Acoustic and Perceptual Assessment of Stop Consonants Produced by Normal and Dysarthric Speakers

by

Kelly Lynn Poort

B.S.E., University of Iowa (1991)
S.M., Massachusetts Institute of Technology (1995)
E.E., Massachusetts Institute of Technology (1996)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2000

Author .....
Harvard-MIT Division of Health Sciences and Technology
June 2, 2000

Certified by .....
Kenneth N. Stevens, Sc.D.
Clarence J. LeBel Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Martha L. Gray, Ph.D.
Edward Hood Taplin Professor of Medical and Electrical Engineering
Co-director, Harvard-M.I.T. Division of Health Sciences and Technology

Acoustic and Perceptual Assessment of Stop Consonants Produced by Normal and Dysarthric Speakers

by

Kelly Lynn Poort

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on June 2, 2000, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This research is a first step toward characterizing motor control and coordination difficulties of dysarthric speakers through the development of acoustic measures which reflect articulatory movements. Aspects of dysarthric stop-consonant production, including primary articulator placement and rate of movement, laryngeal function and the respiratory system, are assessed using perceptual and acoustic data.

Acoustic data were obtained from eight adults (4M,4F) with dysarthria (etiologies cerebral palsy, cerebellar ataxia and paralysis) and eight adults (4M,4F) with normal speech and hearing. Subjects recorded isolated, single-syllable utterances containing a word-initial stop followed by a vowel. Auditory-perceptual evaluations of type of voicing, place and manner of articulation, presence of a precursor, and production quality were collected. Visual-perceptual spectrogram assessment was performed and ratings assigned to the following spectrographic attributes: precursor, prevoicing, abruptness of release, time course of release, voice onset time (VOT), and $F1$ and $F2$ transitions. Acoustic measures examine stop burst spectral tilt, initial $F2$ value, $F1$ and $F2$ transitions, multiple stop bursts, prevoicing, VOT, $F0$, and airway pressure control (intraoral and lung).

Perceptual data yield a stop "goodness" score for each speaker, reflecting accuracy and quality of stop production. Poorer spectrographic attribute ratings are correlated with poorer stop goodness scores. The attributes most highly correlated with stop goodness for voiceless stops: time course of release (TCR) and VOT; for voiced stops: precursor, abruptness of release, TCR and time course of $F2$ rise. These dysarthric speakers often generated excessive noise near the release. This noise may be attributed to prolonged frication or aspiration, or faulty velopharyngeal port manipulation. Acoustic measures of prevoicing correspond to auditory-perceptual precursor, and VOT to type of voicing. Airway pressure control difficulties may be due to formation of ejective rather than pulmonary releases and/or difficulty maintaining subglottal pressure. In summary, qualitative and quantitative acoustic correlates of perception could be identified in the speech of dysarthric speakers, and hypotheses were drawn regarding articulatory difficulties. This research has implications for

diagnosis and remediation of disordered speech production. The range of natural variability in the normal baseline has application to speech recognition and synthesis.

Thesis Supervisor: Kenneth N. Stevens, Sc.D.
Title: Clarence J. LeBel Professor of Electrical Engineering

Acknowledgments

My deepest gratitude and appreciation go to my thesis advisor, Professor Ken Stevens. It has been a great joy and honor to work with him over the past eight years. By example, Ken has shared not only his love of speech research, but also his philosophy on how to collaborate with colleagues and conduct oneself in the academic environment.

I would like to thank my committee members, Bill Peake and Joe Perkell, for their guidance, helpful suggestions, and encouragement during the course of my research. I would also like to acknowledge the contributions of Tom McMahon, a committee member until the time of his death. His assistance in hypothesis development for my research proved invaluable at the initiation of this thesis. I am deeply appreciative of the existence of the Harvard-MIT HST/MEMP program, within which I can pursue my simultaneous interests in medicine and electrical engineering.

A special thank you to the subjects, particularly the dysarthric speakers, for their time and willingness to have their speech recorded. Also, thank you to Hwa-Ping Chang for recording the dysarthric subjects' speech. Thank you to Hale Ozsoy, Mengkiat Goh, Adrienne Prahler, and Dameon Harrell for their assistance with developing and utilizing the data collection and digitizing processes for this data.

Krishna Govindarajan has been invaluable for his computer support, particularly with the *xkl* software program. Thank you to Melanie Matthies for her last-minute, heroic statistical analysis efforts on my behalf. A special thank you to Arlene Wint, who, through the years, has been a source of support and guidance through the maze of MIT requirements. Thank you from the bottom of my heart to Jane Wozniak and Majid Zandipour, for their wonderful support as friends and fellow first-time parents, and to their daughter, Sofia, for being my son Nathaniel's first friend. Thank you to Marilyn Chen for always being there for last-minute photocopying, babysitting or trips to the library. Last, but not least, thank you to all the members of the Speech Communication Group, past and present, for making the laboratory such an excellent place to work: Abeer Alwan, Suzanne Boyce, Helen Chen, Marilyn Chen, Karen Chenausky, Harold Cheyne, Elizabeth Choi, Laura Dilley, John Gould, David Gow, Heather Gunter, Helen Hanson, Mark Hasegawa-Johnson, David Horowitz, Wil Howitt, Jeff Kuo, Sharlene Liu, Sharon Manuel, Noel Massey, Jay Moody, Mike O'Connoll, Stefanie Shattuck-Hufnagel, Janet Slifka, Walter Sun, Jennell Vick, and Lorin Wilde.

To my dad, whom I have always wanted to emulate by being called "Dr. Poort" too, and to my mom, who passed on the love of learning and the discipline required to achieve this goal; I am forever grateful. Also, thank you to my sister and her husband, their three children, and my grandparents for their unwavering support and belief in me through the years.

My final, and biggest, thank you is to my husband, Brian Sperry. Through everything from taking care of Nathan to creating plots and typing my thesis, he has made this thesis possible. I am extraordinarily fortunate that he is part of my world. We've had a busy past couple of years, filled with two doctoral dissertations and a baby. I look forward to what the future will bring for as a family. A final thank you

to my wonderful son, Nathaniel, who, by example, reminds me daily that the most important things in life can be summed up by seeing your child's smile.

I am grateful that this work is supported in part by NIH/NIDCD, HST and a PEO Scholar Award.

Biographical Note

Kelly Lynn Poort was born in Sedalia, Missouri, on July 22, 1968. Her parents are Stephen Milton Poort and Donna Marie Dunlap Poort. She spent the majority of her childhood in Ottumwa, Iowa, and graduated from Ottumwa High School in 1986. She received a Bachelor of Science in Electrical Engineering, a Bachelor of Science in Biomedical Engineering, and a Minor in Chemistry from The University of Iowa, Iowa City, Iowa, in May, 1991. She married Brian John Sperry in August, 1991. She received the S.M. degree in Electrical Engineering and Computer Science in September, 1995, and the Electrical Engineer degree in June, 1996, from the Massachusetts Institute of Technology, Cambridge, Massachusetts. On June 4, 1999, she and her husband became the parents of Nathaniel Poort Sperry. She received a Doctor of Philosophy degree in Electrical and Medical Engineering in June, 2000, from the Harvard University - Massachusetts Institute of Technology Division of Health Sciences and Technology Medical Engineering Medical Physics (HST/MEMP) Program. She also satisfied the doctoral requirements of the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

To my husband,

Brian John Sperry

And our dissertation baby,

Nathaniel Poort Sperry

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Dysarthria comprises a group of speech disorders resulting from disturbances in muscular control. These disorders are caused by damage to the central or peripheral nervous system and are characterized by slow, weak, imprecise, and/or uncoordinated movements of the speech musculature regulating speech breathing, voicing, articulation and nasality (Darley et al., 1975). The acoustic speech signal is a very important source of information for objective, quantitative description of certain aspects of speech movement control in dysarthria. From analysis of a dysarthric patient's speech, the motions of the articulators (tongue, lips, lower jaw, larynx, and respiratory system) can be inferred from acoustic measures such as segmental durations or shifts in frequencies of spectral prominences. Rapid changes in manner of articulation are often reflected by clear boundaries in acoustic waveform and spectrographic records. These boundary delineations make it practical to obtain objective measures of speech segment durations in dysarthria (Lehiste, 1965, and others). A wide range of acoustic parameters related to laryngeal control may be extracted by means of computer-based analysis. For example, fundamental frequency range, glottal amplitude and period perturbation (Ludlow and Bassich, 1984), and harmonic-to-noise ratio (Yumoto et al., 1984). It is also possible to approximate the temporal and spatial aspects of vocal tract area in dysarthric speech patterns from measures of vowel formant frequency

(Kent et al., 1979) or fricative-consonant spectral pattern (Weismer, 1984).

Acoustic analysis is appealing clinically because acoustic data can be obtained simply, noninvasively, and relatively inexpensively. Acoustic analysis of the speech of neurologically-impaired patients may be useful in a variety of ways: (1) facilitating early detection of neurologic damage and identifying subclinical manifestations of neurologic disease (Ramig et al., 1988); (2) contributing to the differential diagnosis of disease of various neural subsystems; (3) quantifying a dysarthric individual's intelligibility, i.e., measuring how well the patient's speech would be recognized by a listener (Kent et al., 1989); (4) focusing the treatment plan in order to develop effective and efficient rehabilitation programs (Ansel and Kent, 1992); (5) enabling longitudinal comparison of a patient's speech, in order to assess improvement due to therapy or to document progressive degeneration, for example that which is attributable to particular neurologic diseases or the use of specific medications; and (6) utilizing the acoustic measurements in a device which would act as a "translator", recognizing the patient's speech then either synthesizing speech sounds which are more readily understood by the listener or enabling operation of various devices, such as computers, upon verbal command. Although the perceptual skills of the speech pathologist contribute significantly to these goals, it may be possible to develop acoustic and physiologic analyses that provide more sensitive and quantitative data on the functioning of the speech motor system, which would then supplement information provided by the speech pathologist.

Quantitative acoustic analysis becomes more challenging to perform as the severity of the dysarthria increases, since the speech tends to contain more and more idio-syncratic features and within-subject variability. In order to perform almost all quantitative acoustic analyses of dysarthria, they must be restricted to virtually error-free (fluent) utterances to facilitate making acoustic measurements (Weismer and Liss, 1991). Consequently, important information about the nature of the more severe dysarthrias is lost. (It bears pointing out that this problem is for instrumental measurements in general, and is not specific to acoustic analysis.) To circumvent this problem when analyzing more severe dysarthric speech, an appropriate strategy

might be to first utilize a coarser grain of analysis (i.e., more qualitative than quantitative, such as visual inspection of spectrographic characteristics) which might reveal immediately accessible characteristics of dysarthria as well as point to quantitative analyses that might be useful (Weismer and Liss, 1991) (Refer to Fig. 1-1).

Perceptual approaches are particularly useful for providing integrated measures of overall speech disability such as intelligibility, naturalness, rate, and general articulatory adequacy (Yorkston et al., 1988). However, perceptual measures do have some notable disadvantages, such as (taken in part from Rosenbek and LaPointe (1985)): (1) trained judges are required; (2) perceptual measures are subjective, since they are based on the judge's interpretation of what he/she heard; (3) it is difficult to separate premorbid characteristics (age, medical and social history) from those that are related to the neurologic problem; (4) perceptual characteristics may be present in some patient and environmental conditions and not others; (5) certain symptoms influence others (i.e., severe articulation problems may influence judgments of hypernasality); and (6) a single perceptual end-product may be the result of any of a number of underlying physiological events.

Primarily due to the last point in the previous paragraph, but also to a lesser extent due to the other listed disadvantages related to perceptual measures, speech scientists caution against making inferences about physiological phenomena from perceptual measurements alone (Duffy, 1995). Since both the diagnosis and the remediation of dysarthria involve determining the incorrect physiologic movements of the articulators, there is strong argument for incorporating instrumental measures (of which acoustic analysis is a subset) into the evaluation of a dysarthric patient, supplementing the information obtained from standard perceptual measures. The instrumental measures would aid in describing breakdown in speech subsystems and guide dysarthric management (Gerratt et al., 1991).

Instrumental measures include acoustic, aerodynamic and physiologic measures. The role of the instruments is not to measure integrative activities, but rather to "bring us closer to events in the peripheral speech mechanism... [and] leave us guessing less about the neuromuscular deficits underlying the perceptual symptoms" (Rosen-

bek and LaPointe, 1985, p. 112). Instrumental measures tend to be more sensitive, quantitative and objective than perceptual measures. On the other hand, instrumental measures can be expensive, often require specialized training, may be invasive, and may have limited application (Zeplin and Kent, 1996).

In an attempt to elicit the motor control and coordination difficulties of dysarthric speakers, the speech sound selected for investigation in this research is one characterized by its dynamic, not static, nature. Stop consonants have been chosen as the focus of this study, since they contain both sequential and simultaneous production events. Stop consonants are produced by closing off the oral cavity, blocking (or "stopping") the flow of air through the mouth for a period of time. Simultaneously, the velopharyngeal port is elevated, preventing airflow through the nasal cavity. These articulatory gestures are the only gestures required to produce a postvocalic stop consonant. Prevocalic and intervocalic stops also require that pressure build up behind the oral closure until a rapid opening of the closure releases the intraoral pressure, creating a sudden, brief flow of air. The closure or complete constriction that is formed to block the airflow is made at a point between the lips and the pharynx. In English (as well as in many other languages), there are three places of articulation where the constriction can be located: the lips, the tongue tip against the alveolar ridge and the body of the tongue against the palate. Stop consonants are further distinguished by whether they are voiced or voiceless. Several cues are utilized by the listener to identify stops as voiced rather than voiceless: the presence of vocal-fold vibration well into the closure interval, a shorter VOT (voice onset time, which is the time between the release of a stop closure and the onset of voicing for the following vowel), lengthening of the vowel preceding the stop, and a lower final value for the first resonant or formant frequency ($F1$) of the preceding vowel (Ohde and Sharf, 1992). The relative importance of each cue varies with the phonetic environment. A summary of the classification of English stop consonants appears in Table 1.1.

The production of an intervocalic stop consonant can be considered to consist of four consecutive phases (based on physiologic events): the onset of closure, when one articulator is approaching the other; the closure, when the articulators are held

Place of Articulation	Voiced	Voiceless
Labial	/b/	/p/
Alveolar	/d/	/t/
Velar	/g/	/k/

Table 1.1: Classification of English stop consonants by place of articulation and voicing.

together, completely obstructing the airflow and creating a pressure buildup behind the constriction; the offset of closure, initiated by the rapid release of the articulator that formed the constriction; and the subsequent movement of the articulators (particularly the tongue body) toward configurations appropriate for the following vowel. Production of a prevocalic stop primarily involves the latter three phases, and production of a postvocalic stop requires only the first two phases, sans pressure buildup. Depending upon the voicing characteristics of the particular stop consonant, various adjustments in the glottal opening, vocal-fold stiffness, and vocal-tract wall stiffness accompany the actions of the lips, tongue blade and/or tongue body.

Acoustic analysis of the speech of individuals with dysarthria is appealing to speech scientists because vast literature already exists on the normal aspects of speech acoustics, to which the dysarthric acoustic data can be compared. Relevant to this research, theoretical models have been developed in the past to describe the articulatory, aerodynamic and corresponding acoustic events occurring during each phase of normal stop-consonant production (stop-consonant production by individuals with normal speech and hearing). The models can be classified according to the frequency ranges involved. The low-frequency model accounts for the vocal-tract pressures and airflows generated by the relatively slow-moving articulators. The high-frequency models account for the filtering of the acoustic signal by the vocal tract and the resultant acoustics produced.

1.2 Literature Survey

Dysarthria was initially characterized by physicians, who viewed it as a sign or symptom of disease. As long ago as 1877, Charcot described "scanning speech" as one of

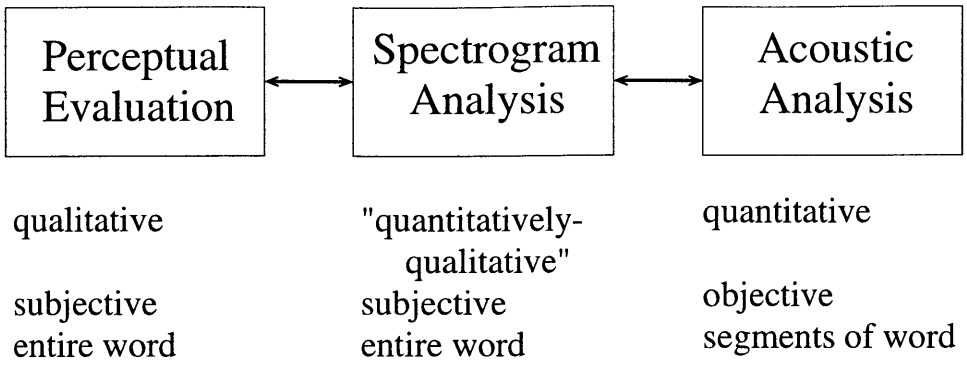

Figure 1-1 : Methods of assessing the acoustic speech signal: auditory-perceptual, visual spectrogram analysis and objective acoustic analysis.

a triad of symptoms in his multiple sclerosis patients (Charcot, 1877). The illness or disease model, frequently employed in the medical field, has traditionally been applied to dysarthrias. According to the illness model, the severity of the dysarthria is associated with the severity of the illness or disease process, and the dysarthria is managed by treating the disease. Thus, dysarthria has been used as an index of disease severity in the past, but little attention was focused on remediation of the speech disorder itself (Yorkston et al., 1988).

Then, in the late 1960s, Darley and colleagues (1969a,b; 1975) at the Mayo Clinic made perhaps the single most important contribution to the study of dysarthrias to date by determining the perceptual speech characteristics associated with a wide variety of neurological conditions. This work demonstrated that major forms of dysarthria could be distinguished by their auditory-perceptual characteristics and that, therefore, the nature of the speech disturbances could be used to infer the site of the lesion. The perceptual characteristics could also be used to guide therapy aimed at improving various aspects of the speech. The perceptual ratings developed by Darley and colleagues remain the primary basis for clinical categorization, rating of severity of the dysarthria, and choice of therapeutic intervention of dysarthrias today (Gerratt et al., 1991; Zeplin and Kent, 1996).

The use of acoustic analysis to evaluate dysarthric speech has a fairly long history. One of the first, if not the first, studies to apply acoustic analysis to the speech of dysarthric speakers was performed by Lehiste (1965). This study is quantitative at

the feature level, recording the number of times speakers made errors, such as nasalization of non-nasal consonants, within a given word list. The study, however, does not attempt to quantify deviations in acoustic measures, such as formant-frequency transitions, from normal speech. It also does not attempt to relate the feature-level observations to the corresponding articulatory movements.

By the time the mid- to late-1980's arrived, a comprehensive list of acoustic measures and associated word intelligibility[1] tests had been developed to evaluate dysarthric speech (Kent et al., 1989). The word intelligibility test designed by Kent et al. for mildly- to moderately-dysarthric individuals examines "19 acoustic-phonetic contrasts that are likely to (a) be sensitive to dysarthric impairment and (b) contribute significantly to speech intelligibility". The test is a multiple-choice single-word close-set (forced-choice) test. It is based on a list of 70 words, appearing in alphabetical order in Appendix A. The test investigates the production of a single word (one of the words from the 70-word list) by placing that target word in a random ordering with three other words, or foils, in each row of the test. The foils differed from the target word by one, or occasionally two, phonetic features. Then, the listeners were asked to circle which of the four words in each row best represented what they heard the speaker to say. The test consists of 70 rows, one row for each word from the corpus.

Chang (1995) utilized this word intelligibility test (after modifying two of the foils) to assess word intelligibility of the eight dysarthric speakers used in the present thesis. Chang recorded the 70-word corpus spoken 8–10 times by each speaker. Details of the recording process, including how it was modified for two of the speakers with dyslexia, are in Section 3.2 of Chang (1995) and summarized in Section 4.1.3 of the present thesis. Descriptions of the eight dysarthric speakers appear in Chapter 2, Section 2.2, of the present thesis. Five listeners, native English speakers not familiar with the speech of dysarthric speakers, performed the word intelligibility test for one repetition per word per dysarthric speaker[2]. The results, shown in Figure 1-2 indicate

[1]Kent et al. (1989) defines intelligibility as "the degree to which the speaker's intended message is recovered by the listener".

[2]The author utilized the same dysarthric speakers as Chang (1995), a subset of words selected

the number of words identified correctly out of a total of 350 words (5 listeners × 70 words/listener) for each dysarthric speaker, expressed as the percent correct. The dysarthric speakers are in order of decreasing intelligibility, from left to right, and are assigned identifiers indicating this order, within sex. It is observed that the dysarthric speakers can be divided into two groups based upon the results of this word intelligibility test. The first group, which could be considered to be more mildly dysarthric, is comprised of the four speakers on the left (DF1, DM1, DF2 and DM2), having word intelligibility percentages of 97, 95, 89 and 82%, respectively. The second group, considered to be moderately dysarthric, is comprised of the four speakers on the right (DF3, DF4, DM3 and DM4), having word intelligibility percentages of 64, 61, 60 and 57%, respectively.

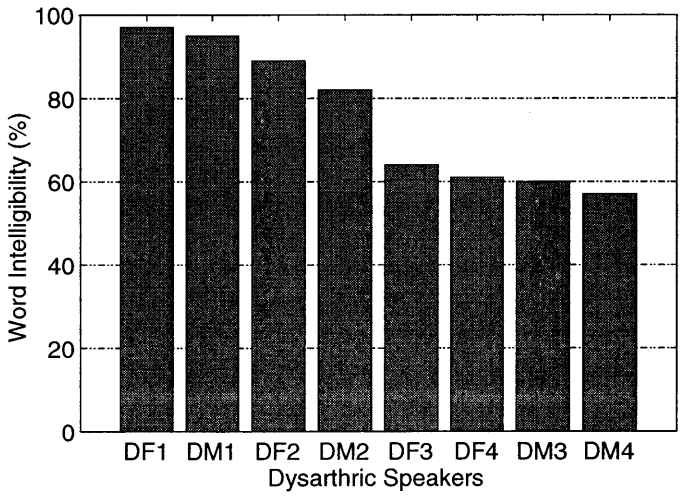

Figure 1-2 : Word intelligibility data for the eight dysarthric speakers (4M, 4F) from Chang (1995), Table 4.1. The data are expressed as the percent identified correctly out of a total of 350 words (5 listeners × 70 words/listener). Speakers are organized from left to right in order of decreasing word intelligibility and are assigned identifiers to indicate this ordering numerically, within sex. For example, DF2 = the Dysarthric Female speaker with the second-highest word intelligibility among the four female dysarthric speakers. These eight dysarthric speakers, saying a subset of these utterances (although not these particular repetitions), are also utilized in the present thesis.

A thorough literature search identified only one study in the past decade which addressed clinicians' use of acoustic analysis in the management of dysarthric patients. The study was performed by Gerratt et al. (1991). The study consisted of compi-

from the 70-word corpus, and different word repetitions than Chang, to examine in Chapters 4–8.

lation and interpretation of a questionnaire distributed to clinicians in each United States Department of Veterans Affairs Medical Center with a Speech Pathology Service. Through the questionnaire, the investigators sought knowledge of the volume of clinical services provided to dysarthric patients, methods employed, instrumental resources, and attitudes of the clinicians about methods for speech assessment. For the portion of the questionnaire related to the use of acoustic analysis, the clinicians were asked to rate, on a 5-point scale, the *clinical value*, *frequency of use*, and, *if currently unavailable in the clinic, predicted use*, of various acoustic measures. The acoustic measures included oscilloscopic, spectrographic, and computer analysis methods for measurement of articulation, voice and prosody, as well as special purpose devices such as Visi-Pitch or the PM Pitch Analyzer for measures of voice, and nasalence measurement of nasal resonance.

Questionnaire results indicated that instrumental measures (including acoustic measures) were judged lower in clinical value, and were used less often, than auditory-perceptual measures. However, when instruments were used, Visi-Pitch was one of the two instruments used most often. The general lack of instrument use is thought to be due to a combination of: (1) scarcity of instrumentation; (2) inability to use the instrument; and (3) clinician preference. Because ratings of *clinical value* and *if currently unavailable, predicted use* exceed *frequency of use* for each acoustic measure, it appears that lack of instrumentation is the most important reason for the infrequent use of acoustic measures. Consistent with this hypothesis, the questionnaire revealed that computer resources are generally poor in the clinics, with most clinics having only one or two computers. Also, fewer than 1 in 10 clinics had an analog-to-digital converter necessary for computer processing of speech signals. (Computer interfaces built into single-purpose devices such as Visi-Pitch were not counted since they are inaccessible for general purpose computer processing.) In addition to the problem of lack of instrumentation, clinicians may possess a limited understanding of the relevancy of instrumentally-acquired data (Coelho et al., 1994) or may perceive instrumental measures as not justified in the management of dysarthric patients because they are indirect measures whose predictive value has not been established

(McNeil, 1986).

Although most of the clinical applications of acoustic analysis referred to in the Gerratt et al. (1991) study are in the area of diagnosis of the dysarthrias, a very important therapeutic application has also recently emerged. When the results of acoustic analysis are displayed on a computer monitor, they can be useful for visual biofeedback. This real-time biofeedback involves the patient attempting to make aspects of his/her speech match various aspects of an acoustic waveform, such as its appearance or duration, the pitch contour, or the loudness level. This type of biofeedback program, in which the patient receives instantaneous and continuous information about his/her neuromotor behavior, may be the most desirable for shaping behavior toward a desired goal (Berry and Goshorn, 1983). A clinical example of acoustic analysis utilized in biofeedback is found in Hodge and Hall (1994). They reported that an 11 year old male, with dysarthria secondary to near-drowning, successfully interpreted the real-time visual feedback of acoustic waveform duration and amplitude displayed on a computer monitor. He then was able to use that biofeedback to help him modify his speech to meet specified requirements, i.e., to shorten the duration of certain sounds.

As a final note, advances of any type that would further the understanding of the speech of individuals with dysarthria have been hindered by the lack of substantial amounts of research in this area. Strand and Yorkston (1994) conducted a review of the dysarthric literature published from 1982 to 1991 and concluded that there is a striking paucity of articles related to dysarthria, compared to studies conducted on other communication disorders. With the exception of editions of proceedings of biennial clinical dysarthria conferences, only 45 data-based articles appeared in the literature during those years. When proceedings are also included, the number of manuscripts reaches a final total of only 86. Fewer than half (43%) of those manuscripts report acoustic data of any kind. Even in those manuscripts which do report acoustic data there is no consistency in reporting the data. It is reported primarily as dependent variables in the studies and rarely is used in the description of a subject or as a criterion for group selection.

1.3 Statement of Purpose

On a fundamental level, this thesis takes an initial step toward addressing the question, "What are the differences between stop consonants produced well and those produced poorly?" This question begins to be addressed by the thesis objectives described in the following paragraphs.

One goal of this thesis is to refine the theoretical models of stop-consonant production so that these models can specify the range of articulatory inputs and of acoustic outputs that are produced by adult speakers with no known speech or hearing disorders. For the most part, these models have been developed previously, with the aid of articulatory, aerodynamic and acoustic data. The high-frequency models will be extended through acoustic analysis of a series of utterances produced by a number of normal speakers. Some of the acoustic variability naturally occurring in the stop-consonant production of these speakers will be characterized by determining the ranges for several of the high-frequency model parameters. Normal variability in articulatory movements will then be inferred from examination of the acoustic variability.

A second goal of this thesis is to characterize motor control and coordination difficulties of dysarthric speakers through the development of acoustic measures which reflect articulatory movements. The model parameter ranges established for normal speakers will provide a baseline against which stop production by individual dysarthric speakers is evaluated. Hypotheses of incorrect articulatory movements will be developed to explain some of the deviations from normal observed in the acoustic measures.

In the context of this second goal, strategies to quantify the differences between normal and dysarthric stop-consonant production will be pursued. Quantification of these differences could supplement auditory-perceptual assessment, aiding clinicians in the determination of how a particular production deviates from the norm in terms of articulatory, laryngeal and respiratory movements. A quantitative baseline of an individual's speech production could be established, facilitating longitudinal compar-

ison in order to assess stability, therapeutic improvement, or deterioration due to progressive neurological disease or the use of specific medications. A final application of quantifying these differences is to enable visual biofeedback, as a therapeutic aid.

These thesis objectives are only a first step in the diagnosis and remediation of dysarthric speech production. The information gathered in this thesis, as well as further research in this area, must be combined with additional medical information from sources such as the patient's medical history, auditory-perceptual evaluations from a speech-language pathologist, and neurological examinations ascertaining lesion location (potentially with the aid of imaging modalities), to make a diagnosis of type and severity of the dysarthria.

1.4 Thesis Outline

In Chapter 2, the dysarthric speakers utilized in this study are introduced. Brief medical histories are provided for each subject. The three primary types of dysarthria exhibited by these individuals are discussed further. For each of these types of dysarthria, lists of deviant speech characteristics are given.

In Chapter 3, the existing theoretical models of stop-consonant production are presented. The models can be classified according to the frequency ranges involved. The low-frequency model accounts for the vocal-tract pressures and airflows generated by the relatively slow-moving articulators. The high-frequency models account for the filtering of the acoustic signal by the vocal tract and the resultant acoustics produced.

In Chapter 4, the perceptual experiment and results are presented. Several aspects of stop production were evaluated by the listeners, including the presence of a precursor (a subject-generated sound prior to the stop release); voicing, place and manner of articulation of the stop; and the quality of the stop production. Results are presented for each of the aspects individually as well as in combination, and measures of stop intelligibility and stop goodness (an assessment of how well the correctly-identified stop are produced) are developed.

In Chapter 5, a visual-perceptual assessment of spectrograms is performed. Seven

attributes were formulated to characterize various aspects of stop production. The attributes are precursor, prevoicing (vocal-fold vibration prior to the release), abruptness of release, time course of release, voice onset time (VOT), time course of $F1$ rise, and time course of $F2$ change. Judges rated these attributes for each spectrogram of the normal and dysarthric speakers. Rating results are correlated with the stop goodness measure of Chapter 4. (Chronologically, Chapter 5 occurred after Chapter 6.)

In Chapter 6, acoustic measures are developed, based on parameters of the high-frequency acoustic models. The acoustic measures assess certain aspects of the speech system during stop production, including the placement of the primary articulator, the rate of movement of the primary articulator, the laryngeal system, and the respiratory system. The results of the acoustic measures applied to normal speech serve as a baseline for comparison with the speech of dysarthric individuals. Results for both normal and dysarthric speakers are interpreted in terms of the information they reveal about articulator control and coordination. (Chronologically, Chapter 6 occurred before Chapter 5.)

In Chapter 7, the results are considered for each individual dysarthric speaker. Perceptual evaluations, spectrogram attribute ratings, and acoustic measure results are interrelated on a speaker-by-speaker basis.

In Chapter 8, the results are summarized, contributions are indicated, and suggestions are given for future research.

Chapter 2

Speaker Dysarthrias

One of the goals of this thesis is to determine the acoustic-to-articulatory mapping that describes the relationship of articulatory movements to resultant acoustic signals produced by dysarthric speakers. In this context, this chapter presents deviant speech characteristics for the three distinct types of dysarthria known to be exhibited by the speakers of this study. These types of dysarthria are spastic, ataxic and athetoid. The manner in which the dysarthric speakers deviate from normal in perceptual, acoustic and physiologic speech characteristics guides the experimental protocol in this thesis as a whole, including the types of questions asked during the perceptual experiment of Chapter 4, the design of the attributes in the spectrogram analysis of Chapter 5, and the development of the quantitative acoustic measures in Chapter 6. The experiments and measures in this thesis were not specifically designed to diagnose type of dysarthria, discriminate between different types of dysarthria, discriminate between different types of dysarthria, nor identify the location of the neurologic lesion; however, the results of the experiments may guide future work in these areas.

Section 2.1 contains descriptions of each of the three types of dysarthria the subjects in this study are known to exhibit. Deviations from normal with regard to respiration, the laryngeal system, and articulation are noted for each type of dysarthria. Section 2.2 consists of all that is known about the medical history, speech characteristics and overall motor involvement for each dysarthric subject.

Type	Localization	Neuromotor basis
Flaccid	Lower motor neuron (final common pathway, motor unit)	Weakness
Spastic	Bilat. upper motor neuron (direct & indirect activation pathways)	Spasticity
Ataxic	Cerebellum (cerebellar control circuit)	Incoordination
Hypokinetic	Basal ganglia control circuit (extrapyramidal)	Rigidity/reduced range of movement
Hyperkinetic	Basal ganglia control circuit (extrapyramidal)	Involuntary movements
Unilateral upper motor neuron	Unilateral upper motor neuron	Weakness/ ? incoordination
Mixed	More than one	More than one

Table 2.1 : Major types of dysarthrias. Localization of the neuroanatomic site of the lesion and the neuromotor basis of the disease are indicated for each type of dysarthria. Adapted from Duffy (1995, Table 1-1), and Darley et al. (1969a,b, 1975).

2.1 Types of Dysarthria

The definition of dysarthria that is widely accepted by speech-language pathologists comes from the work of Darley, Aronson and Brown (1969a,b, 1975). They defined dysarthria as "a collective name for a group of speech disorders resulting from disturbances in muscular control over the speech mechanism due to damage of the central or peripheral nervous system. It designates problems in oral communication due to paralysis, weakness, or incoordination of the speech musculature. It differentiates such problems from disorders of higher centers related to the faulty programming of movements and sequences of movements (apraxia of speech) and to the inefficient processing of linguistic units (aphasia) (Darley et al., 1969a, p. 246). A classification scheme for the dysarthrias was also developed by Darley, Aronson and Brown. This classification scheme divides the dysarthrias into seven types, as shown in Table 2.1. The neuroanatomic site of the lesion and the neuromotor basis of the disease are also shown for each type of dysarthria.

Seven of the subjects in this study have four of the seven types of dysarthria listed in Table 2.1: spastic, ataxic, hyperkinetic (athetoid), and mixed (spastic-athetoid).[1]

[1]The type of dysarthria is not known for the eighth dysarthric subject.

A diagnosis of mixed spastic-athetoid dysarthria indicates that there is perceptual evidence for both types of dysarthria in the subject's speech. These types of dysarthria will be discussed in more detail in the following subsections. The information contained in each of the subsections is a compilation of material, including research reviews, from Darley et al. (1969a,b, 1975), Love (1992), Duffy (1995) and Kent et al. (1998).

2.1.1 Spastic Dysarthria

Spastic dysarthria is associated with damage to the direct and indirect activation pathways of the central nervous system (part of the upper motor neuron system), bilaterally. It may be manifest in any or all of the respiratory, phonatory, resonatory, and articulatory components of speech, but it is generally not confined to a single component. In spastic dysarthria, weakness and spasticity combine to slow the muscle movements as well as to reduce their range and force. This type of dysarthria derives its name from the excessive muscle tone or spasticity that is a feature of the disorder.

Three tables are provided to describe various aspects of spastic dysarthria, Tables 2.2–2.4. Although the findings reported in these tables primarily reflect acquired, not congenital, dysarthria, it is believed that these two types of spastic dysarthria are similar enough in adults for the purposes of this thesis that these tables are still relevant. The tables have been adapted to reflect those aspects of spastic dysarthria most likely to influence stop-consonant production. Table 2.2 lists the most deviant speech characteristics encountered in this type of dysarthria by Darley et al. (1969a). Table 2.3 summarizes the primary characteristics that distinguish between spastic dysarthria and other types of dysarthria. Also included in Table 2.3 are the common oral mechanism findings and the patient complaints encountered in spastic dysarthria. Table 2.4 shows a summary of acoustic and physiologic findings in studies of spastic dysarthria.

Dimension	Speech Component
Imprecise consonants*	Articulatory
Monopitch	Laryngeal
Reduced stress	Prosodic
Harshness*	Laryngeal
Monoloudness	Laryngeal-respiratory
Low pitch*	Laryngeal
Slow rate*	Articulatory-prosodic
Hypernasality	Velopharyngeal
Strained-strangled quality*	Laryngeal
Distorted vowels	Articulatory
Pitch breaks*	Laryngeal
Breathy voice (continuous)	Laryngeal
Excess and equal stress	Prosodic

Table 2.2 : The most deviant speech dimensions encountered in spastic dysarthria by Darley, Aronson, and Brown (1969a), listed in order from most to least severe. Also listed is the component of the speech system associated with the deviant speech characteristics. The component "prosodic" is listed when several components of the speech system may contribute to the dimension. The * indicates those dimensions which tend to be distinctive, or more severely impaired, in spastic dysarthria than any other single dysarthria type. Adapted from Duffy (1995, Table 5-4).

Perceptual

Phonation

Strained-strangled voice quality

Articulation-prosody

Slow rate

Physical

Drooling

Weak face & tongue

Patient Complaints

Slow speech rate

Increased effort to speak

Fatigue when swallowing

Table 2.3 : Primary distinguishing speech and speech-related findings in spastic dysarthria. Adapted from Duffy (1995, Table 5-5).

Speech component	Acoustic or physiologic observation
Respiratory (or respiratory/ laryngeal) (based on studies of spastic cerebral palsy)	Reduced: Inhalatory & exhalatory volumes (shallow breathing) Respiratory intake Vital capacity Rate of amplitude variations
Laryngeal	Decreased: Vocal cord abduction during respiration Fundamental frequency variability Hyperadduction of true & false cords during speech
Velopharyngeal	Increased pharyngeal constriction Slow, sluggish velopharyngeal movement Incomplete velopharyngeal closure
Articulatory/rate/prosody	Reduced: Completeness of articulatory contacts Completeness of consonant clusters Speed and range of tongue movement Range of jaw movement Acceleration & deceleration of articulators Tongue strength Articulatory effort for final word stress Frequency & intensity increases for initial word stress SPL contrasts in consonants Voice-onset-time for stops Amplitude of release bursts for stops Overall speech rate Increased: Syllable & word duration Duration of nonphonated intervals Spirantization during stops Prolonged phonemes Slow phoneme-to-phoneme transitions Centralization of vowel formants Voicing of voiceless stops

Table 2.4 : Summary of acoustic and physiologic findings in studies of spastic dysarthria. Adapted from Duffy (1995, Table 5-6).

2.1.2 Ataxic Dysarthria

Ataxic dysarthria is associated with damage to the cerebellar control circuit. It may be evident in any or all of the respiratory, phonatory, resonatory, and articulatory levels of speech, but its characteristics are most evident in articulation and prosody. Its speech characteristics reflect the effects of incoordination and reduced muscle tone on speech, the results of which are slowness and inaccuracy in the force, range, timing, and direction of speech movements. This type of dysarthria reflects a breakdown in motor organization and control, with poorly controlled or coordinated movements, rather than the muscle weakness, resistance to movement or restriction of movement seen in most other dysarthria types.

Three tables are provided to describe various aspects of ataxic dysarthria. Although the findings reported in these tables primarily reflect acquired, not congenital, dysarthria, it is believed that these two types of ataxic dysarthria are similar enough in adults for the purposes of this thesis that these tables are still relevant. Table 2.5 summarizes the most deviant speech dimensions found by Darley et al. (1969a). Table 2.6 summarizes the primary distinguishing speech characteristics and patient complaints associated with this type of dysarthria. Table 2.7 contains general observations derived from acoustic and physiologic studies.

2.1.3 Athetoid Dysarthria

Athetoid dysarthria is associated with damage to the basal ganglia control circuit. Impairments are often identified in every major component of the speech mechanism. Respiratory dysfunction may contribute to limitations in pitch and loudness due to increased subglottal air pressure. Fundamental frequency is raised with increased subglottal pressure. An attempt to conserve respiratory effort may result in substitution of voiced consonants for their voiceless cognates. Laryngeal dysfunction may lead to weak vocal intensity; a voice low in pitch, monotonous, or possessing inappropriate pitch variation; and a forced or breathy voice quality, accompanying an inability to adduct the vocal folds to the midline of the glottis or insufficient tension in the

Dimension	Speech Component
Imprecise consonants	Articulatory
Excess and equal stress*	Prosodic
Irregular articulatory breakdowns*	Articulatory
Distorted vowels*	Articulatory-prosodic
Harsh voice quality	Phonatory
Prolonged phonemes*	Articulatory-prosodic
Monopitch	Phonatory-Prosodic
Monoloudness	Phonatory-Prosodic
Slow rate	Prosodic
Other	
Excess loudness variations*	Respiratory-phonatory-prosodic
Voice tremor	Phonatory

Table 2.5 : The most deviant speech dimensions encountered in ataxic dysarthria by Darley et al. (1969a), listed in order from most to least severe. Also listed is the component of the speech system associated with each characteristic. The component "prosodic" is listed when several components of the speech system may contribute to the dimensions. Characteristics listed under "other" include features not among the most deviant but which were judged deviant in a number of subjects and are not typical of most other dysarthria types. The * indicates those dimensions which tend to be distinctive, or more severely impaired, in spastic dysarthria than any other single dysarthria type. Adapted from Duffy (1995, Table 6-4).

