in the opposite order as AOS1, although YF was still paired with medium- and extra long-duration responses (now Phase I) and LCF was still paired with short- and long-duration responses (now Phase II). AOS2 requested feedback on 30% of acceptable trials during Phase II LCF treatment, and during Phase I YF treatment, received feedback on trials that were yoked to AOS1’s LCF-condition feedback requests. During Phase I YF treatment, there was no difference between feedback and no-feedback trials in AE (feedback: \( M = 115 \text{ ms}, SD = 117 \text{ ms} \); no-feedback: \( M = 118 \text{ ms}, SD = 117 \text{ ms} \)) or RE (feedback \( M = 0.352, SD = 0.172 \); no-feedback: \( M = 0.353; SD = 0.170 \)). Likewise, during Phase II LCF treatment there was no difference between
feedback and no-feedback trials in RE (feedback $M = 0.207$, $SD = 0.129$; no-feedback: $M = 0.202$; $SD = 0.112$), although an apparent slight difference in AE was noted (feedback: $M = 134$ ms, $SD = 155$ ms; no-feedback: $M = 104$ ms, $SD = 104$ ms). Participant AOS2 gave the same answers to questions about his feedback schedules after treatment in both conditions. That is, he reported that he requested (or thought he received) feedback on a random selection of trials, he did not request (or feel that he received) feedback on “neither good nor bad” trials, and he thought that the feedback helped him improve his durations “somewhat” in both conditions.

**Acquisition:** Performance during acquisition was assessed by comparing the slopes of Phase I and Phase II treatment session data (see open symbols in top and bottom panels of Figures 7-9 and 7-10). Absolute error decreased more reliably in Phase I YF treatment (slope = -4.23, $r^2 = 0.44$) than in Phase II LCF treatment (slope = -2.47, $r^2 = 0.07$). A similar pattern was found for Relative Error, which decreased reliably in Phase I YF treatment (slope = -1.77, $r^2 = 0.68$), but not in Phase II LCF treatment (slope = 0.23, $r^2 = 0.06$).
Retention: Retention of responses trained under blocked practice conditions was examined by comparing average performance on the blocked retention probe (see filled gray symbols in Figures 7-9 and 7-10) to overall practice performance. Relative to practice performance, AOS2 demonstrated more (46%) AE in blocked retention during Phase I YF treatment, and less (-17%) AE in blocked retention during Phase II LCF treatment, although baseline performance was slightly better in the LCF condition. There was less relative timing error (RE) in blocked retention probes than
in practice performance, but the difference between the Phase I YF condition (-21\%) and the Phase II LCF condition (-16\%) was negligible when average performance was compared.

In addition, because AOS2 completed sufficient baseline probes of the blocked retention set, it was possible to examine effect sizes comparing performance on this probe pre-treatment, during the treatment phase, and after the withdrawal of treatment. There were very large treatment effects for blocked retention probes of both sets of behaviors, and the effect size of Phase II LCF treatment on AE ($d_{1T} = 18.81$) was double the effect size of Phase I YF treatment ($d_{1T} = 9.17$). The same pattern was demonstrated for AE for the single 4-day delayed retention test (LCF $d_{1W} = 20.10$; YF $d_{1W} = 9.50$). Relative error data indicated the opposite pattern, as Phase I YF treatment had a large effect on blocked retention probes ($d_{1T} = 9.47$, $d_{1W} = 6.91$), whereas Phase II LCF treatment had only a very small effect ($d_{1T} = 1.52$, $d_{1W} = 1.63$).

Transfer: Transfer of the target responses trained in blocked order to a random-order probe task was assessed across all phases of the experiment for both stimulus sets. During the first treatment phase (“pa”), there were small effects of YF treatment on AE ($d_{1T} = 2.64$; Figure 7-9, top panel) and RE ($d_{1T} = 2.43$; Figure 7-10, top panel) for the random transfer probes in that condition. However, there were also medium and small effects of Phase I YF treatment (“pa”) on AE ($d_{1P} = 3.89$; Figure 7-9, bottom panel) and RE ($d_{1P} = 2.13$; Figure 7-10, bottom panel), respectively, for the untrained “tee” stimuli that were to be held in baseline until Phase II treatment began, demonstrating that training medium- and extra long-duration “pa” syllables and
patterns with YF treatment may have influenced performance on short- and long-duration “tee” syllables and patterns that had not yet been treated. This result did not appear to be consistent with any trends observable during baseline (see Figures 7-9 and 7-10, bottom panels, sessions B2-B4), further suggesting that these behaviors were not independent and that experimental control was compromised for AOS2. There was nevertheless a small measurable effect of Phase II LCF treatment on AE ($d_{IT} = 3.51$; Figure 7-9, bottom panel, sessions P6-P9), although the effect on RE ($d_{IT} = 0.85$; Figure 7-10, bottom panel, sessions P6-P9) for “tee” responses was negligible. For AE, despite the small improvements made before Phase II LCF treatment began, AOS2 demonstrated a larger effect of Phase II LCF treatment on “tee” responses ($d_{IT} = 3.51$) than of Phase I YF treatment on “pa” responses ($d_{IT} = 2.64$) in random transfer probes. The opposite was true for relative error (LCF $d_{IT} = 0.85$, YF $d_{IT} = 2.43$).

To measure the persistence of these small treatment effects shortly after practice conditions were withdrawn, performance on the 4-day delayed probe (Figures 7-9 and 7-10, sessions P4 and P9) was compared to baselines for the random transfer probe for each behavior set. For AE, this comparison indicated small effects with a slight advantage of Phase II LCF treatment on random-order probes ($d_{IW} = 3.73$) compared to Phase I YF treatment ($d_{IW} = 2.83$). For RE, negligible effect sizes were obtained for both Phase I (YF) ($d_{IW} = 0.85$) and Phase II (LCF) ($d_{IW} = 0.96$). The continued collection of “pa” random transfer probes throughout Phase II LCF treatment (Figures 7-9 and 7-10, top panels, sessions P6-P9) demonstrated that the Phase I YF treatment gains with “pa” were well-maintained once treatment for “tee”
began, as effect sizes remained stable for AE ($d_{1M} = 2.80$, compared to $d_{1T} = 2.64$) and RE ($d_{1M} = 2.32$, compared to $d_{1T} = 2.43$).
Figure 7-8. Stimuli target durations (in ms, solid lines) and actual response durations (in ms, broken lines) for AOS2 for random transfer probes during each probe session (Top panel: “pa” responses; Bottom panel: “tee” responses). See Table 7-3 for stimulus cue definitions. LCF = Learner-Controlled Feedback; YF = Yoked Feedback Tx = treatment.
Figure 7-9. Participant AOS2 average absolute error (AE) for responses during treatment sessions and retention and transfer probes. Top panel = Phase I targets; Bottom panel = Phase II targets. B = Baseline, T = Treatment, P = probe, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Figure 7-10. Participant AOS2 average relative error (RE) for responses during treatment sessions and retention and transfer probes. Top panel = Phase I targets; Bottom panel = Phase II targets. B = Baseline, T = Treatment, P = probe, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Motor Programming

To assess the effects of the different feedback conditions on pre-movement motor programming in AOS2, study times (STs; Figure 7-11, left panels) and reaction times (RTs; Figure 7-11, right panels) for one-, two-, and four-syllable responses were examined within the context of random-order probes across all phases of the experiment. Average ST and RT decreased for “pa” and “tee” responses across all phases of the experiment, except for ST for the 4M “pa” response, which increased after the withdrawal of Phase I YF treatment. Error measures for “pa” response execution also stopped improving after the withdrawal of Phase I YF treatment.

During LCF treatment, ST was slightly longer for feedback than no-feedback trials (feedback: $M = 5264$ ms, $SD = 2073$ ms; no-feedback: $M = 4025$ ms, $SD = 1543$ ms); however a smaller difference was apparent for RT (feedback $M = 350$ ms, $SD = 149$ ms; no-feedback: $M = 305$ ms; $SD = 179$ ms). During YF treatment (in which AOS3 did not request feedback at all), as would be expected, there was little difference between feedback and no-feedback trials in ST (feedback: $M = 5369$ ms, $SD = 3006$ ms; no-feedback: $M = 5408$ ms, $SD = 2535$ ms) and RT (feedback $M = 641$ ms, $SD = 280$ ms; no-feedback: $M = 734$ ms; $SD = 434$ ms).

Study Time (INT process) Based on the INT/SEQ model, ST was expected to increase as the target durations of the single-syllable responses lengthened. As shown in Figure 7-11 (left panels), on average, there was no appreciable difference in average ST between the extra-long-duration ($M = 4545$ ms, $SD = 1583$ ms) and the medium-duration ($M = 4508$ ms, $SD = 1254$ ms) “pa” (YF-treated) responses (top left panel),
except for during baseline, when ST for 1X was longer than for 1M by an average of 605 ms. For “tee” (LCF-treated) responses, STs for long syllables ($M = 4728$ ms, $SD = 1479$ ms) were on average 959 ms longer than STs for short syllables ($M = 3769$ ms, $SD = 1283$ ms). Across individual sessions, the duration effect was slightly more reliable for “tee” responses, although it did fluctuate in both (i.e., “pa” and “tee”) response sets. It is worth noting that like AOS1, in baseline, AOS2’s STs were related to target durations such that the shortest average ST was observed for target 1S ($M = 5103$ ms, $SD = 1669$ ms), followed by 1M ($M = 5684$ ms, $SD = 1156$ ms), then 1L ($M = 5942$ ms, $SD = 552$ ms), and finally 1X ($M = 6288$ ms, $SD = 526$ ms), although this pattern fluctuated as different treatments were applied.