Perceptual
Phonation-respiration
Excessive loudness variations
Articulation-prosody
Irregular articulatory breakdowns
Distorted vowels
Excess and equal stress
Prolonged phonemes
Patient complaints
"Drunk"/intoxicated speech
Stumbles over words
Bites tongue/cheek when speaking or eating
Poor coordination of breathing with speech

Table 2.6 : Primary distinguishing speech and speech-related findings in ataxic dysarthria. Adapted from Duffy (1995, Table 6-5).

Speech component	Acoustic or physiologic observation
Respiratory/laryngeal	Abnormal and paradoxical rib cage and abdominal movements
	Reduced vital capacity (probably secondary to incoordination)
	Increased variability of F_0 (fundamental frequency) and intensity during vowel prolongation
Articulation, rate, & prosody	Reduced rate:
	Increased syllable duration
	Increased duration of formant transitions
	Longer voice onset time (but sometimes shorter)
	Lengthened vowel nuclei
	Difficulty initiating purposeful movement
	Slow lip, tongue, & jaw movements
	Increased variability, inconsistency, or instability of:
	Segment duration
	Rate
	Intensity
	F_0
	Range & velocity of articulatory movements
	Increased instability of force & static position control in lip, tongue, & jaw on nonspeech tasks
	Inconsistent velopharyngeal closure
	Reduced variability or restriction of:
	Anterior-posterior tongue movements during vowel production
	Syllable duration
	Occasional failure of articulatory contact for consonants

Table 2.7 : Summary of acoustic and physiologic findings in studies of ataxic dysarthria. Note that many of these observations are based on studies of only one or a few speakers, and that not all speakers with ataxic dysarthria will exhibit these features. Note also that these characteristics are not necessarily unique to ataxic dysarthria: some may also be characteristics of other motor speech disorders, or non-neurologic conditions. Adapted from Duffy (1995, Table 6-6).

folds. There may also be a lack of phonation resulting from either hyperadduction of the vocal folds or generalized hypertonic muscle contraction immobilizing the entire vocal mechanism. When phonation does occur in this situation, the voice will have a strained quality with initial audible glottal attack accompanied by an inability to sustain phonation. The most frequent oral articulatory abnormalities were (1) large ranges of jaw movement; (2) inappropriate positioning of the tongue for phonetic segments (particularly anterior-posterior positioning) because of a reduced range of tongue movement; (3) inability to finely shape the tongue for consonant articulation; (4) instability of velar elevation (difficulty in achieving velopharyngeal closure and in maintaining velar position); (5) prolonged transition times between articulatory movements; and (6) retrusion of the lower lip.

2.2 Dysarthric Speakers Involved in the Study

The dysarthric speakers utilized in this thesis were originally recruited by Hwa-Ping Chang for his 1995 doctoral dissertation entitled "Speech Input for Dysarthric Computer Users," completed in the Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology. The author is deeply indebted to Chang for recruiting these speakers, recording their speech, and kindly permitting the author to utilize the data recordings in the present thesis.

According to Chang (1995), seven of the eight speakers have both dysarthria and cerebral palsy. Cerebral palsy is defined as a non-progressive disorder of motion and posture due to brain insult or injury occurring in the period of early brain growth, generally under three years of age (Lord, 1984). The categories of cerebral palsy represented in this speaker group include the three major clinical types: spastic, athetotic and ataxic. Some of the speakers exhibit signs and symptoms of more than one type of cerebral palsy, as well. Although information is available from the subjects regarding their type of cerebral palsy, no clinical diagnoses of type of dysarthria are available. In lieu of specific clinical diagnoses, the assumption has been made by the author that the type of dysarthria is the same as the type of cerebral palsy. According

to Love (1992), three major types of dysarthria are generally recognized in cerebral palsy: (1) spastic, (2) dyskinetic (athetoid), and (3) ataxic. Love states that no universal classification system exists for the clinical types of cerebral palsy, therefore many experts currently accept the same major categories for cerebral palsy. Since, in this case, the types of cerebral palsy are known, and agree in name with the three major types discussed in Love (1992), it seems reasonable to assume that the types of dysarthria correspond to the types of cerebral palsy. The exception to this assumption is speaker DF4, with spastic cerebral palsy. It is known from Chang (1995) that she had no speech deficits until ten years prior to the time of the recording, when she had surgery to remove an acoustic neuroma. Her dysarthria results from paralysis of the left side of her face, left side of her tongue, and left vocal fold, secondary to the surgery. Due to a lack of clinical diagnosis, her type of dysarthria will be considered "Unclassified". The eighth subject, DM2, was diagnosed with cerebellar ataxia and ataxic dysarthria.

Cerebral palsy is often associated with many other sequelae that can affect speech production, in addition to dysarthria. Such dysfunctions include disturbances in cognition, perception, sensation, language, hearing, emotional behavior, feeding and seizure control (Love, 1992). The eight subjects in this study were specially selected by Chang (1995) and are presently utilized by the author because their speech production difficulties are purely motor in nature, arising from disturbances in the muscular control of their speech mechanisms. Their cognitive and linguistic abilities are intact, with no evidence of apraxia or aphasia.

This section of Chapter 2 contains all that is known about the medical histories for the eight dysarthric speakers. The subsections, one for each speaker, are taken directly from Chang (1995), with minor changes in wording. The type of dysarthria has been included in each subject's history. In his thesis, Chang investigated the use of speech recognition as a computer interface for dysarthric individuals who have difficulty using a keyboard. Consequently, the medical histories include some information about the typing abilities of each subject. This information is also useful for placing the speech deficits within the context of other motor involvements.

Dysarthric Subject	Sex	Age	Highest Level of Education	Word Int.(%)	Type of Disorder	Type of Dysarthria
DM1	M	61	High school	95	CP	Spastic
DM2	M	38	B.S. degree	82	Ataxia	Ataxic
DM3	M	48	Undergraduate	60	CP	Athetoid
DM4	M	45	M.S. degree	57	CP	Spast. - Ath.
DF1	F	61	B.S. degree	97	CP	Spastic
DF2	F	24	Undergraduate	89	CP	Spastic
DF3	F	22	Undergraduate	64	CP	Spast. - Ath.
DF4	F	62	Fifth grade	61	CP + Para.	Unclassified

Table 2.8 : Dysarthric speaker summary. From left to right, columns contain the following information: Subject identifier; Sex; Age; Highest level of formal education; Word Intelligibility (%); Type of disorder (CP = cerebral palsy; Ataxia = cerebellar ataxia; Para. = paralysis of left side of face, left side of tongue, left ear, and left vocal fold secondary to surgery); Type of dysarthria (Spastic, Ataxic, Athetoid, Spast. - Ath. = mixed Spastic-Athetoid, and Unclassified). Adapted from Chang (1995, Table 1-1).

In addition to the subsections for each speaker appearing below, the dysarthric speakers are summarized in Table 2.8. Within sex, the speakers are ordered from highest to lowest word intelligibility, per the results of a perceptual test conducted by Chang (1995) and reported in Chapter 1, Section 1.2, of the present thesis.

2.2.1 Subject DM1

DM1 is a 61-year-old male with spastic dysarthria. He has earned a high school diploma. His mother gave birth to him at home and had difficulty in childbirth. During his birth, the doctor devoted more attention to saving his mother's life and less attention to taking care of him. Three days later, when his mother gave him a bath, she discovered that DM1 moved abnormally. His neuromotor condition is characteristic of spastic cerebral palsy. His muscles are stiff and his movements are awkward [sic]. His muscles have increased tone with heightened deep tendon reflexes (Dorland's Illustrated Medical Dictionary, 1981). His hands and his legs move inward more than outward [sic]. His neck has involuntary movements. He can type only by using his left index finger, while his right hand holds his left hand steady. Subject DM1's speech is more normal sounding and less throaty than the speech of most of the other subjects.

2.2.2 Subject DM2

DM2 is a 38-year-old male with ataxic dysarthria. He has earned a bachelor's degree. Subject DM2's motor control was not observed to be atypical until he was 1 1/2 years old. When he attempted to walk, his parents discovered that he could not keep his balance. He has a lack of muscular coordination and an irregularity of muscular action consistent with cerebellar ataxia. He requires a T-board and must incline his body forward to stably support his right and left palms while he types at the computer. Otherwise, because of tremors and involuntary movements of his hands, he cannot type accurately. Furthermore, because of the inclination of his body and head, he cannot watch the monitor and keyboard simultaneously. He can use all of his fingers to type, but feels pain and is easily fatigued in typing or programming tasks. His speech is typical of ataxic dysarthria with: (1) intermittent disintegration of articulation and irregularities of pitch and loudness, (2) altered prosody involving prolongation of sound, equalization of syllabic stress (by undue stress on usually unstressed words and syllables), and (3) prolongation of intervals between syllables and words (Yorkston, 1988). However, his lip-jaw coordination is essentially normal (similar to the subject in Abbs et al., 1982).

2.2.3 Subject DM3

DM3 is a 48-year-old male with athetoid dysarthria. He is studying for a bachelor's degree. At birth, his umbilical cord was wrapped around his neck. His respiration ceased for approximately 5 minutes, causing damage to the portion of the cerebellum [sic] controlling motor and speech coordination. His motor control is characteristic of athetoid cerebral palsy: a derangement marked by ceaseless occurrence of slow, sinuous, writhing movements, especially severe in the hands and performed involuntarily (Dorland's Illustrated Medical Dictionary, 1981). Because of tremors and involuntary movements of his hands, he cannot type or use a mouse (or joystick) easily. He uses his nose to type his reports and do analysis jobs with the computer. His speech impairment is indicative of poor respiratory control, exhibiting a forced, throaty voice

quality. He also has a large range [sic] of jaw movements. This subject's speech is nonfunctional for oral communication due to the combined effect of severely reduced oral-articulatory abilities, severely reduced vocal loudness, breathiness, whispered and hoarse phonations, intermittent aphonia, and throaty noise.

2.2.4 Subject DM4

DM4 is a 45-year-old male with mixed spastic-athetoid dysarthria. He has earned a Master's degree. His mother had difficulty during childbirth. Her lung was collapsed for ten minutes. Following birth, subject DM4 had brain damage; however his twin brother was healthy. Subject DM4 had evidence of spastic and athetoid cerebral palsy. His arm and leg muscles move involuntarily. His jaw muscle control is impaired and spastic, causing his upper and lower teeth to grind together. As a result, his teeth are ground down. He can only use his index fingers to type or program on the computer. His speech is very disordered sounding to the unfamiliar listener. His speech is less throaty than the speech of subject DM3. His speech impairment is indicative of poor respiratory control, exhibiting a forced, throaty voice quality. He also has a large range [sic] of jaw and head movements. Some of his words are abruptly terminated by unexpected movements of the larynx or respiratory system. His speech is particularly time variant. Both his speech pattern and his speech rate greatly change from one utterance to the next.

2.2.5 Subject DF1

DF1 is a 61-year-old female with spastic dysarthria. She has earned a bachelor's degree. She has had spastic cerebral palsy from the time of her birth. Her muscles are weak, move sluggishly through a limited range of motion, and have stiff movements. The muscles have increased tone with heightened deep tendon reflexes. However, she can still ambulate by herself. All of her fingers are constringent. She uses her right index finger to type on the keyboard. Her speech is slow and seems to emerge with difficulty. She has airflow and lung vital capacity control problems. After talking for

a period of time, her speech becomes weak and decays in amplitude. Therefore, her speech is quite clear and intelligible in isolated utterances (such as the utterances in this study), but not in continuous communication.

2.2.6 Subject DF2

DF2 is a 24-year-old female with spastic dysarthria. She is studying for a bachelor's degree. At birth, DF2 had evidence of cerebral palsy. Her neuromotor condition is characteristic of spastic cerebral palsy: the muscles are stiff and the movements awkward. Her muscles have increased tone with heightened deep tendon reflexes. DF2's speech is very weak sounding to the unfamiliar listener and less throaty than the speech of DM4. Her speech and muscle movements are similar to those of DF4. To type or program on the computer, she can only use a pencil grasped by her left or right fingers. She also has dyslexia.

2.2.7 Subject DF3

Subject DF3 is a 22-year-old female with mixed spastic-athetoid dysarthria. She is studying for a bachelor's degree. At birth, DF3 exhibited some evidence of both spastic and athetoid cerebral palsy, with most of her symptoms consistent with spastic cerebral palsy. In particular, her neuromotor condition is more characteristic of spastic cerebral palsy: her muscles are stiff and her movements are awkward. Her muscles have increased tone with heightened deep tendon reflexes. She also has contraction of her fingers and rotation of her wrists. Moreover, the involuntary movements of the articulatory and pharyngeal muscles indicate that she should be characterized as both dysarthric and dysphagic (Brain, 1969). She primarily utilizes her right thumb, at times accompanied by her left index finger, to type on the keyboard.

Because of her involuntary and jerky body movements, her speech sometimes becomes discontinuous. Her speech mixes spasticity with athetosis: the grimaces of her face and the involuntary movements of her tongue interfere with articulation, and irregular spasmodic contractions of the diaphragm and other respiratory muscles

give the voice a curiously jerky character due to sudden changes in her airflow during speech. Her slow, rasping, and labored speech is generated with a large range of jaw movement, and each word is prolonged. Her speech is weak sounding to the unfamiliar listener and less throaty than the speech of DM3.

2.2.8 Subject DF4

DF4 is a 62-year-old female with an unclassified type of dysarthria. She has a fifth-grade education. At birth, DF4 had apparent spastic cerebral palsy. Her neuromotor condition is like DF3's: her muscles are stiff, her movements are awkward [sic] with heightened deep tendon reflexes, and she has contraction of her fingers and rotation of her wrists. She can use only her right index finger for typing. However, her speech was intact (unaffected by the spastic cerebral palsy) until ten years ago when she had surgery to remove an acoustic neuroma. Following this operation, the left side of her face, the left side of her tongue, her left ear [sic], and her left vocal fold were paralyzed. Her vocal fold and vocal tract nerves and muscles were damaged and her speech became abnormal and lisping. Her speech has especially poor aspiration control. Subject DF4's speech is very weak sounding to the unfamiliar listener and more throaty than the speech of other subjects. Some of the utterances are generated with very breathy and explosive noise. When producing speech, her face grimaces, as though the sounds are produced against considerable resistance. She also has dyslexia. A clinician's diagnosis of the type of dysarthria she has as a result of her paralysis is not available to the author. Consequently, her type of dysarthria is listed as "Unclassified".

Chapter 3

Stop-Consonant Production Models

This chapter describes existing theoretical models of stop-consonant production. These models map the articulatory movements to the resultant acoustic output. The intention of this chapter is to provide a review of speech production theory as it pertains to stop consonants. For a more thorough discussion of stop-production modeling, as well as speech production theory as a whole, the reader is referred to Stevens (1998). The experiments and analysis of Chapters 4–6 are partially motivated by the modeling described in the present chapter. In particular, the acoustic measures developed and applied in Chapter 6 have their basis in these models. In Chapter 6, Section 6.2.1, the range of variability of several of the model parameters is characterized for a group of eight speakers with normal speech and hearing. Establishing the parameter range of variability across a group of normal speakers contributes to the thesis goal of refining and expanding the existing stop-consonant production models.

Section 3.1 discusses several aspects of a low-frequency mechanical model which portrays vocal-tract movements and their associated airflow and pressure changes. This low-frequency circuit model, consisting of lumped-element parameters, is valid for frequencies up to approximately 30–40 Hz. Section 3.2 considers several models of the sequence of sound sources and the corresponding vocal-tract filtering effects. These high-frequency models are useful for describing events that occur at frequencies

above approximately 250–300 Hz, when lumped-element parameters generally can no longer provide reasonable estimates of the vocal-tract's behavior. Each of Sections 3.1 and 3.2 are divided into subsections that examine several of the model parameters and acoustic outputs in greater detail.

3.1 Low-Frequency Model of the Mechanical and Aerodynamic System

A theoretical, low-frequency model which examines vocal-tract movements, airflows and pressures occurring during stop-consonant production has been proposed by Stevens (1993). This circuit model is valid for frequencies up to about 30–40 Hz and is similar to those developed by Rothenberg (1968), Westbury (1979), and Müller and Brown (1980). Based on physiological information about the vocal tract and knowledge of the articulator movements, the model predicts the time average pressures and airflows generated in the vocal tract throughout stop production. Stevens (1993) determined that the pressures and flows within the vocal tract can be estimated by modeling the vocal tract during consonant production as a tube with two constrictions, one at the glottis and one formed by articulator(s) within the vocal tract, as shown in Figure 3-1(a). A corresponding circuit diagram of the system is given in Figure 3-1(b).

The variables shown in Figure 3-1 are defined as:

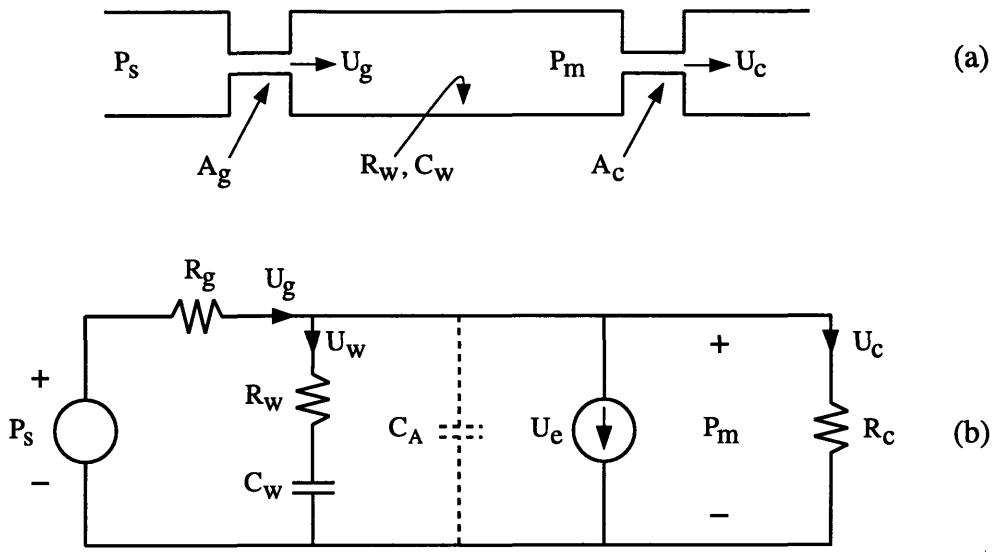

Figure 3-1 : (a) Structural model for estimating average airflows and pressures during consonant production. (b) Equivalent circuit model. (Adapted from Stevens (1993, Fig. 2)).

Variable	Definition
P_s	Subglottal pressure source
R_g	Acoustic glottal resistance
A_g	Cross-sectional area of the glottis
U_g	Glottal airflow
R_w	Acoustic resistance of the vocal-tract walls
C_w	Acoustic compliance of the vocal-tract walls
U_w	Airflow due to inward and outward passive movement of the vocal-tract walls
C_A	Acoustic compliance of the vocal-tract air volume
U_e	Volume velocity source for active muscular contraction and expansion of vocal-tract walls
P_m	Pressure in the mouth
R_c	Acoustic constriction resistance
A_c	Cross-sectional area of the constriction
U_c	Airflow through the constriction

In Figure 3-1(b), the branch containing the acoustic compliance of the vocal-tract air volume, C_A, is shown using dashed lines to indicate that C_A has an effect during only a brief time period following the stop release. The transient sound produced by the discharge of C_A is initiated coincident with the release and has a duration of approximately 1 ms. For a male with a closed vocal-tract volume of about 60 cm^3, C_A is estimated to be 4×10^{-5} cm^5/dyne (Stevens, 1998). This value for C_A is one

to two orders of magnitude smaller than the typical value for C_w. The acoustics of the transient will be investigated in Section 3.2.1.

With minor adjustments, the models in Figure 3-1 are capable of representing three phases of stop-consonant production: the onset of closure, when one articulator is approaching the other; the closure, when the articulators are held together, completely obstructing the airflow and creating a pressure buildup behind the constriction; and the offset of closure, initiated by the rapid release of the articulator that formed the constriction. In this thesis, the main focus of the mechanical modeling is on the time period following release of a prevocalic stop, when the articulators are moving rapidly, accompanied by rapid changes in the acoustic waveform. Several of the model parameters are discussed in greater detail in the subsections below.

3.1.1 Subglottal Pressure and the Respiratory System

The subglottal pressure, P_s, is the principal driving source for the airflow in the vocal tract. A typical range for P_s during normal speech production is 5–10 cm H_20. At the beginning of an expiration (such as immediately prior to release of a prevocalic stop in word-initial position of an isolated utterance) a supraglottal constriction is formed, and the pressure in the lungs is typically about 8 cm H_20 (Stevens, 1998). This value of P_s is believed to be maintained at a fairly constant level throughout production of the entire utterance. Most of the energy for creating the pressure buildup and sustaining that pressure during the utterance comes from the energy stored in the expanded thorax or depressed diaphragm during the previous inspiration. If that inspiration is not sufficient to provide the necessary pressure, then the respiratory musculature must be recruited to provide the airflow needed (Stevens, 1998).

The assumption of a constant P_s throughout the entire production of a stop consonant, however, may not be completely accurate. For example, during production of /p/ in the isolated nonsense syllable /pɑp/, there appears to be a tendency at times for the subglottal pressure to increase as the closure interval progresses (Isshiki and Ringel, 1964). Similar results were found by Hertegård (1994) for the production of /p/ both in repeated /pɑ/ syllables and in three /pɑ/ syllables embedded in a

carrier phrase (without interruptions between syllables). After reaching a maximum value near the end of the closure time period, the subglottal pressure then begins to decrease. This decrease may be initiated immediately prior to or upon release, and is probably associated with the fairly open glottal position required for a voiceless, aspirated stop release. Although the airflow, U_c, is zero during closure, it increases abruptly at the time of the release, becoming quite large (often $>$ 1 l/s) for a brief time interval following release (Isshiki and Ringel, 1964). Part of the rapid airflow can be attributed to expelling the portion of the vocal-tract air volume that expanded during the closure; however, a significant part is thought to be due to the airflow from the lungs, U_g. The decrease in subglottal pressure around the time of the release can be represented by a pressure drop across a linear acoustic resistance, R_s. In Figure 3-1(b), the acoustic resistor R_s would be placed in series with the subglottal pressure, P_s, between P_s and the acoustic glottal resistance, R_g. The value of R_s has been estimated to be somewhere in the range of 1-4 cm H_2O/l/s (Ladefoged, 1963; Rothenberg, 1968, and others).

3.1.2 Acoustic Glottal Resistance

The acoustic glottal resistance, R_g, is the resistance to the flow of air through the glottis. An expression for this resistance appears in Equation 3.1,

$$R_g = \frac{12\mu h}{ld^3} + k\frac{\rho U_g}{2A_g^2} \tag{3.1}$$

where μ is the viscosity of air, h the thickness of the glottal slit, l the length of the glottis, d the glottal width, ρ the density of air, U_g the volume velocity of the airflow through the glottis, A_g the average area of the glottis ($A_g = l \times d$), and k a proportionality constant (Stevens, 1998). The first term of R_g accounts for viscous losses of the air and the second term represents the kinetic resistance due to losses caused by eddy formation at each end of the glottal constriction.

The pressure drop across the glottis can be represented by Equation 3.2,

$$\Delta P = R_g U_g + M_g \frac{dU_g}{dt} \tag{3.2}$$

where the acoustic mass of the air in the glottis is given in Equation 3.3.

$$M_g = \frac{\rho h}{A_g} \tag{3.3}$$

The glottal constriction cross-sectional area, A_g, is the area of the opening between the vocal folds. When the vocal folds are vibrating, A_g represents an average of the glottal opening area created during a given cycle of vocal-fold vibration. The average area, A_g, remains time varying over longer durations, however, as the glottal adjustments necessary for aspiration, voicing, etc., occur during stop-consonant production. The value of A_g is dependent upon the choice of stop and its phonetic environment, as well as the particular speaker. A time period in which the value of A_g is typically changing is in the vicinity of the release of the supraglottal constriction. Prior to the release, as pressure builds up during the closure, outward forces exerted on the upper edges of the vocal folds are believed to cause a passive increase in the glottal area. The average glottal area during this time period can be represented by Equation 3.4,

$$A_g = A_{go} + 2lC_{vf}d_{vf}P_m \tag{3.4}$$

where A_{go} is the average glottal area that would exist if there were no intraoral pressure, C_{vf} the mechanical compliance per unit length of one upper edge of a vocal fold, d_{vf} the effective vertical depth of one vocal-fold edge and P_m the intraoral pressure (Stevens, 1998). As P_m diminishes rapidly following stop release, a passive decrease in the glottal area is thought to occur since the outward forces holding the vocal folds open are no longer present. In addition to these passive forces on the vocal folds, it is possible to have active adjustment of the glottal configuration during stop-consonant production. Some examples of active positioning of the vocal folds include adjustments required to sustain vocal-fold vibrations during the closure interval of a voiced stop; spreading the vocal folds far enough apart to prevent vocal-

fold vibrations during the aspiration noise interval of a voiceless stop, but not so far apart that turbulent noise is not generated; and actively moving the vocal folds closer together to initiate vocal-fold vibrations for the onset of a vowel following stop release. The normal range for glottal area is 0.05 cm^2 (on average) during the modal vocal-fold vibrations that occur in the following vowel and 0.1–0.4 cm^2 during aspiration or breathy voicing.

3.1.3 Supraglottal Cavity Volume

The supraglottal cavity is the region of the vocal tract between the constriction created by the glottis and a supraglottal constriction formed by one or more articulators. Adjustments in the supraglottal cavity volume can be made via passive and/or active movement of the non-rigid vocal-tract walls. Passive movement of the vocal-tract walls occurs in response to changes in pressure within the vocal tract. Active movement is made by the activation of muscle(s) in the walls of the vocal tract. The term "vocal-tract walls" refers to such structures as the inner surfaces of the cheeks and lips, the dorsal surface of the tongue, the floor of the mandible, the inner surface of the velum, and the inner walls of the pharynx. The larynx also has the ability to raise or lower, changing supraglottal cavity dimensions.

The passive movement of the walls of the vocal tract can be represented by an impedance in the circuit model. At low frequencies (up to 30–40 Hz), the impedance of the walls can be approximated by an acoustic resistance R_w in series with an acoustic compliance C_w (Stevens, 1993). Average values are estimated to be R_w = 10 dyne-sec/cm^5 and $C_w = 10^{-3}$ cm^5/dyne for labial and alveolar stop consonants, in which the total surface area of the vocal-tract walls posterior to the incisors is approximately 100 cm^2. For velar stops, average values are R_w = 15 dyne-sec/cm^5 and $C_w = 8 \times 10^{-4}$ cm^5/dyne, where the wall surface area posterior to the velar constriction is believed to be closer to 70 cm^2. These element values are estimated from the data of Rothenberg (1968), Ishizaka et al. (1975), and Glass (1986). (The passive effects of the non-rigid vocal-tract walls at higher frequencies are discussed in Section 3.2.3.) The active movement of the walls of the vocal tract causes voluntary

expansion or contraction of the supraglottal cavity volume. The effect of this volume change is represented in the circuit model by the volume velocity source U_e, which is positive if there is an active expansion and negative if there is an active contraction of the volume.

To produce a prevocalic (or intervocalic) stop consonant, intraoral pressure must build up during the closure interval. Voiced and voiceless stop consonants require different articulatory adjustments in order to achieve this pressure buildup. To sustain vocal-fold vibrations during the closure interval of a voiced stop, a transglottal pressure difference must be maintained. The mechanism used to maintain this pressure differential may be active enlargement of the supraglottal vocal tract and/or relaxation of the supraglottal musculature resulting in a passive expansion of the supraglottal cavity (Svirsky et al., 1997). For voiceless stops, the objective following closure is to quickly terminate glottal vibrations via spreading the glottis and stiffening both the vocal folds and the vocal-tract walls. The mechanism for spreading the glottis and stiffening the vocal folds is believed to have both a passive component, due to the intraoral pressure pushing the vocal folds apart, and an active component, due to activation of the vocal-fold musculature. The increased wall stiffness is thought to be achieved through active involvement of the supraglottal musculature, inhibiting outward displacement of the vocal-tract walls (Svirsky et al., 1997).

3.1.4 Acoustic Constriction Resistance

The acoustic constriction resistance, R_c, is the resistance to the flow of air through a supraglottal constriction. In English stop consonants, the constriction can occur at any one of three possible locations in the vocal tract: the lips, the tongue tip against the alveolar ridge and the body of the tongue against the palate. Since the shape of the constriction immediately following the release is not known, two different shapes will be considered, circular and rectangular. The resistance R_c consists of two parts, a viscous component and a kinetic component. The formula for the viscous component depends upon the shape of the constriction. If the constriction is assumed to be rectangular, the viscous component is given by Equation 3.5,

$$R_{viscous} = \frac{12\mu l_c}{bd^3} \tag{3.5}$$

where μ is the viscosity of air, l_c the length of the constriction, b the larger dimension of the rectangular constriction, and d the smaller dimension. (This equation assumes $d \ll b$.) If a circular constriction is assumed instead, then the formula for the viscous component is given by Equation 3.6,

$$R_{viscous} = \frac{128\mu l_c}{\pi D^4} \tag{3.6}$$

where D is the diameter of the circular cross section.

The kinetic component of the resistance represents energy losses due to the transitions from narrow to wide vocal-tract cross-sectional dimensions at each end of the constriction. This kinetic resistance is shown as the second term in Equation 3.7 for the overall resistance, R_c:

$$R_c = R_{viscous} + k\frac{\rho U_c}{2A_c^2} \tag{3.7}$$

where ρ is the density of air, U_c the volume velocity of the airflow through the supraglottal constriction, A_c the supraglottal constriction cross-sectional area, and k a proportionality constant.

In order to determine the pressure drop across the constriction following the stop release, the acoustic mass of the air within the constriction should be taken into account. The drop in pressure across the constriction is shown in Equation 3.8,

$$\Delta P_c = R_c U_c + \frac{d}{dt}\left(\frac{\rho l_c U_c}{A_c(t)}\right) \tag{3.8}$$

where the time dependence of the supraglottal constriction cross-sectional area after the release, A_c, is explicitly denoted (Massey, 1994; Stevens, 1998).

The constriction cross-sectional area, A_c, is time-varying following the release and depends upon the stop produced, as well as its phonetic environment. A method has been developed to estimate a linear rate of increase for A_c from acoustic data with the aid of models (Poort, 1995). The method is outlined as follows: (1) An initial

linear rate of increase for A_c is estimated. This initial estimate could be taken from a number of sources, including articulation data; (2) The initial estimate for A_c is used as a parameter in the expression for R_c. (Refer to Equations 3.7 and 3.8. In Poort (1995), the $R_{viscous}$ and acoustic mass terms were neglected. Additionally, k was set equal to 1.0 based on Stevens (1998).) Then R_c is a parameter in the circuit model of Figure 3-1(b), from which average pressures and airflows in the vocal tract are calculated; (3) Utilizing A_c and the calculated airflow U_c, the amplitude of the frication noise source following the release can be computed. The source amplitude is approximately proportional to $U_c^3 A_c^{-2.5}$, based on empirical data with some theoretical support (Fant, 1960; Stevens, 1971; Shadle, 1985; Pastel, 1987, and others). The model predicts that the noise burst amplitude rises to a peak within the initial few milliseconds following the release, then decreases rapidly as A_c continues to increase. The duration of the burst is measured as the time interval during which the amplitude of the noise continuously remains within 10 dB of the maximum noise amplitude; (4) The linear rate of increase for A_c is adjusted, and this series of steps is repeated, until the duration of the modeled noise burst is equal to the duration of the noise burst measured from the experimental acoustic data (to the nearest millisecond). For a more detailed discussion of this procedure, including a description of how the noise burst duration was measured from the acoustic data, refer to Poort (1995). When this method was used to determine A_c for /p/ in spot spoken by one speaker (Subject 1 in Poort (1995)), the resulting model output is shown in Figure 3-2. A table of some linear rates for A_c following the release, as determined for several speakers and utterances, appears in Table 3.1. (These linear rates represent averages, since the rate of release is probably not linear for the first few milliseconds following release.) For Figure 3-2 and Table 3.1, the value of A_g decreases linearly from 0.1 to 0.05 cm^2 for the first 40 ms following the stop release, then remains 0.05 cm^2 thereafter. The change in value of A_g during this time period reflects the transition in vocal-fold configuration from the position required for the relatively unaspirated stop consonant to the position required for the following vowel.

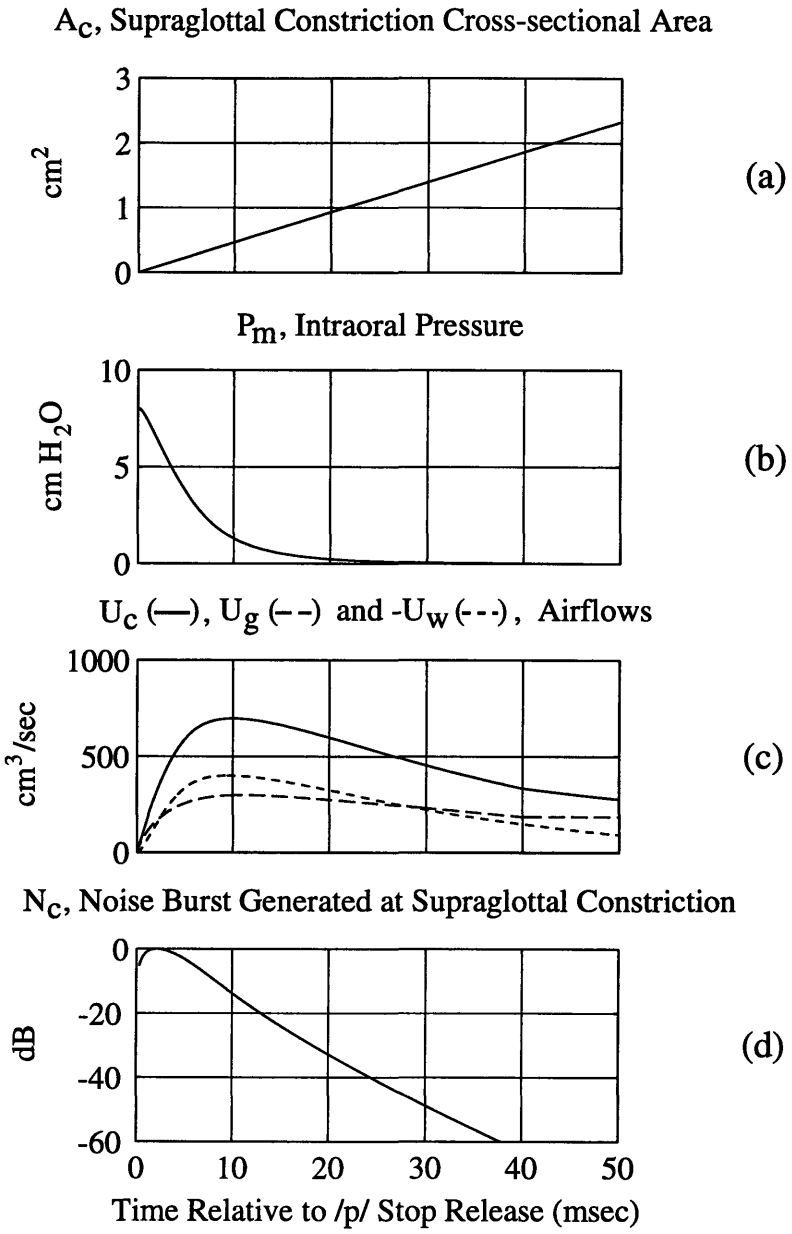

Figure 3-2 : The relatively unaspirated stop consonant /p/ upon release of closure in the utterance "Say spot again" spoken by Subject 1: (a) Linear rate of increase in lip-opening constriction cross-sectional area, A_c (47 cm^2/s); (b) Pressure within the mouth, P_m; (c) Airflow through the lip-opening constriction, U_c (solid line), airflow through the glottis, U_g (dashed line), and airflow generated by the inward displacement of the vocal-tract walls, $-U_w$ (dotted line) (the negative sign indicates the direction of displacement of U_w is inward); (d) Relative amplitude of frication noise burst, N_c. Time zero is the instant of stop release. Reprinted from Poort (1995).

Stop Consonant	Constriction Cross-sectional Area Linear Rate of Increase (cm^2/s)		
	Subject 1	Subject 2	Subject 3
/p/ in spot	47	38	39
/p/ in speet	53	40	35
initial /t/ in stot	25	–	26
initial /t/ in steet	30	–	20

Table 3.1 : Linear rates of increase for the constriction cross-sectional area, A_c, following labial and alveolar unaspirated stop-consonant releases. Rates were not determined for the initial /t/ in utterances spoken by Subject 2. Table is adapted from Poort (1995, Table 4.2).

3.2 High-Frequency Models of the Generation and Filtering of Vocal-Tract Sources

Theoretical, high-frequency models of stop-consonant production have been developed to account for the generation and filtering of the acoustic signal by the vocal tract and the resultant acoustics produced. The high-frequency models are particularly useful for modeling events that occur at frequencies above approximately 250 - 300 Hz, when lumped-element parameters generally can no longer provide reasonable estimates of the vocal tract's behavior. In this thesis, the focus of the high-frequency modeling is on describing events that occur during times when rapid articulator movements are made, corresponding to rapid changes in the acoustic waveform. These time periods, which include the few tens of milliseconds after the release, are known to contain acoustic information important to the perception of stops (Cooper et al., 1952). The primary focus of the models is on events occurring upon release of the pressure buildup in the supraglottal cavity, following the closure interval of a prevocalic stop consonant. As a consequence of the changing airflows and pressures after the release, various types of sound sources are generated in the vocal tract. The current theoretical model proposes the existence of a sequence of four different types of sources following the release. The first is the transient sound as the compressed air in the vocal tract is expelled, the second is the frication noise burst generated at the supraglottal constriction, the third is the aspiration noise which arises from turbulence near the glottis, causing transitions to become apparent in the formants, and the fourth is

the vocal-fold vibrations generated during a voiced stop or succeeding vowel (Fant, 1973; Stevens, 1993). A schematic representation of the four types of sound sources following release appears in Figure 3-3. For a given stop-consonant release, not all of these sources may be present. These sound sources are filtered by the vocal tract, resulting in spectra with unique characteristics that depend upon the type and location of each source, as well as the shape of the vocal tract downstream from the source. Vocal-tract filter models exist to describe the resonances or formant-frequency transitions occurring during the time period following the release. These sound sources and vocal-tract filter models will be discussed in greater detail in the subsections below.