A further expectation for ST was that multi-element or multi-unit responses would incur longer STs than single syllables. This pattern was generally confirmed in that, on average, two- and four-syllable patterns had longer STs than one-syllable responses across “pa” and “tee” responses in each stage of the experiment, with the exception of the 4L “tee” response having a slightly shorter ST ($M = 3760$ ms, $SD = 1605$ ms) than the 1L “tee” response ($M = 3835$ ms, $SD = 1266$ ms) during Phase II (LCF) treatment. While all of the sequence length differences (2 vs. 1, 4 vs. 2, and 4 vs. 1) attenuated across time for the “tee” (LCF) condition, the opposite was true for the 4 versus 2 and the 4 versus 1 comparisons in the “pa” (YF) condition. That is, the discrepancy between four- and two-syllable “pa” patterns changed directions and grew from baseline (-3492 ms; 2 > 4) to Phase I (YF) treatment (-507 ms) to the withdrawal of treatment (3048 ms; 4 > 2), and the ST discrepancy between four- and one-syllable
“pa” responses grew from Phase I (YF) treatment (1668 ms) to the post-treatment phase (4964 ms). This finding can be accounted for by the substantial reduction in ST for the 2X response coincident with Phase I YF treatment and the substantial increase in ST for the 4M response following the withdrawal of this treatment.

**Reaction Time (SEQ process):** Based on the INT/SEQ model, it was expected that increasing the number of syllables in a response would increase RT if the syllables were programmed as separate units. The only indication of this pattern for AOS2 was observed during baseline, when the polysyllabic responses had on average longer RTs (by 98 to 168 ms) than the monosyllabic responses, although the 1L and 1X RTs fluctuated. There was not an appreciable difference in RTs for two- versus four- syllable responses in baseline. After baseline, a systematic relationship in RT between polysyllabic and monosyllabic responses was not found, and in many cases, under both feedback conditions, two- and/or four-syllable responses were initiated on average slightly faster than one-syllable responses.
Figure 7-11. Participant AOS2 study time (ST, left panels) and reaction time (RT, right panels) for treated responses elicited as random-order probes during baseline and each treatment phase. Top panels = Phase I targets (top legend); Bottom panels = Phase II targets (bottom legend). See Table 7-3 for stimulus cue definitions. B = Baseline, Tx = Treatment, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Summary: AOS2

To summarize, AOS2 exhibited longer ST and greater AE on trials for which he requested feedback, although he did not report noticing a difference between feedback and no-feedback trials in either treatment condition. During acquisition, YF treatment was associated with improvement in both AE and RE, whereas LCF treatment was not. The blocked retention and random transfer probes indexed greater retention and transfer of absolute timing in the LCF condition, and greater retention and transfer of relative timing in the YF condition. The “tee” (LCF) baseline was influenced by “pa” (YF) treatment. In programming responses during random transfer probes, AOS2 demonstrated a slightly more-reliable complexity (duration) effect on average ST for single syllable “tee” responses, than for “pa” responses. A complexity effect was also evident in RT for both sets of responses. Sequence length effects on RT were only evident during baseline; however, they were observed on ST across all phases.

7.2.3 AOS3

Response execution: AE and RE

Participant AOS3 was able to reliably differentiate the four different response durations in her productions during acquisition and probes, as shown in Figures 7-12 and 7-13, respectively. Average absolute and relative error data relating to these productions are displayed in Figures 7-14 and 7-15, respectively. Each data point in Figures 7-14 and 7-15 represents average performance for all response types (e.g., 1S,
1L, 2S, 4L) administered in each particular task. Because AOS3 could not produce the “pa” syllable, the syllable “la” was used instead.

Baseline: Like the previous participants, AOS3 also did not demonstrate very stable baseline performance on either set (“la” or “chee”) of target behaviors (see Figures 7-14 and 7-15, first four sessions); however a clear linear trend was also not apparent, except in the case of RE for “chee” random transfer probes. Phase I treatment commenced after four baseline sessions. Despite the initial instability, treatment effects were observed and measured with the effect size statistic, accounting for baseline variability.

Feedback during acquisition: Like AOS2, participant AOS3 was given treatment conditions in the opposite order as AOS1, but unlike AOS2, YF was paired with short- and long-duration responses (Phase I) and LCF was paired with medium- and extra-long-duration responses (Phase II). AOS3 requested feedback on 30% of acceptable trials during Phase II LCF treatment, and during Phase I YF treatment, received feedback on trials that were yoked to AOS1’s LCF-condition feedback requests. During Phase I YF treatment, there was no substantial difference between feedback and no-feedback trials in AE (feedback: $M = 326 \text{ ms}, SD = 452 \text{ ms}$; no-feedback: $M = 285 \text{ ms}, SD = 405 \text{ ms}$) or RE (feedback $M = 0.353, SD = 0.116$; no-feedback: $M = 0.353; SD = 0.104$). Likewise, during Phase II LCF treatment there was no appreciable difference between feedback and no-feedback trials in AE (feedback: $M = 320 \text{ ms}, SD = 380 \text{ ms}$; no-feedback: $M = 342 \text{ ms}, SD = 434 \text{ ms}$) or RE (feedback $M = 0.277, SD = 0.134$; no-feedback: $M = 0.278; SD = 0.137$). After Phase I YF
treatment, AOS3 reported that she thought that she had received feedback on good and bad trials equally, and had *not* received feedback on good trials. However, after Phase II LCF treatment, she reported that she requested feedback on good and bad trials equally, and chose not to request feedback on “neither good nor bad” trials. AOS3 thought that both feedback conditions helped her improve her durations “somewhat.”

**Acquisition:** Performance during acquisition was assessed by comparing the slopes of Phase I and Phase II treatment session data (see open symbols in Figure 7-14 and 7-15, top and bottom panels, respectively). Phase I YF treatment was strongly associated with a decrease in AE (slope = -31.07, $r^2 = 0.94$), whereas no reliable change was evident for Phase II LCF treatment (slope = -3.92, $r^2 = 0.002$). Conversely, Phase I YF treatment did not appear to influence relative error (slope = -0.05, $r^2 = 0.03$), while Phase II LCF treatment was actually associated with an increase in RE (slope = 1.05, $r^2 = 0.40$).
Retention: Retention of responses trained under blocked practice conditions was examined by comparing average performance on the blocked retention probe (see filled gray symbols in Figures 7-14 and 7-15) to overall practice performance. Relative to practice performance, AOS3 demonstrated less AE in blocked retention, and the difference was similar for Phase I YF treatment (-17%), and Phase II LCF treatment (-20%). AOS3 demonstrated slightly more (3%) relative timing error (RE) in blocked retention probes than in practice performance for the Phase I YF treatment behaviors.

Figure 7-12. Stimuli target durations (in ms, solid lines) and actual response durations (in ms, broken lines) for AOS3 for each practice block during acquisition (Phase I on left; Phase II on right). See Table 7-3 for stimulus cue definitions. Session labels X.1-X.4 represent the four practice blocks within the X practice session. LCF = Learner-Controlled Feedback; YF = Yoked Feedback.
but less (-13%) RE for blocked retention probes in the Phase II LCF condition. Caution should be applied in interpreting these results, as baselines across the two phases were not comparable.

**Transfer:**

Transfer of the target responses trained in blocked order to a random-order probe task was assessed across all phases of the experiment for both stimulus sets (see filled black symbols in Figures 7-14 and 7-15). During the first treatment phase (“la”), there was a very small positive effect of YF treatment on AE ($d_{1T} = 1.32$) and a negligible negative effect on RE ($d_{1T} = -0.76$). However, Phase I YF treatment of short- and long-duration “la” targets had even greater positive effects on untreated medium- and extra long-duration “chee” targets as measured by AE ($d_{1P} = 4.52$) and RE ($d_{1P} = 1.01$), which were to be held in baseline until Phase II treatment began (Figures 7-14 and 7-15, bottom panels, P1-P4). This demonstrated that the YF treatment of short and long “la” syllables may have influenced the learning of medium and extra-long “chee” syllables that had not yet been treated. Another possibility is that the error reduction observed in “chee” responses during Phase I treatment for “la” was simply an extension of the unstable baseline for “chee” responses. This seems plausible for RE, given the slightly decreasing trend during baseline, but it seems less likely that the medium-sized effect ($d_{1P} = 4.52$) on AE was due to its unstable baseline which did not show a decreasing trend. Thus, it is probable that experimental control was compromised by a lack of independence of the two sets of behaviors under Phase I YF treatment. There was no substantial effect of Phase II LCF treatment on AE ($d_{1T}$...
= 0.05) for “chee” random transfer responses (Figure 7-14, bottom panel, P6-P9), and RE for these probes increased very slightly with treatment, compared to pre-treatment probes ($d_{IT} = -0.15$) (Figure 7-15, bottom panel, P6-P9). Thus, for AE, AOS3 demonstrated a negligible effect of Phase II LCF treatment ($d_{IT} = 0.05$) and a very small effect of Phase I YF treatment ($d_{IT} = 1.32$) on random transfer probes, and for RE, AOS3 demonstrated negligible negative effects during both treatment phases (LCF $d_{IT} = -0.15$; YF $d_{IT} = -0.76$). Only AE for the “chee” responses underwent considerable change, and only when the “pa” responses were treated ($d_{IP} = 4.52$).

To measure the persistence of these treatment effects shortly after practice conditions were withdrawn, performance on the 4-day delayed probe (Figures 7-14 and 7-15, top panels session P5; bottom panels session P9) was compared to baselines for the random transfer probe for each behavior set. For AE, this comparison indicated a very small positive effect of Phase I YF treatment on random-order probes ($d_{IW} = 1.60$) and a negligible negative effect of Phase II LCF treatment ($d_{IW} = -.20$). For RE, very small or negligible negative effect sizes were obtained for both Phase I (YF) ($d_{IW} = -1.16$) and Phase II (LCF) ($d_{IW} = -0.26$), which, combined with the treatment phase random probe data, suggested that AOS3 had slightly greater relative error for both sets of behaviors during and after treatment than she did before treatment began. The continued collection of “pa” random transfer probes throughout Phase II LCF treatment (Figures 7-14 and 7-15, top panels, sessions P6-P9) demonstrated that the very small Phase I YF treatment gains for AE were maintained once treatment for “chee” began; as effect sizes remained stable ($d_{IM} = 1.34$, compared to $d_{IT} = 1.32$),
however, relative timing performance continued to worsen ($d_{IM} = -1.25$, compared to $d_{IT} = -0.76$).
Figure 7-13. Stimuli target durations (in ms, solid lines) and actual response durations (in ms, broken lines) for AOS3 for each probe session (Top panel: “la” responses; Bottom panel: “chee” responses). See Table 7-3 for stimulus cue definitions. LCF = Learner-Controlled Feedback; YF = Yoked Feedback Tx = treatment.
Figure 7-14. Participant AOS3 average absolute error (AE) for responses during treatment sessions and retention and transfer probes. Top panel = Phase I targets; Bottom panel = Phase II targets. B = Baseline, T = Treatment, P = probe, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Figure 7-15. Participant AOS3 average relative error (RE) for responses during treatment sessions and retention and transfer probes. Top panel = Phase I targets; Bottom panel = Phase II targets. B = Baseline, T = Treatment, P = probe, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Motor Programming

To assess the effects of the different feedback conditions on pre-movement motor programming in AOS3, study times (STs) and reaction times (RTs) for one-, two-, and four-syllable responses were examined within the context of random-order probes across all phases of the experiment (see Figure 7-16). Average ST (left panels) and RT (right panels) decreased for “la” responses only between Baseline and Phase I (YF) treatment of those responses. For “chee” responses that were subjected to LCF treatment in Phase II, ST and RT showed a decreasing trend between each stage of the experiment.