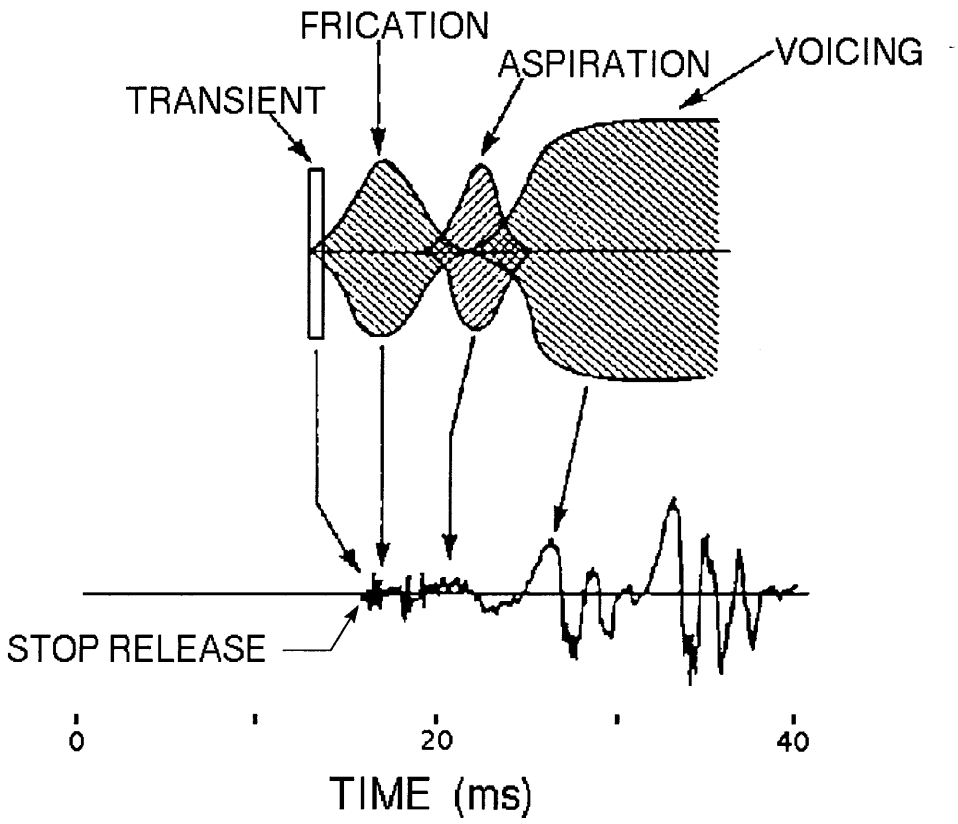

Figure 3-3 : Schematic representation of the sequence of events occurring upon release of a voiceless, relatively unaspirated stop consonant. A typical acoustic waveform (with time scale) is shown at the bottom. The stop-consonant release and the onset of the transient sound occur simultaneously, at approximately 16 ms on the time scale in this schematic depiction. Adapted from Stevens (1993).

3.2.1 Transient and Frication Noise

The transient sound is generated as the air that has been compressed in the vocal tract discharges through the constriction immediately following the stop-consonant release. The transient source occurs before the frication noise source reaches its maximum amplitude. The transient is a significant component of the sound at the release of a stop consonant only if the rate of change in cross-sectional area of the constriction is sufficiently rapid and if the length of the constriction is sufficiently short, creating an abrupt increase in airflow at the release. In terms of the equivalent circuit in Figure 3-1(b), this initial transient is represented by the flow from the acoustic compliance, C_A. The duration of the transient is brief, typically less than 1 ms. The transient flow through the constriction at release can be modeled as a volume-velocity source located at the constriction. The amplitude and spectrum of this volume-velocity transient are determined by the intraoral pressure built up during closure and by the rate of increase in constriction cross-sectional area following release. On occasion, the generation of multiple transients has been observed when the constriction length under static conditions is relatively long ($\geq$ 1 cm), for example during the production of a velar stop. This series of transients is thought to be caused by repeated vibration of the tongue surface against the palate due to the Bernoulli effect, as the tongue is being displaced from the closed position following release. The rate of constriction opening is typically slower for a velar stop than for a labial or an alveolar stop due to the larger muscle mass and greater inertia of the tongue body. This slower rate for velar stops, coupled with a longer constriction length, may result in the occurrence of two or more of these vibrations before the separation becomes too great to permit further vibration. The spacing between multiple transients is only a few milliseconds. In a series of transients, the one which is considered to be a significant component of the sound at the stop-consonant release is the first transient (burst) for which the waveform amplitude following the transient does not return to the background noise level.

Following the transient, rapid airflow through the narrow supraglottal constriction

results in turbulence, creating a frication noise source. The turbulence is generated at a surface or obstacle downstream from the constriction, and may be concentrated primarily in a narrow region of the vocal tract (such as the lower incisors), or may be distributed over a region of a centimeter or more. The frication noise is typically represented as concentrated near an obstacle downstream from the constriction and is modeled as a sound-pressure source in series with the acoustic tube (Fant, 1960), (Stevens, 1998). In some instances, there may be fluctuations in the flow through the constriction, giving rise to an additional volume-velocity or monopole source. If the turbulence is distributed over a region, then modeling the source as a single, lumped element may be inappropriate. A more suitable model would be a distributed source, which may, in turn, be approximated by several lumped sources, where each source may have different amplitude and spectral characteristics.

The amplitude and spectrum of the single sound-pressure source typically used to represent the frication noise source can be estimated approximately (Stevens, 1993), based on the work of Fant (1960), Stevens (1971), Shadle (1985), Pastel (1987), and others. As discussed in Section 3.1.4, the source amplitude is modeled as approximately proportional to $U_c^3 A_c^{-2.5}$, where U_c is the airflow through the constriction and A_c the constriction cross-sectional area. Based on this model, the amplitude is predicted to rise to a peak within the initial few milliseconds following the release, then to decrease rapidly as A_c continues to increase. The spectrum of the sound-pressure source tends to have a broad peak at a frequency proportional to u/d, where u is the velocity of the airstream and d the cross-dimension of the constriction.

3.2.2 Aspiration Noise

As the supraglottal constriction cross-sectional area of the supraglottal constriction increases following release of the stop consonant, the level of the frication noise source decreases and one of two events occurs, depending upon the configuration of the glottis. Either generation of turbulence noise occurs at the glottis, or there is initiation of vocal-fold vibration. After the release, the cross-sectional area of the glottal opening is decreasing from a relatively abducted configuration, and the event that occurs

depends upon how quickly the glottal opening narrows. For a voiced stop, the glottal opening area decreases fairly quickly, resulting in vocal-fold vibration immediately following the frication noise. For a voiceless stop, the glottal area decreases more gradually, and turbulence noise is generated in the vicinity of the glottis prior to the initiation of vocal-fold vibration for the succeeding vowel. The turbulence noise that is generated by rapid airflow through a relatively open glottal constriction is referred to as "aspiration" noise. Aspiration noise is generated as the airflow through the glottis impinges on the surfaces of the vocal tract downstream from the constriction, including the false vocal folds and the epiglottis. The quantity of aspiration noise present depends upon the phonetic environment of the voiceless stop. The aspiration noise source is believed to be distributed throughout a 2–3 cm region above the glottis, and can be modeled as a distributed sound-pressure source. The random fluctuation of the airflow through the glottis may give rise to a monopole noise source as well (Stevens, 1998). To a rough first approximation, a single sound-pressure source can be substituted for the distributed source, in which case the amplitude and spectrum of this single source can be estimated using the same formulas as were used for the amplitude and spectrum of the frication noise source in Section 3.2.1.

The aspiration noise source is filtered by the supraglottal vocal tract, with some modifications by the subglottal system which is at least weakly coupled to the vocal tract through the relatively open glottis. When the vowel following the stop consonant is produced with a relatively narrow airway constriction, having a cross-sectional area comparable to the area of the glottal opening, significant turbulence noise can be generated near the vocalic constriction in addition to the laryngeal region. That is, the aspiration noise is mixed with frication noise that is a consequence of turbulent airflow at the vocalic constriction (Stevens, 1998). The contribution of this frication noise to the sound output can dominate the spectrum, and the filtering of the noise is then determined primarily by the part of the vocal tract downstream from the vocalic constriction. The vocalic constriction also causes a reduction in airflow and consequently a reduction in the amplitude of the aspiration noise source. This effect of supraglottal turbulence noise during a spread glottal configuration will be

especially evident for high vowels, for which there is a narrowing of the oral cavity, and sometimes for low back vowels, for which there is a narrow constriction in the pharyngeal region (Stevens, 1998).

3.2.3 Voicing

The fourth and final source following the stop-consonant release is the voicing source generated during a voiced stop or a succeeding vowel. The voicing source is produced by varying the airflow through quasiperiodic lateral movements of the vocal folds, creating a periodic modulation of the glottal area. For a voiced stop, vocal-fold vibration is initiated immediately following the frication noise burst. For a voiceless stop, the glottal area decreases more slowly following the release, and aspiration noise is generated in the vicinity of the glottis for an interval of time after the frication noise and prior to the onset of glottal vibrations. Since the acoustic impedance of the glottis is usually large compared with the impedance of the supra- and subglottal cavities, at least over most of the glottal cycle and over most of the frequency range of interest for speech, the vocal-fold vibrations can be modeled to a first approximation by a periodic volume-velocity source (Stevens, 1998). The spectrum of this modeled source is a line spectrum, where the amplitudes of individual components are proportional to the Fourier transform of the single pulse. These components occur at multiples of the fundamental frequency, $F0$. During voicing, the volume-velocity waveform forms the excitation for the formant frequencies of the vocal tract.

3.2.4 Vocal-Tract Filter Models

The configuration of the vocal tract is continuously changing following the release of a stop consonant. The sound sources described in Sections 3.2.1 - 3.2.3 excite formant frequencies in the vocal tract downstream from the sources. The vocal tract acts as a filter, influencing the shape of the resultant spectrum at the mouth opening.

The filtering effects of the vocal tract can be modeled via a set of concatenated tubes having varying cross-sectional areas, similar to the tube model shown in Fig-

ure 3-1(a). These theoretical, high-frequency vocal-tract filter models have equivalent circuit models that utilize transmission-line theory, as opposed to the lumped-element parameters appearing in the low-frequency circuit model of Figure 3-1(b). The high-frequency models are particularly useful for frequencies above approximately 250–300 Hz, when lumped-element parameters generally can no longer provide reasonable estimates of the vocal tract's behavior. Based on knowledge of the cross-sectional areas of the tubes, including the variation of A_g and A_c with time, these models can predict the formant-frequency transitions occurring in the acoustic signal following the stop-consonant closure interval. The converse is also true, whereby knowledge of the formant-frequency transitions and use of the models can lead to information about the tube, or cavity, cross-sectional areas.

Idealized vocal-tract filter models for each of the three places of articulation are shown in Figures 3-4, 3-5 and 3-6. The arrows on these diagrams indicate the direction of expansion (or contraction) of the various cavities upon release of the stop, as the articulators move toward configurations appropriate for a following schwa vowel, /ə/. (If the direction of the arrows is reversed, the transition from a schwa to the closure interval of the stop consonant would be modeled.) Cavity dimensions, timing of articulator movements, and rates of movements depend upon the specific phonetic environment and speaker. For example, for a velar stop the location of the constriction (and, therefore, the lengths of the cavities anterior and posterior to the constriction) varies with the choice of following vowel. In particular, if the velar stop is followed by a front vowel, the constriction is more anterior than when it is followed by a back vowel.

To determine the formant-frequency transitions from the vocal-tract tube filter models, the wave equation must be solved. These tube models have an arbitrary area function $A(x)$, in which the cross-sectional area of the tube can vary with position x along the length of the tube. One of the strategies for solving the wave equation under these circumstances is to partition the vocal tract into several short, juxtaposed tubes of constant cross-sectional area. The wave equation is solved for each short tube, subject to the boundary conditions at both ends of the short tube. The length of

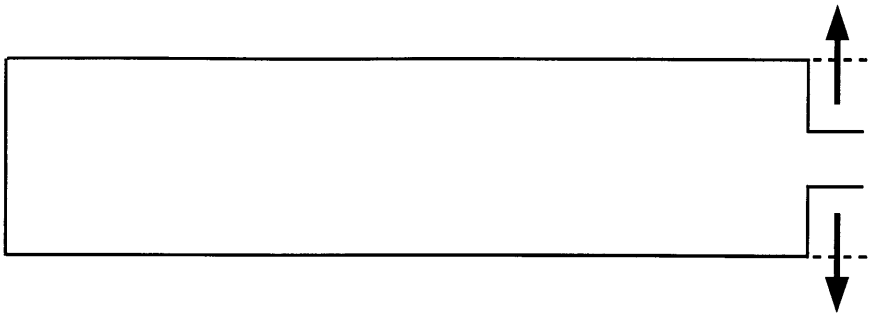

Figure 3-4 : Labial stop-consonant vocal-tract filter model. The glottis is on the left, modeled as closed, and the lips are on the right, modeled by a short front cavity (solid lines). As the lips open following the stop-consonant release, the direction of their movement is indicated by the arrows, with the final configuration being a uniform vocal tract, appropriate for the schwa vowel (dashed lines). Reprinted with permission from Stevens (1998).

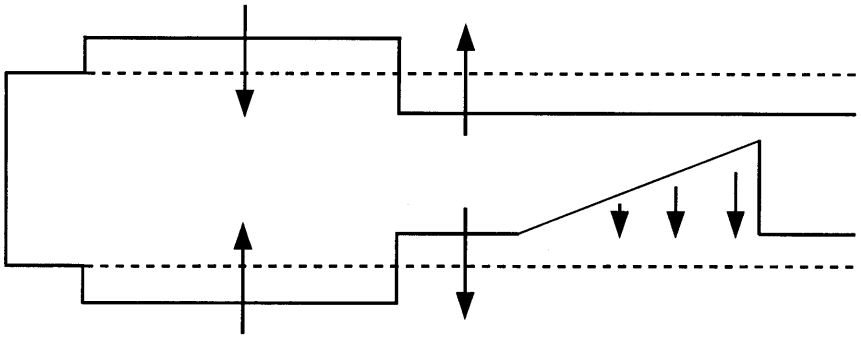

Figure 3-5 : Alveolar stop-consonant vocal-tract filter model. From left to right, the glottis is modeled as closed, the pharyngeal region is expanded because the tongue root is in a forward position, and the tongue-tip constriction is shown with a tapering of the cross-sectional area behind the constriction (solid lines). As the stop consonant is released, the articulators are moving (denoted by the arrows) toward a final configuration of a uniform vocal tract, as for the schwa vowel (dashed lines). Reprinted with permission from Stevens (1998).

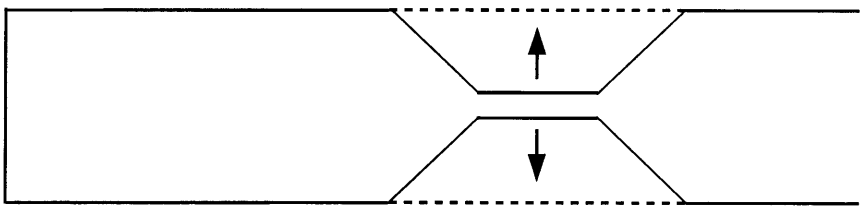

Figure 3-6 : Velar stop-consonant vocal-tract filter model. The glottis is on the left, modeled as closed, and the tongue-body constriction is modeled with tapering cross-sectional area on both sides of the constriction (solid lines). As the stop consonant is released, the movement of the tongue body (denoted by the arrows) is toward a final configuration of a uniform vocal tract, appropriate for the schwa vowel (dashed lines). Reprinted with permission from Stevens (1998).

each short tube is arbitrary. For a given $A(x)$, the solution to the wave equation is assumed to be quasistatic, i.e., the rate of change of the vocal-tract shape is slow compared to the rate of change of the natural frequencies. When $A(x)$ is considered for several consecutive instants in time following the stop release, the transitions in the formant frequencies can be calculated.

A number of adjustments may need to be made to the formant-frequency transitions calculated via the wave equation. The formant values will be affected by several sources of loss in the vocal tract and at the glottis. The radiation impedance at the mouth opening results in a slight shift in the formant frequencies, typically less than 5% (Stevens, 1998), except for short front-cavity resonances. The vocal-tract walls are non-rigid, having a finite impedance that can be modeled by an acoustic resistance in series with an acoustic mass, for frequencies in the range of approximately 100–300 Hz. (Refer to Section 3.1.3 for a discussion of the passive effects of the non-rigid vocal-tract walls at lower frequencies.) The mass reactance portion of the impedance causes a significant shift in $F1$ when a supraglottal constriction is present (Fant, 1972). The amount of this shift is greatest for a completely closed vocal tract, such as during a voiced stop, shifting $F1$ from 0 to approximately 180 Hz. As the constriction opens and $F1$ increases above 180 Hz, the non-rigid walls affect the value of $F1$ less and less. If the glottis is fairly open, as for a voiceless aspirated stop and, to a lesser extent, a voiceless unaspirated stop, the glottal impedance can no longer be modeled as infinite. The reactive part of the finite impedance causes an upward shift in the formant frequencies (Stevens, 1998). The relative shift is greatest for $F1$, and becomes progressively smaller for higher frequencies. The amount of shift corresponds to the degree of glottal opening. A more open glottis also allows coupling to occur between the subglottal and supraglottal cavities. The subglottal impedance may also have a reactive part which results in a shift in the formant frequencies. Additionally, the coupling may result in excitation of the subglottal resonances and a shift in the natural frequencies of the coupled resonators relative to those of the tubes in isolation. Finally, the supraglottal constriction cross-sectional area trajectory may be modified during the initial 5–10 ms following the release, due to the influence of the intraoral

pressure. The modification is expected to be greatest for a velar stop, in which the release is slower and the constriction longer than for the labial and alveolar stops. The intraoral pressure-induced slowing of the increase in A_c following the release, while not a source of loss in the vocal tract, does correspond to a temporary slowing in the rate of increase of $F1$, and may have similar rate-slowing effects on the higher formant frequencies. The sources of loss in the vocal tract and glottis affect not only the frequencies of the vocal-tract resonances, but also their bandwidths, thus affecting the overall shape of the spectrum produced by various sources and configurations.

3.3 Summary

This chapter provides a review of speech production theory as it pertains to stop consonants. In particular, the focus is on modeling prevocalic stop consonants in word-initial position of isolated utterances. Two types of stop-consonant production models are discussed. In Section 3.1, a low-frequency mechanoaerodynamic model is described which portrays vocal-tract movements and their associated airflow and pressure changes. In Section 3.2, a set of high-frequency models of sound sources and the corresponding vocal-tract filter models are discussed. These models serve as a basis for the experiments and analysis of Chapters 4–6. In particular, the acoustic measures developed and applied in Chapter 6 have their foundation in these models. Some of the model parameter ranges are characterized in Section 6.2.1 for a group of normal speakers. The models are also used to help develop hypotheses regarding the incorrect articulatory movements of dysarthric speakers in Section 6.2.2.

Chapter 4

Perceptual Evaluations

A perceptual experiment was designed to assess the production of word-initial stops in a series of utterances spoken by normal and dysarthric speakers. Several aspects of stop production were evaluated by the listeners, including the presence or absence of a precursor (a speaker(subject)-generated sound preceding the stop), voicing of the stop, place and manner of articulation, and the "quality" of the stop. Although the listeners heard the word-initial stop in the context of the entire single-syllable utterance, they were instructed to make these evaluations based solely on the production of the stop. This test attempts, in part, to assess "stop intelligibility" (not word intelligibility, as was performed by Chang (1995) and discussed in Section 1.3.3), by examining aspects of stop production which may contribute to the listeners' correct identification of the intended stop. Additionally, the test assesses how well the correctly-identified stops are produced. The combination of these two assessments provides an overall measure of "stop goodness".

This chapter is divided into three sections. In Section 4.1 the perceptual experiment protocol is discussed, including the corpus, speakers, recording method, listeners and test procedure. Section 4.2 contains the results and discussion. Then the perceptual analysis is summarized in Section 4.4.

4.1 Experiment

4.1.1 Corpus

The entire word list, or corpus, consists of the 70 words shown in Appendix A. This corpus was designed by Kent et al. (1989) in the context of developing a word intelligibility test for use in the clinical evaluation of dysarthric speakers. In the present perceptual experiment, the focus is on the production of stop consonants. The 13 words with word-initial stops (bad, beat, bill, bunch, dock, dug, geese, pat, pit, tile, cake, cash, coat), in which the stop consonant is normally released, were the only words examined. Both the dysarthric and the normal speakers spoke the same set of 13 words. The perceptual experiment results for eight of the words (bad, bunch, dock, dug, geese, pat, tile, and coat) are discussed in the present chapter, since this subset was also utilized for the data analyses in the remainder of the thesis. (The spectrogram and acoustic analyses of Chapters 5 and 6, respectively, were limited to eight utterances due to the number of measurements to be made by hand across 16 subjects. For the perceptual evaluations, which are less time consuming, all 13 utterances were included.) The perceptual results for the five utterances beat, bill, pit, cake and cash are briefly mentioned in Section 4.2, and the experiment responses are included in Appendix F, Section F.1.

4.1.2 Speakers

The dysarthric speakers were originally recruited by Chang (1995) for his doctoral dissertation entitled "Speech Input for Dysarthric Computer Users", completed while a member of the Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology. There are eight dysarthric speakers, consisting of four female and four male adults ranging in age from 22 to 62. These subjects exhibited one or more of three different types of dysarthria: spastic, athetoid (hyperkinetic) and ataxic. A detailed discussion of these eight subjects as well as an overview of these three types of dysarthria appear in Chapter 2.

The normal speakers, recruited by the author, were individuals with no known speech or hearing disorders. There are eight normal speakers, consisting of four female and four male adults ranging in age from 21 to 74.

4.1.3 Recording Method

This section describes the methods utilized to record the speech of the dysarthric and the normal speakers. The dysarthric speakers were prompted by displaying the desired utterances (words) on a video monitor. The spoken utterances were recorded to an audio cassette tape, and then the speech was digitized with the aid of a VAX computer system. The desired utterances for the normal speakers were displayed on paper. The spoken utterances were digitally recorded to a DAT tape and then downsampled with the aid of a UNIX computer system. The details of each of these two recording methods are discussed in the following paragraphs.

The speech of the dysarthric speakers was originally recorded by Chang (1995). The corpus appears in Appendix A and is discussed in Section 4.1.1. Details of the recording methods and data processing are provided here, and are taken in part from Chang. The speakers or subjects were asked to use their normal speaking voices for the recordings. Prior to the recording sessions, the subjects could practice saying the utterances until they were comfortable with them. The dysarthric subjects recorded ten repetitions of the corpus, with the order for each repetition being randomized, for a total of 700 recorded word repetitions (70 words x 10 repetitions/word). (It bears observing at this point that these 10 repetitions of the corpus are *in addition* to the initial single version of the corpus, a total of 70 words, that Chang (1995) used for the word intelligibility test discussed in Section 1.2 of this thesis. In other words, word intelligibility for these subjects was assessed using different repetitions of the words than the ones utilized by the author in this chapter, as well as in future chapters of this thesis.) The recordings were always made in a quiet room, although only some of the recordings (it is unknown by the author which ones) were made in a soundproof booth. Occasionally, due to the subjects' transportation limitations, it was prudent on the part of Chang and his assistants to make recordings in alternative locations,

such as in the subjects' homes. To record the speech, an omnidirectional microphone was located 10 cm from the subject's mouth. The mouth-to-microphone distance did vary, however, depending upon movements made by the subject.

The utterances were presented one at a time on a computer monitor placed in front of the subject. The subjects were allowed to choose font size on the computer screen in order to reduce visual errors. Two of the subjects have dyslexia: DF2 and DF4. To accommodate this learning disability, an assistant read the words from the computer monitor, and then the words were presented to these two subjects via headphones. Although these subjects pronounced the words immediately after hearing them spoken, they were instructed to pronounce the utterances as they normally would. During the recording session, the subjects were permitted to repeat or bypass any words they found difficult to pronounce. Also, subjects were occasionally asked to repeat words when extraneous noises (such as coughs, environmental noises, etc.) interfered with the recording process.

It is evident from listening to the tape recordings that some noises produced by the subjects (such as saliva noises, audible breathing, sounds indicating the subject is too close to the microphone, and wheelchair noises generated by involuntary body movements), as well as background noises (such as computer-generated beeps, keyboard clicks, room noises, and conversations between researchers) could not be completely eliminated from the recording sessions. Consideration should be given to the fact that these data were simultaneously being recorded by a second, head-mounted microphone for use by a speech recognizer (Chang, 1995). Due to the nature of the precise timing required by the input to the recognizer, the use of multiple researchers to manage the recording setup, and the considerable effort required on the part of the dysarthric subjects to record their speech as cleanly as possible, it is understandable why some subject and non-subject extraneous noises remained in the final recordings. The 700 words were recorded in two or three different sessions per subject, with at least one to two weeks between consecutive sessions. Chang, in the context of utilizing a speech recognizer in his research, devised this recording schedule as a way to take into account variations in the speech patterns of the dysarthric subjects over

time. For the purposes of the analysis described in the present thesis, this recording schedule helps prevent speaker fatigue as well.

The speech was recorded to cassette tapes, using an analog tape recorder. Then, the instructions in Appendix B were followed for digitizing the data using the VAX computer system and storing it on a UNIX computer. A sampling rate of 16 kHz was chosen due to the high frequency content often present in the speech of dysarthric speakers. The lowpass filter had a cutoff frequency of 7.5 kHz. The gain (amplitude) of the dysarthric data was effectively normalized during this digitizing process, as discussed in Section B.2 of Appendix B. The decision was made to normalize these dysarthric recordings because the recording environment was not well controlled between recording sessions of the same subject (i.e., distance between the microphone and the subject could vary) and subjects also exhibited large volume changes due to poor respiratory control. From the 700-word data set for each speaker, three repetitions of each of the 13 words containing word-initial stops were manually extracted with the aid of laboratory computer software for further analysis.

The speech of the normal speakers was digitally recorded and processed using the instructions in Appendices C, D and E. Appendix C, Section C.1, provides guidelines for the composition of the word list, or corpus. As discussed in Section 4.1.1, the corpus is the same one used by Chang (1995) to record the speech of the dysarthric speakers. No utterance padding was performed, although the normal speakers were asked to read over the corpus and practice saying some of the words prior to the actual recording session. Additionally, they were instructed to try to avoid changes in the $F0$ pattern as they reached the end of each set of words on the list. The speakers were asked to speak as they would normally; however, no other attempts to calibrate, monitor or control SPL were made. Additional instructions given to the speakers appear in Appendix C, Section C.2. The normal speakers recorded ten repetitions of the corpus, each repetition randomized, for a total of 700 recorded word repetitions (70 words $\times$ 10 repetitions/word). These 700 words were recorded in one recording session per speaker, in which the speakers took breaks and drank water as often as they desired.

A DAT (Digital Audio Tape) player was utilized to digitally record the normal speech to a DAT tape, according to the instructions contained in Appendix D. Then, the instructions in Appendix E were used to transfer the data from the DAT tape to the UNIX computer system in the laboratory via the MacIntosh computer system. As detailed in Appendix E, once on the UNIX system, the data were upsampled, lowpass filtered, and downsampled to achieve the desired sampling rate. A final sampling rate of 16 kHz was selected, to facilitate comparison with the dysarthric data. Three repetitions of each of the 13 words containing word-initial stops were then manually extracted from each speaker's data with the aid of laboratory computer software for further analysis.

4.1.4 Listeners

Four adult listeners, members of the Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, participated in the experiment. Through research experience in this laboratory as well as in the field of speech communication in general, the listeners had had prior experience making judgments of the kind required for this experiment. Additionally, their experience increased the likelihood that they would respond to questions about a particular utterance (word repetition) without being unduly influenced by the utterances heard preceding that one, therefore reducing the amount of bias that might otherwise affect such an experiment. During the experiment, the listeners wore headphones and were permitted to adjust the volume to the sound level they personally desired. (Refer to Section 4.1.5 for details regarding the experiment question format and procedure.)

4.1.5 Procedure

This perceptual experiment involved assessment of production of the word-initial stops in a randomized ordering of three repetitions of each of the 13 words containing word-initial stops, spoken by all 16 speakers (8 dysarthric and 8 normal). The experiment was divided into three sessions, each about an hour long, with 208 utterances

(word repetitions) per session. The sessions were conducted at least one day apart for a given listener, in an effort to alleviate listener fatigue.

The experiment was conducted with the aid of a computer interface, within which the listener could request either to listen to a given utterance as many times as s/he wished or to advance to the next utterance. Returning to previous utterances was not permitted. A series of questions was asked regarding the production of the initial sound in each utterance. Listeners responded by selecting buttons on the computer screen (Fig. 4-1).

Question 1 (Q1) was, "Is the initial sound a vowel, a consonant with a precursor or a consonant without a precursor?" If the listener answered "vowel", then the computer program automatically advanced to the next utterance. If the listener answered "consonant (with or without precursor)", then the program asked a series of three more questions. Question 2 (Q2) was, "What is the type of voicing (voiced or voiceless) of the consonant?" Question 3 (Q3) was, "What is the place of articulation (labial, labiodental, dental, alveolar, palatal, velar or glottal) of the consonant?" Question 4 (Q4) asked, "What is the manner of articulation (fricative, glide, nasal, liquid (/l/ or /r/), affricate or stop) of the consonant?" If the listener responded to Q4 by selecting a choice other than "stop", then the computer program automatically advanced to the next utterance. If the listener responded "stop" to Q4, then Question 5 (Q5) was asked as follows, "How well was the stop produced?" Listeners were to judge the quality of the stop production utilizing the classifications "good", "fair" and "poor".

This perceptual experiment is a forced-choice test, in that at each of the three stages of questioning (Q1, Q2–Q4, and Q5), prior to advancing to the next stage of questioning or the next utterance, the listener *must* make a selection from among the answers given. A flow chart outlining the question progression and the possible responses to each question is shown in Figure 4-1. The listeners were provided with a set of written instructions in addition to the questions and answers described above. This set of "Additional Instructions for the Listeners" appears below. Besides these instructions, the listeners were given no additional information to assist them in

responding to the questions. Of particular note, the listeners were not provided with definitions of "good", "fair" or "poor" quality in Q5, but rather were to use their own internal models of stop production quality.

Additional Instructions for the Listeners

You will be listening to a series of utterances spoken by normal and dysarthric speakers. The speakers intend to be producing monosyllabic words that begin with a singleton consonant.

Your task is to answer a series of questions about the initial sound of each utterance. You must reply to each question before advancing to the next stage of questioning or to the next utterance (an error message will appear otherwise and you will be unable to advance).

Some specific instructions are here:

1. Ignore preceding or simultaneous beeps/static/background noises/sounds indicating subject too close to microphone/etc.

2. Q1: Precursor is defined to be any unnatural sound *generated by the speaker (subject)* which precedes the initial consonant in the monosyllabic word. Examples include excessive prevoicing, audible breathing, etc.

3. Q2–4: Use the following table for assistance:

Place of Articulation	Manner of articulation					
	Stop	Fricative	Glide	Liquid	Nasal	Affricate
Labial	p b		w		m	
Labiodental		f v				
Dental		θ ð				
Alveolar	t d	s z	y	l r	n	
Palatal		š ž				č ǰ
Velar	k g				ŋ	
Glottal	ʔ	h				

Table 4.1 : Sounds heard in the English language that most closely correspond to the choices for place and manner of articulation in the perceptual experiment. When two columns appear for a particular manner of articulation, the entries on the left-hand side are for voiceless sounds and the entries on the right-hand side are for voiced sounds.

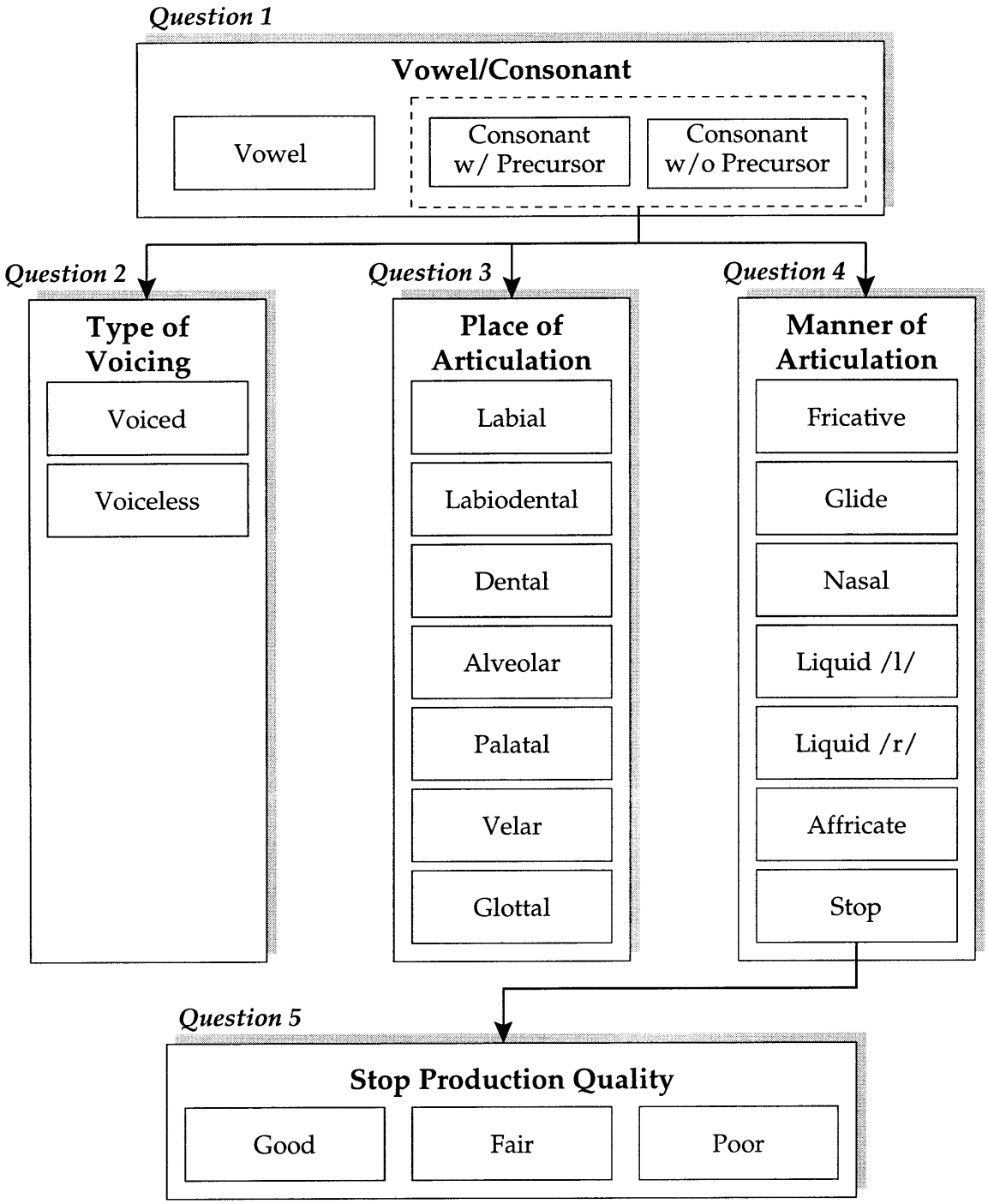

Figure 4-1 : Flow chart of perceptual experiment, showing question progression and possible responses for each question.

4. Q5: Do allow presence/absence of precursor to influence response to this question, but do not allow sounds in Instruction (1) above to do so.

5. Please feel free to make notes on the additional piece of paper provided if you feel your response to a particular utterance or set of utterances is not how you would have liked to answer (in other words, you would have liked to be able to select different answers than were available), or if you adopted any particular convention in your responses, not adequately captured by the responses alone. Please note utterance number next to any notes you make about a particular utterance.

4.2 Results and Discussion

4.2.1 Stop Goodness Score

The results of this perceptual experiment led to the idea of a measure of "stop goodness" which combines the listener responses from Q1–Q5. The responses to each of the five experiment questions are believed to provide important pieces of information relevant to the perception of the stop on one of two levels. The first level is the detection of the intended stop. This level can be viewed as an assessment of "stop intelligibility", or the degree to which a given speaker's intended word-initial stop consonant is recovered by the listener. For each stop correctly identified by the listener, the second level provides an assessment of how well that stop is produced. It is asserted that combining information from both levels, in an overall measure of "stop goodness", provides a more complete picture of stop production than the use of the first level alone. In particular, inclusion of the second level may enhance comparison of these perceptual data to the word intelligibility (Chang, 1995), the spectrogram analysis (Chapter 5), and the acoustic analysis (Chapter 6).

The first level of stop perception in this perceptual experiment (identification of the stop itself) is addressed in two parts. In the first part, Q1 (the portion of that question which assesses perception of a consonant versus a vowel) and Q4 (manner

of articulation of the consonant) determine whether a stop or another obstruent, sonorant or vowel is heard. Then, in the second part, Q2 and Q3 address which specific stop is heard. The portion of Q1 assessing the presence or absence of a precursor may affect both the detection of the following stop (Level 1) as well as the impression of how well that stop is produced (Level 2), so this question spans both levels. Then the second level of stop perception is addressed by Q5, rating the quality of the stop production.

Tests currently in clinical use to assess intelligibility in adult dysarthric speakers include the Assessment of Intelligibility in Dysarthric Speakers (AIDS) (Yorkston and Beukelman, 1981) and the Frenchay Dysarthria Assessment (FDA) (Enderby, 1983). A third assessment, consisting of two word intelligibility tests developed by Kent et al. (1989), is not in clinical use at this time but bears mentioning since the word intelligibility test conducted by Chang (1995) on the dysarthric subjects in this thesis is based on one of these two tests. Each of the AIDS, FDA and Kent et al. tests includes minimal-pair contrasts in their assessments of word intelligibility, determining information similar to Q1–Q4 of the perceptual experiment in this thesis. This perceptual experiment includes two components not found in standard clinical assessments, however. The first component is identification of a precursor preceding the stop release (in Q1). Since the stops in this study are not only word-initial, but also utterance-initial, the presence of a precursor may be partially indicative of how the speaker initiates an utterance as well as how the speaker produces a stop consonant. The second component is the inclusion of the stop production quality judgment (Q5).

The listener responses to Q1–Q5 are combined in Figure 4-2 for the eight utterance subset (bad, bunch, dock, dug, geese, pat, tile, and coat). Word repetitions in which the listener correctly identified the stop consonant (including voicing and place of articulation) were quantified according to the response to Q5: "Good" = 3, "Fair" = 2 and "Poor" = 1. This weighting scheme favored those speakers who produced their stop consonants well. Word repetitions in which the initial sound was identified as a vowel, or the initial consonant was incorrectly identified with regard to voicing,

place or manner of articulation, were given a value of 0. Listeners were instructed to allow the presence or absence of a precursor to influence their responses to Q5. Consequently, stop consonants judged to have precursors were not automatically assigned a value of 0, even though a precursor would not normally be present prior to the stop release. Instead, if the stop was otherwise correctly identified, then a value was assigned according to the response to Q5. The results were averaged across utterances, repetitions, and listeners, providing one score per speaker. In the case of normal speakers, the results were averaged across speakers as well, providing one score overall.[1] The resultant scores were organized left to right in order of decreasing "stop goodness" in Figure 4-2.

The use of the combined, weighted listener responses to Q1–Q5 (Fig. 4-2) as a measure of stop goodness can be contrasted with the use of just Q1–Q4 (Fig. 4-3) and the use of Q1–Q5 without the application of weighting (Fig. 4-4). In Figure 4-3, the resultant measure reflects strictly the correct identification of the stop consonant, with no incorporation of stop quality perception. This measure, derived solely from Q1–Q4, is considered to be an assessment of "stop intelligibility". Although this measure does suggest a reordering of the three dysarthric speakers with highest goodness scores (DF1 now has a higher ranking than DM2 and DM1), it does not discriminate as well between these three speakers and the normal speakers as does the stop goodness score derived from Figure 4-2. A t-test (significance level $\alpha = 0.05$) was performed on the data in each of Figures 4-2 and 4-3. The t-test results indicate that the normal and dysarthric speakers have significantly different means in Figure 4-2 (i.e., the null hypothesis that the dysarthric speakers are all members of the same normal speaker group was rejected). The t-test results for the data in Figure 4-3 indicate, however, that it is not possible to separate any of the first three dysarthric speakers (DM2, DM1 and DF1) from the normal speakers using significance level $\alpha = 0.05$ (i.e., it was not possible to reject the aforementioned null hypothesis for these speakers). Adding information regarding stop production quality to the measure of stop

[1]The normal speakers' results were so similar to one another that it was deemed not useful to report their scores individually.