ST and RT did not appear to be related to whether or not AOS3 requested feedback on a given trial. During YF treatment (in which the AOS3 did not request feedback at all), there was little difference between feedback and no-feedback trials in ST (feedback: $M = 1824$ ms, $SD = 1869$ ms; no-feedback: $M = 2000$ ms, $SD = 1605$ ms), and RT (feedback $M = 188$ ms, $SD = 71$ ms; no-feedback: $M = 179$ ms; $SD = 44$ ms). During LCF treatment, there were also no appreciable differences between feedback and no-feedback trials in ST (feedback: $M = 2335$ ms, $SD = 2438$ ms; no-feedback: $M = 2302$ ms, $SD = 1804$ ms) or RT (feedback $M = 179$ ms, $SD = 73$ ms; no-feedback: $M = 182$ ms; $SD = 78$ ms).

Study Time (INT process): Based on the INT/SEQ model, ST was expected to increase as the target durations of the single-syllable responses lengthened. As can be seen in Figure 7-16 (left panels), on average, this was not the case. In fact, in all but the baseline stage for “chee” responses, shorter syllables on average were associated
with longer STs than longer syllables. It did not appear that this reversal of expected ST differences was directly related to INT processing effectiveness, as there was not an appreciable difference in absolute timing error between the two targets.

A further expectation for ST was that polysyllabic responses would incur longer STs than single syllables. While this was generally the case on average for the “chee” responses and some of the “la” responses, the pattern was far from consistent across sessions (Figure 7-16, left panels), and further, the 4L “la” response actually had shorter average STs during baseline than any of the other “la” targets. Interestingly, when the two-syllable responses were compared to the 4-syllable responses, average STs were longer for the two-syllable responses in both conditions until the responses were treated (Phase I YF for “pa” and Phase II LCF for “chee”), at which point the relative STs reversed to the predicted direction (i.e., 4 > 2).

**Reaction Time (SEQ process):** Based on the INT/SEQ model, it was predicted that increasing the number of syllables in a response would increase RT if those syllables are programmed as separate units. This pattern was only observed for the “la” responses in the first baseline session, and for the “chee” responses during Phase II LCF treatment. Interestingly, when RT for the 2S “la” response decreased dramatically by 65% (376 ms) after the first baseline (Figure 7-16, top right panel, sessions B1 and B2), a coincident 124% (4683 ms) increase was observed in ST for that response (top left panel, same sessions). Moreover, a large decrease in AE (from 1295 ms to 384 ms) and a large increase in RE (from 0.19 to 0.34) also coincided with these ST and RT shifts. The increase in ST appeared to be associated with faster
response initiation and reduced absolute error. Although it is possible to interpret a
ST/RT pattern such as this as evidence of multi-unit integration, given the coincident
increase in relative error, the transitory nature of the ST increase, and the lack of a
systematic relationship between ST and RT for the remainder of the experiment and
for other responses, that is not likely an appropriate interpretation here. It is possible
that for this one point in time, AOS3 tried to program the entire sequence as a single
unit during INT, but was not successful in executing a motor program with the correct
relative pattern.
Figure 7-16. Participant AOS3 study time (ST, left panels) and reaction time (RT, right panels) for treated responses elicited as random-order probes during baseline, treatment, and retention phases. Top panels = Phase I targets (top legend); Bottom panels = Phase II targets (bottom legend). See Table 7-3 for stimulus cue definitions. B = Baseline, Tx = Treatment, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Summary: AOS3

To summarize, AOS3 did not exhibit a difference between feedback and no-feedback trials for any dependent variables in either treatment condition, although she thought that she did not receive feedback on relatively good trials during YF treatment. During acquisition, YF treatment was strongly associated with improvement in AE and was not linked to a reliable change in RE. LCF treatment did not appear to affect acquisition AE, but it was associated with worsening RE. The random transfer probes indexed a very small advantage of YF treatment to reduction in absolute timing error and a very small advantage of LCF treatment to minimization of increases in relative timing error upon treatment withdrawal. Blocked retention probe performance did not yield any reliable insights about differences between conditions. Baseline random transfer probes for “chee” were more affected by Phase I YF treatment for “la” than “la” probes were, although likely influenced by differences in true baseline. In programming responses during random transfer probes, AOS3 did not demonstrate a reliable complexity (duration) effect in ST or RT. Sequence length effects were more evident, although not completely consistent, in ST than in RT.

7.2.4 AOS4

Response execution: AE and RE

Participant AOS4 was able to reliably differentiate the four different response durations in his productions during acquisition (Figure 7-17) and with a few exceptions during probes (Figure 7-18). Average absolute and relative error data
relating to these productions are displayed in Figures 7-19 and 7-20, respectively. Each data point in Figures 7-19 and 7-20 represents average performance for all response types (e.g., 1S, 1L, 2S, 4L) administered in each particular task.

**Baseline:** Like the other participants, AOS4 also did not demonstrate very stable baseline performance on either set (i.e., “pa” or “chee”) of target behaviors (see Figures 7-19 and 7-20, first four sessions); however a clear linear trend was also not apparent, except in the case of RE for “chee” random transfer probes (Figure 7-20, bottom panel, sessions B1-B4), which increased across baseline sessions. Phase I treatment commenced after four baseline sessions. Despite the initial instability, treatment effects in the random transfer probe were observed and measured with the effect size statistic, taking into account baseline variability.

**Feedback during acquisition:** AOS4 was given treatment conditions in the same order as AOS1, but LCF was paired with medium- and extra-long-duration responses (Phase I) and YF was paired with short- and long-duration responses (Phase II). The syllables were trained in the same order, that is, with “pa” in Phase I and “chee” in Phase II. AOS4 requested feedback on 30% of acceptable trials during LCF treatment (Phase I) and during YF treatment (Phase II), received feedback on trials that were yoked to AOS1’s Phase I feedback requests. During LCF treatment, there was no difference between feedback and no-feedback trials in AE (feedback: $M = 142$ ms, $SD = 143$ ms; no-feedback: $M = 142$ ms, $SD = 174$ ms) or RE (feedback $M = 0.159$, $SD = 0.068$; no-feedback: $M = 0.165$; $SD = 0.076$). Likewise, during YF treatment there was no appreciable difference between feedback and no-feedback
trials in AE (feedback: $M = 74$ ms, $SD = 68$ ms; no-feedback: $M = 80$ ms, $SD = 48$ ms) or RE (feedback $M = 0.146$, $SD = 0.075$; no-feedback: $M = 0.144$; $SD = 0.064$). Despite this, AOS4 reported that during Phase I LCF treatment, he had requested feedback mostly on good trials, and had chosen not to request feedback on “neither good nor bad” trials. He thought that the feedback given in the LCF condition helped him improve his durations “a lot.” His responses were the same for Phase II YF treatment, except that he reported receiving feedback on good and bad trials equally.

**Acquisition:** Performance during acquisition was assessed by comparing the slopes of Phase I and Phase II treatment session data (see open symbols in Figure 7-19 and 7-20, top and bottom panels, respectively). Absolute error did not reliably decrease during Phase I LCF treatment (slope = -2.39, $r^2 = 0.09$) and was weakly associated with a small decrease during Phase II YF treatment (slope = -1.84, $r^2 = 0.26$). Relative Error did not decrease appreciably in Phase I LCF treatment (slope = -0.28, $r^2 = 0.23$), but increased slightly with Phase II YF treatment (slope = 0.32, $r^2 = 0.38$).
Retention: Retention of responses trained under blocked practice conditions was examined by comparing average performance on the blocked retention probe (see filled gray symbols in Figures 7-19 and 7-20) to overall practice performance. Relative to practice performance, AOS4 demonstrated more AE in blocked retention, and the differences were similar for Phase I (LCF) (38%), and Phase II (YF) (42%). AOS4 also demonstrated more relative timing error (RE) in blocked retention probes than in

Figure 7-17. Stimuli target durations (in ms, solid lines) and actual response durations (in ms, broken lines) for AOS4 for each practice block during acquisition (Phase I on left; Phase II on right). See Table 7-3 for stimulus cue definitions. Session labels X.1-X.4 represent the four practice blocks within the X practice session. LCF = Learner-Controlled Feedback; YF = Yoked Feedback.
practice performance, with a greater treatment-to-probe disparity for the Phase II YF treatment behaviors (93%), than the Phase I LCF condition (48%).

**Transfer:** Transfer of the target responses trained in blocked order to a random-order probe task was assessed across all phases of the experiment for both stimulus sets (see filled black symbols in Figures 7-19 and 7-20). During the first treatment phase ("pa"), there was a small positive effect of LCF treatment on AE ($d_{IT} = 2.34$) and a negligible negative effect on RE ($d_{IT} = -0.32$). However, Phase I LCF treatment of medium- and extra-long-duration "pa" targets also appeared to have had a small positive effect on AE ($d_{IP} = 2.03$) and a medium-sized negative effect on RE ($d_{IP} = -3.91$) for untreated short- and long-duration "chee" random probe targets, which were to be held in baseline until Phase II treatment began (Figures 7-19 and 7-20, bottom panels, P1-P4). This suggested that the LCF treatment of medium and extra-long "pa" syllables might have influenced the learning of short and long "chee" syllables that had not yet been treated. Another possibility was that, particularly in the case of RE, the "pre-treatment" effect reflected a continuation of an unstable and increasing baseline. This explanation is slightly less tenable for AE, which was less consistent in baseline, and thus it is necessary to consider that these behavior sets were not independent under Phase I LCF treatment, and experimental control was compromised for AOS4. Despite this limitation, there was nonetheless a measurable a medium-sized effect of Phase II YF treatment on AE ($d_{IT} = 5.10$) for "chee" random transfer responses (Figure 7-19, bottom panel, P5-P8). RE for "chee" responses was unchanged with the application of Phase I YF treatment ($d_{IT} = 0.04$) (Figure 7-20,
bottom panel, P5-P8). Thus, for AE, AOS4 demonstrated a smaller effect of Phase I LCF treatment ($d_{IT} = 2.34$) than of Phase II YF treatment ($d_{IT} = 5.10$) on random transfer probes, and for RE, AOS4 demonstrated negligible treatment effects for random transfer probes during treatment phases (LCF $d_{IT} = -0.32$; YF $d_{IT} = 0.04$).