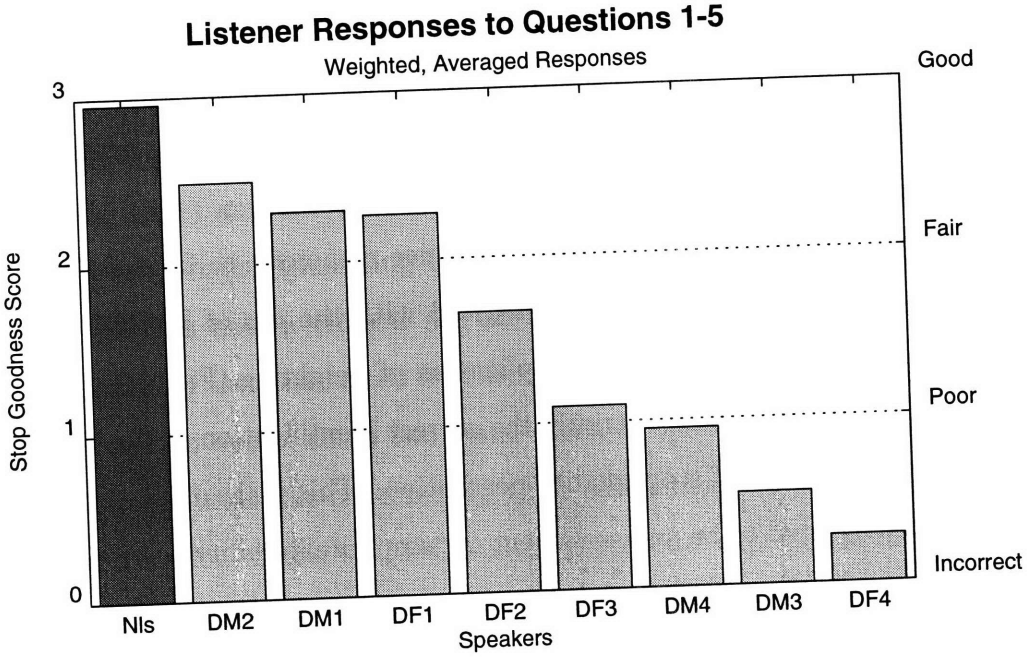

Figure 4-2 : Combined, weighted listener responses to Q1–Q5 provide a measure of "stop goodness". Word repetitions in which the listener correctly identified the presence of a consonant (with or without precursor), the type of voicing, and the place and manner of articulation for the consonant were quantified according to the response to Q5: Good = 3, Fair = 2 and Poor = 1. Repetitions in which the initial sound was identified to be a vowel, or the initial consonant was incorrectly identified with regard to voicing, place or manner of articulation, were given a value of 0 (Incorrect). Scores were then averaged across 8 utterances, 3 repetitions/utterance and 4 listeners to generate one value reflecting stop goodness for a given speaker. In the case of normal speakers (Nls), the scores were also averaged across all 8 speakers. The normal and dysarthric (DF1–DF4, DM1–DM4) speakers are organized from left to right in order of decreasing stop goodness score.

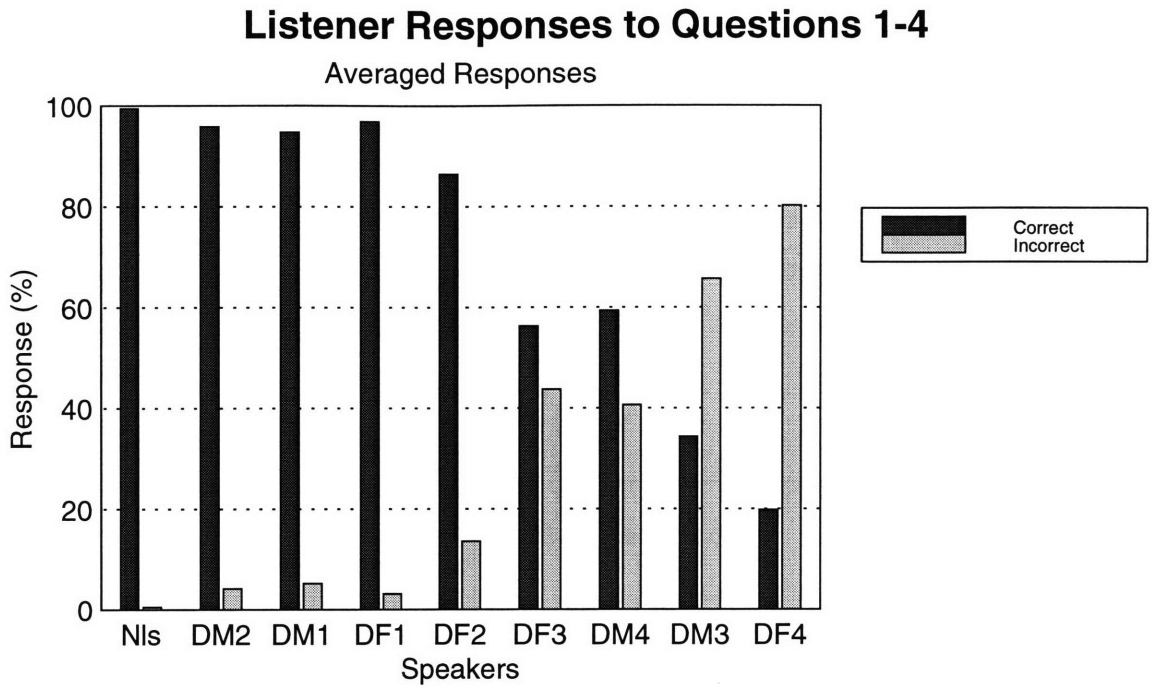

Figure 4-3 : Combined listener responses (%) to Q1–Q4. The category "Correct" contains all word repetitions in which the listener correctly identified the presence of a consonant (with or without precursor), the type of voicing, and the place and manner of articulation of the consonant. The category "Incorrect" contains all remaining word repetitions. For each speaker, responses shown averaged across 8 utterances, 3 repetitions/utterance and 4 listeners. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness, as determined in Figure 4-2.

intelligibility reveals that a statistically-significant difference exists in stop production between the normal and the dysarthric speakers. Furthermore, it demonstrates that the production quality of DM2 is better than that of DM1 or DF1.

In Figure 4-4 the responses to Q1–Q5 are divided into the four stop production quality scores (Good, Fair, Poor and Incorrect) for each speaker. From this figure it can be appreciated that normal speakers are judged to have good quality stops the vast majority of the time, mildly dysarthric speakers (DM2, DM1, DF1, and DF2) have fair quality stops more often than normals, and moderately dysarthric speakers (DF3, DM4, DM3, and DF4) have a predominance of "Incorrect" productions, in which the stop consonant was not produced correctly. (The designations "mildly" and "moderately" dysarthric refer to the word intelligibility results of Chang (1995), as discussed in Section 1.2.) In order to determine a speaker order for this plot, however,

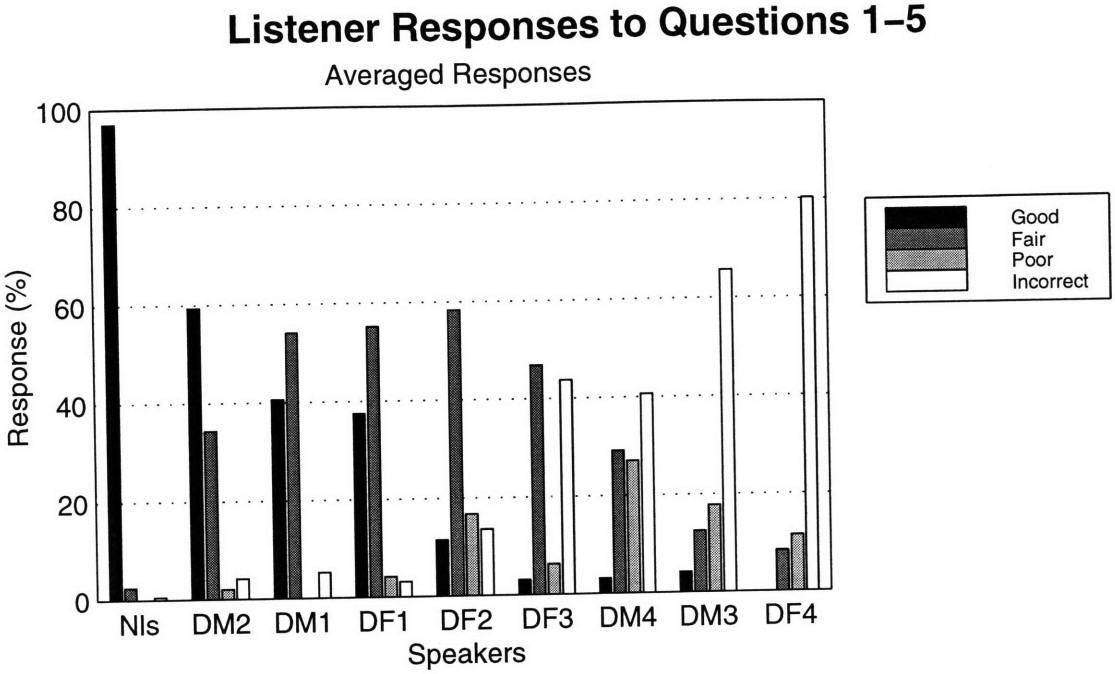

Figure 4-4 : Combined listener responses (%) to Q1–Q5. Word repetitions in which the listener correctly identified the presence of a consonant (with or without precursor), the type of voicing, and the place and manner of articulation of the consonant are divided into Good, Fair and Poor ratings according to the responses to Q5. The category "Incorrect" contains all remaining word repetitions. For each speaker, responses shown averaged across 8 utterances, 3 repetitions/utterance and 4 listeners. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

some type of weighting scheme must be applied. (The speaker order shown in the plot comes from Figure 4-2.) In addition to determining a speaker order, the weighting scheme of Figure 4-2 also has the advantage of providing a more convenient measure of stop goodness to reflect a given dysarthric speaker's stop production, rather than four values per speaker as portrayed in Figure 4-4.

A closer examination of Figure 4-2 reveals that there is a wide range in stop goodness scores for the dysarthric speakers involved in this study. Some speakers (DM2, DM1 and DF1) are close to, although still significantly different from, normal speakers, while other dysarthric speakers (such as DF3, DM4, DM3, and DF4) have quite low stop goodness scores. Word intelligibility results are available for these dysarthric speakers (Chang, 1995), as discussed in Chapter 1, Section 1.2, including Figure 1-2. The stop goodness scores can be compared to the word intelligibility

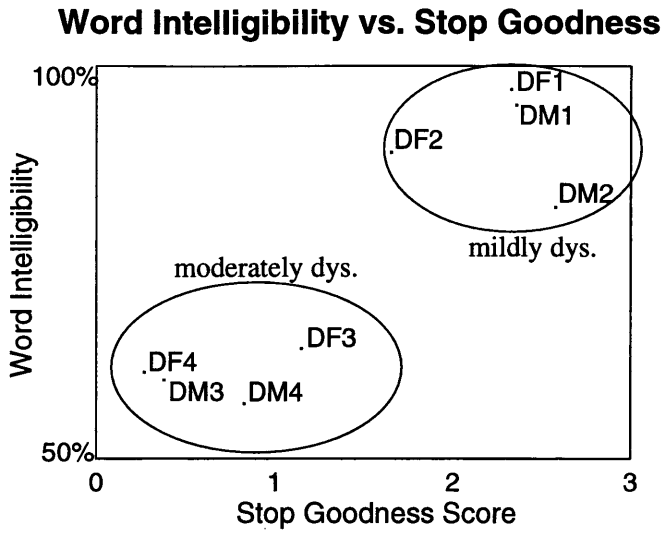

Figure 4-5 : Comparison of stop goodness scores and word intelligibility results (Chang, 1995) for the dysarthric speakers (DF1–DF4,DM1–DM4). The mildly- and moderately-dysarthric speaker groupings, based on the word intelligibility results, are maintained by the stop goodness scores.

results as long as the reader keeps in mind that the goodness scores are based on four experienced listeners' judgments of word-initial stop production in three repetitions/word for eight words, whereas word intelligibility is based on five naive listeners' judgments of production of the entire word, one repetition/word (a *different* repetition than was utilized to assess stop goodness), for all 70 words in the corpus (refer to Appendix A). A graph of the comparison of stop goodness to word intelligibility is shown in Figure 4-5. From this figure it is observed that there is some speaker-order shuffling within each of the mildly and moderately dysarthric groupings, but no speakers transfer from one group to the other. This finding is appealing, since it is consistent with stop goodness begin a partial predictor of word intelligibility. Later in this thesis, these stop goodness scores will be compared with data obtained from spectrogram analysis and acoustic analysis in order to develop a more complete picture of how these speakers produce stop consonants.

Figures 4-2 to 4-4 show the results for the eight-utterance subset (bad, bunch, dock, dug, geese, pat, tile, and coat) of the 13 words containing word-initial stops. Comparable figures showing the results for all 13 utterances are in Appendix F, Figures F-1 to F-3, respectively. When the full 13 utterance set is considered (the

additional utterances are beat, bill, pit, cake, and cash), the results are not noticeably different, in general, from the eight-utterance subset. The combined, weighted listener responses to Q1–Q5 shown in Figure F-1 for the 13 utterances produce the same speaker order and same general distribution of stop goodness scores as was seen in Figure 4-2. From Figures F-2 and F-3, it can be seen that the small increase in stop goodness score for DM2 in Figure F-1 compared to Figure 4-2 can be attributed for the most part to a proportionate increase in "Good" responses to Q5, and the small decrease in stop goodness scores for DM4 and DM3 is attributable to proportionate increases in "Incorrect" or "Incorrect" responses. These small changes in goodness scores do not impact the speaker ordering, however.

4.2.2 Responses to Individual Perceptual Test Questions

The listener responses can be considered on a question-by-question basis for Questions 1–4, allowing a more in-depth examination of the precursor (when present), type of voicing, and place and manner of articulation. Figure 4-6 shows the listener responses to Q1. From this figure it is observed that a few of the speakers (DF2, DM4, DF4, and, to a lesser extent, DF1) tend to produce a precursor prior to the stop release. (A precursor is a sound generated by the speaker.) Different speakers may generate different types of sounds in this precursor time interval. By listening to the acoustic signals, the author inferred that DF2 tends to have air leaking out her nose during this time interval, attributed to difficulty appropriately controlling her velopharyngeal port opening. Speakers DM4 and DF4 tend to vary the positions of their articulators and vocal folds to produce a variety of precursor sounds. These sounds tend to be somewhat dependent upon the following stop, such as noise production preceding an intended voiceless stop or inadvertent vowel generation or excessive prevoicing preceding an intended voiced stop. Speaker DF1 has abnormally long and loud prevoicing prior to some of her voiced stops. It is also observed in Figure 4-6 that the two speakers with poorest stop goodness scores, DM3 and DF4, were judged to omit their stops entirely approximately 10–15% of the time. The omission may be attributed to either deletion of the stop or such a prolonged duration between the

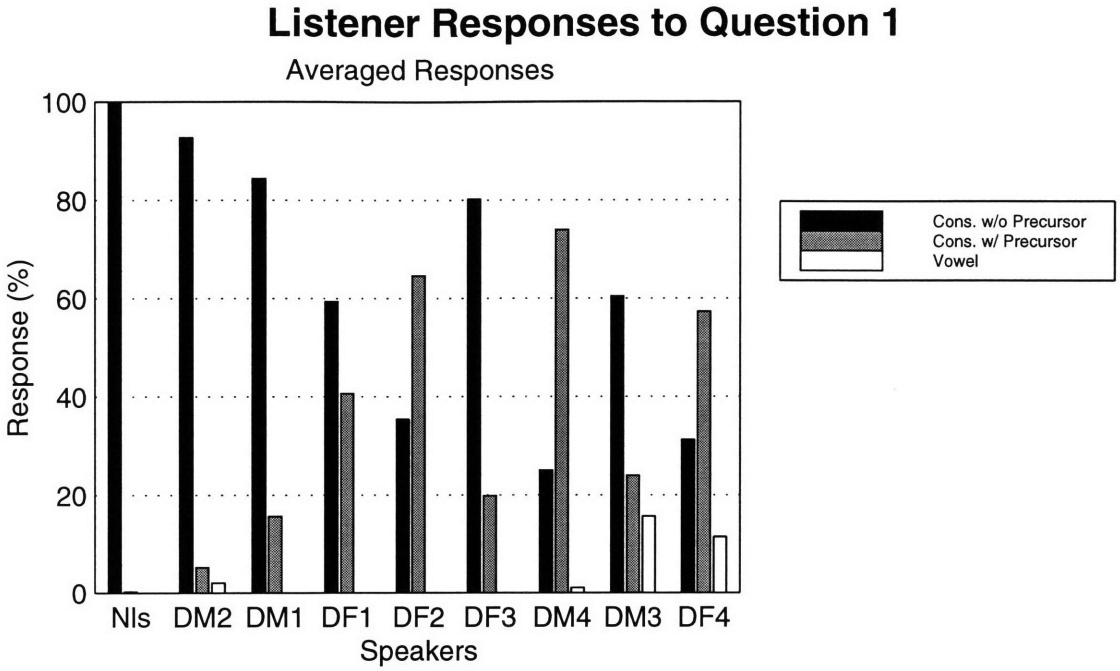

Figure 4-6 : Listener responses (%) to Q1, identifying the initial sound in the utterance as a vowel, a consonant with a precursor or a consonant without a precursor. Responses shown averaged across 8 utterances, 3 repetitions/utterance and 4 listeners for each speaker. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

stop release and the onset of the following vowel that the listener judged the stop to be deleted and the preceding stop release to be a precursor.

The listener responses to Q2 are shown in Figure 4-7, by voicing of the intended stop. From this figure it can be seen that, on average, the dysarthric speakers tend to voice their voiced stops correctly more often than their voiceless stops. Two speakers in particular have difficulty properly voicing their voiceless stops, DM4 and DM3. From the acoustic signal and Figure 4-7, it is observed that speaker DM4 tends to shorten the VOT (voice onset time) to the extent that his voiceless stops are judged to be voiced. Speaker DM3 tends to either produce a voiced consonant instead of a voiceless consonant, or, more often, to omit the voiceless stop entirely, such that the initial sound of the utterance is judged to be a vowel.

The responses to Q3 are summarized by place of articulation in Figure 4-8 and by individual stop in Table 4.2. These data show several individual speaker differences for the four moderately-dysarthric speakers. Those speakers will be considered one

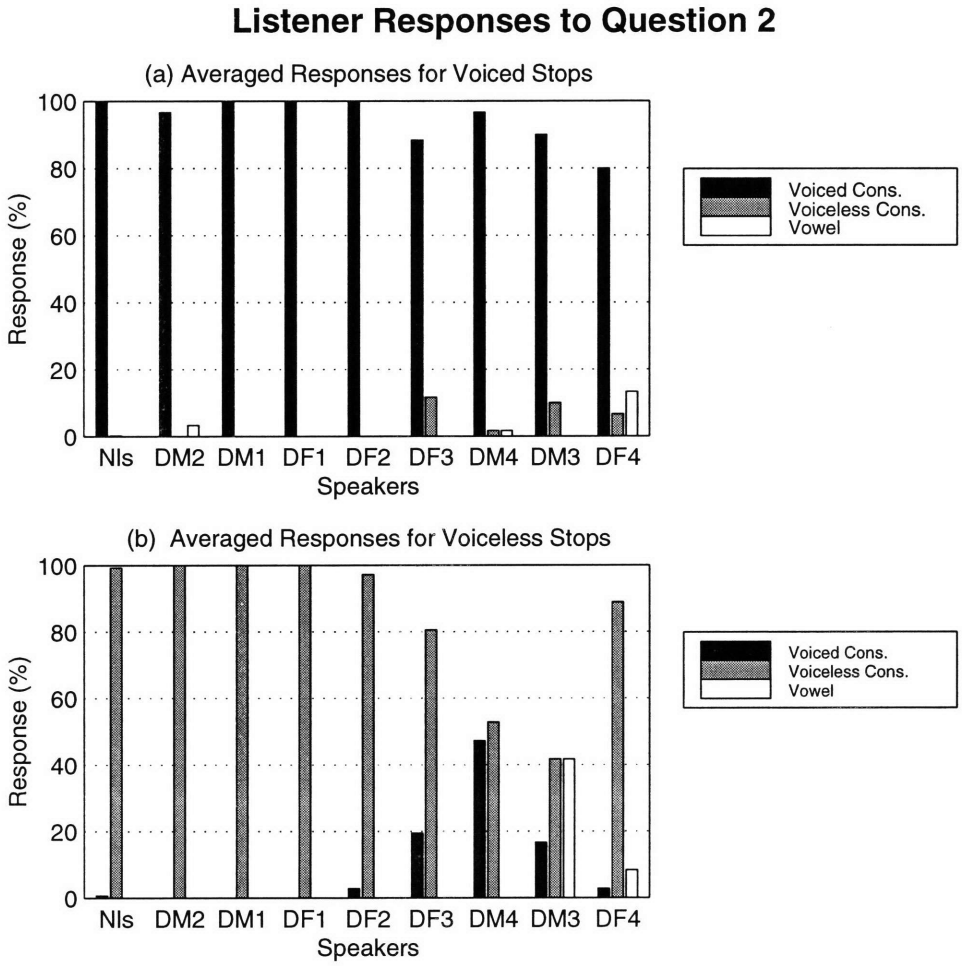

Figure 4-7 : Listener responses (%) to Q2, identifying the type of voicing (voiced or voiceless) of the consonant. Instances in which the initial sound was identified as a vowel are also indicated. For each speaker, responses shown averaged across 4 listeners, 3 repetitions/utterance, and (a) 5 utterances containing intended word-initial voiced stops or (b) 3 utterances containing intended word-initial voiceless stops. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

at a time, beginning with DF3. From Figure 4-8, speaker DF3 is judged to produce labial and velar places of articulation well, but alveolar places poorly. A closer look at her alveolar stop production in Table 4.2 reveals that her alveolar stops are typically judged to have a velar place of articulation. One possible explanation for these judgments is if she makes the alveolar closure with her tongue body instead of her tongue tip, placing it further back along the palate so that the front cavity has a length more similar to a velar than to an alveolar stop. Speaker DM4 is something of the converse of DF3, producing alveolar place correctly and having more difficulty with labial and velar places. Other than to note the variability in place of articulation for intended labial and velar stops, no particular pattern emerges from a study of the data for DM4 in Table 4.2. In that table, speaker DM3 is noted to have more difficulty with place of articulation for voiceless stop production than for voiced stops. Voiceless stops are judged to be glottal stops or vowels more often than they are judged to have the correct place of articulation. This observation is in agreement with the findings for DM3 in Figure 4-7. Finally, speaker DF4 has more trouble producing alveolar and velar places of articulation than labial places of articulation. Velar stops and, to a lesser extent, alveolar stops are typically judged to be glottal stops or vowels. Since both alveolar and velar stops are produced by movements of the tongue, it is reasonable to hypothesize that she has difficulty positioning her tongue during stop production.

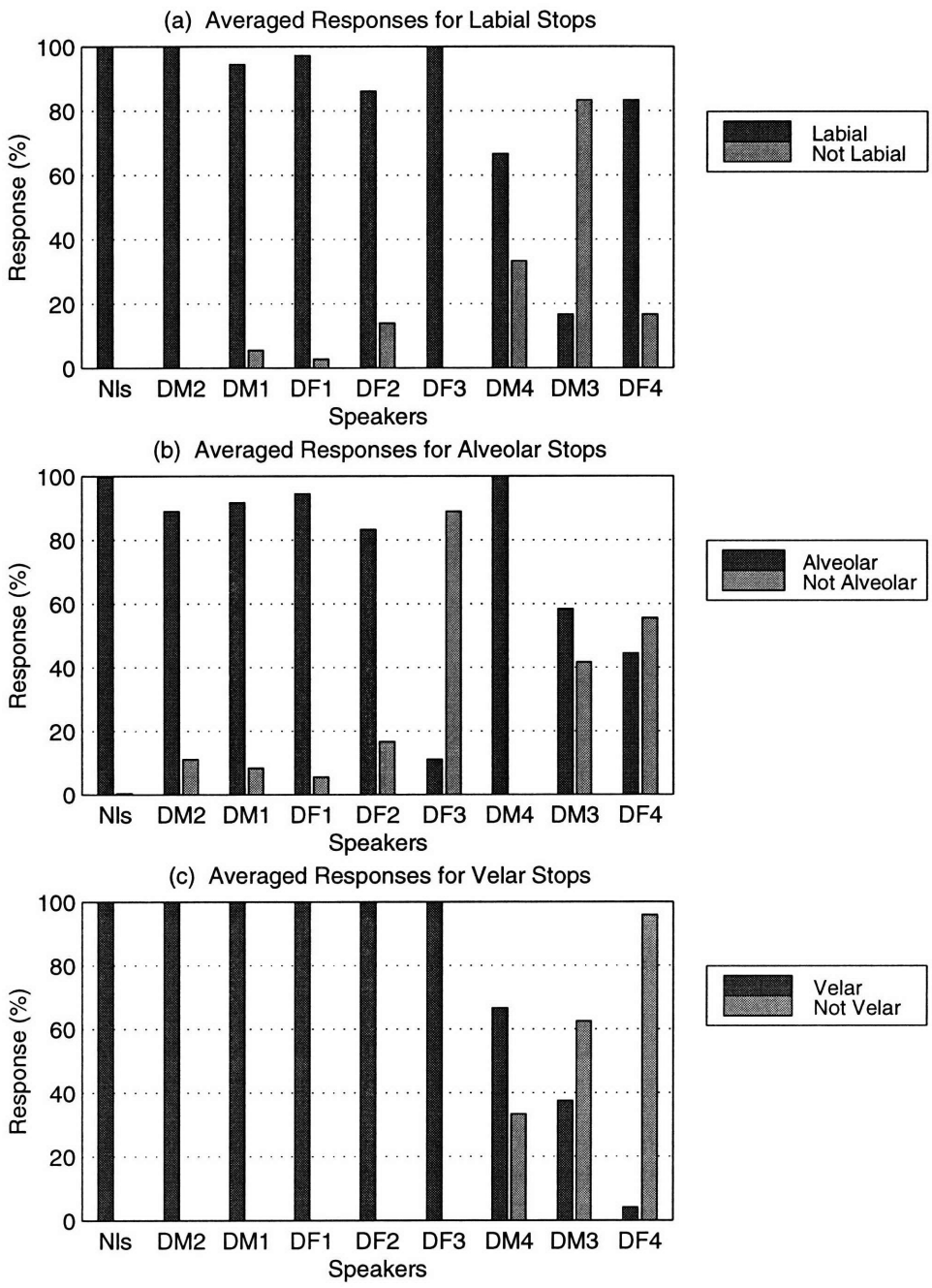

Figure 4-8 : Listener responses (%) to Q3, identifying the place of articulation of the consonant. Instances in which the initial sound was identified as a vowel are included in the category "Not [place of articulation]" in each subplot. For each speaker, responses shown averaged across 4 listeners, 3 repetitions/utterance and (a) 3 utterances containing intended word-initial labial stops or (b) 3 utterances containing intended word-initial alveolar stops or (c) 2 utterances containing intended word-initial velar stops. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

Word-Initial Stop	Labial	Alveolar	Velar	GS or V	Other
Normals					
/b/ (Avg. of 2 utts.)	100.0	0.0	0.0	0.0	0.0
/p/	100.0	0.0	0.0	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	99.5	0.0	0.0	0.5
/t/	0.0	100.0	0.0	0.0	0.0
/g/	0.0	0.0	100.0	0.0	0.0
/k/	0.0	0.0	100.0	0.0	0.0
DM2					
/b/ (Avg. of 2 utts.)	100.0	0.0	0.0	0.0	0.0
/p/	100.0	0.0	0.0	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	83.3	0.0	8.3	8.3
/t/	0.0	100.0	0.0	0.0	0.0
/g/	0.0	0.0	100.0	0.0	0.0
/k/	0.0	0.0	100.0	0.0	0.0
DM1					
/b/ (Avg. of 2 utts.)	100.0	0.0	0.0	0.0	0.0
/p/	83.3	0.0	16.7	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	87.5	4.2	0.0	8.3
/t/	0.0	100.0	0.0	0.0	0.0
/g/	0.0	0.0	100.0	0.0	0.0
/k/	0.0	0.0	100.0	0.0	0.0
DF1					
/b/ (Avg. of 2 utts.)	95.8	4.2	0.0	0.0	0.0
/p/	100.0	0.0	0.0	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	95.8	0.0	0.0	4.2
/t/	0.0	91.7	0.0	0.0	8.3
/g/	0.0	0.0	100.0	0.0	0.0
/k/	0.0	0.0	100.0	0.0	0.0

Table 4.2 : Confusion matrices containing listener responses (%) to Q3, identifying the place of articulation for each stop. The rows represent the intended word-initial stop and the columns the listeners' responses, where GS or V = Glottal Stop or Vowel and Other = Labiodental, Dental or Palatal. For each speaker, responses shown averaged across 4 listeners, 3 repetitions/utterance and one utterance/word-initial stop (unless otherwise indicated). For normal speakers, responses also averaged across all 8 speakers. The confusion matrices are in order of decreasing stop goodness for the normal and dysarthric (DF1–DF4, DM1–DM4) speakers. Table is continued.

Word-Initial Stop	Labial	Alveolar	Velar	GS or V	Other
DF2					
/b/ (Avg. of 2 utts.)	87.5	8.3	0.0	0.0	4.2
/p/	83.3	0.0	8.3	0.0	8.3
/d/ (Avg. of 2 utts.)	4.2	87.5	0.0	0.0	8.3
/t/	0.0	75.0	8.3	8.3	8.3
/g/	0.0	0.0	100.0	0.0	0.0
/k/	0.0	0.0	100.0	0.0	0.0
DF3					
/b/ (Avg. of 2 utts.)	100.0	0.0	0.0	0.0	0.0
/p/	100.0	0.0	0.0	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	12.5	87.5	0.0	0.0
/t/	0.0	8.3	83.3	0.0	8.3
/g/	0.0	0.0	100.0	0.0	0.0
/k/	0.0	0.0	100.0	0.0	0.0
DM4					
/b/ (Avg. of 2 utts.)	75.0	12.5	12.5	0.0	0.0
/p/	50.0	25.0	16.7	0.0	8.3
/d/ (Avg. of 2 utts.)	0.0	100.0	0.0	0.0	0.0
/t/	0.0	100.0	0.0	0.0	0.0
/g/	0.0	50.0	33.3	8.3	8.3
/k/	0.0	0.0	100.0	0.0	0.0
DM3					
/b/ (Avg. of 2 utts.)	25.0	62.5	8.3	0.0	4.2
/p/	0.0	0.0	0.0	100.0	0.0
/d/ (Avg. of 2 utts.)	0.0	70.8	16.7	0.0	12.5
/t/	0.0	33.3	0.0	66.7	0.0
/g/	25.0	0.0	50.0	0.0	25.0
/k/	0.0	0.0	25.0	75.0	0.0
DF4					
/b/ (Avg. of 2 utts.)	87.5	8.3	0.0	4.2	0.0
/p/	75.0	8.3	0.0	8.3	8.3
/d/ (Avg. of 2 utts.)	8.3	50.0	0.0	8.3	33.3
/t/	0.0	33.3	8.3	58.3	0.0
/g/	0.0	8.3	8.3	83.3	0.0
/k/	0.0	0.0	0.0	100.0	0.0

Table 4.2 : (continued) Confusion matrices containing listener responses (%) to Q3, identifying the place of articulation for each stop. The rows represent the intended word-initial stop and the columns the listeners' responses, where GS or V = Glottal Stop or Vowel and Other = Labiodental, Dental or Palatal. For each speaker, responses shown averaged across 4 listeners, 3 repetitions/utterance, and one utterance/word-initial stop (unless otherwise indicated). For normal speakers, responses also averaged across all 8 speakers. The confusion matrices are in order of decreasing stop goodness for the normal and dysarthric (DF1–DF4, DM1–DM4) speakers.

The responses to the last question, Q4, are shown in Figure 4-9 and Table 4.3. As might be anticipated, Figure 4-9 exhibits the trend that, in general, the number of times the initial sound is not judged to be a stop increases as the speaker's stop goodness scores decrease. Table 4.3 divides the "Not a Stop" category into three components: Other Obstruent, Sonorant and Vowel. The intended word-initial stops are divided into Voiced and Voiceless stops. Only speakers DM3 and DF4 show a large difference from normal. As was observed in the previous questions, DM3 produces voiceless stop consonants that are frequently judged to be vowels. For speaker DF4, her voiced stops are most often judged to be sonorants. She is correctly voicing these stops for the most part, but is either not forming a complete constriction or is not closing the velopharyngeal port completely. Her voiceless stops are most often judged to be obstruents other than stops. Consequently, during part or all of the stop-release time period, the constriction remains in a narrow configuration, permitting the generation of turbulence noise over a longer time period than would ordinarily be generated during a stop release.

	Stop	Other Obstruent	Sonorant	Vowel
Normals				
Voiced	100.0	0.0	0.0	0.0
Voiceless	100.0	0.0	0.0	0.0
DM2				
Voiced	96.7	0.0	0.0	3.3
Voiceless	100.0	0.0	0.0	0.0
DM1				
Voiced	100.0	0.0	0.0	0.0
Voiceless	100.0	0.0	0.0	0.0
DF1				
Voiced	100.0	0.0	0.0	0.0
Voiceless	100.0	0.0	0.0	0.0
DF2				
Voiced	93.3	1.7	5.0	0.0
Voiceless	91.7	5.6	2.8	0.0
DF3				
Voiced	95.0	0.0	5.0	0.0
Voiceless	94.4	5.6	0.0	0.0
DM4				
Voiced	85.0	1.7	11.7	1.7
Voiceless	94.4	2.8	2.8	0.0
DM3				
Voiced	95.0	5.0	0.0	0.0
Voiceless	58.3	0.0	0.0	41.7
DF4				
Voiced	21.7	0.0	65.0	13.3
Voiceless	41.7	50.0	0.0	8.3

Table 4.3 : Confusion matrices containing listener responses (%) to Q4, identifying the manner of articulation of the stop consonants. The rows indicate the intended type of voicing, and the columns are the listeners' responses. For each speaker, responses shown averaged across 4 listeners, 3 repetitions/utterance and 5 utterances containing intended word-initial voiced stops (first row) or 3 utterances containing intended word-initial voiceless stops (second row). For normal speakers, responses also averaged across all 8 speakers. The confusion matrices are shown in order of decreasing stop goodness for the normal and dysarthric (DF1–DF4, DM1–DM4) speakers.

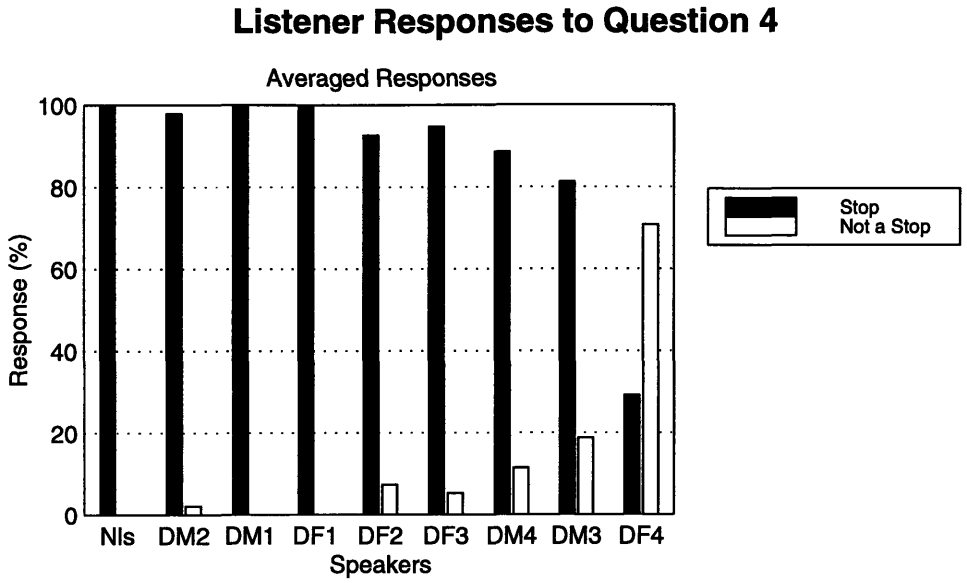

Figure 4-9 : Listener responses (%) to Q4, identifying the initial sound in the utterance as a stop consonant or not. Instances in which the initial sound was identified as a vowel are included in the category "Not a Stop". Responses shown averaged across 8 utterances, 3 repetitions/utterance and 4 listeners for each speaker. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

4.3 Conclusions

A summary of the individual-speaker observations for listener responses to Q1–Q4 is as follows. Speaker DF1 has a precursor some of the time, attributable to excessive prevoicing prior to some of her voiced stops. Speaker DF2 also has a precursor some of the time, attributable to a faulty velopharyngeal port. Speaker DF3 produces alveolar stops like velar stops, which may be attributable to use of the tongue body to form the constriction, rather than the tongue tip or blade. Speaker DM4 has a precursor the majority of the time, attributable to the production of a variety of voiced and voiceless sounds prior to the stop release. Noise production tends to precede intended voiceless stops and vocalizations tend to precede intended voiced stops. Speaker DM4 also tends to shorten the VOT of voiceless stops such that they are judged to be voiced, and has difficulty correctly producing labial and velar stops. Instead of producing a voiceless stop, speaker DM3 tends to produce a voiced stop (largely in the form of a glottal stop) or omit the stop entirely. Finally, speaker DF4

has a precursor the majority of the time, attributable to reasons similar to those for DM4. Speaker DF4 also has difficulty positioning her tongue to produce an alveolar or velar stop, difficulty forming a complete closure in the vocal tract during stop-consonant production, and may have difficulty with closing the velopharyngeal port and/or moving the primary articulator rapidly following the release.

Listener responses for Q1–Q4 are included in Appendix F, Section F.1, for the set of 13 words containing word-initial stops. Although the listener responses may vary slightly from the eight-utterance subset to the 13-utterance set, the individual-speaker observations discussed above do not change. The complete dataset for this perceptual experiment is provided in Appendix F, Section F.2.

A few observations can be made across speakers from the listener responses to Q1–Q4. From Q1 (Fig. 4-6), a precursor tends to be generated more frequently by speakers with lower stop goodness scores. (The stop "goodness" score, which provides an overall assessment of a speaker's ability to produce stop consonants, will be discussed further in the following paragraphs.) The mildly-dysarthric speakers do not make many voicing errors; the moderately-dysarthric speakers tend to make more voicing errors for voiceless stops than for voiced stops, with the exception of speaker DF4 (Fig. 4-7). Place of articulation errors tend not to be as common for mildly-dysarthric than for moderately-dysarthric speakers (Fig. 4-8). For the moderately-dysarthric speakers, place errors are primarily speaker-dependent. In Figure 4-9, in general fewer stops are heard as stops by the listeners as the goodness score decreases, consistent with observations made from Figure 4-3.

A single number, the stop "goodness" score, was developed for each speaker from the listeners responses to Q1–Q5 (Fig. 4-2). This score consolidates the listeners' impressions of stop production for a given speaker into one number, which can be more readily compared to the results in Chapters 5 and 6 than a set of numbers per speaker. The stop goodness scores will be utilized as the x-axis in some of the results figures of Chapters 5 and 6 to facilitate comparison of the results from different types of data. In particular, the comparison of acoustic data results to the stop goodness score will aid in identification of acoustic correlates of perception.

The listener responses to Q1–Q5 (Fig. 4-2) were compared to the listener responses to Q1–Q4 (Fig. 4-3). The responses to Q1–Q4 were not able to differentiate all dysarthric speakers from normal, whereas inclusion of the additional quality judgments (Q5) to Q1–Q4 did enable this differentiation. Due to this finding, it is determined that an assessment based strictly on voicing, place and manner of articulation captured only some of the existing differences in production between normal and dysarthric speakers. The results indicate that, at least for some dysarthric speakers, there are aspects of stop production which still are not normal even when the stop consonant itself is identified correctly by the listeners. The quality ratings may indicate evidence of articulatory difficulties mildly- to moderately-dysarthric speakers are having even when their stops are otherwise intelligible.

The stop goodness score results derived from Q1–Q5 (Fig. 4-2) were also compared to the word intelligibility results of Chang (1995) (Fig. 1-2), as shown in Figure 4-5. Although the speaker ordering changes slightly from the word intelligibility test results to the stop goodness scores, the same four speakers remain within each of the mildly- and moderately-dysarthric speaker groupings established by the word intelligibility test results. This finding is appealing since the intelligibility of the word-initial stop (stop goodness is partially based upon stop intelligibility) should be partially predictive of the intelligibility of the entire word.

4.4 Summary

Section 4.1 discusses the corpus, speakers, recording method, listeners and procedure for the perceptual experiment. In Section 4.2, the listener responses for each of Q1–Q4 are considered. Observations are made regarding the dysarthric speakers' abilities to produce various aspects of stop consonants, such as voicing, place and manner of articulation. The results for the eight-utterance subset of words containing word-initial stops utilized in this thesis are compared to the results for the full set of 13 available utterances.

Also in Section 4.2, an overall measure of stop goodness is developed from the

listener responses to the five perceptual test questions. This measure is reflective not only of the type of voicing, place and manner of articulation of the stop consonant, but also incorporates an assessment of the quality of the stop production. This quality assessment indicates that dysarthric stop production differs significantly from normal, on average, even when the stop consonant is intelligible with regard to voicing, place and manner of articulation. The stop goodness measure, or score, was also found to be a partial predictor of word intelligibility, as expected, since a portion of the stop goodness score consists of stop intelligibility, and the stop is the first phoneme of the word. The stop goodness score provides a single number per dysarthric speaker, which will be useful when attempting to identify acoustic correlates of perception in Chapters 5 and 6.

Chapter 5

Spectrogram Analysis

Spectrogram Analysis (SA) is the visual assessment of spectrograms in order to characterize several attributes of stop-consonant production near the time of the release, assigning ratings on a scale from 1 (Good) to 3 (Poor). SA is included in this thesis to fill a niche between perceptual and acoustic analysis. SA is more similar to perceptual experiments than to acoustic analysis in its subjective and qualitative nature, yet it enables a more "quantitatively-qualitative" approach than perceptual analysis via the assignment of a numerical rating system to the assessment of how well several individual attributes of the stop were produced. SA also enables use of the visual system (as opposed to the auditory system) to evaluate the stop within the context of production of the entire word. Production over at least several hundreds of ms both before and after the stop release can readily be examined in some detail, as opposed to either listening to the entire word (as in perceptual analysis) or closely examining only a few ms at a time (up to a hundred ms or so), as is typical of the acoustic analysis performed later in this thesis. Examination of stop production via SA provides information of a kind that is not readily accessible via either auditory-perceptual or acoustic analysis.

SA was performed for the stop consonants produced by the normal and dysarthric speakers involved in this study. Section 5.1 contains the general guidelines developed for attribute evaluation. Section 5.2 contains the results and discussion. Then the SA is summarized in Section 5.4.

5.1 Experiment

5.1.1 Corpus, Speakers, Recording Method, and Judges

Spectrograms were examined from eight words with word-initial stops: bad, bunch, dock, dug, geese, pat, tile, and coat. This dataset is the that utilized for the perceptual evaluations in Chapter 4 (refer to Sections 4.1.1 and 4.1.3). The 16 speakers (8 normal and 8 dysarthric) have been discussed in Section 4.1.2 and Chapter 2. There were two judges, speech researchers from our laboratory, the Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology.

5.1.2 General Guidelines for Attribute Evaluation

Spectrograms from three repetitions per word per speaker were included in the experiment, for a total of 384 utterances (word repetitions). The broadband spectrograms were created by an algorithm that placed a 6.4-msec Hamming window every 1 ms throughout each repetition, generating a 256-point DFT at each window placement, and normalizing the resultant spectrogram to the maximum amplitude on a per repetition basis.

The spectrograms were judged on seven attributes: presence of precursor, presence of prevoicing, abruptness of release, time course of release, voice onset time (VOT), time course of first formant frequency ($F1$) rise, and time course of second formant frequency ($F2$) change. The judges rated the production of these aspects of each utterance, utilizing a scale of 1 (Good), 2 (Fair) and 3 (Poor). More details regarding how the judgments were made for the individual attributes are given in the subsections below. A similar rating scale has been applied by Klatt and Klatt (1990). They established a four-step scale to quantify the presence or absence of random noise over the course of vowel production in the acoustic time waveform.

The judges calibrated their ratings schemes to one another by taking a subset of spectrograms from both normal and dysarthric speakers (approximately 40 of the total of 384), rated them independently of one another and then met to confer on

the results. When there was disagreement by 2 points (in other words, when one of the judges awarded a 1 and the other a 3), then a discussion ensued until the judges were in more agreement regarding the details of how to judge that particular attribute, and one of the judges would then change his/her value by 1–2 points to be in closer agreement with the other judge's value. The judges evaluated the remaining spectrograms (approximately 340) independently of one another.

A few remarks about the manner in which the spectrograms were rated should be made prior to a discussion of the attributes themselves. First and foremost, the guidelines appearing in the subsections below should be viewed as general, and are not meant to provide a comprehensive discussion of all situations encountered in the dysarthric speakers' spectrograms. The judges' experience and interpretation were relied upon for assessment of individual spectrograms. The judges found it helpful throughout the rating process to compare dysarthric speakers' spectrograms to a baseline established by the spectrograms of normal speakers producing the same words, in order to determine how to assign the ratings. The judges have also had prior experience in reading normal spectrograms, and, in the discussion of the attributes below, it is presumed that the reader is familiar with how to read normal spectrograms as well. The discussion of attributes is focused on how the dysarthric speakers' spectrograms deviate from normal. One of the judges, with less training in spectrogram reading than the other judge, occasionally supplemented information from the spectrograms with information from both perception and the time waveform, solely to determine the location of the stop release. With more training in spectrogram reading in the future, it is hoped that this step could be avoided. It is also important to realize that these seven attributes are not all mutually exclusive. For example, at the stop release itself the Time Course of Release attribute and the Abruptness of Release attribute overlap in their assessment of the quality of stop release production. In the instances in which the stop was not produced at all, attribute judgments were made in the vicinity of where the stop release should have occurred, i.e., the transition from precursor to vowel.

Rating	Description of Events Prior to Stop-Consonant Release
1	No precursor present (no noise or phonation other than normal prevoicing)
2	Small amount of noise present *and/or* phonation > 200 ms prior to release *and/or* phonation ends several tens of ms prior to release
3	Large amount of noise present, with or without phonation > 200 ms prior to release

Table 5.1: Precursor Attribute Assessment

Precursor

In the context of SA, "precursor" is defined to be any noise or phonation other than normal prevoicing (refer to Prevoicing below), occurring prior to the stop release. This definition of "precursor" does not differentiate between sounds generated by the speaker (except normal prevoicing) and background sounds (including, but not limited to, room noises, wheelchair noises, computer beeps and researchers' conversations with one another) in this time interval. It does allow for separation of prevoicing from most of the rest of the sounds occurring prior to the release, based on the anticipated low frequency range of the prevoicing. This definition can be contrasted with the definition of "precursor" used in the perceptual and acoustic analyses. In perceptual analysis (Chap. 4), "precursor" is defined to be any abnormal sound generated by the speaker in the time period preceding the stop release, including but not limited to abnormally long or loud prevoicing, audible breathing, etc. Listeners were instructed to ignore any noises such as computer beeps, static, sounds indicating the speaker was too close to the microphone or any other background noises not generated by the speaker's vocal tract, either preceding or during production of the utterance. In acoustic analysis (Chap. 6), the "precursor" time interval is defined to be the 100 ms immediately prior to the stop release (placement of the Hamming window never overlapped the stop release itself). From the average spectrum created over that time interval, only the amplitude of the peak in the 0–500 Hz range was examined, as a measure of prevoicing. The rating scale for the precursor in the context of SA is in Table 5.1.

Rating	Description of Events Prior to Stop-Consonant Release
1	Prevoicing duration $\leq$ 100 ms, relatively low intensity, ends $\leq$ 20 ms prior to release
2	Prevoicing duration $\leq$ 100 ms and relatively high intensity, *or* 100 ms $<$ prevoicing duration $\leq$ 200 ms and relatively low intensity *or* Prevoicing ends several tens of ms prior to release, while otherwise satisfying the duration and intensity requirements of Rating 1
3	Prevoicing duration $>$ 100 ms and relatively high intensity, *or* Prevoicing duration $>$ 200 ms irrespective of intensity, *or* Prevoicing ends several tens of ms prior to release, while otherwise satisfying the duration and intensity requirements of Rating 2

Table 5.2: Prevoicing Attribute Assessment in Voiced Stop Production

Rating	Description of Events Prior to Stop-Consonant Release
1	No prevoicing present
2	(unassigned)
3	Prevoicing present

Table 5.3: Prevoicing Attribute Assessment in Voiceless Stop Production

Prevoicing

Prevoicing is the vibration of the vocal folds immediately prior to the stop-consonant release. It appears in the spectrogram as periodic excitation of the glottal source in the 0–500 Hz range. Normal speakers may or may not prevoice prior to voiced stop production, in anticipation of the short VOT following the voiced stop. When normal speakers do prevoice, the prevoicing is short in duration (typically $\leq$ 100 ms) and relatively low in intensity. Normal speakers are not expected to prevoice prior to voiceless stop-consonant production, in which the VOT is much longer. In the dysarthric speakers' spectrograms of voiced stops, when phonation occurs more than 200 ms preceding the stop release or ends several tens of ms prior to the release, it is considered to overlap with the Precursor attribute (refer to Precursor above). The rating scales for prevoicing are in Table 5.2 for voiced stops and Table 5.3 for voiceless stops.

Rating	Description of Events at Stop-Consonant Release
1	Distinct, obvious, rapid release time identified
2	Release time a little unclear, blurred or "fuzzy", in which formant frequencies are not all excited simultaneously but rather are excited in a "staggered" fashion prior to vowel; Double burst may be evident at release for labial or alveolar stops
3	Release time very unclear/blurred/"fuzzy", such that it is difficult or impossible to identify a stop release; Triple or higher-order burst may be evident at release

Table 5.4: Abruptness of Release Attribute Assessment

Abruptness of Release

The Abruptness of Release attribute characterizes the nature of the stop-consonant release itself, how readily the release time is identified and how instantaneously the release occurs. This attribute is a detailed examination of only the stop release characteristics, over the course of approximately 10–20 ms surrounding the time of the stop release. In contrast, the Time Course of Release attribute described below examines a 200 to 300 ms time period, encompassing the release time as well as a period of time both before and after the release. The rating scale for abruptness of release is in Table 5.4.

Time Course of Release

The Time Course of Release attribute attempts to broadly characterize the transition from the stop closure interval through the stop release and into the following vowel (a 200 to 300 ms time period). This attribute does not provide a detailed accounting of only one aspect of stop production, but rather determines whether a series of aspects is produced well. In the time period prior to the release, this attribute focuses on the presence or absence of noise (typically mid- to high-frequency noise, $>$ 2 kHz) *immediately* prior to the release. (In contrast, the Precursor attribute focuses on noise throughout several hundreds of ms prior to the release.) During and after the release, the focus is on: (a) when visible, appropriate excitation of vocal-tract formants in the frication noise (and aspiration noise for voiceless stops) at and after the stop release;

Rating	Description of Events Near Stop-Consonant Release
1	No or very little noise present before release and after vowel onset; When visible, appropriate excitation of formants in frication noise, aspiration noise (for voiceless stops), and the higher formants ($\geq F3$) at vowel onset; No additional formants and no dropouts in spectral energy in the $F1$ and $F2$ transition region from stop to vowel steady state
2	A small amount of noise is present before and/or after release, but it does not obscure visible formant excitation (formant excitation still may not be visible on the spectrogram due to low intensity); Formants appear to be excited appropriately, although there may be small fluctuations in intensities and frequencies; No additional formants and up to only one dropout in spectral energy may be present in the $F1$ and $F2$ transition region from stop to vowel steady state
3	Enough noise is present around the time of the stop release to make detection of the release difficult or impossible; Formant excitation may not be appropriate, with large fluctuations in intensities and frequencies; Additional formants and dropouts in spectral energy may be present in the $F1$ and $F2$ transition region from stop to vowel steady state

Table 5.5: Time Course of Release Attribute Assessment

(b) the presence or absence of noise at vowel onset (typically in the frequency range $> F2$); (c) when visible, the appropriate excitation of the higher formants ($\geq F3$) at vowel onset; and (d) the existence of dropouts in spectral energy (time periods during which spectral energy is first present, then absent, then present once again) or additional formants in the $F1$ and $F2$ transition region from stop to vowel steady state. The rating scale for time course of release is in Table 5.5.

VOT

For the purposes of SA, Voice Onset Time (VOT) is the time duration between the stop-consonant release and the onset of voicing in the following vowel. The onset of voicing is defined to be the first pitch period of the vowel. (This definition of VOT is *different* from the definition utilized in the acoustic analysis, in which the onset of voicing is defined to be at a time typically slightly later in the utterance, at the start of the first glottal pulse in which the peak amplitude is $\frac{1}{4}$ of the maximum amplitude occurring during vowel steady state.) The VOT attribute is typically a more useful

Rating	Description of Events After Stop-Consonant Release
1	VOT within normal range for the voiced stop
2	VOT $\approx$ 10 ms or so longer than normal *or* large fluctuations in intensity of the glottal pulses during the first 100 ms or so of the vowel, making it difficult to determine the onset of voicing
3	VOT $\geq$ 20 ms longer than normal

Table 5.6: VOT Attribute Assessment for Voiced Stop Production

Rating	Description of Events After Stop-Consonant Release
1	VOT within normal range for the voiceless stop
2	VOT slightly too short or long ($\leq$ 15 ms outside normal range)
3	VOT quite short or long ($\geq$ 15 ms outside normal range)

Table 5.7: VOT Attribute Assessment for Voiceless Stop Production.

measure in the context of a voiceless stop than a voiced stop, since it is rare for the VOT to be too long in the voiced stops produced by the normal and dysarthric speakers involved in this study. The rating scale for VOT are in Table 5.6 for voiced stops and Table 5.7 for voiceless stops.

Time Course of F1 Rise

The Time Course of $F1$ Rise attribute characterizes the formant values, transition rate and transition direction of the first formant frequency, $F1$, from the first glottal pulse of the vowel to a time approximately 100 ms or so later in the vowel. For normal speakers, this attribute is typically a more meaningful measure in the context of voiced stops, rather than voiceless, aspirated stops. The $F1$ transition following a voiceless, aspirated stop is largely complete by the time of vowel onset. Additionally, within the voiced stops produced by normal speakers, the Time Course of $F1$ Rise attribute is a more useful measure for stops preceding low vowels, since the final value of $F1$ is higher for a low vowel, resulting in a greater transition in frequency for $F1$ with more of the transition likely to occur during the glottal pulses of the vowel rather than preceding vowel initiation. Therefore, for the voiceless, aspirated stops and voiced stops preceding high vowels in this study (utterances pat, tile, coat, and geese), most

Rating	Description of Events After Stop-Consonant Release
1	$F1$ in first 1 to 2 pitch periods < approx. 80% of $F1$ in vowel steady state; $F1$ at end of 100-ms interval within ±200 Hz of normal
2	$F1$ in first 3 to 4 pitch periods < approx. 80% of $F1$ in vowel steady state; *and/or* $F1$ at end of 100-ms interval more than 200 Hz but less than 400 Hz different from normal
3	$F1$ in the first 5 or more pitch periods < approx. 80% of $F1$ in vowel steady state; *and/or* $F1$ transition falls, instead of rises; *and/or* One or more dropouts in spectral energy exist in $F1$ within the 100 ms following vowel onset; *and/or* $F1$ at end of 100-ms interval more than 400 Hz different from normal

Table 5.8 : Time Course of $F1$ Rise Attribute Assessment in Voiced Stop Production. When the $F1$ rise is not visible in the spectrogram, this assessment is based upon only the $F1$ value at the end of 100-ms interval, and is therefore less meaningful.

Rating	Description of Events After Stop-Consonant Release
1	$F1$ in first 1 to 2 pitch periods ≤ $F1$ in vowel steady state; $F1$ at end of 100-ms interval within ±200 Hz of normal
2	$F1$ in first 3 to 6 pitch periods < $F1$ in vowel steady state; *and/or* $F1$ at end of 100-ms interval more than 200 Hz but less than 400 Hz different from normal
3	$F1$ in the first 7 or more pitch periods < $F1$ in vowel steady state; *and/or* $F1$ transition falls, instead of rises; *and/or* One or more dropouts in spectral energy exist within the 100 ms following vowel onset; *and/or* $F1$ at end of 100-ms interval more than 400 Hz different from normal

Table 5.9 : Time Course of $F1$ Rise Attribute Assessment in Voiceless Stop Production. When the $F1$ rise is not visible in the spectrogram, this assessment is based upon only the $F1$ value at the end of 100-ms interval, and is therefore less meaningful.

or all of the $F1$ transition is frequently not visible in the normal spectrograms as well as in some of the dysarthric spectrograms. When the rise is visible, it is possible to evaluate it based upon all the information contained in Tables 5.8 and 5.9. When the rise is not visible, the Time Course of $F1$ Rise attribute is not as meaningful. The judgments then become based solely upon whether the steady-state formant frequency values in the vowel are correct or not, which is not a reflection of the transition itself, and, furthermore, can be difficult to assess from the spectrograms alone due to their poor frequency resolution. The rating scales for time course of $F1$ rise are in Table 5.8 for voiced stops and Table 5.9 for voiceless stops.

Rating	Description of Events After Stop-Consonant Release
1	Initial frequency of $F2$ trajectory within $\pm$200 Hz of correct; Rate of $F2$ transition can be only at most slightly incorrect and direction of $F2$ transition (increasing, decreasing or constant) must be correct; $F2$ at end of 100-ms interval within $\pm$200 Hz of normal
2	Initial frequency of $F2$ trajectory *not* within $\pm$200 Hz of correct; *or* $F2$ transition rate noticeably slower than normal; *or* $F2$ transitions in incorrect direction; *or* $F2$ at end of 100-ms interval more than 200 Hz but less than 500 Hz different from normal
3	More than one of the items listed in Rating 2 is present; *and/or* Dropout(s) in spectral energy exist during $F2$ transition; *and/or* $F2$ at end of 100-ms interval more than 500 Hz different from normal

Table 5.10 : Time Course of $F2$ Change Attribute Assessment. When the $F2$ transition is not visible in the spectrogram, this attribute assessment is based upon only the $F2$ value at the end of 100-ms interval, and is therefore less meaningful.

Time Course of F2 Change

The Time Course of $F2$ Change attribute characterizes the formant values, transition rate and transition direction of the second formant frequency, $F2$, from the first glottal pulse of the vowel to a time approximately 100 ms or so later in the vowel. Similarly to the Time Course of $F1$ Rise attribute for normal speakers, the Time Course of $F2$ Change attribute is typically a more useful measure for voiced stops than for aspirated, voiceless stops. This attribute can be one of the more difficult attributes to assess, since the F2 trajectory can vary considerably depending upon the choice of stop and following vowel. As an aid to correct identification of the start of the F2 trajectory in the vowel, excitation of F2 may be visible in the preceding frication noise (in the case of voiced stops) or frication and aspiration noise (in the case of voiceless stops). While keeping in mind that not all of the transition may be visible for the voiceless stops in this study, it is possible to apply the rating scale appearing in Table 5.10 for time course of $F2$ change.

Spectrograms are included from six speakers to demonstrate attribute assessment for normal, mildly- and moderately-dysarthric speakers. These spectrograms, along with attribute assignments, are shown in Figures 5-1 to 5-6.

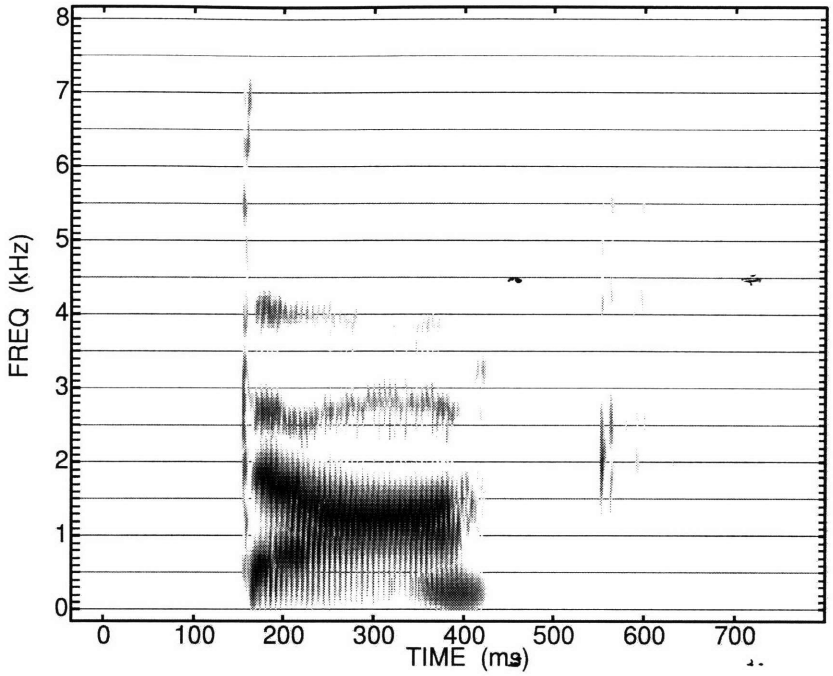

Figure 5-1 : Spectrogram for normal female speaker (NF3) saying the word dock. Spectrogram calculated using a 6.4-msec Hamming window to generate a 256-point DFT spectrum every 1 msec. All seven attributes are assigned a value of 1 (averaged across the two judges).

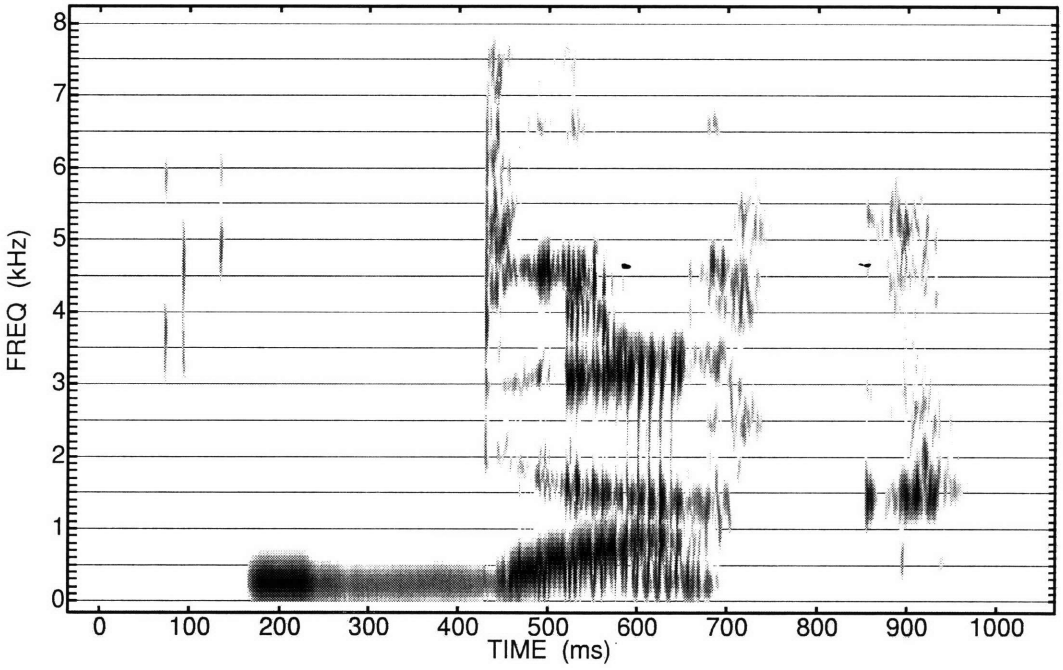

Figure 5-2 : Spectrogram for dysarthric female speaker (DF1) saying the word dock. Spectrogram calculated using a 6.4-msec Hamming window to generate a 256-point DFT spectrum every 1 msec. The attribute assignment (averaged across the two judges) is as follows: Precursor (1.5), Prevoicing (2.5), Abruptness of Release (1), Time Course of Release (2), VOT (1), Time Course of F1 Rise (2), and Time Course of F2 Change (1).

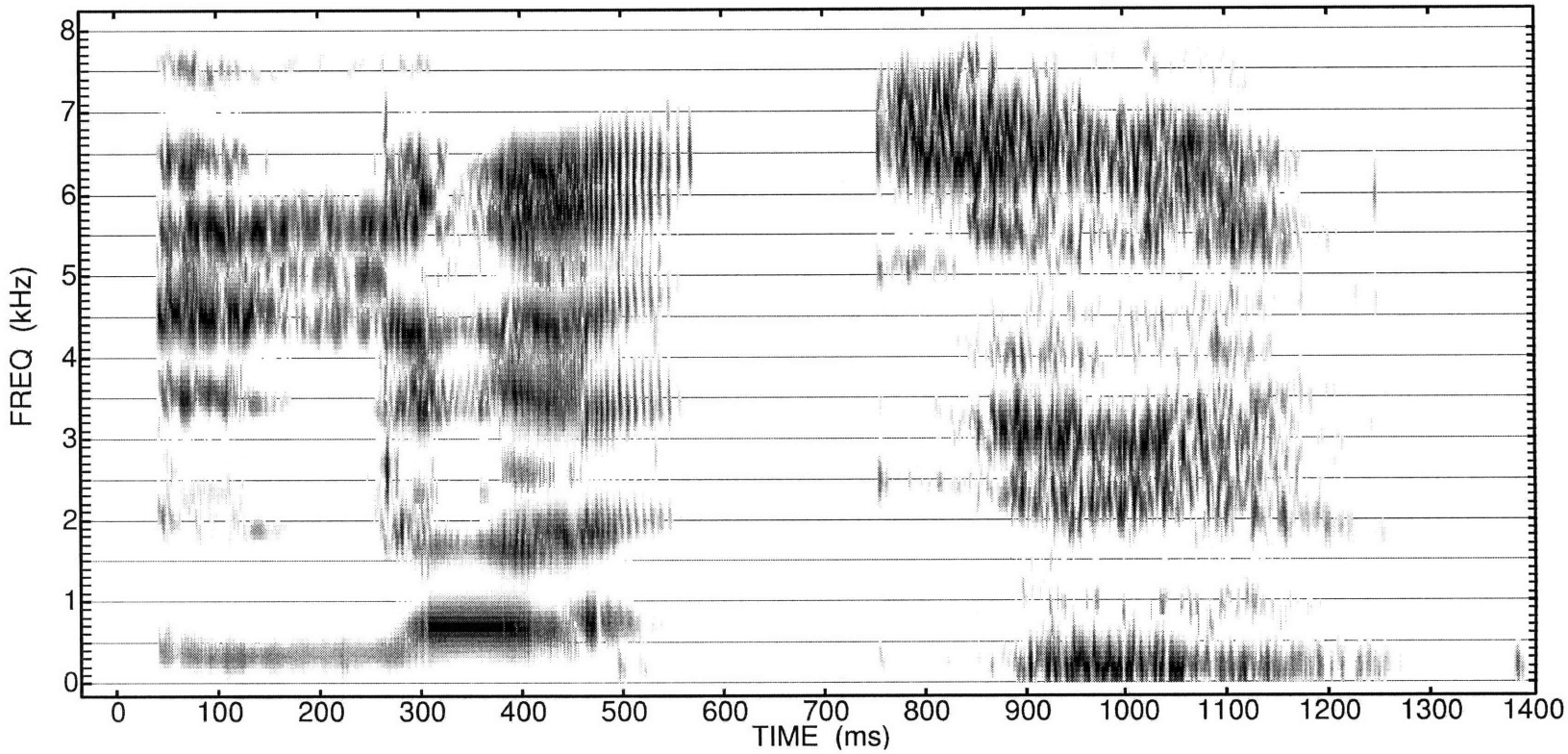

Figure 5-3 : Spectrogram for dysarthric female speaker (DF4) saying the word dock. Spectrogram calculated using a 6.4-msec Hamming window to generate a 256-point DFT spectrum every 1 msec. The attribute assignment (averaged across the two judges) is as follows: Precursor (3), Prevoicing (2), Abruptness of Release (2), Time Course of Release (2.5), VOT (1), Time Course of F1 Rise (1.5), and Time Course of F2 Change (2).

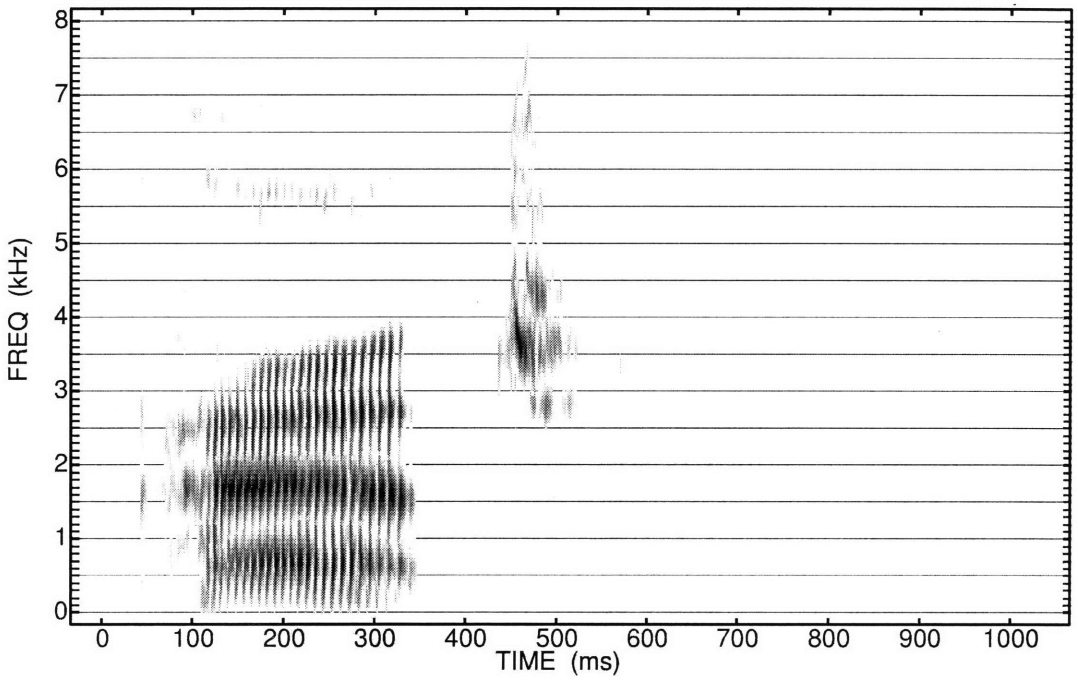

Figure 5-4 : Spectrogram for normal male speaker NM3 saying the word pat. Spectrogram calculated using a 6.4-msec Hamming window to generate a 256-point DFT spectrum every 1 msec. All seven attributes are assigned a value of 1 (averaged across the two judges).

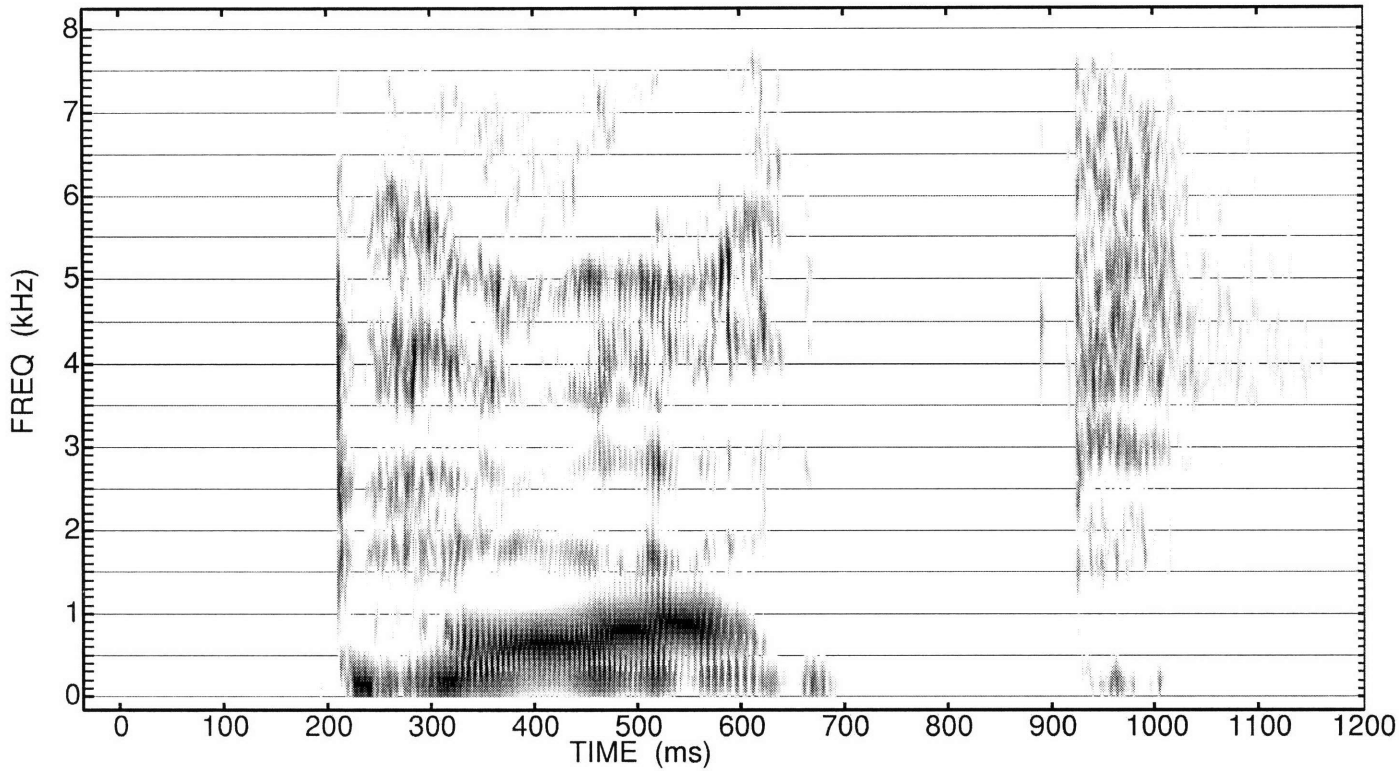

Figure 5-5 : Spectrogram for dysarthric male speaker (DM1) saying the word pat. Spectrogram calculated using a 6.4-msec Hamming window to generate a 256-point DFT spectrum every 1 msec. The attribute assignment (averaged across the two judges) is as follows: Precursor (1), Prevoicing (1), Abruptness of Release (1), Time Course of Release (2), VOT (2.5), Time Course of F1 Rise (2.5), and Time Course of F2 Change (1).

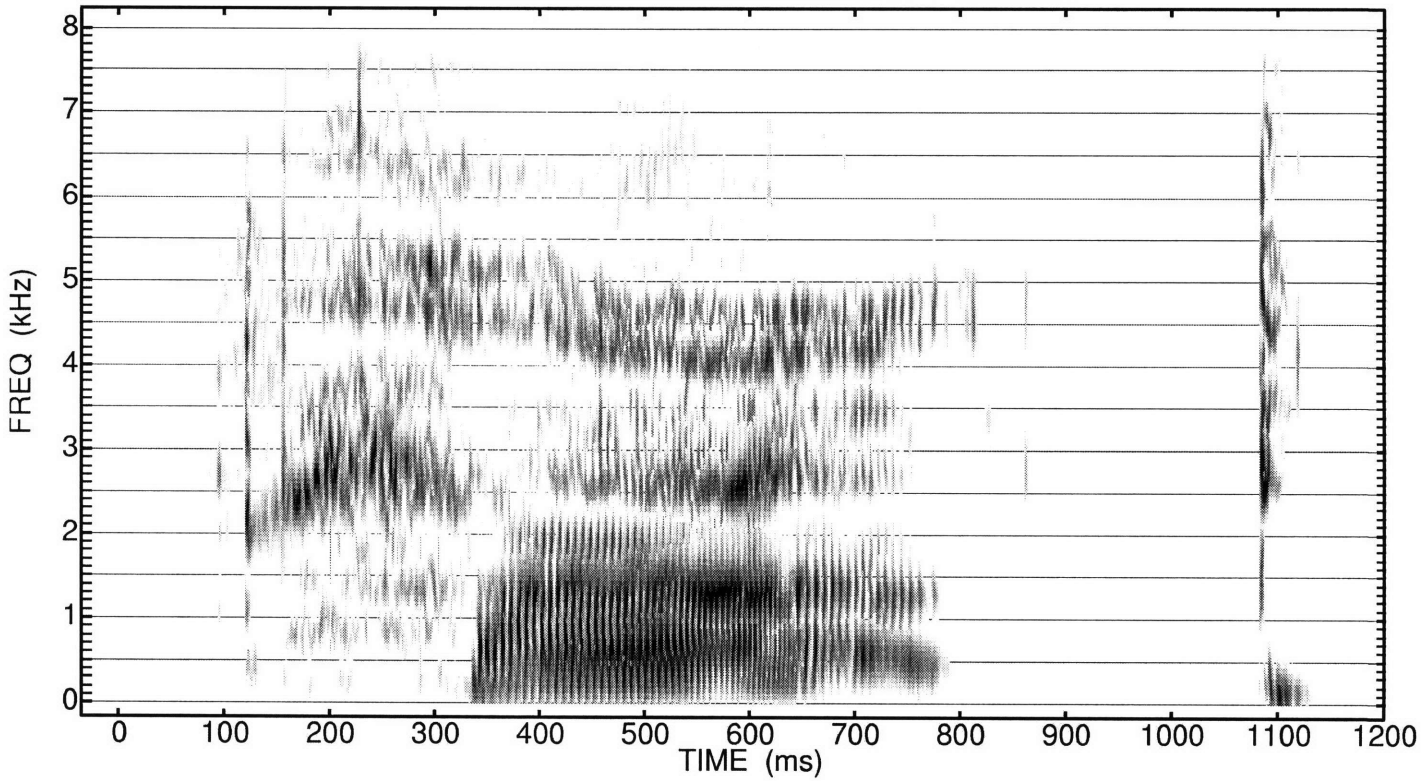

Figure 5-6 : Spectrogram for dysarthric male speaker (DM4) saying the word pat. Spectrogram calculated using a 6.4-msec Hamming window to generate a 256-point DFT spectrum every 1 msec. The attribute assignment (averaged across the two judges) is as follows: Precursor (1.5), Prevoicing (1), Abruptness of Release (1), Time Course of Release (2.5), VOT (3), Time Course of F1 Rise (2.5), and Time Course of F2 Change (2).

5.2 Results and Discussion

The results for each of the seven attributes in the Qualitative Spectrogram Analysis (SA) are shown in Figures 5-8 to 5-14. The results are averaged across all utterances (except where noted), repetitions and judges, providing one rating per speaker. For normal speakers, the results are averaged across speakers as well, providing one rating overall.[1] The speaker order appearing in the figures is that of the stop goodness score developed in Chapter 4. Additional SA attribute data appear in Appendix G.

Prior to a discussion of the results for each attribute, it is important to give special consideration to one dysarthric speaker in particular, DF2. In contrast to the other speakers, DF2 has difficulty appropriately controlling her velopharyngeal port opening. Consequently, air leaks out through her nose preceding and throughout almost all of her utterances. This audible air leakage appears in her spectrograms as broadband noise in the mid- to high-frequency range, typically 2–8 kHz, but occasionally as low as 1 kHz. A sample spectrogram of her speech appears in Figure 5-7. The exact characteristics of this noise production do vary with the sounds this speaker generates, but the virtually constant presence of noise in at least some of the speech frequencies has an effect across 4 of the 7 attributes. The attributes affected are those attributes which examine events in the 1–8 kHz frequency range. Only three attributes (Prevoicing, VOT and Time Course of F1 Rise) focus exclusively on events in the 0–1 kHz range, and therefore remain unaffected by this noise. Although listeners can fairly readily distinguish between this noise and the underlying speech signal most of the time (as shown in the perceptual experiment results of Chap. 4), the distinction is much more difficult to make in the visual spectrogram evaluation performed in SA.

The Precursor attribute results are shown in Figure 5-8, averaged across all 8 utterances. In general, the presence of a precursor is associated with a lower stop goodness score. Due to air leaking out her nose prior to the stop release, DF2 has a particularly poor precursor score compared to her stop-goodness speaker ranking. As

[1]The normal speakers' results are so similar to one another that it was deemed not necessary to report their ratings individually.