To measure the persistence of these treatment effects shortly after practice conditions were withdrawn, the 4-day delayed probe (Figures 7-19 and 7-20, top panels session P4; bottom panels session P8) was compared to baselines for the random transfer probe for each behavior set. For AE, this comparison indicated a small positive effect of Phase I LCF treatment on random-order probes ($d_{IW} = 2.80$) and a medium-sized effect of Phase II YF treatment ($d_{IW} = 4.69$). For RE, a negligible negative effect size was obtained for Phase I (LCF) ($d_{IW} = -0.24$) and a negligible positive effect size was obtained for Phase II (YF) ($d_{IW} = 0.27$). The continued collection of “pa” random transfer probes throughout Phase II YF treatment (Figures 7-19 and 7-20, top panels, sessions P5-P8) demonstrated that the small Phase I LCF treatment gains for AE were well-maintained once treatment for “chee” began, as the effect size remained stable ($d_{IM} = 2.61$, compared to $d_{IT} = 2.34$), and relative timing performance did not change considerably ($d_{IM} = 0.28$, compared to $d_{IT} = -0.32$).
Figure 7-18. Stimuli target durations (in ms, solid lines) and actual response durations (in ms, broken lines) for AOS4 for each random transfer probe (Top panel: “pa” responses; Bottom panel: “chee” responses). See Table 7-3 for stimulus cue definitions. LCF = Learner-Controlled Feedback; YF = Yoked Feedback Tx = treatment.
Figure 7-19. Participant AOS4 average absolute error (AE) for responses during treatment sessions and retention and transfer probes. Top panel = Phase I targets; Bottom panel = Phase II targets. B = Baseline, T = Treatment, P = probe, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Figure 7-20. Participant AOS4 average relative error (RE) for responses during treatment sessions and retention and transfer probes. Top panel = Phase I targets; Bottom panel = Phase II targets. B = Baseline, T = Treatment, P = probe, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Motor Programming

To assess the effects of the different feedback conditions on pre-movement motor programming in AOS4, study times (STs) and reaction times (RTs) for one-, two-, and four-syllable responses were examined within the context of random-order probes across all phases of the experiment (see Figure 7-21). Although fluctuations were observed across sessions, average STs for one-syllable “pa” responses decreased across all stages of the experiment, whereas STs for multi-syllable “pa” responses decreased with Phase I LCF treatment, but then increased upon the withdrawal of treatment. Average STs for all of the “chee” responses decreased across the experiment, with the exception of ST for the 2S response, which was at its longest when it was undergoing treatment during the Phase II YF treatment stage. These results were not coincident with any changes in response execution error. Average RTs for all responses tended to decrease across each stage of the experiment.

ST and RT did not appear to be related to whether or not AOS4 requested feedback on a given trial. During LCF treatment, there was no difference between feedback and no-feedback trials in ST (feedback: $M = 1480$ ms, $SD = 1021$ ms; no-feedback: $M = 1488$ ms, $SD = 943$ ms), and a very small difference in RT (feedback $M = 356$ ms, $SD = 150$ ms; no-feedback: $M = 375$ ms; $SD = 150$ ms). During YF treatment, there was a small difference between feedback and no-feedback trials in ST (feedback: $M = 1373$ ms, $SD = 736$ ms; no-feedback: $M = 1303$ ms, $SD = 625$ ms) and a negligible difference in RT (feedback $M = 268$ ms, $SD = 130$ ms; no-feedback: $M = 259$ ms; $SD = 121$ ms).
Study Time (INT process): Based on the INT/SEQ model, it was predicted that ST would be influenced by the duration of the single-syllable responses. As can be seen in Figure 7-21 (left panels), STs were, on average, longer for the extra-long-duration ($M = 2727 \text{ ms}$, $SD = 1174 \text{ ms}$) than the medium-duration ($M = 2376 \text{ ms}$, $SD = 1039 \text{ ms}$) “pa” (LCF-treated) response (top left panel), and also longer for the long-duration ($M = 2472 \text{ ms}$, $SD = 1220 \text{ ms}$) than the short-duration ($M = 1885 \text{ ms}$, $SD = 752 \text{ ms}$) “chee” (YF-treated) response (bottom left panel). These patterns of average ST were reliable across every stage of the experiment (i.e., Baseline, Phase I, Phase II) for AOS4, although fluctuations were observed in some sessions. It is also of interest to note that overall, and particularly during the second treatment phase, average ST was related to target duration such that AOS4 displayed the shortest average STs for target 1S (target duration = 150 ms), followed by 1M (target duration = 275 ms), then 1L (target duration = 450 ms), and finally 1X (target duration = 575 ms), although this pattern was not consistent across all stages of the experiment.

A further expectation for ST was that polysyllabic responses would incur longer STs than single syllables. This was only the case for the post-treatment stage of “pa” responses, in which STs for multi-syllable responses (average = 3410 ms) were longer than STs for single-syllable responses (average = 1913 ms). There was also a trend in this direction for “pa” responses for the first baseline session, but it disappeared by the next session. Otherwise, a systematic influence of sequence length was not observed on ST.
**Reaction Time (SEQ process):** For RT, it was expected, based on the INT/SEQ model, that increasing the number of syllables in a response would increase the time required to initiate the movement if those syllables are programmed as separate units. While no consistent sequence length effect was noted across sessions, average phase data suggest that this pattern seems to have emerged for “pa” responses after Phase I (LCF) treatment, but for “chee” responses, this pattern was attenuated during Phase II (YF) treatment. Single-syllable complexity (duration) effects on RT were evident except for the “pa” responses during Phase I treatment, and the “chee” responses during true baseline.
Figure 7-21. Participant AOS4 study time (ST, left panels) and reaction time (RT, right panels) for treated responses elicited as random-order probes during baseline, treatment, and retention phases. Top panels = Phase I targets (top legend); Bottom panels = Phase II targets (bottom legend). See Table 7-3 for stimulus cue definitions. B = Baseline, Tx = Treatment, LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long.
Summary: AOS4

To summarize, AOS4 exhibited no substantial differences between feedback and no-feedback practice trials in either treatment condition, although he thought that he had requested feedback on primarily good trials during the LCF treatment. During acquisition, YF treatment was associated with slightly more improvement in AE than LCF treatment, but YF treatment was also associated with a slight increase in RE. Relative error performance on the blocked retention probe was more similar to acquisition performance in the LCF condition, compared to the YF condition. The random transfer probes indexed greater transfer of absolute timing in the YF condition, and negligible relative timing transfer in both conditions. Absolute timing in the “chee” (YF) baseline was positively influenced by “pa” (LCF) treatment, whereas relative timing was negatively influenced. In programming responses during random transfer probes, AOS4 demonstrated reliable complexity (duration) effects for single-syllable responses during each experimental phase for average ST, and during most phases for average RT. Sequence length effects were observed intermittently in both ST and RT, but did not appear to be related.

7.3. Discussion

The primary purpose of Experiment 4 was to determine whether giving learners self-control over their feedback schedules would enhance their ability to learn speech motor timing behaviors. The data from Experiment 4 cannot be discussed without first acknowledging the apparent lack of independence of the two behavior
sets. That is, Phase I treatment for “pa” targets was always associated with a change on the “chee” set that was to be held in baseline prior to Phase II treatment. While this loss of experimental control is problematic for the interpretation of the present results, Experiment 4 was nevertheless conducted in its entirety with four participants. Individuals with AOS and aphasia are notoriously diverse and variable, and have been known to demonstrate both within- and between-participant discrepancies on measures of motor programming (e.g., Clark and Robin, 1998; Maas, Robin, Wright et al., in press). No two participants in the present experiment, who ranged from mildly- to severely-impaired, presented with similar speech impairment profiles. Therefore, it was important to sample a range of AOS severities before drawing conclusions regarding the appropriateness of the experimental methods, particularly given their novelty to single subject research and the exploratory nature of the project. It was also worthwhile to carry out the protocol to its conclusion, even after the first hints of compromised experimental control were available, so that the extent and characteristics of the cross-condition interactions could be fully examined in each of the counterbalanced treatment orders. These data may provide important information for designing better-controlled experiments to test these hypotheses in the future.

Another serious concern for the present experiment was the degree of instability of the error measures present during baseline testing. Because there is not a literature on the performance of these particular tasks by individuals with AOS (cf. Maas, 2006 for the closest approximation), it was not possible to know a priori what constituted an acceptable degree of variability in baseline. In most cases, a linear trend
of decreasing AE or RE was not present in baseline data. When a linear-appearing trend was observed, it was typically only present in RE, but not AE data. Because of this, and because of restricted resources and limited participant availability, it was decided to proceed with the experiment after the maximum number of baseline sessions (four to five) permitted by time were collected. The analyses employed were the most appropriate way to examine the data in light of this baseline variability and the number of sessions.

7.3.1. Feedback Requests

Recent reports have provided evidence that learners prefer to receive feedback on relatively “good” trials (Chiviacowsky & Wulf, 2002), and that they learn better when given “good” feedback (Chiviacowsky & Wulf, 2002, 2007). Participants in the present experiment determined their own 30% feedback schedules during LCF treatment; however, it was not clear what they based their feedback decisions on. Descriptive comparisons of AE, RE, ST, and RT on feedback trials versus the first three no-feedback trials per practice block were not particularly informative about differences between the types of trials. One problem with this approach is that without the availability of appropriate statistical methods for comparative purposes, it is unclear which differences are meaningful.