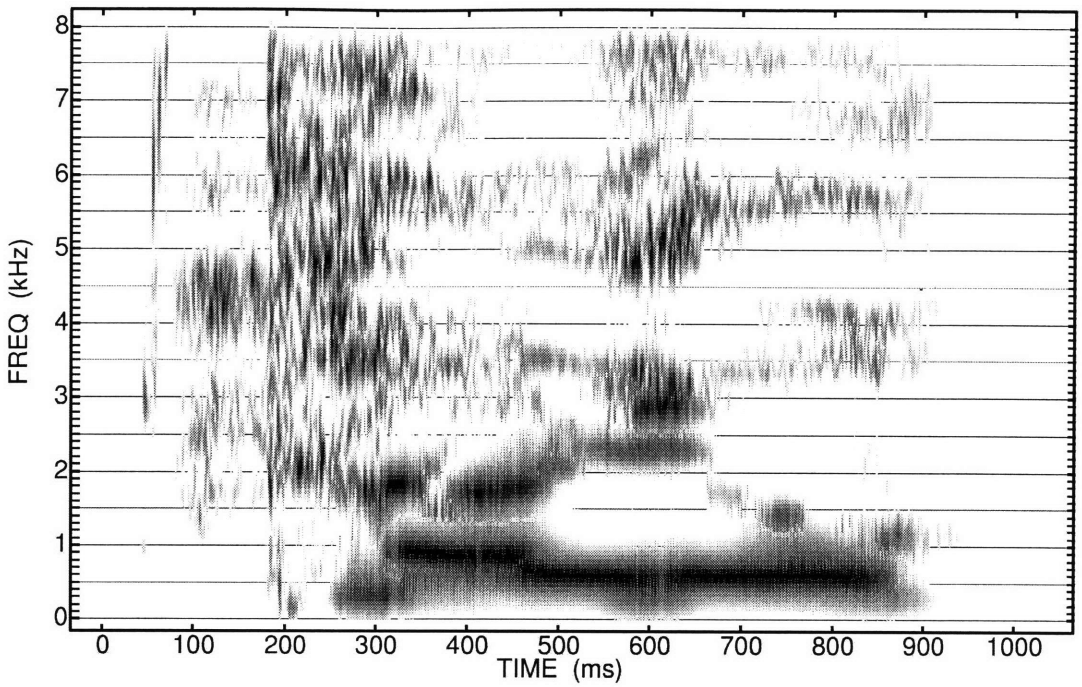

Figure 5-7 : Spectrogram for dysarthric female speaker (DF2) saying the word tile. Spectrogram calculated using a 6.4-msec Hamming window to generate a 256-point DFT spectrum every 1 msec. Air leakage through the velopharyngeal port appears as broadband noise, generally in the 2–8 kHz range, but occasionally as low as 1 kHz.

discussed in Chapter 4, speakers DM4 and DF4 tend to produce a variety of precursor sounds by varying the positions of their articulators and vocal folds. The sounds range from abnormally long prevoicing to inadvertent vowel generation and noise production. The precursor for speaker DM3 can partly be attributed to background noise in the recording environment. Although these noises are not speaker-generated, they remain difficult to separate from speech sounds by visual examination of the spectrogram data alone.

The Prevoicing attribute results are shown in Figure 5-9(a) for voiced stops and Figure 5-9(b) for voiceless stops. Since normal speakers occasionally prevoice prior to voiced stops, it is anticipated that prevoicing will be more common prior to the voiced stop production of dysarthric speakers as well. Indeed, that trend can be observed by comparing Figures 5-9(a) and (b). The presence or absence of prevoicing appears to be speaker dependent to a certain degree. While some of the dysarthric speakers prevoice prior to voiced stops but do not do so prior to voiceless stops, the three dysarthric speakers who tend to prevoice prior to voiceless stops (DF2, DM4 and

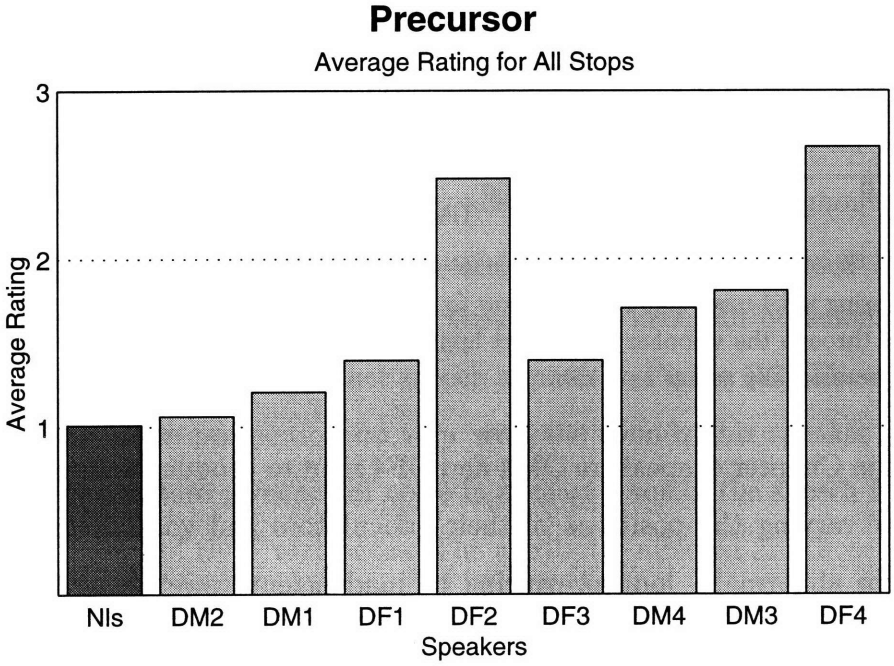

Figure 5-8 : Precursor attribute results from Spectrogram Analysis. Ratings averaged across 8 utterances, 3 repetitions/utterance and 2 judges per speaker. The normal speakers' ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

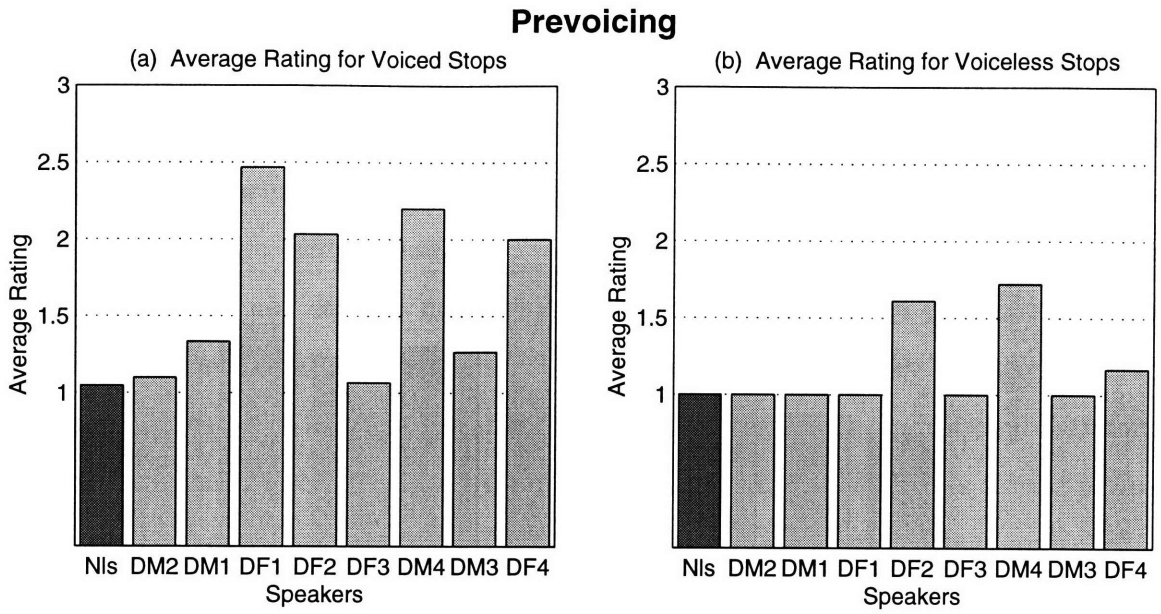

Figure 5-9 : Prevoicing attribute results from Spectrogram Analysis. Ratings averaged across 2 judges, 3 repetitions/utterance and (a) 5 utterances containing intended word-initial voiced stops or (b) 3 utterances containing intended word-initial voiceless stops. For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness scores, as determined in Chapter 4.

DF4) also tend to prevoice prior to voiced stops. For these three dysarthric speakers, the presence of abnormally long prevoicing, unnaturally loud prevoicing, or prevoicing that ends several tens of milliseconds prior to the release (the prevoicing may actually be excitation of $F1$ in the context of inadvertent vowel generations), likely contributes to the listener judging these stops to have precursors in Q1 of Chapter 4 (Fig. 4-6, page 87). (In Chap. 4, Q1, the presence of abnormal prevoicing was included in the judgment of presence of precursor.) Speaker DF1 tends to have unnaturally long and loud prevoicing prior to her voiced stops but not the voiceless ones. She is apparently anticipating the need for vocal-fold vibrations at or shortly after the release of a voiced stop by building up subglottal pressure, approximating the vocal folds, and initiating vocal-fold vibrations prior to the release. She differs from normals in that she builds up too much subglottal pressure and initiates vocal-fold vibrations too early.

The Abruptness of Release attribute is shown in Figure 5-10, averaged across all 8 utterances. In general, as the release time becomes slower and less easily identified,

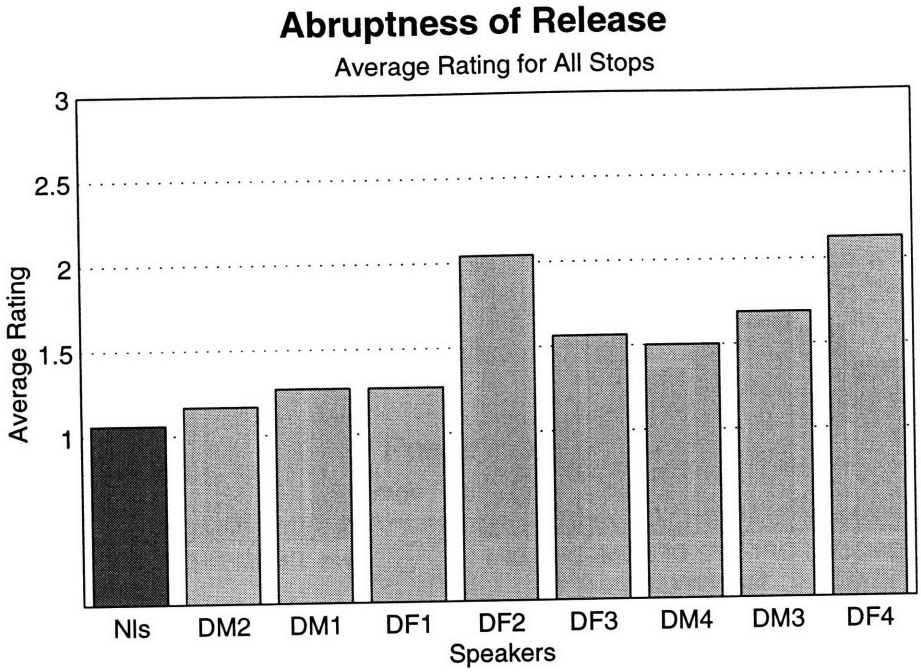

Figure 5-10 : Abruptness of Release attribute results from Spectrogram Analysis. Ratings averaged across 8 utterances, 3 repetitions/utterance and 2 judges per speaker. For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

the stop goodness score becomes poorer as well. With a leaky velopharyngeal port, speaker DF2 cannot build up adequate intraoral pressure prior to the release. This poor pressure buildup, combined with the air leaking through her nose, leads to weaker bursts and formant frequencies obscured by noise. For the remaining mildly dysarthric speakers (DM2, DM1 and DF1), the release tends to be quite abrupt and is comparable to the release for the normal speakers. For the four moderately dysarthric speakers (DF3, DM4, DM3 and DF4), the number of times that the release is judged to be unclear, blurred, "fuzzy" or difficult to identify increases, indicating that these speakers, on average, have more difficulty moving the primary articulator rapidly at the time of the release.

The Time Course of Release attribute is shown in Figure 5-11, averaged across all 8 utterances. Although speaker ratings are slightly more variable, there does seem to be a general trend toward poorer ratings as the stop goodness score decreases. This attribute is the one which shows the largest difference (about 0.7) between normal

speakers and the speaker with the best average ratings of the dysarthric speakers. The Time Course of Release attribute examines whether a series of aspects of the stop is produced well, from the stop closure through the release and into the following vowel. Speakers whose mean ratings are in the vicinity of a 2 (such as DM2, DM1, DF1, and, perhaps DM4) have, on average, a small amount of noise in this time period, although not enough noise to obscure formant-frequency excitation. They also tend to have small fluctuations in the intensity and frequency of their first two formant frequencies. In their $F1$ and $F2$ transition regions, no additional formants appear; however, there may be a dropout in spectral energy (defined as a time period when spectral energy is momentarily absent). As the speakers' ratings approach 3 on average (such as speakers DM3, DF4, and, to a lesser extent, DF2 and DF3) the quantity of noise increases, tending to make detection of the release difficult or impossible. Additionally, formant excitation is more likely to be characterized by large fluctuations in intensities and frequencies, the presence of additional formants, and the existence of one or more dropouts in spectral energy. (For DF2, at least some of the noise is attributable to air leaking out her nose. There is also the presence of a nasal-cavity resonance due to this air leakage.)

The Voice Onset Time (VOT) attribute results are shown in Figure 5-12(a) for voiced stops and Figure 5-12(b) for voiceless stops. The results in Figure 5-12(a) reflect when the VOT is too long in voiced-stop production, and the results in Figure 5-12(b) reflect when the VOT deviates from normal in voiceless-stop production. From Figure 5-12(a), it is rare for the VOT to be too long in the voiced-stop production of either the normal or the dysarthric speakers in this study. A comparison of Figures 5-9(a) and 5-12(a) reveals that speakers DF1, DF2, DM4 and DF4 abnormally prevoice much more frequently than they lengthen the VOT for the voiced stops. In other words, if one of these dysarthric speakers is going to err in the voicing aspect of voiced-stop production, s/he tends to initiate vocal-fold vibration too early rather than too late. The average rating for the VOT of the voiceless stops corresponds well to the stop goodness score for all of the speakers. Although this VOT average rating can indicate that the VOT is judged to be either too short or too long, the typical

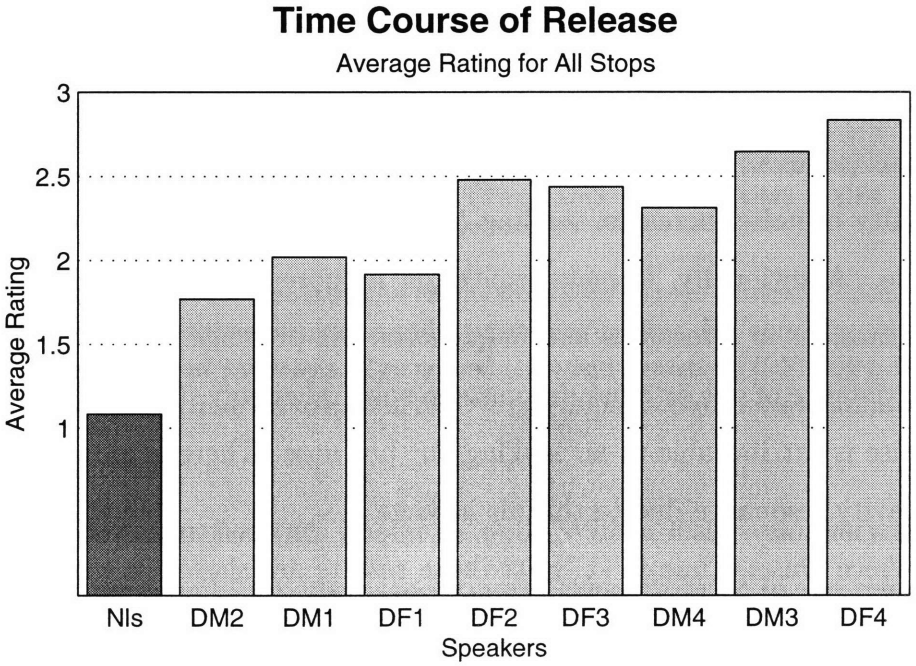

Figure 5-11 : Time Course of Release attribute results from Spectrogram Analysis. Ratings averaged across 8 utterances, 3 repetitions/utterance, and 2 judges per speaker. For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

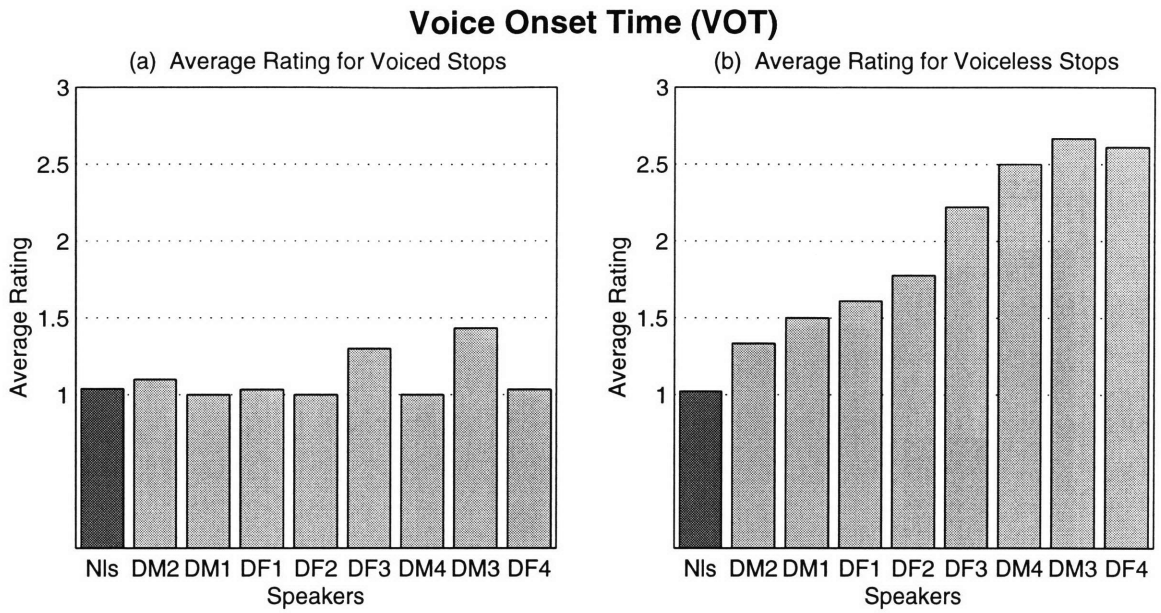

Figure 5-12 : Voice Onset Time (VOT) attribute results from Spectrogram Analysis. For each speaker, ratings averaged across 2 judges, 3 repetitions/utterance and (a) 5 utterances containing intended word-initial voiced stops, or (b) 3 utterances containing intended word-initial voiceless stops. For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

manner in which the dysarthric speakers err is to lengthen the VOT of the voiceless stops.

The results for the VOT attribute in Figure 5-12 can be compared to the listener responses to Q2 in Figure 4-7 (page 88), identifying the type of voicing of the word-initial sound. The listener responses agree with the finding that it is more common for the dysarthric speakers to deviate from normal VOT duration for voiceless stops than for voiced stops. Figure 4-7(a) examines, in essence, when the duration of the VOT is too short, such that a voiceless stop is identified to be voiced. Speakers DM3, DM4, and, to a lesser extent DF3, all have VOTs for voiceless stops that are too short, as indicated both in Figure 4-7(a) and Figure 5-12(b). Speaker DF4, also judged to have a deviant VOT in Figure 5-12(b), produces the majority of her intended voiceless stop consonants as voiceless (from Fig. 4-7(a)); therefore, the deviation must be in the direction of a prolonged VOT.

The Time Course of $F1$ Rise attribute results are shown in Figure 5-13. Results

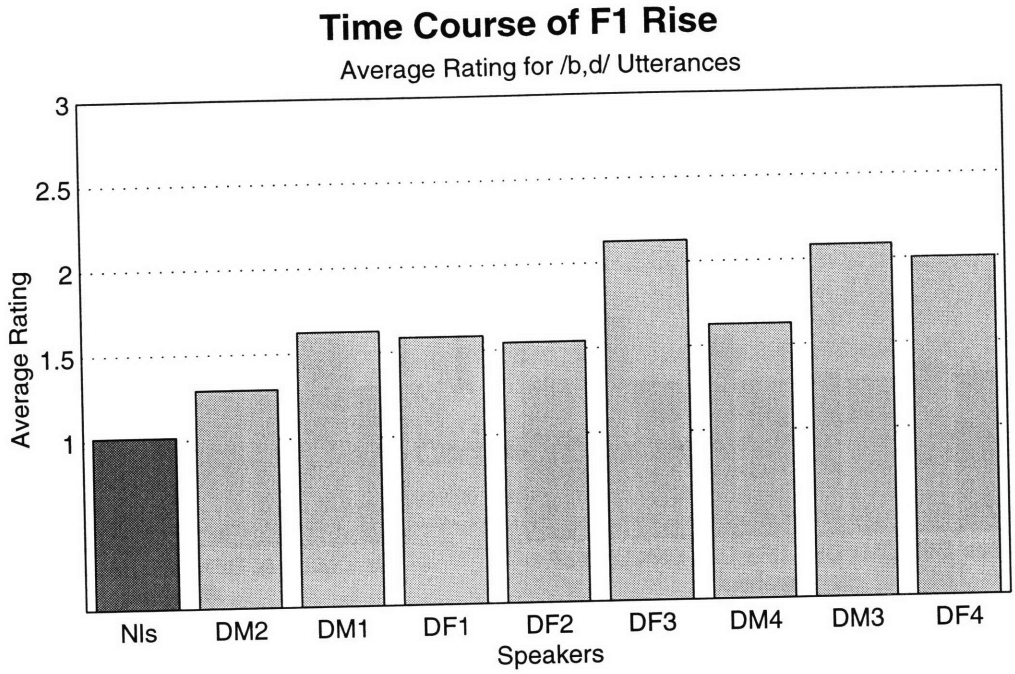

Figure 5-13 : Time Course of $F1$ Rise attribute results from Spectrogram Analysis. For each speaker, ratings averaged across 2 judges, 3 repetitions/utterance, and the 4 utterances containing either intended word-initial /b/ or /d/ followed by a low vowel. For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

are only shown averaged across the four utterances containing voiced stops followed by low vowels, since these utterances are the only ones for which the $F1$ rise is consistently visible, and therefore measurable. In this dataset, these utterances are bad, bunch, dock, and dug. From this figure, it appears that there is a general trend toward poorer time course of $F1$ rise as stop goodness scores decrease, although there is some interspeaker variability. Compared to normal, a poorer time course of $F1$ rise is associated with one or more of the following: slower transition, incorrect transition direction, incorrect value for $F1$ 100 milliseconds into the vowel, and presence of one or more dropouts in spectral energy during the transition.

The Time Course of $F2$ Change attribute results are shown in Figure 5-14. Results are only shown averaged across the five utterances containing voiced stops, since these utterances are the only ones for which the $F2$ trajectory is consistently visible, and therefore measurable. Similar to the Time Course of $F1$ Rise results, these results

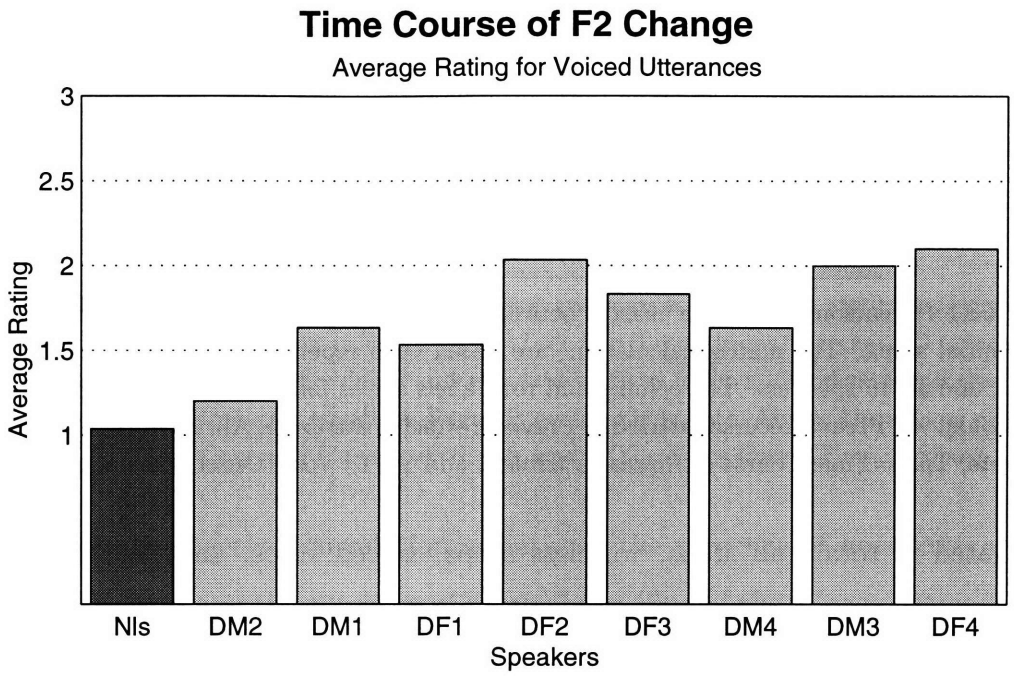

Figure 5-14 : Time Course of F2 Change attribute results from Spectrogram Analysis. For each speaker, ratings averaged across 2 judges, 3 repetitions/utterance, and 5 utterances containing intended word-initial voiced stops. For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

also reveal a general trend toward an association between poorer time course of $F2$ change and lower stop goodness scores, although again there is variability between speakers. Compared to normal, a poorer time course of $F2$ change is associated with one or more of the following: incorrect initial value of $F2$, slower rate of transition, incorrect transition direction, incorrect value for $F2$ 100 milliseconds into the vowel, and presence of one or more dropouts in spectral energy during the transition. Similar to the other attributes which examine events in frequency regions $>$ 1 kHz, the rating for the time course of $F2$ change may be influenced for speaker DF2 by air leaking out her nose.

The relationship between the seven attributes and the stop goodness score was explored through the calculation of a series of Pearson r correlation matrices. Taking into consideration that two of the attributes, Time Course of $F1$ Rise (TCF1) and Time Course of $F2$ Change (TCF2), are only measurable for subsets of the stop consonants (/b,d/ and voiced word-initial stops, respectively), four matrices were

	Good	Prec	Prev	Abru	TCR	VOT
Good	1.000					
Prec	-0.727	1.000				
Prev	-0.467	0.646	1.000			
Abru	-0.642	**0.822**	0.487	1.000		
TCR	**-0.874**	0.756	0.443	0.706	1.000	
VOT	-0.621	0.330	-0.047	0.357	0.672	1.000

Table 5.11 : Pearson correlation matrix between stop goodness score and SA attributes for all word-initial stops. The matrix calculations are based on 3 repetitions/utterance, 8 utterances, 2 judges, and all 16 speakers. The column and row labels are as follows: Good=stop goodness score (from Chap. 4), Prec=Precursor attribute, Prev=Prevoicing attribute, Abru=Abruptness of Release attribute, TCR=Time Course of Release attribute, and VOT=Voice Onset Time attribute.

calculated in total: all stops, voiceless stops, voiced stops, and /b,d/ stops. The resultant matrices are shown in Tables 5.11–5.14, respectively. In each matrix, the first column, labeled "Good", indicates how well the individual attributes are able to predict the stop goodness score. A negative sign on an r value indicates that the relationship between the goodness scores and the attribute is negative. The remaining columns provide information about the relationships between the various attributes. A correlation will be considered "high" if the magnitude of the correlation coefficient is in the range 0.8–1.0. Since the attributes are not mutually exclusive, correlation between some of the attributes is to be expected.

Across all stops, the only attribute highly predictive of stop goodness is the Time Course of Release (TCR in Table 5.11). This attribute examines, in part, the amount of noise present over several hundreds of msec near the stop release, irrespective of the voicing or place of the stop. There is a strong relationship between the presence of that noise and poorer stop goodness scores. That noise is attributed to one or more of the following events: prolonged frication noise, prolonged or inappropriate (in the case of voiced stops) aspiration noise, and air leaking through a faulty velopharyngeal port. A high correlation is also observed in Table 5.11 between the presence of a precursor (Prec) and the abruptness of the release (Abru). The Precursor attribute examines both phonation and noise production prior to the release. Of the two, only noise production is examined by this attribute *immediately* prior to the release. (The Prevoicing (Prev) attribute examines phonation immediately prior to the release.) It

	Good	Prec	Prev	Abru	TCR	VOT
Good	1.000					
Prec	-0.630	1.000				
Prev	-0.462	0.568	1.000			
Abru	-0.486	0.754	0.488	1.000		
TCR	**-0.850**	0.713	0.417	0.650	1.000	
VOT	**-0.906**	0.598	0.326	0.495	**0.891**	1.000

Table 5.12 : Pearson correlation matrix between stop goodness score and SA attributes for all word-initial voiceless stops. The matrix calculations are based on 3 repetitions/utterance, 3 utterances, 2 judges, and all 16 speakers. The column and row labels are as follows: Good=stop goodness score (from Chap. 4), Prec=Precursor attribute, Prev=Prevoicing attribute, Abru=Abruptness of Release attribute, TCR=Time Course of Release attribute, and VOT=Voice Onset Time attribute.

is likely that the observed correlation is between the presence of noise immediately prior to the release and the presence of noise at the time of the release, worsening the release abruptness.

For the voiceless stops, VOT becomes a strong predictor of stop goodness, in addition to the Time Course of Release (Table 5.12). The VOT attribute for voiceless stops reflects the deviation of VOT from normal. A high correlation exists between poorer stop goodness scores and increasing VOT deviation from normal, either toward a shorter VOT (more similar to voiced stops) or a longer VOT (likely due to increasing the aspiration noise interval). It should be observed, however, that these two attributes are also highly correlated with each other. Part of the variability in stop goodness score that is explained by Time Course of Release is also explained by VOT.

When only voiced stops are considered, an additional attribute becomes applicable, Time Course of $F2$ Change. From Table 5.13, it is observed that four attributes are highly correlated with stop goodness scores: Precursor, Abruptness of Release, Time Course of Release and Time Course of $F2$ Change. Inadvertent vowel generation is more likely to occur preceding voiced stops than voiceless stops. When considered along with noise during that pre-release time period, the presence of a precursor becomes more strongly predictive of the goodness score for voiced stops than for voiceless stops. The Abruptness of Release attribute also is more strongly predictive of voiced than voiceless stop goodness scores. Given that the Precursor and Abruptness of Re-

lease attributes are highly correlated, this finding is not surprising (refer to discussion of these two attributes for all stops). The Time Course of Release attribute is highly correlated with stop goodness, both due to the evaluation of noise, as discussed earlier, and also due to the assessment of formant-frequency appearance for $F1$ and $F2$, which are more visible for voiced stops. Precursor and Time Course of Release are highly correlated. This finding is understandable, given that both attributes assess noise immediately prior to the release. Additionally, dysarthric speakers who generate noise during the precursor time period also tend to generate noise during and after the stop release as well. The additional attribute, Time Course of $F2$ Change, is strongly predictive of stop goodness. It is also highly correlated with Precursor and Time Course of Release. Since Time Course of Release evaluates certain aspects of the formant frequencies, this overlap is not surprising. It is a more important finding that Time Course of $F2$ Change and Precursor are highly correlated, since these two attributes are assessed over very different time periods and aspects of the stop production. This finding indicates that when the dysarthric speakers produce one aspect of the stop poorly, they tend to produce another, unrelated aspect of the stop poorly as well. The final observation regarding the correlation matrix for voiced stops is that VOT is not highly correlated with either the goodness score or the Time Course of Release (unlike for voiceless stops). Since it is rare for the VOT to deviate from normal for voiced stops, this finding seems reasonable.

In the final correlation matrix, for /b,d/ stops, there is an additional attribute, Time Course of $F1$ Rise. Comparing this matrix to the one for voiced stops, all the same observations can be made regarding correlations. Additionally, Time Course of $F1$ Rise is highly correlated with stop goodness, Time Course of Release and Time Course of $F2$ Change. Since all three time course attributes examine aspects of the formant frequencies, this observation is understandable.

A single measure reflecting overall stop production can be generated by averaging across all attributes, utterances, word repetitions, and judges. (For the Time Course of $F1$ Rise attribute, only the utterances with intended word-initial /b,d/ are included; for the Time Course of $F2$ Change attribute, only the utterances with

	Good	Prec	Prev	Abru	TCR	VOT	TCF2
Good	1.000						
Prec	**-0.825**	1.000					
Prev	-0.559	0.731	1.000				
Abru	**-0.804**	**0.895**	0.594	1.000			
TCR	**-0.914**	**0.846**	0.624	0.781	1.000		
VOT	-0.340	0.115	-0.154	0.238	0.279	1.000	
TCF2	**-0.859**	**0.802**	0.551	0.712	**0.892**	0.337	1.000

Table 5.13 : Pearson correlation matrix between stop goodness score and SA attributes for all word-initial voiced stops. The matrix calculations are based on 3 repetitions/utterance, 5 utterances, 2 judges, and all 16 speakers. The column and row labels are as follows: Good=stop goodness score (from Chap. 4), Prec=Precursor attribute, Prev=Prevoicing attribute, Abru=Abruptness of Release attribute, TCR=Time Course of Release attribute, VOT=Voice Onset Time attribute, and TCF2=Time Course of $F2$ Change attribute.

	Good	Prec	Prev	Abru	TCR	VOT	TCF1	TCF2
Good	1.000							
Prec	**-0.829**	1.000						
Prev	-0.512	0.633	1.000					
Abru	**-0.855**	**0.858**	0.459	1.000				
TCR	**-0.947**	**0.837**	0.533	**0.831**	1.000			
VOT	-0.372	0.013	-0.232	0.175	0.349	1.000		
TCF1	**-0.854**	0.702	0.451	0.756	**0.886**	0.498	1.000	
TCF2	**-0.873**	**0.860**	0.624	0.772	**0.944**	0.252	**0.877**	1.000

Table 5.14 : Pearson correlation matrix between stop goodness score and SA attributes for all word-initial /b,d/ stops. The matrix calculations are based on 3 repetitions/utterance, 4 utterances, 2 judges, and all 16 speakers. The column and row labels are as follows: Good=stop goodness score (from Chap. 4), Prec=Precursor attribute, Prev=Prevoicing attribute, Abru=Abruptness of Release attribute, TCR=Time Course of Release attribute, VOT=Voice Onset Time attribute, TCF1=Time Course of $F1$ Rise attribute, and TCF2=Time Course of $F2$ Change attribute.

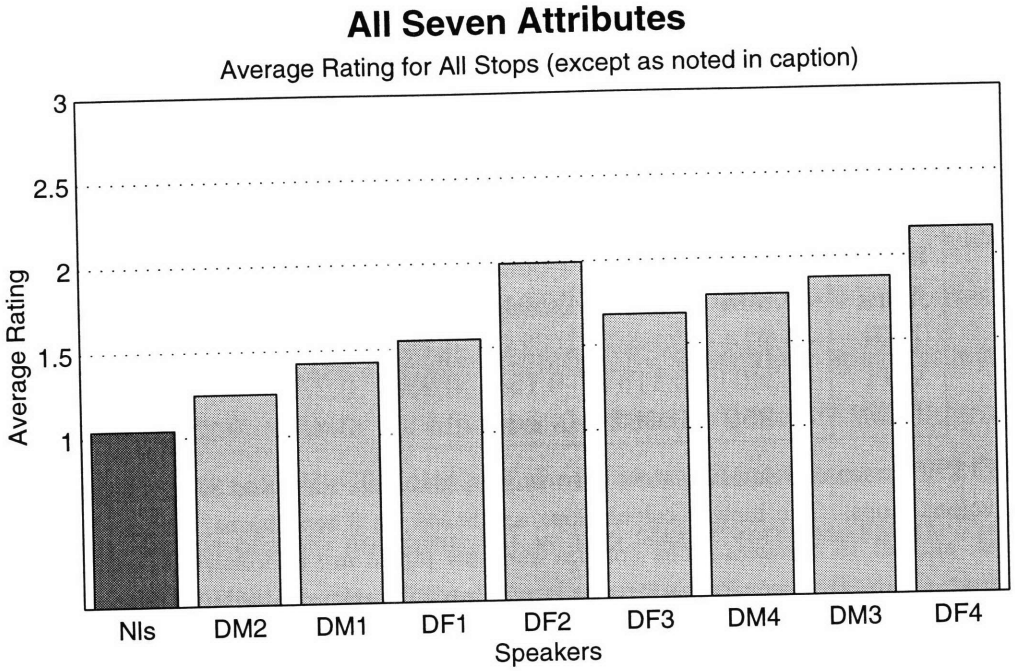

Figure 5-15 : Average across all seven attribute results from Spectrogram Analysis. Ratings averaged across all attributes, utterances, word repetitions, and judges. (For the Time Course of $F1$ Rise attribute, only utterances with intended word-initial /b,d/ are included; for the Time Course of $F2$ Change attribute, only the utterances with intended word-initial voiced stops are included.) For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

intended word-initial voiced stops are included.) The results for this measure are shown in Figure 5-15. This figure reveals a nice correspondence between poorer average attribute ratings and decreasing stop goodness scores across all speakers. (DF2 is considered a special case, for reasons discussed earlier.) The relationship between this average attribute measure and stop goodness could have been anticipated from the results for the individual attributes. As shown in the correlation matrices of Tables 5.11–5.14, all attributes were negatively correlated to some extent with the stop goodness score. The existence of such a relationship is appealing in that it indicates agreement between the perceptual evaluations and the qualitative spectrogram analysis of this data. At least in part, SA has been able to capture and quantify what the listeners indicate they perceive in the speech of these speakers.

As a measure of consistency in rating schemes across the two judges, a chi-square

	Normal			Dysarthric		
Rating	1	2	3	1	2	3
1	93.2	4.1	0	36.4	11.2	3.2
2	2.1	0.6	0	7.1	16.3	13.1
3	0	0	0	0.7	3.1	8.8

Table 5.15 : Chi-Square Test for interjudge agreement. The test was performed on the judges' scores for normal and dysarthric speakers separately. Ratings assigned by Judge 1 form the rows, and the ratings from Judge 2 form the columns. For normal speakers, $\alpha = 0.01$, $p = 0$. For dysarthric speakers, $\alpha = 0.01$, $p = 4.5\text{e-}7$. Tabulated values are given as percentages.

test was performed. The results of this test are shown in Table 5.15 for normal speakers on the left and dysarthric speakers on the right. For each speaker group, the results of the test are significant ($\alpha = 0.01$). The p-value of zero for normal speakers and very close to zero ($p = 4.5 \times 10^{-7}$) for dysarthric speakers indicates that the rows and columns are not likely to be independent. In other words, the rating schemes are essentially the same between the two judges.

For normal speakers, the two judges gave the same rating 93.8% of the time and differed by one in their ratings only 6.2% of the time. The judges never awarded a rating of three to normal speech, consequently they never differed by two in their ratings. For dysarthric speakers, the two judges gave the same rating 61.5% of the time, differed by one 34.5% of the time and differed by two 3.9% of the time. (The total differs by 0.1% from 100% due to rounding.) It is observed that Judge 2 tended to award more twos and threes than Judge 1. (Judge 1 awarded 50.8% ones, 36.5% twos, and 12.6% threes; Judge 2 awarded 44.2% ones, 30.6% twos, and 25.1% threes.) This tendency is attributed to the use of slightly different mappings for the ratings. Judge 1 tended to place the dysarthric spectrograms of this study into the broader context of disordered speech in general, resulting in a ratings assignment that reflected the mild-to-moderate dysarthric nature of the speakers. Judge 2 tended to more frequently apply the full range of ratings to this particular set of dysarthric spectrograms, assigning threes to the worst productions of every attribute. While observable in the table, these tendencies have minimal to no affect on the significance of the overall result that the ratings schemes are essentially identical.

5.3 Conclusions

A summary of the individual-speaker observations across attributes follows. The mildly dysarthric speakers will be considered first. Speaker DM2 is the speaker most similar to normals across all attributes. He differs from normals most noticeably in the time course of his release. Speaker DM1 is most noticeably different from normal in the time course of release, time course of $F1$ rise and time course of $F2$ change. The $F1$ and $F2$ transitions for this speaker will be examined in more detail in Chapter 6. Speaker DF1 is notable for prevoicing excessively prior to voiced stops. She differs from normals in that she builds up too much subglottal pressure and initiates vocal-fold vibrations too early compared to normal prevoicing. Speaker DF2 has been discussed extensively due to the effect her faulty velopharyngeal port opening has on many of her attributes. The air that is almost continually leaking through her nose appears in the spectrogram as broadband noise in the 1–8 kHz range. This noise tends to lead to the presence of a precursor, as well as noisy time periods at and after the stop release, affecting both the release characteristics and the following $F2$ transition. The inability to build up sufficient intraoral pressure during the stop closure interval results in weaker, less clear bursts. The nasal-cavity resonance is an additional formant present throughout most of the stop production. Speaker DF2 is also judged to have unnatural prevoicing compared to normal. She is judged to prevoice at least some of the time prior to both voiced and voiceless stops. In voiced stop production she tends to err by initiating vocal-fold vibration too early, in the form of prevoicing, rather than too late, in the form of a prolonged VOT. The judgment of the presence of a precursor for DF2 in Q1 of Chapter 4 (refer to Fig. 4-6, page 87) is associated with *both* the presence of a precursor (as defined in this chapter, including the possibility of inadvertent vowel generation) and the presence of prevoicing in the SA.

The moderately-dysarthric speakers will be considered next, on a speaker-by-speaker basis across attributes. Speaker DF3 is most notable for occasionally lengthening her VOT for voiced stops, occasionally shortening her VOT for voiceless stops,

and for having a deviant time course of $F1$ rise. Speakers DM4 and DF4 have very similar observations from their attribute ratings, although DF4 typically has the worse rating of the two (except, perhaps, for the prevoicing attribute). Each of these two speakers tends to produce a variety of precursor sounds. They also tend to prevoice prior to voiced stops and, to a lesser extent, prior to voiceless stops. Their releases are less abrupt and the time course is noticeably poorer than normal. Their VOT for voiceless stops tends to be too long. Their $F1$ and $F2$ transitions also deviate from normal. For speaker DM3, the presence of a precursor is partly attributed to background (nonspeaker-generated) noises. This speaker tends to lengthen his VOT from some of his voiced-stop productions, compared to normals, and tends to shorten his VOT for voiceless stops. His remaining attribute ratings are commensurate with his stop goodness score.

Across-speaker observations can also be made. The attributes and the stop goodness scores were always found to be negatively correlated, using Pearson r correlation matrices. This finding indicates a correlation between higher attribute ratings and poorer stop goodness scores. Across all stops, Time Course of Release was found to be highly correlated with the stop goodness score. For voiceless stops, Time Course of Release and VOT were highly correlated with the goodness score. A high correlation was observed between goodness and Precursor, Abruptness of Release, Time Course of Release and Time Course of $F2$ Change for voiced stops. When velars are no longer under consideration in the voiced stops, the same group of attributes is found to be highly correlated to stop goodness, along with the additional attribute Time Course of $F1$ Rise. This observation culminates in the development of a single measure, averaged across all seven attributes (Time Course of $F1$ Rise and Time Course of $F2$ Change are only considered over the subsets of utterances for which these attributes are meaningful), reflecting overall stop production.

The results of the spectrogram analysis (SA) reveal that, at least in part, SA has been able to capture and quantify what listeners perceive in the speech of the normal and dysarthric speakers. The use of spectrogram analysis may have clinical value. For example, clinicians could receive training in how to assign attribute ratings, then

compare the results for a given speaker to an established norm to assist in diagnosis and/or remediation.

5.4 Summary

In Section 5.1 the corpus, speakers, recording method, and judges utilized in the Spectrogram Analysis (SA) are discussed. Also, the General Guidelines for Attribute Evaluation are presented, complete with tables of rating scales and their descriptions for the seven attributes assessed from the spectrograms of the normal and dysarthric speakers. These seven attributes are as follows: Precursor, Prevoicing, Abruptness of Release, Time Course of Release, VOT, Time Course of $F1$ Rise, and Time Course of $F2$ Change.

Section 5.2 contains the SA results and discussion. The attributes and the stop goodness scores were always found to be negatively correlated, using Pearson r correlation matrices. This finding indicates a correlation between higher attribute ratings and poorer stop goodness scores. Across all stops, Time Course of Release was found to be highly correlated with the stop goodness score. For voiceless stops, Time Course of Release and VOT were highly correlated with the goodness score. A high correlation was observed between goodness and Precursor, Abruptness of Release, Time Course of Release and Time Course of $F2$ Change for voiced stops. When velars are no longer under consideration in the voiced stops, the same group of attributes is found to be highly correlated to stop goodness, along with the additional attribute Time Course of $F1$ Rise. This observation culminates in the development of a single measure, averaged across all seven attributes (Time Course of $F1$ Rise and Time Course of $F2$ Change are only considered over the subsets of utterances for which these attributes are meaningful), reflecting overall stop production. Examination of this single measure along with the correlation matrices reveals that, at least in part, SA has been able to capture and quantify what the listeners perceive in the speech of these normal and dysarthric speakers. The spectrogram attributes better capture the differences between the speech of the normal and the dysarthric speakers than the

acoustic measures of Chapter 6. This finding suggests that a better strategy (than the one in this thesis) would be to devise acoustic measures based on SA findings, rather than based on measures of normal speech. Spectrogram analysis may have clinical applications in diagnosis and remediation of disordered speech.

Chapter 6

Acoustic Analysis

Acoustic analysis was performed to provide objective, quantitative measures of stop consonants produced by the normal and dysarthric speakers involved in this study. Acoustic measures were developed to assess certain aspects of the speech system during stop production. These aspects are the placement of the primary articulator, the rate of movement of the primary articulator, the laryngeal system, and the respiratory system. The development of the acoustic measures is discussed in Section 6.1.

The results and discussion of the acoustic measures applied to normal stop-consonant production are presented in Section 6.2.1. These normal data were collected primarily to serve as a baseline for comparison with the speech of individuals who have dysarthria. These data also contribute to knowledge of the range of variability naturally occurring in the speech of normal speakers, for potential future speech recognition or synthesis applications. Section 6.2.2 contains the results and discussion of the acoustic measures performed on stop consonants produced by individuals with dysarthria. The dysarthric data results are compared to the baseline provided by the results of the normal speakers. The results for both Sections 6.2.1 and 6.2.2 are interpreted in terms of the information they reveal about articulator control and coordination. Section 6.4 summarizes the results of the acoustic analysis.

6.1 Data Acquisition and Processing

6.1.1 Corpus, Speakers and Recording Method

Acoustic analysis was performed on eight words with word-initial stops: bad, bunch, dock, dug, geese, pat, tile, and coat. This dataset is the same as was recorded and utilized for the perceptual evaluations in Chapter 4 (refer to Sections 4.1.1 and 4.1.3) and the spectrogram analysis in Chapter 5. The 16 speakers (8 normal and 8 dysarthric) have been discussed in Chapter 2 and Section 4.1.2.

6.1.2 Signal Processing

The signal processing software program 'xkl' utilized to process the acoustic data was developed in our laboratory, the Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, for use with UNIX- and LINUX-based computer systems. This software is based on the signal processing software program KLSPEC developed by Dennis H. Klatt (also from our laboratory) for use with a VAX-based computer system.

As a first step in the development of the acoustic measures of Section 6.1.3, the acoustic signal must be pre-processed in both the time and frequency domains. The required signal processing is described in the following three subsections. The first subsection contains identification of the stop-consonant release and the vowel onset in the acoustic time waveform. The second subsection describes the set of three average spectra created before, during and after the stop release. The third and final subsection contains a description of a second set of three spectra generated at and after vowel onset.

SRT and VIT Identification

This subsection describes the identification of two specific times in the acoustic time waveform. These times will be useful as reference points for the calculation of spectra in later subsections of the present section (Section 6.1.2) and in the determination of

the acoustic measures in Section 6.1.3. The first time is the stop-consonant release and the second time is the onset of the vowel.

The Stop Release Time (SRT) is the time in the acoustic waveform (to the nearest tenth of a ms) when release of the stop consonant occurs. Specifically, the SRT is defined to be the time in the vicinity of stop production when the waveform amplitude transitions from background noise, prevoicing or other speaker- and/or nonspeaker-generated sounds prior to the stop release (generally sounds of low frequency and low amplitude) to the (generally) higher frequencies and higher amplitudes associated with the rapid movement of the primary articulator away from closure and the decrease in intraoral pressure at the initiation of the stop burst or transient. The primary articulator is the articulator responsible for making the oral closure in the vocal tract during the stop closure interval preceding release. An example of the SRT is shown in Figure 6-1. The SRT is identified from the time waveform for each of the 3 repetitions $\times$ 8 utterances $\times$ 16 speakers by visually examining the time waveform, listening to the acoustic signal and utilizing the perceptual experiment results for Questions 1 and 3.

There are a few special situations to be considered when identifying the SRT: (1) If multiple stop bursts (transients) are present, the waveform amplitude between bursts may either return to the background noise level (occurs more often between the first few successive bursts), or may be greater than the background noise level, indicating either that the constriction is remaining wide enough to excite the front cavity resonances on a continuing basis or that the formants in the oral cavity behind the constriction are excited (these two events occur more often between the last few successive bursts). The SRT is defined to be the initiation of the first burst for which the waveform amplitude does *not* return to the background noise level following that burst. This definition of SRT is based on the burst being detected by a listener when the waveform amplitude does not return to background noise level. (2) Occasionally, instead of generating the intended stop consonant, dysarthric speakers may generate a glottal stop or omit the stop consonant altogether. Although these two events can appear somewhat similar in the time waveform, it is possible to distinguish a

glottal stop from the absence of a stop by the sudden presence of high frequencies of high amplitude (relative to the background noise) as voicing starts abruptly following the glottal stop. Additionally, some irregularities may be present in the first and/or second glottal pulse of the vowel following a glottal stop.

The Vowel Initiation Time (VIT) is the time in the acoustic waveform (to the nearest tenth of a ms) when the vowel begins. The VIT occurs at the transition between production of the stop and the following vowel, and is defined to be the time following stop release corresponding to the start (positive or negative zero crossing) of the first complete glottal pulse in which the maximum waveform amplitude is at least $\frac{1}{4}$ of the maximum amplitude of the glottal pulses in vowel steady state (the "$\frac{1}{4}$-rule"). This definition of VIT is partially motivated by a desire to locate the point in the acoustic waveform when the vowel onset is likely to begin to be audible, and partially motivated by a desire to identify the VIT using a technique that could lend itself to automation in the future, such as for speech recognition applications. A glottal pulse is not "complete" if it overlaps part of the noise produced during the stop, or if it is too short in duration *and* does not have a shape resembling the glottal pulses produced during the steady-state portion of the vowel. Although rare, a glottal pulse may be "incomplete" even if its amplitude satisfies the $\frac{1}{4}$-rule. Consequently, the first complete glottal pulse may be several pitch periods after the stop release (more common for dysarthric speakers than for normal speakers). The author's judgment was required to make the distinction between a "complete" and an "incomplete" glottal pulse. If there was no stop present, the VIT was still chosen to satisfy the $\frac{1}{4}$-rule. An example of the VIT is shown in Figure 6-1. The VIT is identified from the time waveform for each of the 3 repetitions $\times$ 8 utterances $\times$ 16 speakers by visually examining the time waveform and listening to the acoustic signal. It is important to note that VIT is not the same as the "voice onset time" (VOT), which is standard terminology for the duration between the stop release and the onset of the vowel. (Refer to Section 6.1.3, Laryngeal and Respiratory Systems subsection, for the use of the VOT in this thesis.)

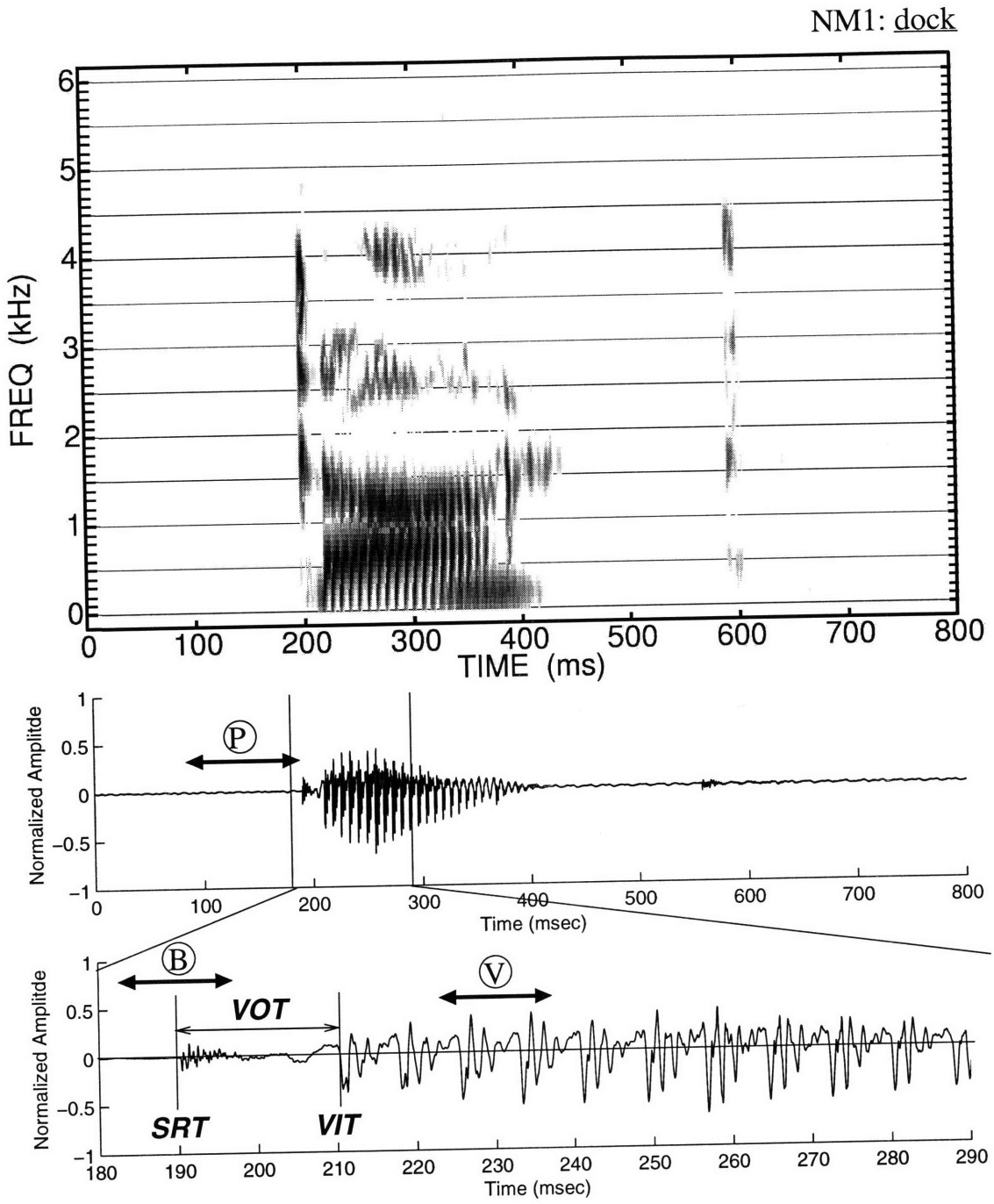

Figure 6-1 : Normal male speaker (NM1) saying the word dock. Spectrogram (top) calculated using a 6.4-msec Hamming window to generate a 256-point DFT spectrum every 1 msec. Acoustic time waveform (middle) and magnified time waveform (bottom). Waveform amplitudes proportional to sound pressure recorded at the microphone. Vertical lines in the middle waveform indicate the time period of magnification shown in the bottom waveform. Durations P, B and V indicate averaging intervals for the Precursor, Burst and Vowel average spectra, respectively. SRT is the stop release time, VIT is the vowel initiation time, and VOT (Voice Onset Time) = VIT-SRT.

Spectra for Relative Amplitude Measures

This subsection describes the creation of three average spectra before, during and after the stop release. These spectra are utilized for acoustic measures involving relative amplitudes. A 6.4 ms Hamming window was used to generate each individual 512-point DFT spectrum, from which the average spectra were then calculated. The spectra were averaged in order to smooth out irregularities attributed to variability in the individual spectra. Three average spectra were calculated, one for the time period prior to the release (Precursor Average Spectrum), one for the time period during release (Burst Average Spectrum), and one for the time period after the release (Vowel Average Spectrum). The generation of each averaged spectrum is discussed in more detail below. Average spectra were created for each of the 3 repetitions $\times$ 8 utterances $\times$ 16 speakers.

The Precursor Average Spectrum is generated as follows. First, a Hamming window is placed to the left of the SRT, so that the right edge of the window is immediately prior to the SRT. This time becomes the end of the precursor spectral averaging interval. Next, 100 ms is subtracted from the end time. This earlier time becomes the start of the averaging interval. (Exceptions to these starting and ending times are listed in the next paragraph.) Spectra are generated every millisecond from the beginning to the end of this 100-ms interval, then the spectra are averaged together to generate the Precursor Average Spectrum. An example of the averaging interval is indicated by the letter P in Figure 6-1, and the resultant spectrum is shown in Figure 6-2. It is observed that the averaging interval may contain sounds produced by the speaker (e.g., prevoicing, air audibly leaking from the nose) as well as background noises (e.g., wheelchair squeaking, noises in the speakers' homes, and conversations between the researchers).

Occasionally, there may be an exception to the start and/or end time(s) of the precursor spectral averaging interval, resulting in a shorter time interval over which the average spectrum is calculated. The exceptions are as follows: (1) Exception to the start time: If the time period prior to the burst, as recorded in the data file for the

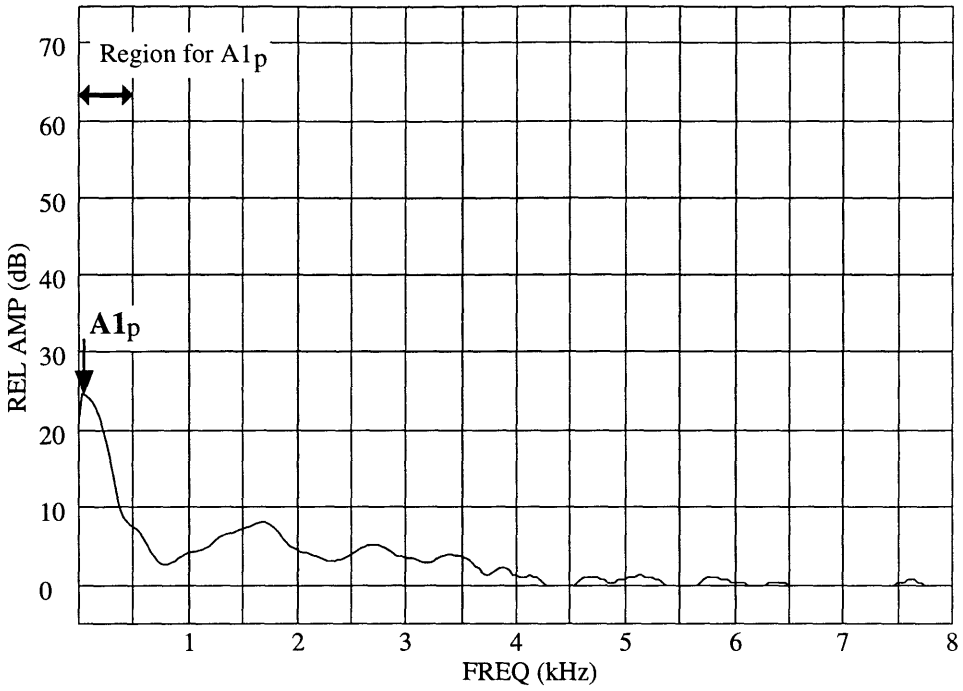

Figure 6-2 : Precursor Average Spectrum for dock spoken by a normal male speaker (NM1). This spectrum represents the average of spectra placed 1 msec apart throughout the precursor spectral averaging interval. This interval is denoted by P in Figure 6-1. For details of how this interval was determined, refer to the text. The peak amplitude $A1_p$ in the 0–500 Hz region is indicated.

particular repetition, is less than 100 ms, place the start of the averaging interval 4 ms after the start of the data file (4 ms represents half the window duration, rounded up to the nearest ms due to software restrictions). This choice of window placement aligns the left edge of the window with the beginning of the data file. (2) Exception to the end time: If multiple bursts (transients) are present, place the initial window (the window which determines the *end* of the averaging interval) so that its right edge is immediately prior to the *very first* burst, irregardless of whether the waveform amplitude between bursts returns to the background noise level. This time will now become the end of the averaging interval.

To generate the Burst Average Spectrum, the right edge of the Hamming window is initially placed at the VIT. Then, the window is shifted to 7 ms earlier in the acoustic signal. If the window is now on or prior to the SRT, then the window is in its final position. If the window position is not early enough in time (far enough to the left) to precede or coincide with the SRT, then the SRT itself becomes the final

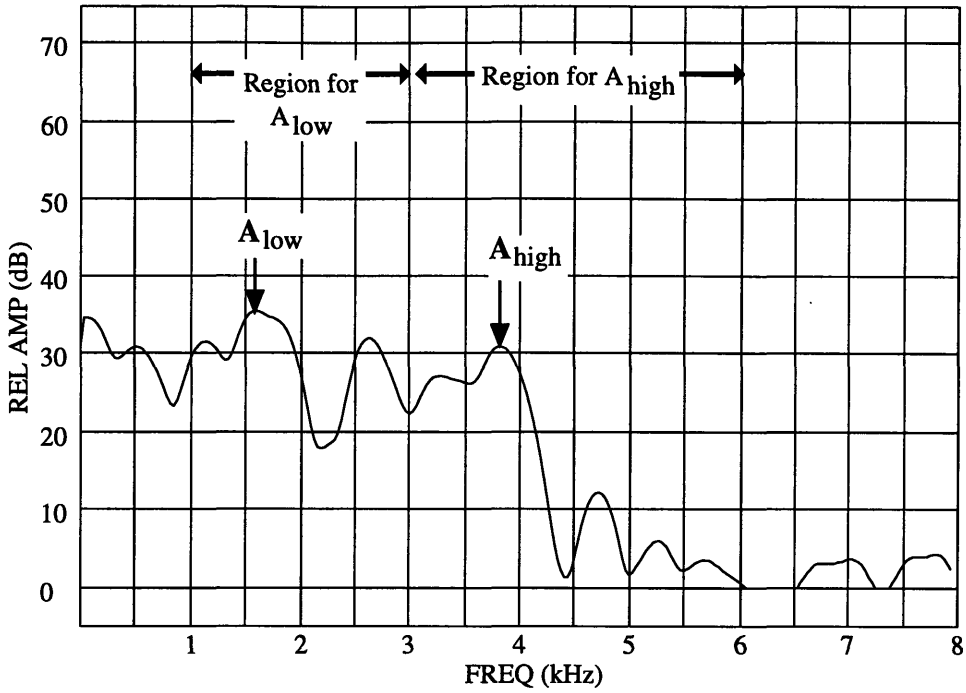

Figure 6-3 : Burst Average Spectrum for dock spoken by a normal male speaker (NM1). This spectrum represents the average of spectra placed 1 msec apart throughout the 15-msec burst spectral averaging interval. This interval is denoted by B in Figure 6-1. For details of how this interval was determined, refer to the text. The peak amplitudes A_{low} and A_{high} in the frequency regions 1–3 kHz and 3–6 kHz, respectively, are indicated. For female speakers, the frequency regions within which to identify A_{low} and A_{high} become 1–3.5 kHz and 3.5–7kHz, respectively.

window position. Spectra are created every ms from 7 ms preceding to 7 ms following the final window placement (a total of 15 ms). These spectra are averaged together to generate the Burst Average Spectrum. The 15-msec time interval over which the spectra are averaged contains both the transient and the frication noise. Calculations have shown that the transient and frication noise spectra have similar shapes for a given stop and phonetic environment (Stevens, 1998), so averaging across these two types of spectra is considered reasonable. The 15-msec time interval may also contain background noise or prevoicing prior to the stop release, but the effect of these sounds is considered negligible in the frequency range of interest (> 1 kHz). An example of the averaging interval is indicated by the letter B in Figure 6-1, and the resultant spectrum is shown in Figure 6-3.

To generate the Vowel Average Spectrum, the Hamming window is placed 20 ms after the VIT. Spectra are created every ms from 7 ms preceding to 7 ms following

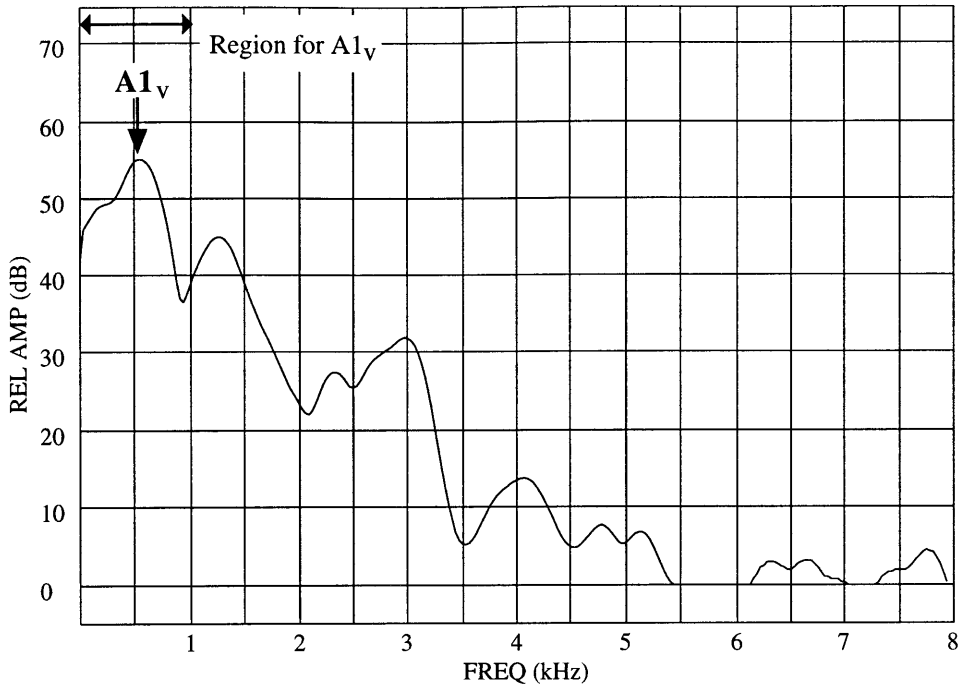

Figure 6-4 : Vowel Average Spectrum for dock spoken by a normal male speaker (NM1). This spectrum represents the average of spectra placed 1 msec apart throughout the 15-msec vowel spectral averaging interval. This interval is denoted by V in Figure 6-1. For details of how this interval was determined, refer to the text. The peak amplitude $A1_v$ corresponds to $F1$ in the 0–1 kHz region indicated.

this window placement (a total of 15 ms). These spectra are averaged together to generate the Vowel Average Spectrum. The 15-msec time interval over which the spectra are averaged is long enough to contain at least one complete pitch period for male or female speakers. An example of the averaging interval is indicated by the letter V in Figure 6-1, and the resultant spectrum is shown in Figure 6-4.

Spectra for Formant-Frequency Transitions

This subsection describes a series of three individual spectra generated at and after vowel onset. These spectra are utilized for acoustic measures involving formant-frequency transitions. A 6.4 ms Hamming window was used to create each 512-point DFT spectrum. During creation of each spectrum the first difference was calculated, in order to apply pre-emphasis. Pre-emphasis was utilized in an attempt to suppress the contribution of $F0$ and the "glottal shoulder" to the lower frequencies in the spectrum. The generation of each specific spectrum is discussed in the next paragraph.

Spectra were created for the 3 repetitions of bad, bunch, dock, and dug spoken by three of the dysarthric speakers (DF1, DM1 and DM2) and all of the normal speakers.

The set of three spectra was initially created by centering the Hamming window over the first part of each of the following glottal pulses: the glottal pulse identified by the VIT (the glottal pulse which begins at the VIT), the glottal pulse closest to 20 ms following the time of the initial spectrum, and the glottal pulse closest to 40 ms following the time of the initial spectrum. The first-differenced 512-point DFT spectrum was then calculated for each window position. With the aid of the spectrogram, the window position for each spectrum was shifted slightly in time as needed within the first part of the glottal pulse until the final choice of spectrum contained peaks at values similar to the peaks seen in the spectrogram. These final formant-frequency transition spectra will be referred to as Spectra A, B and C in the text below.

6.1.3 Acoustic Measures

Several acoustic measures were developed to assess certain aspects of the articulatory system. These aspects are the placement of the primary articulator, the rate of movement of the primary articulator, the laryngeal system, and, to some extent, the respiratory system. The primary articulator is responsible for forming the oral closure in the vocal tract and is anatomically anchored to the lower mandible. For labial stop consonants, the primary articulator is the lips, for alveolars it is the tongue tip, and for velars it is the tongue body. The respiratory and laryngeal systems act as secondary articulators, assisting in the production of the stop consonant but not forming the actual closure.

The first step in the development of the acoustic measures, pre-processing of the acoustic signal, was discussed in Section 6.1.2. The second, and final, step in the development of the measures is to make specific duration, frequency and amplitude measurements from the signal, based in part on the acoustic theory presented in Chapter 3. This second step is described in the next three subsections.

Placement of Primary Articulator

The placement of the primary articulator is assessed via two different measures. The first measure examines the tilt of the Burst Average Spectrum for labial and alveolar stop consonants. The second measure examines the value of $F2$ in Spectrum A of the two utterances with word-initial /d/. These two measures are discussed in this subsection.

The first measure assessing primary articulator placement examines the tilt of the Burst Average Spectrum for labial and alveolar stop consonants. During production of a labial stop there is no cavity in front of the vocal-tract constriction. In the absence of a front cavity, the burst spectrum should appear downward sloping toward higher frequencies, according to the vocal-tract models of Section 3.2.4. Production of an alveolar stop consonant involves the placement of the tongue tip against the palate to form the constriction, resulting in the presence of a short front cavity (approximately 2 cm in length) between the constriction and the lips. Models indicate that the lowest resonance of this front cavity is typically in the range 4–5 kHz. Therefore, the burst spectrum should be either uniformly flat across all frequencies or upward sloping toward higher frequencies. As discussed in the next paragraph, the difference between peak amplitudes in low and high frequency regions of the Burst Average Spectrum is calculated as a measure of burst tilt, assessing the degree to which a given Burst Average Spectrum reflects correct placement of the primary articulator.

The peaks for the amplitude difference A_{high} - A_{low} (in dB) are measured from the burst average spectrum for labial and alveolar stop consonants as follows. The amplitude A_{low} is the peak spectral amplitude in the region 1–3 kHz for male speakers and 1–3.5 kHz for female speakers. The amplitude A_{high} is the peak spectral amplitude in the region 3–6 kHz for male speakers and 3.5–7 kHz for female speakers. When selecting the peak within a particular region, the following rules apply. The value of the highest *peak* in the region was chosen, not the highest value in that region (if these two values differed). If there are two peaks of equal amplitude in the region, the peak corresponding to the higher frequency was chosen for A_{low} and the peak

corresponding to the lower frequency for A_{high}. A peak on the lower border (but not the upper border) of a given region is considered to be within that region. Peak amplitudes are accurate to ± 1 dB. Examples of A_{low} and A_{high} peak amplitudes are shown in the burst average spectrum of Figure 6-3.

A second measure assessing primary articulator placement comes from the formant-frequency transition Spectrum A. Spectrum A is the spectrum closest in time to the stop release of the three transition spectra, and therefore is the spectrum most likely to contain some residual information about the position of the primary articulator at the time of the release. This measure is particularly useful for alveolar stops, since the tongue tip position at the time of the release is approximately the same regardless of the following vowel. This consistent tongue tip placement appears in Spectrum A as a similar value of $F2$ across utterances. The value of $F2$ is examined in the Spectrum A of the two utterances with word-initial /d/ as an indicator of correct placement of the tongue tip at the time of the release.

Rate of Primary Articulator Movement

The rate of movement of the primary articulator is assessed via two different measures. The first measure examines formant-frequency transitions $F1$ and $F2$ following the stop release. A second, more qualitative measure infers the rate of primary articulator movement from the number of consecutive stop bursts occurring when the stop consonant is released. These two measures are discussed in this subsection.

The first measure to assess rate of movement of the primary articulator examines formant-frequency transitions $F1$ and $F2$ following the stop release. Formant-frequency transitions contain information about both the stop and the succeeding vowel. Over the course of the transition time period, the resonant frequencies in the vocal tract change from being predominantly influenced by the stop (reflecting, in part, movement of the primary articulator away from the constriction) to being predominantly influenced by the vowel (primarily reflecting jaw movement away from the stop release and tongue body movement toward the vowel steady state). The $F1$ and $F2$ transitions were measured from each of Spectra A, B and C. These three

spectra are considered to be in the early portion of the transition, and therefore can be interpreted in terms of the information they provide about the rate of stop release. In an attempt to visualize as much of the early transition as possible, only voiced stops were examined, in which no aspiration noise is present. Furthermore, only the labial and alveolar voiced stops were studied, due to the brevity of the $F1$ transition as well as the merging of the $F2$ and $F3$ transitions in the case of /g/ preceding the high, front vowel /i/ in geese.

In the course of this research, it was observed that the dysarthric speakers may not always produce the vowels correctly in these utterances. When the vowel is incorrectly produced, the formant-frequency transition rate may be affected, since the transition is now to a different vowel. Thus, an incorrect vowel may confound the ability to compare formant-frequency transitions for normal and dysarthric speakers. In an effort to minimize the effects of such an event, spectra were considered from solely the early part of the transition, the part where the influence of the following vowel is smallest. Additionally, formant frequencies were only measured from the three dysarthric speakers with highest word intelligibility (DF1, DM1 and DM2). These three speakers were believed to be least likely to produce their vowels incorrectly.

A second, more qualitative measure infers the rate of primary articulator movement from the number of consecutive stop bursts (transients) occurring when the stop consonant is released. When the stop consonant is released more slowly, the Bernoulli effect can dominate for a period of several milliseconds. During this time period, the constriction remains narrow long enough that the articulators are drawn together again due to the diminished pressure present within the constriction. This event is followed by the articulators separating again due to pressure buildup behind the constriction. This series of events can occur multiple times, leading to two or more stop bursts in a row, as discussed in Section 3.2.1. For normal speakers this series of events is not uncommon for velar stops, because the tongue body possesses large muscle mass, and therefore moves fairly slowly (compared to the tongue tip or lips), and the constriction length is longer, facilitating the production of consecutive stop bursts. The number of consecutive stop bursts was counted for each word repe-

tition. All bursts are included in the count, irrespective of whether they occur before or after the SRT, or are the SRT itself.

Laryngeal and Respiratory Systems

The function of the laryngeal and, to some extent, respiratory systems is assessed through a series of measures. The first measure of the laryngeal system examines the presence of prevoicing prior to the stop release. The second measure is the Voice Onset Time (VOT), reflecting the time it takes for the vocal folds to begin vibrating following the stop release. The third measure of the laryngeal system is the examination of the fundamental frequency, $F0$, immediately after vowel onset. There are also two measures created to assess changes in air pressure within the respiratory system. (In this case, the term "respiratory system" is interpreted to include not only the lungs and trachea, but also the oral and nasal cavities with respect to the ability to build up intraoral pressure prior to the stop release.) Of these two measures, one assesses labial and alveolar stop consonants and the other assesses velar stops. All of these measures are discussed in this subsection.

The first measure related to the laryngeal system examines the presence of prevoicing prior to the stop release. When conditions are conducive (vocal folds are not too adducted, abducted or stiffened; sufficient transglottal pressure is present), the vocal folds will begin to vibrate before the stop is released. As a measure of prevoicing, the amplitude difference $A1_v$ - $A1_p$ is calculated, where $A1_p$ is the peak amplitude in the 0–500 Hz region of the Precursor Average Spectrum, and $A1_v$ is the peak amplitude in the 0–1 kHz region (the peak corresponding to $F1$) of the Vowel Average Spectrum. When identifying peak values within each spectrum, the peak selection rules discussed earlier in this section (Section 6.1.3) apply. (Frequencies are accurate to $\pm$100 Hz, and peak amplitudes are accurate to $\pm$1 dB in these spectra.) An example of the $A1_p$ peak amplitude is shown in the precursor average spectrum of Figure 6-2, and an example of the $A1_v$ peak amplitude is shown in the vowel average spectrum of Figure 6-4. The peak amplitude $A1_v$ is included as a reference value in this measure, since it remains approximately the same across different vowels for a

given normal speaker. Prevoicing that is brief in duration (typically < 100 ms) and low in amplitude may occur preceding voiced stops for some normal speakers. The amplitude difference is measured for all voiced stops but not for voiceless stops, in which the background noise is of sufficient variation between speakers as to render $A1_p$ of questionable value.

The second measure of laryngeal system function is the Voice Onset Time (VOT). The VOT is the duration between the stop release and the onset of the vowel. In this thesis, VOT is defined to be VIT - SRT. This duration reflects the time it takes for the vocal folds to begin vibrating following the stop release. For normal speakers, the VOT is shorter for voiced stops than for voiceless stops since there is no aspiration noise present in voiced-stop production. VOT is measured for all stops. An example of the VOT is shown in Figure 6-1.

The third measure of the laryngeal system is the examination of the fundamental frequency, $F0$, immediately after vowel onset. During this time period, $F0$ is expected to be slightly higher following a voiceless stop than a voiced stop, based on the acoustic theory of Sections 3.2.2 and 3.2.3. This theory states that the third and fourth sound sources after the release of a voiceless stop in a /CV/ sequence are aspiration noise arising from turbulence generated near the glottis and the voicing source of the following vowel, respectively. In order to generate the turbulence noise, the vocal folds must be held in an intermediate position, far enough apart to prevent voicing but not so far apart that turbulent airflow is not generated. This intermediate vocal-fold position requires that the vocal folds be slightly stiffer for voiceless stops than is necessary during the same time period for voiced stops. At the time of vowel onset, the vocal folds retain some of this stiffness residually, increasing $F0$ for the first few glottal pulses of the vowel. The onset of the vowel also reflects the ability of the respiratory system to maintain sufficient subglottal pressure to initiate and sustain vocal-fold vibration at that time.

$F0$ is measured on a particular waveform by recording the starting time of each pitch period for the first five pitch periods beginning with the VIT. (This strategy of measuring $F0$ beginning with the VIT may mean some earlier pitch periods are

missed.) Then $F0$ is the reciprocal of the difference in time between each consecutive pair of pitch periods. This measure yields four values of $F0$ for each repetition, from which an average value of $F0$ is calculated for the repetition. Four utterances were selected for evaluation: bad, dug, pat, and tile. The average $F0$ values for each repetition of bad and dug were averaged together and, likewise, the average $F0$ values for pat and tile were averaged together, to create $F0_{vcd}$ and $F0_{vcls}$, respectively, for a given speaker. Then, the acoustic measure $F0$ Ratio (mean) $= (F0_{vcls} - F0_{vcd})/F0_{vcls}$, expressed as a percentage. It is also possible to calculate the range of the $F0$ Ratio by considering how the average $F0$ value for each repetition varies across repetitions for the 6 repetitions that compose each of $F0_{vcd}$ and $F0_{vcls}$.

There are two measures designed to reflect air pressure control in the respiratory system. One measure indirectly assesses changes in air pressure for labial and alveolar stops and the other indirectly assesses changes for velar stops. For the purposes of these measurements, the "respiratory system" is interpreted to include the lungs, trachea, oral and nasal airway passageways. For labial and alveolar stops, the measure is $A1_v - A_{high}$. The amplitude $A1_v$ is measured from the vowel average spectrum (Fig. 6-4), and the amplitude A_{high} is measured from the burst average spectrum (Fig. 6-3), as discussed earlier. For normal speakers, this measure is predominantly influenced by the value of A_{high}, which is higher for alveolar than labial stops, as discussed for the acoustic measure of burst tilt, $A_{high} - A_{low}$, reflecting placement of the primary articulator. Within a given place of articulation, through, $A1_v - A_{high}$ is higher for voiced stops than for voiceless stops produced by normal speakers. One possible explanation, related to the control of air pressure, is the existence of intraoral pressure differences between voiceless and voiced stops at the time of the release. If there is prevoicing preceding the voiced stop, then the intraoral pressure must remain lower than the subglottal pressure to maintain a transglottal pressure difference. Also, to initiate voicing immediately following the release, a pressure difference must be present across the glottis. For a voiceless stop, intraoral pressure and subglottal pressure can equilibrate prior to release. Consequently, near the time of release, the lower intraoral pressure for voiced stops can result in a lower value of A_{high}, or a

larger $A1_v - A_{high}$ difference. For normal speakers, the value of $A1_v$ remains fairly consistent across vowel contexts, so it is not likely to be a source of variation in $A1_v - A_{high}$ values for voiced versus voiceless stops. The value of $A1_v$ depends upon the subglottal pressure at the start of the vowel, and the value of A_{high} depends upon intraoral pressure at the time of the burst. These two pressure values are about the same for normal speakers. For dysarthric speakers, however, the two pressure values may vary, as discussed in Section 6.2.2.