Another approach, illustrated in Figure B-1 (Appendix B), is to consider the percent difference in each dependent measure between feedback versus no-feedback trials for all participants in both treatment conditions. Since no difference between
feedback and no-feedback trials was expected for the YF condition, the magnitude of any discrepancies between these trial types for the YF condition can serve as a benchmark for indicating a non-meaningful difference. Applying this criterion to the present treatment data supports the interpretations that were given to the feedback versus no-feedback trial data in the Results section. That is, the only apparent differences between these trial types in LCF treatment were observed for participant AOS2, whose mean ST and AE were over 20% greater for trials on which he requested feedback versus those on which he did not. The fact that AE was greater for feedback versus no-feedback trials might suggest that AOS2 knew when he had performed well, and chose to receive feedback on relatively “poor” trials so that he could use the augmented outcome information to guide his more off-target responses to improve future trials (see Wright, Smith-Munyon, & Sidaway, 1997). The fact that ST was longer for feedback trials as opposed to no-feedback trials could signify that AOS2 wanted to receive extrinsic outcome information on trials for which he had more-completely prepared responses. Or vice-versa, perhaps AOS2 decided in advance of each trial whether or not he would request feedback, and strategically took extra time to program responses when he knew he would seek feedback afterwards. This speculation is supported by the observation that this participant favored receiving feedback on the first three or four trials per block (even though he reported requesting feedback randomly), and thus it is possible that this result represents an order effect, and not a preference for receiving feedback on trials in which he prepared longer or had greater absolute timing error. Further, since AOS2 typically exhausted his
feedback requests before any no-feedback trials were conducted, the three no-feedback trials per block used for this analysis did not usually involve an active feedback decision on his part.

With the exception of ST and AE in AOS2, on the whole, in contrast to Chiviacowsky and Wulf’s finding (2002), feedback-requesting behavior in the present experiment did not appear to be related to any of the dependent measures of response execution or motor programming. This could mean that learners simply did not have a preference about which trials they received feedback on, or that they might have, but just could not evaluate their own performance well enough to reliably seek feedback based on performance. Whatever the case, most participants intuited that there was no difference between their feedback and no-feedback trials, because they reported that they requested or received feedback “randomly” or “on good and bad trials equally.” AOS4 was the only participant who reported that he sought feedback after relatively good trials, which, in light of the lack of difference between his feedback and no-feedback trials, suggests that he was not able to accurately self-evaluate his responses. AOS1 reported that during YF treatment he thought he received feedback on relatively good trials, indicating that he was not a good judge of his response durations in the YF condition. (He did, however, correctly report that during the LCF treatment, he received feedback on good and bad trials equally.)
7.3.2. Response Execution

As to whether or not learner-controlled feedback affected participants’ learning of the speech timing patterns, different results were obtained for acquisition (i.e., performance during practice) and measures of retention and transfer (i.e., performance during probes). An overview of general response execution results is shown in Table 7-5. In the absence of appropriate baselines with which to compare practice session data, acquisition was examined by observing linear trends across practice sessions. A fair amount of inconsistency, both within and across participants, was apparent during treatment sessions. The three observations that suggested the most reliable acquisition effects were all in the YF condition (i.e., AOS2 AE and RE, and AOS3 AE). Two additional situations were observed in which the YF condition was associated, albeit less strongly, with reductions in error (i.e., AOS1 RE and AOS4 AE), and in one case, the YF condition was not associated with error reduction, but prevailed over the LCF condition, which was associated with an increase in error (i.e., AOS3 RE). Of the eight (4 participants x 2 phases) treatment situations, only two provided weak support for the LCF condition (i.e., AOS1 AE and AOS4 RE).

Although YF was associated with error reduction during acquisition more often than LCF was, the pattern was not consistent for AE or RE when the relationships between treatment session and error change were weak (i.e., AOS1 and AOS4). Where the relationships were stronger (i.e., AOS2 and AOS3), YF appeared to have benefited both AE and RE. However, these were also the two participants who received the YF condition first, raising the question of whether an order effect was
present in the data. An order effect was not apparent, however, for the two participants who received the LCF condition first (i.e., AOS1 and AOS4) and demonstrated weaker and more ambiguous acquisition effects. A plausible interpretation is that the YF condition yielded superior acquisition effects, but that due to the cross-condition transfer during Phase I, these effects were masked when participants received the LCF condition first.

These data contrast with earlier work (Chiviacowsky & Wulf, 2002) that did not find a significant difference in acquisition between self-controlled and yoked feedback conditions on a finger movement task. However, numerical data in that study did demonstrate a tendency for the self-controlled feedback group to have less absolute error and faster reduction in relative timing error during acquisition. Although the acquisition results in the present experiment were not unequivocal, they suggested the opposite, that for learners with AOS performing a motor speech timing task, self-controlled feedback does not facilitate acquisition. In fact, it appeared that relieving participants of the responsibility of making their own feedback choices (i.e., providing YF) resulted in greater change during acquisition. The feedback-requesting data discussed in the previous section demonstrated that, for the most part, these participants did not make feedback decisions related to movement outcome, perhaps because they were impaired in their ability to utilize intrinsic feedback sources to self-evaluate their own performance similar to or concomitant with their difficulties in using extrinsic feedback (e.g., Ballard & Robin, 2007). It would stand to reason, then, that requiring speakers with AOS to decide which trials they wanted feedback on,
which presumably asks them to engage in some degree of post-movement evaluative processing, might be disruptive or distracting. Another possibility is that the linguistic demands of requesting feedback interfered with movement processing, given participants’ aphasia. Perhaps for these learners, providing feedback according to a predetermined or independently-generated schedule optimized their ability to use it for reducing performance error on subsequent trials. Or, maybe participants “tried harder” on more trials during YF treatment because they did not know which ones they would receive feedback on. It is not clear, however, what “trying harder” would have meant, particularly since ST and RT (i.e., the intervals during which responses were presumably prepared and initiated) were not typically greater in YF practice conditions than in LCF conditions.

Turning to treatment effects on random transfer, when participants were required to produce the four different responses for each set (i.e., four response types for “pa” or four different response types for “chee”) in random order without feedback during probes, a pattern emerged that contrasted with the acquisition results. The treatment effects on AE during the treatment phase ($d_{1T}$) and upon the withdrawal of treatment ($d_{1W}$) appeared to be related to the durations of the syllables trained. In other words, each participant demonstrated enhanced transfer (in terms of reduced AE) on responses comprised of short (150 ms) and long (450 ms) syllables, as opposed to those comprised of medium (275 ms) and extra-long (575 ms) syllables. There was evidence, however, that the treatment effects associated with short- and long-syllable responses were intensified under LCF conditions. Participants AOS1 and AOS2, who
were trained on short- and long-duration syllables under LCF conditions, demonstrated relatively robust benefits of LCF treatment on AE for random transfer of these responses (both during the treatment phase, $d_{IT}$, and at withdrawal, $d_{IW}$). AOS3, who received training on short- and long-duration syllables under YF conditions, on the other hand, demonstrated only a very minor benefit of YF treatment on AE for random transfer of these responses. AOS4, who also received training on short- and long-duration syllables under YF conditions, did demonstrate a benefit of YF treatment during the treatment phase, but an ambiguous result upon treatment withdrawal. Thus, the results for AE during random transfer appear to slightly favor the LCF condition.

Interestingly, relative timing error for these same random transfer probe responses diverged from the results for AE for some participants. AOS2, who demonstrated a benefit of LCF treatment on AE, evinced a benefit of YF treatment on RE during random transfer probes, although it was not maintained upon treatment withdrawal. AOS3, who demonstrated a small benefit of YF treatment on AE, evinced a very small benefit of LCF treatment on RE. For AOS4, all treatment effects ($d_{IT}$ and $d_{IW}$) on RE for random transfer probes were too small to interpret. AOS1 was the only participant whose RE data were consistent with his AE data, in both cases favoring the LCF treatment applied to short- and long-duration syllables.

Next, although the pre-treatment changes in “chee” responses ($d_{IP}$) that were associated with Phase I treatment for “pa” were undesirable from an experimental control standpoint, they potentially afford another opportunity to inspect transfer
effects related to the two different feedback conditions across participants. Both treatment conditions (i.e., LCF and YF), when employed during Phase I treatment for “pa,” produced small- or medium-sized effects on the untreated “chee” behaviors. At first glance, the strongest beneficial (in terms of reducing error) pre-treatment transfer effects appear to be products of YF treatment (see $d_{IP}$ for AE for AOS2 and AOS3). Further, a demonstrably negative (in terms of increasing error) pre-treatment transfer effect was observed in the context of LCF treatment (see $d_{IP}$ for RE for AOS4). In each of these cases, the effects of Phase I treatment for “pa” were actually stronger on untreated “chee” responses than they were on treated “pa” responses (positively for YF, as in AOS2 and AOS3, and negatively for LCF, as in AOS4). However, on closer inspection, in each of these instances the seemingly-substantial pre-treatment transfer effects can be explained by unequal baseline variance. That is, in these three cases, “chee” responses had much more stable true baselines than “pa” responses, which could account for larger pre-treatment transfer effect sizes. In the absence of homogenous variability in true baseline, it is not possible to draw firm conclusions about differences in pre-treatment transfer under YF versus LCF conditions.

Although these results do not offer a clear or consistent picture of which feedback condition was most beneficial to learning, a number of potentially informative observations can be made. First, as was demonstrated in Experiments 1 and 2 (Chapter 5) and in a host of other motor learning studies, acquisition performance did not predict true learning (here, indexed by random transfer). Even though the probe results were suggestive of a stimulus effect with a hint of positive
influence of LCF treatment, the acquisition data suggested a potential benefit of YF treatment.

Second, there was no clear relationship between the two different error measurements, suggesting that the treatments used here affected absolute and relative temporal aspects of movements differently. Changes in RE were generally not as strong as those in AE, and RE was noted to actually increase very slightly with treatment in some situations. Of course, RE measures were only averages of data for the two- and four-syllable responses, whereas AE measures were averages of all four response types. Given that multi-syllable responses typically had greater AE than single-syllable responses, however, average AE was probably strongly influenced by the same responses that were measured by RE. It is interesting, in any event, that some improvements in RE were observed, given that detailed feedback about relative timing was never provided (participants only knew whether their relative timing patterns were correct or incorrect). This finding signifies that some participants, particularly AOS1 and AOS2, may have been able to develop GMPs governing the temporal relationships in the 4-syllable patterns. For participants who did not demonstrate improvements in RE, or who showed slightly worse RE with practice (i.e., AOS3 and AOS4), it might have been the case that their efforts were primarily devoted to reducing absolute timing error (i.e., the construct that they received detailed feedback about) at the cost of relative timing error. Another possibility is that perhaps these participants, like a subset of the participants with AOS described by Clark and Robin (1998), were relatively severely impaired in their ability to develop new GMPs.
Table 7-5. Overview of Experiment 4 response execution results for each participant in each experimental task. Columns show the feedback condition that was associated with better acquisition, retention, or transfer. Parentheses denote slight differences between conditions and brackets denote results that were confounded by baseline inequalities. P1 = Phase 1, P2 = Phase 2, LCF = learner-controlled feedback; YF = yoked feedback; S = short, M = medium, L = long, X = extra-long, AE = absolute error, RE = relative error.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Acquisition</th>
<th>Blocked Retention</th>
<th>Random Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AE</td>
<td>RE</td>
<td>AE</td>
</tr>
<tr>
<td>AOS1: Mild</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1: LCF S/L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2: YF M/X</td>
<td>(LCF)</td>
<td>YF</td>
<td>[YF]</td>
</tr>
<tr>
<td>AOS2: Severe</td>
<td>YF</td>
<td>YF</td>
<td>LCF</td>
</tr>
<tr>
<td>P1: YF M/X</td>
<td>YF</td>
<td>YF</td>
<td>LCF</td>
</tr>
<tr>
<td>P2: LCF S/L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOS3: Moderate</td>
<td>YF</td>
<td>(YF)</td>
<td>[ND]</td>
</tr>
<tr>
<td>P1: YF S/L</td>
<td>YF</td>
<td>(YF)</td>
<td>[ND]</td>
</tr>
<tr>
<td>P2: LCF M/X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOS4: Moderate</td>
<td>(YF)</td>
<td>(LCF)</td>
<td>[ND]</td>
</tr>
<tr>
<td>P1: LCF M/X</td>
<td>(YF)</td>
<td>(LCF)</td>
<td>[ND]</td>
</tr>
<tr>
<td>P2: YF S/L</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interpretation: YF enhanced acquisition. Supports random transfer results for participant with adequate baseline. LCF slightly enhanced transfer of parameterization (absolute timing).
Table 7-6. Overview of key Experiment 4 motor programming results for the random transfer probe for each participant. Columns describe presence of duration and sequence length effects on ST and RT in each feedback condition. P1 = Phase 1, P2 = Phase 2, LCF = learner-controlled feedback; YF = yoked feedback; S = short, M = medium, L = long, X = extra-long, ST = study time, RT = reaction time.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Duration Effect (1S &lt; 1L)</th>
<th>Sequence Length Effect (4 &gt; 2 &gt; 1 syllable)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LCF</td>
<td>YF</td>
</tr>
<tr>
<td>AOS1: Mild</td>
<td>Yes for ST</td>
<td>Yes for ST</td>
</tr>
<tr>
<td>P1: LCF S/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2: YF M/X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOS2: Severe</td>
<td>Yes for ST (in baseline)</td>
<td>Yes for ST and RT</td>
</tr>
<tr>
<td>P1: YF M/X</td>
<td>and RT</td>
<td></td>
</tr>
<tr>
<td>P2: LCF S/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOS3: Moderate</td>
<td>Not reliable</td>
<td>Not reliable</td>
</tr>
<tr>
<td>P1: YF S/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2: LCF M/X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOS4: Moderate</td>
<td>Yes for ST; yes for RT except during treatment</td>
<td>Yes for ST; yes for RT except true baseline</td>
</tr>
<tr>
<td>P1: LCF M/X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2: YF S/L</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interpretation: Evidence for complexity (duration) effect on INT processing in majority of participants.

7.3.3. Motor Programming

Using the self-selection paradigm in the context of this treatment study permitted the measurement of changes in two motor programming processes across each phase of the experiment. Specifically, ST and RT patterns for each response type
in the random transfer probes were inspected (see Table 7-6 for general overview of key results). Although inconsistencies were found for different response types in individual sessions, for the most part, ST and RT for “pa” responses (which were treated during Phase I) decreased between baseline and Phase I treatment, and ST and RT for “chee” responses (which were treated during Phase II) generally decreased until they were treated in Phase II treatment. In several cases, ST and or RT for “pa” responses stopped decreasing after Phase I treatment (e.g., AOS1 ST, AOS2 ST for 4M, AOS3 ST and RT, and AOS4 ST and RT for sequences), indicating that reductions in these measures were indeed related to treatment and likely not solely due to the trends demonstrated in baseline. Inspection of ST and RT standard deviations supported this pattern of results (see Figure B-2 in Appendix B). That is, ST and RT variability also generally decreased with treatment, except for AOS4, whose ST and RT variability for “chee” increased with Phase II treatment. Further, standard deviations of these measures for the “pa” responses also stopped decreasing (or actually increased again) in several cases following the withdrawal of Phase I treatment for “pa,” corroborating the notion that these changes in ST and RT were indeed related to treatment. Treatment-related changes in ST and RT indicated that speakers with AOS were able to improve their efficiency and consistency of INT and SEQ processing, although these changes did not appear to be enhanced by one feedback condition (i.e., LCF or YF) more than the other.

Unlike Experiment 3 and other recent studies that have failed to find a duration effect on ST for single-syllable speech responses (Mass, Robin, Wright et al., in press;
Wright et al., submitted), three of the four participants in Experiment 4 exhibited longer average STs for 1L than 1S responses, indicating that programming the internal structure of long syllables was more demanding than programming of short syllables. Although this effect has largely been elusive in the speech domain, the present results are consistent with findings from finger movement studies (e.g., Klapp, 1995; Immink & Wright, 2001; Wright et al., 2004), indicating that duration may be a useful operationalization of motor complexity for single speech units in some cases (cf. Wright et al., submitted). Others have speculated that the lack of robust duration effects on ST in speech movements may be reflective of high degrees of automaticity regarding this variable in speech production (e.g., Maas, Robin, Wright et al., in press; Wright et al., submitted). However, the present results suggest that for speakers with AOS, for whom speech production may have become less automatic (e.g., Varley & Whiteside, 2001), the demands associated with assigning temporal features to speech motor programs are high enough that differences between syllables of different durations are measurable in ST. Interestingly, Maas, Robin, Wright and colleagues (in press), who did not find a duration effect on ST for finger or speech movements, tested participants with relatively mild AOS, whereas the current experiment, which did find evidence for a duration effect on ST for speech, studied participants with more moderate or severe impairments, lending support to the notion that the construct of “automaticity” might be related to INT demand in the programming of speech movements.
Average RT results for two participants (i.e., AOS2 and AOS4) also demonstrated the duration effect for single syllable responses such that longer-duration syllables took longer to initiate than shorter-duration syllables. As discussed for a similar finding in Experiment 3, there are several possible reasons for this, including kinematic influences or continuation of INT processing in the RT interval. Given that these two participants did demonstrate duration effects on ST, thus providing evidence that they engaged in INT processing during ST, it is plausible that they either did not complete the response preparation before they pressed the “ready” button at the end of ST and had to continue INT processing during the RT interval, or that the program that they loaded into a motor buffer during the ST interval decayed quickly and had to be re-programmed during RT. Either way, the finding of a duration effect during both ST and RT within-participant provides support for the notion that INT processing is disrupted in AOS (e.g., Maas, Robin, Wright et al., in press).

The examination of sequence length effects (4 > 2 > 1 syllable) on RT was made difficult by a large amount of within-participant variability across sessions. However, two key findings potentially speak to the issue of whether or not participants integrated sequences into single “chunks” or programmed them one syllable at a time. First, participant AOS1’s average RT performance across the three phases of random transfer probes (i.e., baseline, Phase I, Phase II) demonstrated a decreasing sequence length effect between 4- and 1-syllable “pa” responses with LCF treatment, and an increasing sequence length effect for 4- versus 1-syllable “chee” responses with YF treatment. This pattern, in conjunction with the finding that relative error (an index of
GMP cohesion) improved more for these probe responses during LCF treatment than YF treatment, suggests that over time, the elements of the 4L “pa” response became integrated into a single unit, whereas those of the 4M “chee” response did not (or actually “dis-integrated”). While this observation could be taken in support of an advantage of LCF treatment to motor-program assembly and/or stability, one important caveat must be considered. Experiment 3 demonstrated that healthy speakers had more difficulty unitizing the 4S pattern (analogous to the 4M pattern here, except with different absolute scaling) than the 4L pattern. Thus, it is possible that AOS1, who was most mildly impaired, also exhibited this pattern due to stimulus-related factors, and not as a result of feedback condition22.

The second sequence length observation that deserves comment was found for participant AOS2, who appeared to take longer on average to initiate 2- and 4- element responses compared to 1- element responses during baseline, but not once treatment began. Although the pattern was not unequivocal (e.g., see 1X and 1L RTs in baselines 3 and 2, respectively), it does seem to suggest that this participant learned to integrate sequences into single-unit programs. An overall reduction in RE for both sets of responses following baseline is consistent with this view. Participants AOS3 and AOS4 did not demonstrate any reliable RT patterns that were interpretable with regard to sequence length effects.

22 AOS1 also provided evidence in probes for longer average RTs for 2S than 4L (after fourth baseline) responses, yet shorter average RTs for 2X than 4M responses. If RT is taken as an index of the degree of unitization of response sequences, then these results, along with Experiment 3 results, suggest that sequences that began with relatively short syllables were more difficult than sequences that began with relatively long syllables. In fact, anecdotally, this was reported by many learners. This pattern is consistent with standard segmental stress patterns of English and developmental patterns suggesting that it is more difficult for speakers to learn to produce unstressed syllables in the initial position of words (e.g., Allen and Hawkins, 1978; 1979, in Hargrove & McGarr, 1994).
7.3.4. Conclusions

Although the findings in Experiment 4 must be considered in the context of significant caveats regarding baseline variability and cross-condition transfer effects, several important conclusions can be drawn from this novel approach to investigating the role of feedback in speech motor programming and learning in AOS. First, it was evident that the feedback conditions that were compared within-subject across treatment phases had different effects on acquisition versus learning. Thus, this experiment adds to the growing body of research, including the first three experiments in this dissertation, that implores researchers and clinicians to consider multiple measures of performance (e.g., acquisition, retention, transfer) when assessing treatment effects (see Schmidt & Bjork, 1992; Schmidt & Lee, 2005). Second, despite its limitations, Experiment 4 did provide preliminary evidence that extends the principle of learner-controlled feedback both to the speech domain and to impaired learners. Specifically, the results of Experiment 4 suggest that allowing individuals with AOS to choose their own feedback during practice of a speech motor task is detrimental to performance during acquisition, but may confer slight benefits to transfer, particularly with regard to accuracy of absolute temporal parameters. This finding is partially-consistent with Chiviacowsky and Wulf (2002), who found a benefit of “self-controlled” versus “yoked” feedback on absolute error for transfer of trained finger movement patterns, yet no difference during acquisition or for relative error. However, unlike Chiviacowsky and Wulf (2002), participants in the current experiment did not appear to use a strategy of requesting feedback on relatively good
trials, demonstrating that learner-controlled feedback may enhance learning even when learners do not rely on any demonstrable strategy for choosing feedback trials.

Finally, although this experiment was not designed to permit direct comparison to extant data on INT/SEQ processing in AOS (i.e., Maas, Robin, Wright et al., in press), a couple of general observations can be made with respect to motor programming and execution in AOS relative to normal patterns. First, Maas, Robin, Wright and colleagues (in press) have demonstrated that speakers with relatively mild AOS have longer STs than healthy speakers, but comparable RTs. In the current experiment, the overall patterns demonstrated by speakers with AOS suggested that, although STs and RTs appeared abnormally long early in the experiment, participants were generally able to demonstrate STs and RTs that were within the normal range (i.e., as observed in Experiment 3) after extensive treatment, although these measures were highly variable. It should be noted, however, that there was a time limit on ST for Experiment 3 and not for Experiment 4, which may have contributed to this observation. Three participants (AOS2, AOS3, and AOS4) exhibited mean AEs and/or REs during acquisition that appeared to be greater than those of normal speakers, although firm conclusions cannot be drawn due to the considerable methodological differences between these experiments. Related to recent findings suggesting that AOS is related to INT-specific deficits (e.g., Deger & Ziegler, 2002; Maas, Robin, Wright et al., in press), three speakers with AOS in the present experiment did demonstrate a duration effect on ST, which suggests that they did engage in INT processing during that interval. The fact this effect is not robust for healthy speakers and speakers with
mild AOS in this paradigm suggests that the present finding may be an indication of reduced INT efficiency in the present participants, lending support to the INT-deficit hypothesis of AOS.

Appendix B

![Figure B-1. Percent differences of ST, RT, AE, and RE between feedback and no-feedback trials for each participant in the Learner-Controlled Feedback (LCF) treatment condition (left panel) and the Yoked Feedback (YF) treatment condition.](image-url)
Figure B-2. Study time (ST; left panels) and Reaction time (RT; right panels) standard deviations for random transfer probe “pa” (top panels) and “chee” (bottom panels) responses for each participant in Experiment 4. LCF = Learner-controlled feedback, YF = Yoked feedback, S = short, M = medium, L = long, X = extra-long, Ph. = phase, B = baseline, Tx = treatment.
8.0. General Discussion and Conclusions

Although variables associated with augmented feedback have only recently attracted attention in the speech motor learning literature, (e.g., Adams & Page, 2000; Adams, Page, & Jog, 2002; Austermann Hula et al., in press; Ballard et al., 2007; Fossett et al., 2008; McNeil et al., 2007; Steinhauer & Grayhack, 2000), they have long been recognized in the limb motor learning literature as some of the most important factors in learning. According to Schema Theory (e.g., Schmidt, 1975), motor learning is directly related to the availability of outcome information regarding performance. However, because feedback can have both positive and negative effects on learning (e.g., Salmoni, et al., 1984), it is critical that a theory-driven, principled approach govern its application to both normal and disordered systems.

The four experiments in this dissertation represent a program of research that aims to extend principles of motor learning related to augmented feedback to speech motor learning in AOS. Experiments 1 and 2 tested the hypotheses that manipulating the frequency or timing of augmented feedback would impact treatment outcomes of a commonly used treatment for AOS. Specifically, in Experiment 1, two of four participants with AOS demonstrated retention and transfer benefits when they were given Phonetic Placement Therapy (PPT) with feedback on 60% of their responses rather than 100% of responses. Feedback on 100% of trials was not associated with enhanced learning for any participants, although it did promote acquisition for one participant (for whom it also promoted retention and transfer). In Experiment 2, a
different participant demonstrated a benefit of a 5-second feedback delay on acquisition, retention, and transfer of speech skills trained with PPT.

The rationale for reducing the frequency or immediacy of augmented feedback has to do with allowing learners the opportunity to attend to their own intrinsic feedback sources and estimate their errors. According to the guidance hypothesis (Salmoni et al., 1984), augmented feedback during practice of a new skill serves to guide learners to improved performance. However, in order for learners to sustain new levels of performance under conditions in which augmented feedback is unavailable, they must develop an ability to self-evaluate their own performance (Schmidt, 1975; Swinnen et al., 1990). Reducing the availability of augmented feedback allows learners the opportunity to develop that ability instead of relying on extrinsic information sources (e.g., Bruechert et al., 2003). Since self-evaluation is thought to occur, perhaps spontaneously and unavoidably, during the post-movement feedback delay (e.g., Hogan & Yanowitz, 1978; Swinnen, 1990; Swinnen et al., 1990; Liu & Wrisberg, 1997), delaying the provision of augmented feedback affords learners an interval of time in which they can attend to intrinsic information sources uninterrupted. In general, as these manipulations suggest, learning seems to be enhanced when feedback is available, but “difficult” to use (Swinnen et al., 1990). The precise level of “difficulty” that is optimal likely varies across different tasks, practice conditions, and individual skill levels (e.g., Guadagnoli & Lee, 2004).

This notion that the optimal level of difficulty, or “challenge point” in Guadagnoli and Lee’s (2004) terminology, is a function of the interaction between
task factors and individual skill has important implications for the treatment of neurogenically-disordered systems because of the heterogeneity of symptoms and concomitant impairments that often present with these disorders. Apraxia of speech is a prime example of such a disorder because (a) the primary motor impairment in AOS might manifest differently across individuals or within individuals at different times (e.g., Clark & Robin, 1998), (b) it rarely presents in isolation, often being accompanied by significant cognitive-linguistic disruption, (c) it is thought to affect the way that individuals are able to use feedback in learning (Ballad & Robin, 2007), and (d) given its relative rarity, group generalizations about learning are difficult to make and thus individual factors must be given due consideration. Indeed, the AOS treatment literature is rife with examples of experimental manipulations of practice conditions that have affected individuals differently (e.g., Knock et al., 2000; Ballard et al., 2007; Austermann Hula et al., in press; Maas et al., 2002).

Because providing feedback with a hard and fast predetermined structure (e.g., “60% frequency” or “5 second delay”) is unlikely to be equally-beneficial for all learners with AOS, it is important to allow flexibility in feedback dimensions so that individual learners can receive it in a way that is most likely to help them. How can it be determined, though, what the best feedback conditions are for a particular learner? One way to provide feedback that may be optimally useful is to allow learners to determine for themselves when to receive it.

Giving learners self-control of feedback delivery might be beneficial to learning for two possible reasons. First, allowing learners to take charge of regulating
some aspects of their learning environments might increase their feelings of self-efficacy and motivation, causing them to become more actively-involved in their learning (e.g., Boekaerts, 1996; Zimmerman, 1989). Indeed, research on the learning of finger and limb movements has demonstrated that learners who have self-control over their feedback schedules prefer to receive feedback on relatively good trials (Chiviacowsky & Wulf, 2002), and that feedback about good performance enhances learning relative to feedback about poor performance (Chiviacowsky & Wulf, 2007), supporting the notion that self-controlled feedback possibly serves a motivational or rewarding function.

However, this seems contrary to the guidance hypothesis, which postulates that feedback is helpful because it provides learners with guidance to improve performance (Salmoni et al., 1984). Presumably, that guidance would be counterproductive on trials in which performance is good, because it might induce learners to make maladaptive short-term corrections of too-small errors, leading to trial-to-trial instability (Schmidt, 1991). However, if learners possess sufficiently-developed self-evaluation skills for a given task, meaning that they do not require augmented feedback to provide new information for correcting errors, then feedback about good performance likely serves a motivational purpose, confirming learners’ own movement appraisals and encouraging them to repeat successful movements (Chiviacowsky & Wulf, 2007). Thus, as Schmidt and Lee (2005:397) suggest, feedback “has optimal informational [emphasis mine] value when the learner is uncertain about the reliability of his or her inherent sources of information.”
This leads to the second possible benefit of learner-controlled feedback, that perhaps it enhances learning because learners get movement information when they need it, whether on good trials to encourage them to repeat successful performance, or on bad trials to guide them to improved performance (Chiviacowsky & Wulf, 2002). However, even when different groups of learners perform equally-well during practice and get feedback about the same types of trials (e.g., relatively good trials), those who request feedback after performance demonstrate enhanced learning relative to those who request feedback before performance (Chiviacowsky & Wulf, 2005). This finding suggests that it is not necessarily the motivational or informational functions of feedback \textit{per se} that confer the benefit to learner-controlled feedback situations (as these were presumably the same for both groups), but rather that there is something about seeking feedback \textit{after} movement completion that promotes learning. Specifically, learners who make feedback decisions after movement completion have the opportunity to inform their feedback choices with their own performance appraisals. This might encourage them to evaluate their own performance more frequently or extensively, and/or use the information that they generate through self-evaluation differently than learners who make feedback decisions before trials devoid of outcome information about the trial for which they request feedback.

Experiment 3 of this dissertation speaks to the issue of why learner-controlled feedback benefits learning. First, the key finding from the general motor learning literature demonstrating that feedback chosen after movements is more effective than feedback chosen before movements was extended for the first time to the speech
motor learning domain. Evidence for the benefit of the “After” feedback condition, in terms of reduced error and/or increased efficiency of motor programming was found in acquisition, retention, and transfer tasks. The differences between “After” and “Before” feedback were even more pervasive in Experiment 3 than previously shown in the limb literature (Chiviacowsky & Wulf, 2005), possibly due to the greater number of different response types trained in the present experiment (i.e., four), compared to Chiviacowsky and Wulf (2005) (i.e., one response type), even though the number of trials per task was constant (i.e., 60 practice trials per response type). Alternatively, the more prevalent effects in the current experiment could have been related to the experimental paradigm (i.e., the self-selection paradigm), which, by virtue of the additional study time interval, may have promoted more extensive task processing and stronger acquisition and/or learning effects. These possibilities represent empirically-testable hypotheses could be addressed in future studies.

The novel manipulation in Experiment 3 that contributed to the investigational dialogue about why learner-controlled feedback works was the addition of a forced self-evaluation task. It was hypothesized that, if the advantage of self-selecting feedback after responses was conferred by virtue of more extensive or frequent spontaneous self-evaluation, then forcing learners who selected feedback before responses to self-evaluate would confer the same benefit to them. Indeed, some evidence in favor of this hypothesis was generated in Experiment 3. For learners who made feedback decisions before trials, the addition of the self-evaluation task was associated with less absolute error in transfer to a novel temporal scaling; less relative
error in retention and random-order transfer; faster response initiation in acquisition, retention, and transfer; and longer response preparation, particularly in acquisition and random transfer. In contrast, the only significant difference found between groups LCB+E (who requested feedback before performance and were required to self-evaluate) and LCA (who requested feedback after performance) was in ST for the 1L response in the first block of practice.

While these results are consistent with the notion that learner-controlled feedback chosen after performance benefits learning because it promotes spontaneous self-evaluation, data from a fourth group of participants, LCA+E (who requested feedback after performance and completed the self-evaluation task) suggested that the potency of spontaneous self-evaluation for learning can be reduced by the addition of a forced self-evaluation task (or, conversely, that the potency of forced self-evaluation for learning can be reduced by putative spontaneous self-evaluation, e.g., Swinnen et al., 1990). It was predicted that adding the self-evaluation task to the group who chose feedback after trials would not result in an appreciable change in learning, based on the assumption that these participants were already disposed to spontaneously self-evaluate performance for the purpose of informing their feedback choices. However, several differences were found between the two groups, including longer response preparation times, longer response initiation times, and greater relative error for the LCA+E group compared to the LCA group. These discrepancies suggested that adding a self-evaluation task to the condition in which participants were already presumed to spontaneously self-evaluate contributed to a decrement in performance and learning.
Even though increasing the processing demands during the feedback delay interval by introducing another task (i.e., forced self-evaluation response) might logically be expected to reduce performance and learning of the primary task, there were two reasons not to anticipate this in Experiment 3. First, tasks that interfere with processing during the feedback delay interval are typically thought of as those that block maintenance of movement information in short-term memory (e.g., Swinnen, 1990). A self-evaluation task of the type used in Experiment 3, in contrast, might reasonably be expected to allow this movement information to remain active, as it should be compatible with the kind of information processing that typically occurs after movements. Second, even though the addition of the forced self-evaluation task lengthened the duration of the feedback delay interval, this might have been expected to have a positive effect on learning (see Experiment 2). The inter-trial interval, which has been argued to be even more influential for learning (e.g., Salmoni et al, 1984), was constant across groups.

Thus, the fact that adding forced self-evaluation to the “After” feedback condition had a deleterious effect on learning potentially signifies that it caused learners to process movement information or use augmented feedback differently. Motor learning can be viewed as “an increasing function of the degree to which participants use [knowledge-of-results feedback] to test response hypotheses” (Guadagnoli & Kohl, 2001:217). Error estimation, or self-evaluation, is in essence a hypothesis about performance that can be tested by comparing it with augmented feedback. The present results suggest that spontaneous or forced self-evaluation of
performance, but not both, encourages learners to develop and test these hypotheses, and provide support for the hypothesis that learner-controlled feedback, when arranged post-performance, is associated with more frequent and/or extensive spontaneous self-evaluation.

Experiment 4 provided evidence that the benefits of learner-controlled feedback may also apply to speech motor learning in individuals with AOS. Interestingly, during acquisition, yoked feedback (in which participants did not have control over their own feedback schedules) appeared to be a performance-enhancing variable for these individuals, contrary to null acquisition differences reported for these conditions in normal finger-movement learners (Chiviacowsky & Wulf, 2002). That is, during acquisition, participants were generally more accurate in executing responses when they did not have control over their feedback schedule. To the extent that movement appraisal is particularly demanding for participants with AOS, this finding could be a reflection of a processing limit during the feedback delay interval. Relieving participants of this added demand enhanced their performance during acquisition. However, this performance enhancement was temporary, for the learning (i.e., transfer) effects associated with yoked feedback treatment were not superior to learner-controlled feedback treatment when assessed with the random-order probe. Indeed, Experiment 4 provided qualified evidence that in at least two participants, learner-controlled feedback was more effective for learning, and in particular, for learning of absolute timing (i.e., parameterization) control. This finding adds to the body of literature demonstrating differential performance versus learning effects for
other principles of motor learning (e.g., feedback frequency, feedback delay, random practice).

One perplexing finding in Experiments 3 and 4 that was inconsistent with previous work (e.g., Chiviacowsky & Wulf, 2002; 2005) was that learners did not choose to receive feedback about relatively good performance. In fact, results suggested that healthy speakers and one speaker with AOS chose feedback on relatively poor trials. This might indicate that the tasks used in Experiments 3 and 4 created more uncertainty on the part of learners with regard to their performance, which is particularly plausible given that learners in the present experiments practiced four different response types, whereas in previous work, only one pattern was practiced. Greater performance uncertainty may have prompted learners to seek feedback for informational purposes when performance was particularly off-target, as a means of guiding them to better performance.

Another possibility might be that some learners may not have willfully “chosen” feedback at all. A handful of participants in Experiment 3 appeared to request feedback on the same (or close to the same) trials in every block, suggesting that they were following some a priori-decided schedule for requesting feedback and not relating their feedback requests to the primary movement-production task. Perhaps these learners were not actively engaged in the learning process, or perhaps they did not require feedback on 30% of responses, but were constrained by the task instructions to request feedback at times when they did not need it. Whatever the reason, the finding that learners received feedback on relatively poor performance
does present a limitation for comparing the effects of learner-controlled feedback with previous studies, as “good” and “bad” feedback may not affect learning the same way (Chiviacowsky & Wulf, 2007). Future work might address this issue by comparing “good” versus “bad” feedback in the context of speech motor learning.

In addition to the potentially-significant limitations enumerated in Chapter 7 with regard to concerns about experimental control, Experiment 4 was also limited in terms of the extent to which behaviors indicative of learning could be measured. First, retention of trained responses in blocked order was only soundly assessed for one participant (AOS2). Given the very large treatment effects that were observed in this task for AOS2, future studies should endeavor to include a similar measure for every participant. Second, transfer, the other hallmark of true learning, was only assessed up to four days post-treatment. Future studies should consider treatment designs that permit measurement of more long-term effects, as was done in Experiments 1 and 2. Longer-term changes might provide more information about the different effects of learner-controlled and yoked feedback conditions on motor programming and execution in AOS.

Further, although the experimental paradigm used in Experiments 3 and 4 permitted observations to be made regarding treatment-induced changes in both response preparation and execution, the task lacked ecological validity relative to actual treatment for motor speech disorders. Thus, an important next step will be to apply learner-controlled feedback to an extant treatment for AOS, as was done for
other feedback manipulations in Experiments 1 and 2, and observe its contributions to improving word production.

Although the self-selection paradigm is an innovative approach to studying INT and SEQ processing within-subject and within-trial, it is becoming increasingly apparent that its application to speech tasks may require considerations beyond what has been carried over from its application to finger-movement tasks. For example, the single-unit duration manipulation that has been used to demonstrate complexity effects on INT processing via increased study times for finger movements does not appear to directly translate to the speech domain. In addition, the self-selection paradigm is limited in the evidence that it can provide for unit integration, as study time does not appear to be able to differentiate sequences programmed as single units or as multiple separate units the way that choice RT can. Investigations are underway to clarify these issues and further refine the application of the self-selection paradigm to speech (e.g., Wright et al., submitted).

Another relevant direction of investigation will be to study the applicability of principles of motor learning to the treatment of other motor speech disorders that have been argued to disrupt speech motor programming, such as hypokinetic dysarthria secondary to Parkinson’s Disease (e.g., Spencer & Rogers, 2005) and ataxic dysarthria (see Inhoff, Diener, Rafal & Ivry, 1989). In addition, since most of the data concerning the application of learner-controlled feedback has been generated in young adults (e.g., Experiment 3; Chiviacowsky & Wulf, 2002; 2005; cf. Chiviacowsky, Wulf, Laroque de Medeiros, Kaefer, & Wally, 2008, for data from children), and changes in
motor control and learning are known to occur with age (e.g., Krampe, Engbert, & Kliegl, 2001; Swinnen, et al., 1998), it will also be important for future studies to consider teasing apart any age-related versus neurological disorder-related differences in how these variables affect motor learning. Preliminary data on reduced feedback frequency indicate that its beneficial effects also apply to motor learning in the elderly (Rice, 2003).

Motor learning studies typically only assess movement execution. Experiments 1 and 2 of this dissertation extended findings from limb research to speech motor learning in AOS, and thus contributed to our knowledge of treatment factors relevant to managing the disorder, as well as to an understanding of the commonalities between speech and limb motor control. The second half of this dissertation took the relatively novel approach of addressing motor learning during motor programming stages as well, and in so doing demonstrated that feedback variables may have specific influences on particular aspects of motor programming. Future research should endeavor to further establish links between principles of motor learning and their impact on movement execution on the one hand, and theoretical notions about motor programming and the underlying processes that are disrupted in AOS on the other.
References