The second measure of air pressure control in the respiratory system assesses changes in air pressure during velar stop production. This measure, $A1_v - A_{max23b}$, compares the mid-frequency region of the burst to the $F1$ region of the vowel. Velar stops have a front cavity length typically in the 3–5 cm range. With a front cavity of this length, the vocal-tract filter models of Section 3.2.4 predict a spectral prominence in approximately the 2–3 kHz region. This region corresponds to the $F2$–$F3$ region of the following vowel. The peak spectral amplitude A_{max23b} is selected from the frequency region between *and including* F2 and F3 in the burst average spectrum, where the formants are determined via examination of the vowel average spectrum. When selecting the peak within this region, the following selection rules apply. The value of the highest *peak* in the region was chosen, not the highest value in that region (if these two values varied). If there are two peaks of equal amplitude in the region, the peak in the F3 range was chosen for stops preceding front vowels (utterance geese), and the peak in the $F2$ range was chosen for stops preceding back vowels (utterance coat). The amplitude $A1_v$ serves as a reference value and is measured from the vowel average spectrum, as discussed earlier (Fig. 6-4). Frequencies are accurate to ± 100 Hz, and peak amplitudes are accurate to ± 1 dB in these spectra. The $A1_v - A_{max23b}$ measure is designed to reveal similar intraoral pressure differences between voiced and voiceless velar stops as were discussed for the $A1_v - A_{high}$ measure of labial and alveolar stops.

6.2 Results and Discussion

6.2.1 Normal Speakers

This section contains the results of the acoustic measures performed on word-initial stop consonants produced by individuals with no known speech or hearing disorders. These data were collected by the author primarily to serve as a baseline for comparison with the speech of individuals who have dysarthria. These data also contribute to knowledge of the range of variability naturally occurring in the speech of normal speakers, for potential future speech recognition or synthesis applications.

The acoustic measures were developed to assess several aspects of the speech production system: placement of the primary articulator, rate of movement of the primary articulator, the laryngeal system, and, to some extent, the respiratory system. In this section, the results of those measures are presented and interpreted in terms of the information they reveal about normal articulator control and coordination. The data presented are in general agreement with published data for normal speakers. This results and discussion section is divided into three subsections below, reflecting various aspects of the articulatory system.

Placement of Primary Articulator

The acoustic measure $A_{high} - A_{low}$ (measured from the burst average spectrum) is plotted against the measure $A1_v - A_{high}$ (measured from the burst and vowel average spectra) in Figures 6-5 and 6-6, assessing the placement of labial and alveolar stop consonants. In addition to information about place provided by $A_{high} - A_{low}$, information from $A1_v - A_{high}$ is also utilized to separate these stops. Figure 6-5 shows the averages across all speakers, repetitions and, in the case of voiced stops, two utterances, for each of the four stop consonants. Labial stops are well separated from alveolar stops, on average, along both axes. The spectral prominence in the 4–5 kHz frequency range in the burst average spectrum for the alveolar stops (this prominence is due to excitation of a short cavity, approximately 2 cm long, in front of the constriction) results in an increase in A_{high} for alveolars as compared to labials. This

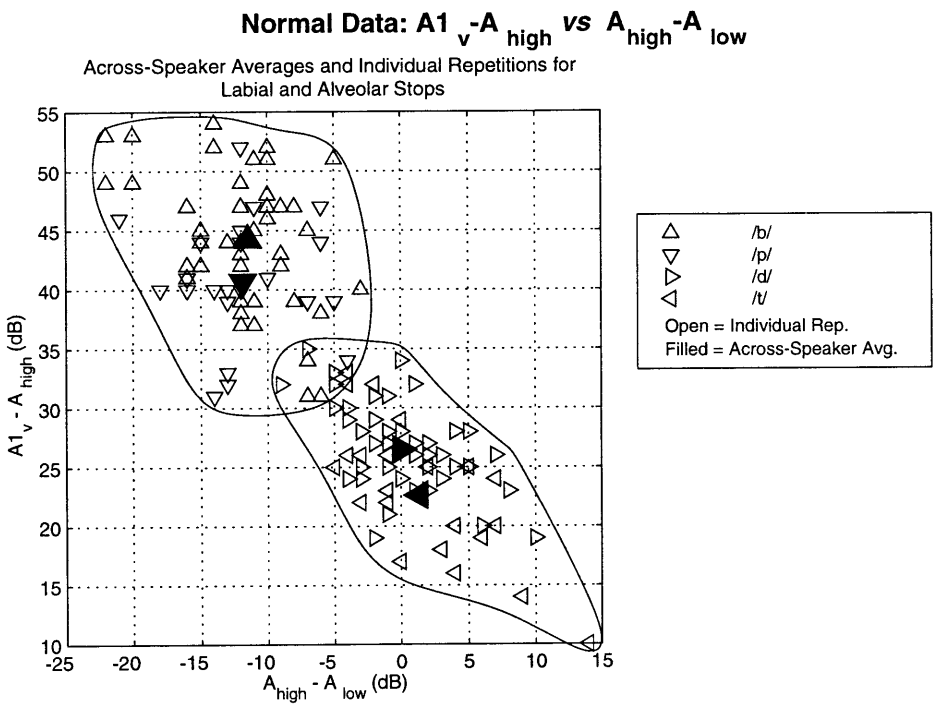

Figure 6-5 : Acoustic Measures $A1_v - A_{high}$ vs. $A_{high} - A_{low}$ for normal speakers. Across-speaker averages and individual word repetitions are shown for word-initial labial and alveolar stop consonants. The amplitude difference $A1_v - A_{high}$ is a measure of air pressure control, and the difference $A_{high} - A_{low}$ is a measure of burst tilt. For details of how these measurements were made, refer to Sections 6.1.2 and 6.1.3. The means, calculated across all 8 speakers, 3 repetitions/utterance and one utterance for the voiceless stops (two utterances for the voiced stops), are shown as filled triangles for each of the four stops. Individual word repetitions are also shown for each stop, and lines circumscribe the range of the data.

increase in A_{high} is reflected in a 12 dB average increase for the value of $A_{high} - A_{low}$ and an 18 dB average decrease for the value of $A1_v - A_{high}$ for alveolars, compared to labials. The finding that the labial burst at high frequencies is about 18 dB weaker, on average, than the alveolar burst agrees well with data from Stevens et al. (1999). Stevens et al. examined syllable-initial stop consonants in the context of sentences. Syllable-initial consonants were defined to be either word-initial consonants or, if they were word-internal, they were prestressed or the final consonant in a cluster. In that study, a similar amplitude difference was measured, and it was observed that labial bursts were about 15 dB weaker at high frequencies than alveolar bursts.

In Figure 6-5, an impression of the range of variability is obtained from the two circumscribed regions containing the individual repetitions for each labial and alveolar

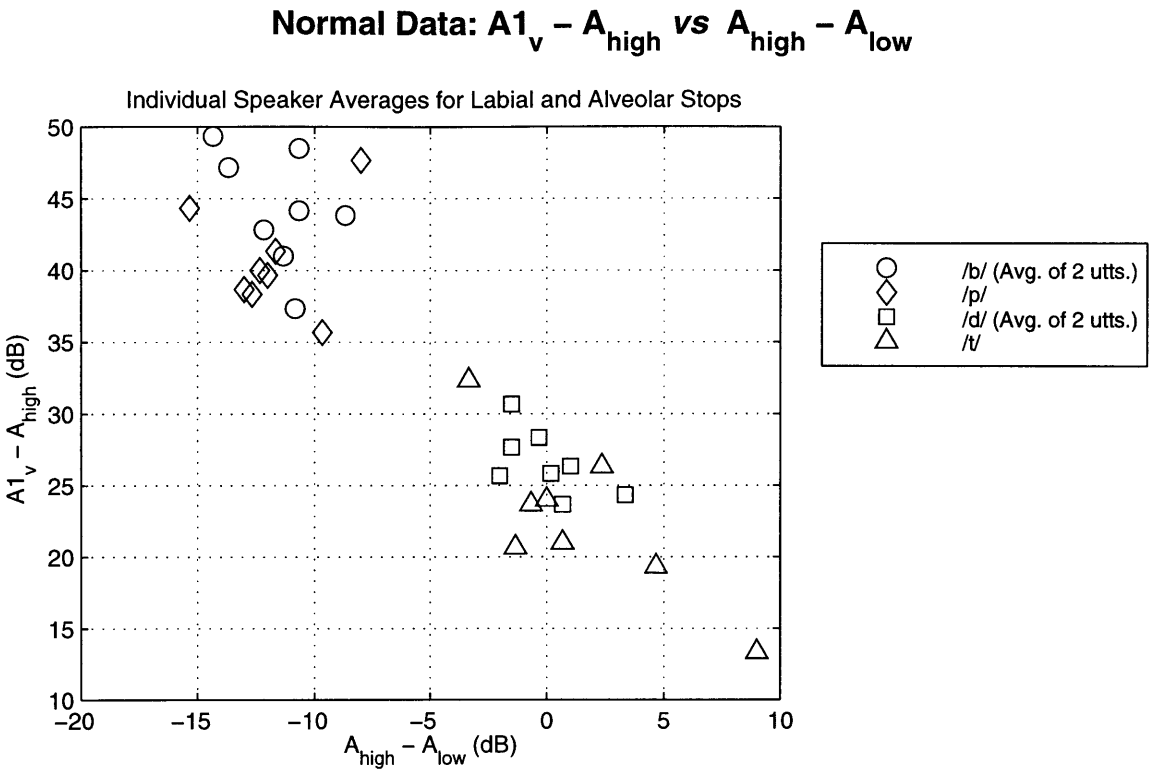

Figure 6-6 : Acoustic Measures $A1_v - A_{high}$ vs. $A_{high} - A_{low}$ for normal speakers. Individual speaker averages shown for word-initial labial and alveolar stop consonants. The amplitude difference $A1_v - A_{high}$ is a measure of burst strength, and the difference $A_{high} - A_{low}$ is a measure of burst tilt. For details of how these measurements were made, refer to Sections 6.1.2 and 6.1.3. Each individual normal speaker is represented by four data points. A data point is the average of 3 repetitions/utterance across one utterance for the voiceless stops, two utterances for the voiced stops.

stop spoken by each speaker. A small amount of overlap is seen in the repetitions of the two regions. Although it cannot be appreciated from this figure, a given normal speaker maintains separation of labial and alveolar stops in the $A1_v - A_{high}$ dimension and has at most a 5 dB overlap in the $A_{high} - A_{low}$ dimension, considering repetitions separately. If the repetitions are averaged together for each speaker, then Figure 6-6 shows that there is no overlap on a per-speaker basis. In other words, on average, the labial stops can be separated from the alveolar stops for each of the eight normal speakers, using this set of two acoustic measures.

Formant-frequency transitions are shown for the normal male speakers in Figure 6-7 and for the normal female speakers in Figure 6-8. The $F1$ and $F2$ trajectories are shown for each of the utterances bad, bunch, dock and dug. For the alveolar /d/, the

value of $F2$ at Time A_{tw} (the time closest to vowel onset) is less variable across vowel contexts than the value of $F2$ for other places of articulation. This initial value of $F2$ reflects, in part, the relative invariance of the constriction location to changes in vowel context. For the normal male speakers, $F2$ is about 1500 Hz for /dɑ/ and 1550 Hz for /dʌ/, on average, in Figure 6-7. For the normal female speakers, $F2$ is about 1800 Hz for /dɑ/ and 1900 Hz for /dʌ/, on average, in Figure 6-8. These values are similar across vowel types (within sex), indicating, as expected, that these normal speakers do not noticeably vary the position of their tongue tip against their palate to produce /d/ in different phonetic environments. (It is noted, however, that these two vowels, /ɑ/ and /ʌ/, have very similar values for $F2$ as well.)

Rate of Primary Articulator Movement

During the time period from the stop release to the following vowel, the vocal tract changes shape due to movements of the primary articulator and jaw away from their required positions for the stop consonant and the movement of the tongue body toward the required position for the vowel. The rates of these movements are reflected in the formant-frequency transitions of Figures 6-7 and 6-8. The means and ranges for these transitions are as expected for normal speakers at vowel onset. By the time of vowel onset, the rate of increase in $F1$ has slowed. The initial, rapid rise in $F1$ attributable to the primary articulator movement away from the release is generally complete by the time of the VIT. Consequently, the $F1$ rise seen in these trajectories is the slower rise attributable to jaw movement away from the release and tongue body movement toward the following vowel.

A second measure infers rate of release from the number of bursts (transients) occurring sequentially in each word repetition during stop-consonant production. The average number of bursts is indicated by the bars in Figure 6-9 for each stop. As shown in that figure, this group of eight normal speakers does not generate multiple bursts when labial or alveolar stops are produced. When velar stop are produced, however, they do occasionally generate more than one burst in a row. As indicated by the range bars in Figure 6-9, the maximum number of sequential velar stop bursts

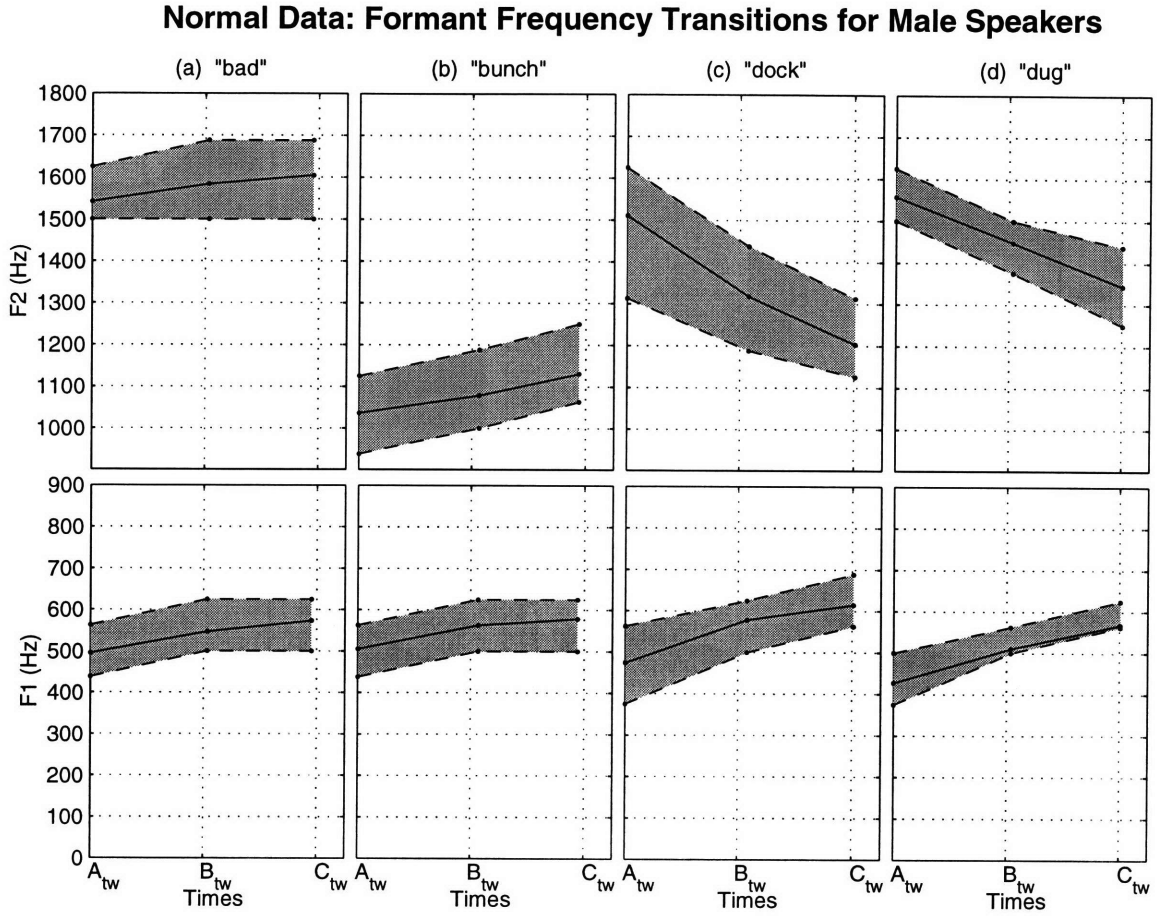

Figure 6-7 : Formant-Frequency Transition Acoustic Measure for normal male speakers. Transitions measured for word-initial labial and alveolar stops and following vowels. Along the x-axis the times have been labeled A_{tw}, B_{tw} and C_{tw}, respectively, as discussed in Section 6.1.2. The subscript tw refers to the time warping that may occur when the formant frequency values are averaged across repetitions. (To a rough approximation, Time A_{tw} can be considered to be at the VIT, Time B_{tw} at VIT + 20 ms and Time C_{tw} at VIT + 40 ms, but, for more accurate times, the reader is referred to the discussion of Section 6.1.2). This measure was averaged across all 4 male speakers and 3 repetitions/utterance, for the utterances (a) bad, (b) bunch, (c) dock and (d) dug. The mean is shown as the solid line and the range extrema are denoted by dashed lines. The full range is shaded gray.

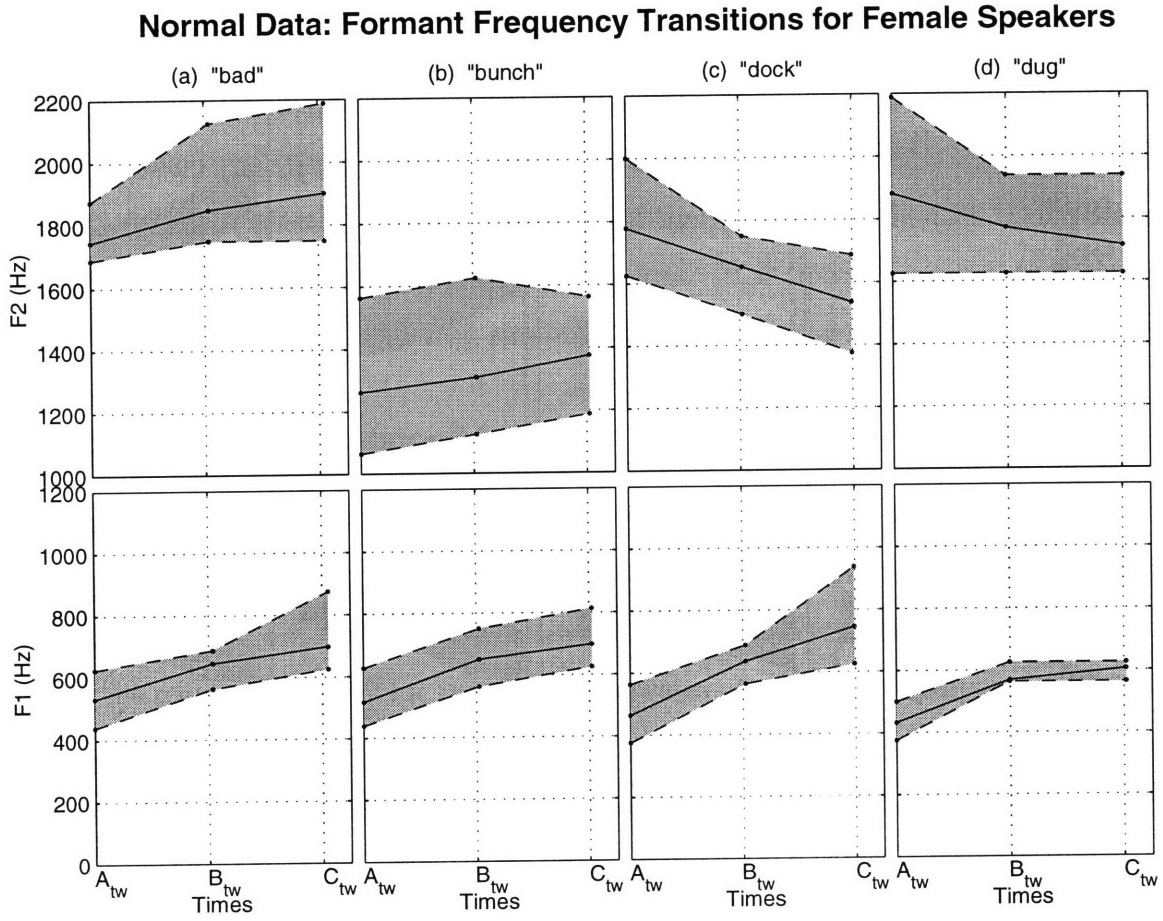

Figure 6-8 : Formant-Frequency Transition Acoustic Measure for normal female speakers. Transitions measured for word-initial labial and alveolar stops and following vowels. Along the x-axis the times have been labeled A_{tw}, B_{tw} and C_{tw}, respectively, as discussed in Section 6.1.2. The subscript tw refers to the time warping that may occur when the formant frequency values are averaged across repetitions. (To a rough approximation, Time A_{tw} can be considered to be at the VIT, Time B_{tw} at VIT + 20 ms and Time C_{tw} at VIT + 40 ms, but, for more accurate times, the reader is referred to the discussion of Section 6.1.2.) This measure was averaged across all 4 female speakers and 3 repetitions/utterance, for the utterances (a) bad, (b) bunch, (c) dock and (d) dug. The mean is shown as the solid line and the range extrema are denoted by dashed lines. The full range is shaded gray.

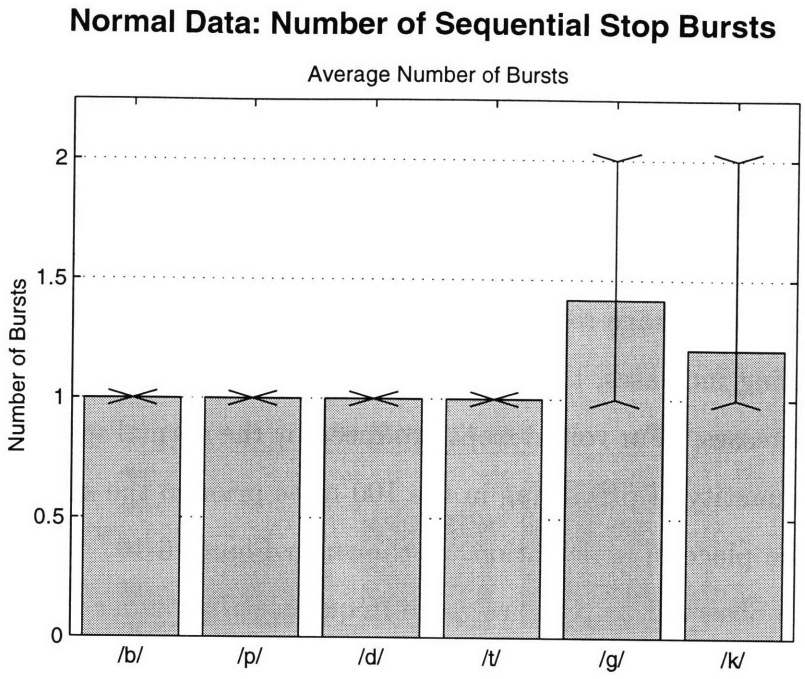

Figure 6-9 : Acoustic Measure of Number of Stop-Consonant Bursts in multiple-burst sequences for normal speakers. Number of bursts per word repetition shown averaged across all 8 normal speakers, 3 repetitions/utterance, and one utterance for each word-initial stop, with the exception of /b/ and /d/, which each contain two utterances. The bars represent the mean, and the error bars are the range extrema. (In the case of labial and alveolar stops, no multiple bursts were observed.)

is two for these particular speakers. Some of the normal speakers were observed to produce two sequential bursts more frequently than other speakers. The presence of two sequential bursts can indicate a slower rate of release. After the first of the two bursts for a velar stop, the tongue body moves toward the palate again, narrowing or closing the constriction. Following the second burst, the tongue body moves away from the palate toward the following vowel. When the SRT is considered to be the first of the two bursts, the overall rate is slowed, since the tongue body does not move away from the palate in a smooth, continuous fashion. Rather, the descent is slowed and reversed for a period of time, resulting in a slower rate overall.

Laryngeal and Respiratory Systems

The acoustic measure $A1_v - A1_p$ (taken from the precursor and vowel average spectra) assesses the duration and amplitude of the prevoicing present prior to the release of the voiced stops. Prevoicing, or vibration of the vocal folds prior to the stop release,

naturally occurs prior to voiced stops (but not voiceless stops) for some normal speakers. In anticipation of voicing the upcoming stop, the vocal folds are approximated. For some speakers, the subglottal pressure is great enough and the supraglottal cavity walls are relaxed or actively expanded enough to permit prevoicing. Figure 6-10 shows the results of this measure for each of the voiced stops separately, and Figure 6-11 shows the average results across all stops. As the duration and/or amplitude of the prevoicing increases, the value of $A1_p$ increases, and the amplitude difference $A1_v - A1_p$ decreases.[1] For voiced stops produced by the normal speakers in this study, the average quantity of prevoicing in the 100 msec prior to the stop release does not depend on the place of articulation, as shown in Figure 6-10. Some of the normal speakers were observed to prevoice more frequently than other speakers. The measure $A1_v - A1_p$ is not reported for voiceless stop consonants because the $A1_p$ value essentially should reflect the absence of prevoicing, but instead it is determined by the background noise level in the recording room.

The duration from the stop release to the onset of the vowel, or the voice onset time (VOT), was measured for these normal speakers. The VOT is a measure of how long it takes for the vocal folds to begin vibrating following release. The results of the VOT measure are reported in Figure 6-12 for each stop separately and in Figure 6-13 by type of voicing. The VOT values in this study are somewhat longer than the standard values reported in the literature, particularly for the voiceless stops (Zue, 1976). In this study, the VOT was defined to be the difference between the SRT and the VIT, where the VIT was defined to be the time corresponding to the start of the first complete glottal pulse in which the maximum waveform amplitude is at least $\frac{1}{4}$ of the maximum amplitude of the glottal pulses in vowel steady state. Satisfying the part of this definition that requires the amplitude to be at least $\frac{1}{4}$ of the maximum steady-state amplitude may at times result in selecting a vowel onset time that is later following the stop release than the vowel onset time utilized in other studies. Additionally, the VIT definition requires a "complete" glottal pulse,

[1] $A1_v$, the amplitude of $F1$ in the vowel, is utilized as a reference value, since it remains approximately the same across different vowels for a given normal speaker.

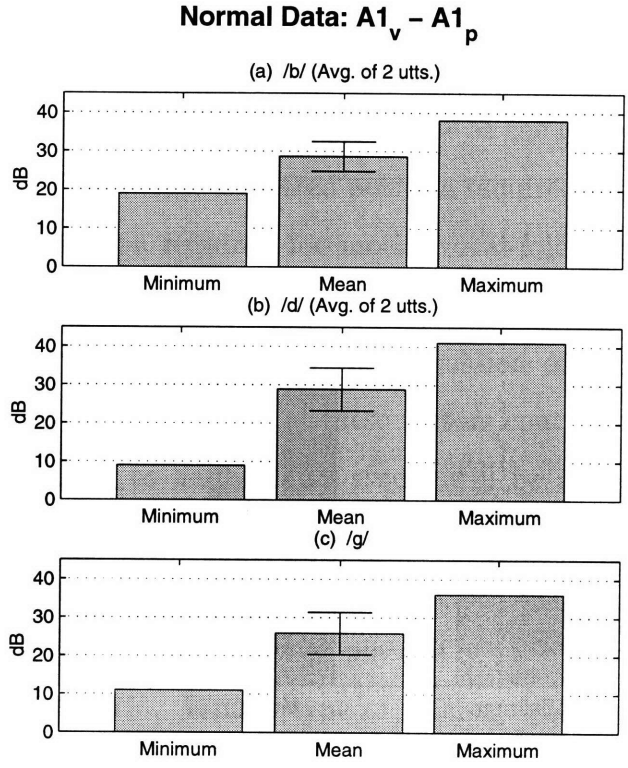

Figure 6-10 : $A1_v - A1_p$ Acoustic Measure by individual word-initial voiced stop for normal speakers. This amplitude difference is a measure of the presence of prevoicing prior to the stop-consonant release. For details of how the measurement was made, refer to Sections 6.1.2 and 6.1.3. The measure was averaged across all 8 normal speakers, 3 repetitions/utterance, and the number of utterances indicated for the word-initial voiced stops in (a)–(c). In each plot, the data shown are characterized by the mean, with a one standard deviation error bar, and the range extrema.

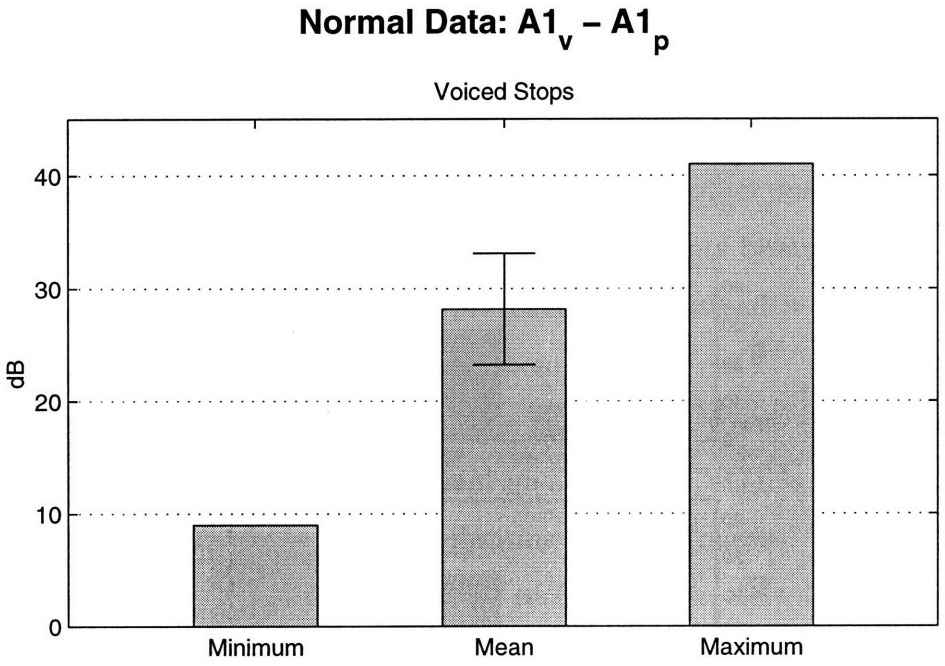

Figure 6-11 : $A1_v - A1_p$ Acoustic Measure across all word-initial voiced stops for normal speakers. This amplitude difference is a measure of the presence of prevoicing prior to the stop-consonant release. For details of how the measurement was made, refer to Sections 6.1.2 and 6.1.3. The measure was averaged across all 8 normal speakers, 3 repetitions/utterance, and 5 utterances containing word-initial voiced stops. In each plot, the data shown are characterized by the mean, with a one standard deviation error bar, and the range extrema.

where an "incomplete" pulse is a pulse which overlaps part of the noise produced during the stop or a pulse which is too short in duration and does not have a shape resembling the glottal pulses produced during vowel steady-state. The first few glottal pulses following a voiceless-stop release are more likely to match the definition of "incomplete" glottal pulses than the initial few pulses following a voiced-stop release, because of the need to generate aspiration noise following the voiceless-stop release. This aspiration noise may overlap the initial glottal pulse. Also, the first two or three glottal pulses following voiceless stop-consonant release may be breathier than later glottal pulses (and therefore not have the same shape as the later pulses), as the vocal folds transition from the stiffened, abducted position required for the aspiration noise to the approximated position required for modal vocal-fold vibration. The $F0$ for the first two or three glottal pulses may be higher as well, resulting in a shortened duration for the glottal pulse. (Refer to the discussion of the $F0$ Ratio acoustic measure in the next paragraph.) Each of these factors contributes to the likelihood that the vowel onset time, or VIT in this study, will be at a later point in time following the SRT than the vowel onset time in other studies, resulting in a longer VOT for the voiceless-stop consonants in this study. Despite the manner in which VIT was defined, the variation of VOT with place of articulation agrees with findings in the literature (Klatt, 1975). Within type of voicing, VOT is shortest for labials and longest for velars, except perhaps for /t/ and /k/ which have approximately the same average values.

The results of the fundamental frequency ($F0$) ratio calculations are shown in Figure 6-14 for male and female speakers. The fundamental frequency is the frequency at which the vocal folds vibrate. Since vocal-fold vibration requires not only appropriate configuration of the glottis and compliance of the vocal folds, but also transglottal pressure, $F0$ is also related to a minor extent to the respiratory system. The positive mean $F0$ ratio seen in Figure 6-14 for normal speakers indicates that the $F0_{vcls}$ value is greater than the $F0_{vcd}$ value, on average. The higher value for $F0_{vcls}$ is attributed to a residual effect of the stiffened vocal-fold position required for aspiration noise production prior to the vowel onset. These findings are consistent

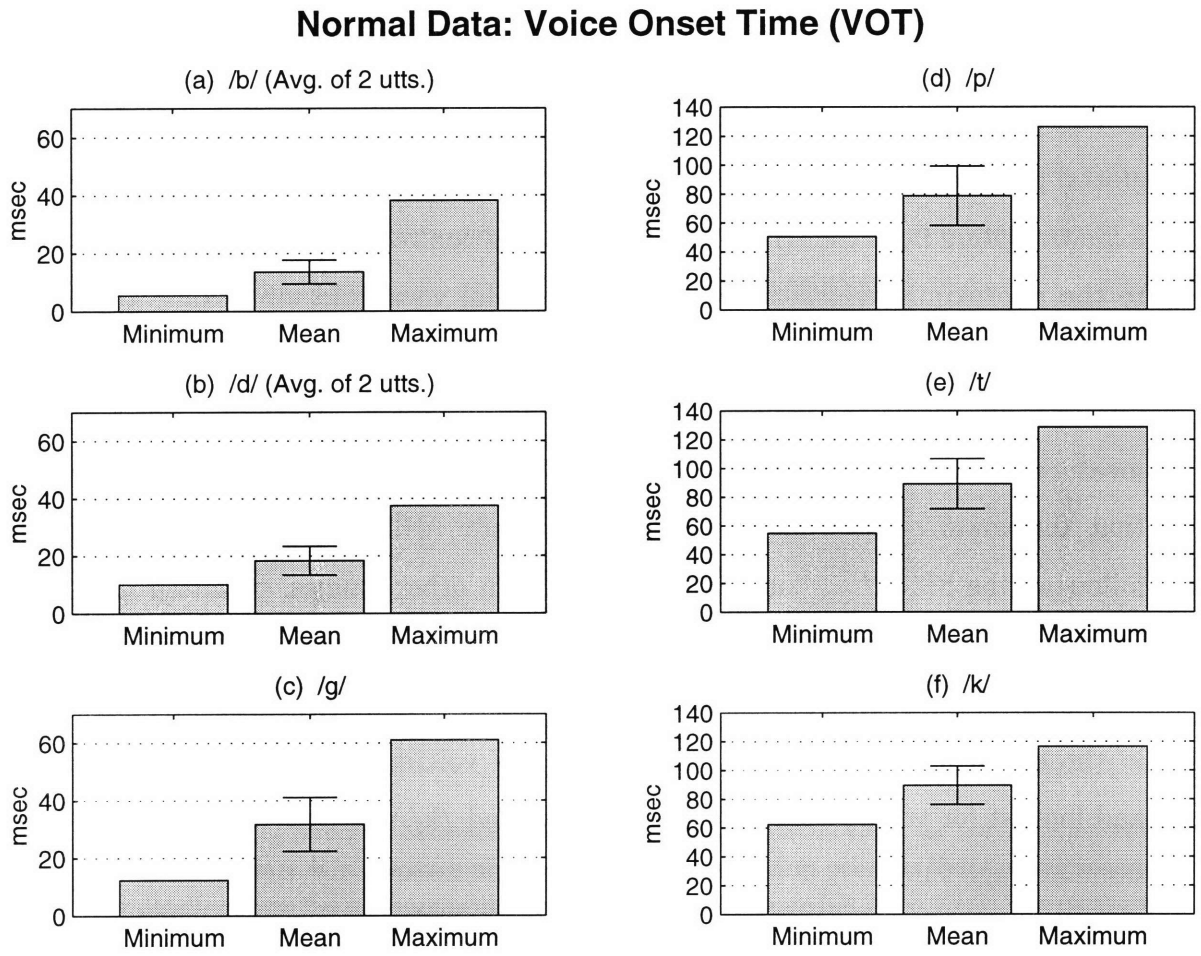

Figure 6-12 : Voice Onset Time (VOT) Acoustic Measure for normal speakers. The VOT was averaged across all 8 normal speakers, 3 repetitions/utterance, and the number of utterances indicated for the individual word-initial stops in (a)–(f). In each plot, the data shown are characterized by the mean, with a one-standard deviation error bar, and the range extrema.

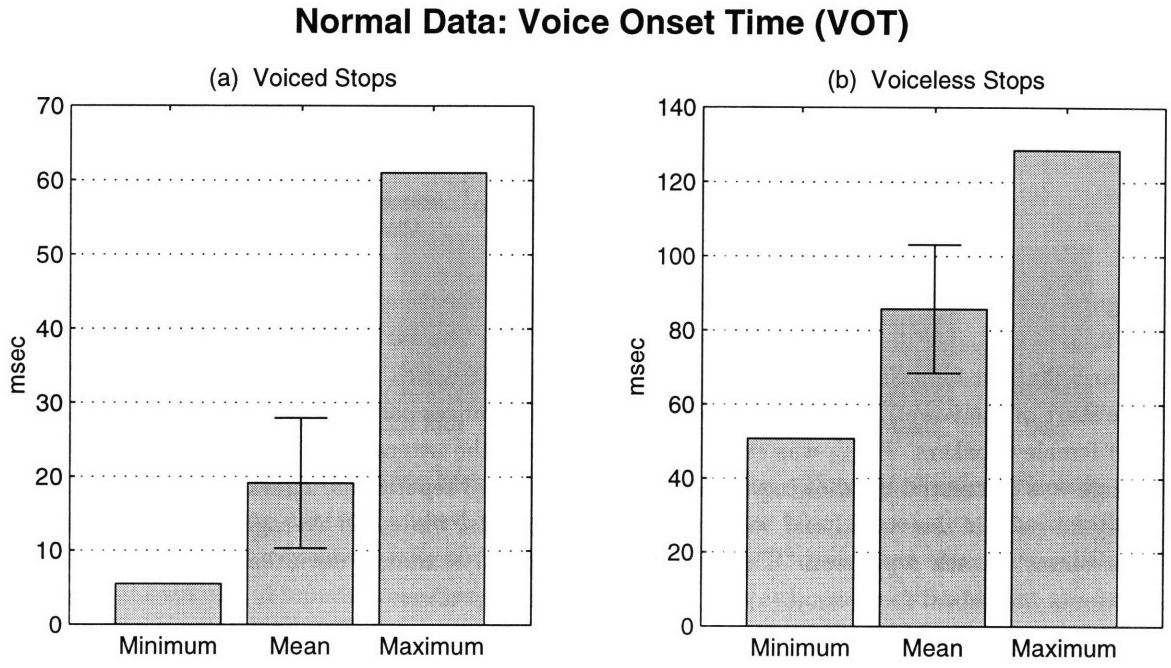

Figure 6-13 : Voice Onset Time (VOT) Acoustic Measure for normal speakers. The VOT was averaged across all 8 normal speakers, 3 repetitions/utterance, and (a) 5 utterances containing word-initial voiced stops or (b) 3 utterances containing word-initial voiceless stops. In each plot, the data shown are characterized by the mean, with a one-standard deviation error bar, and the range extrema.

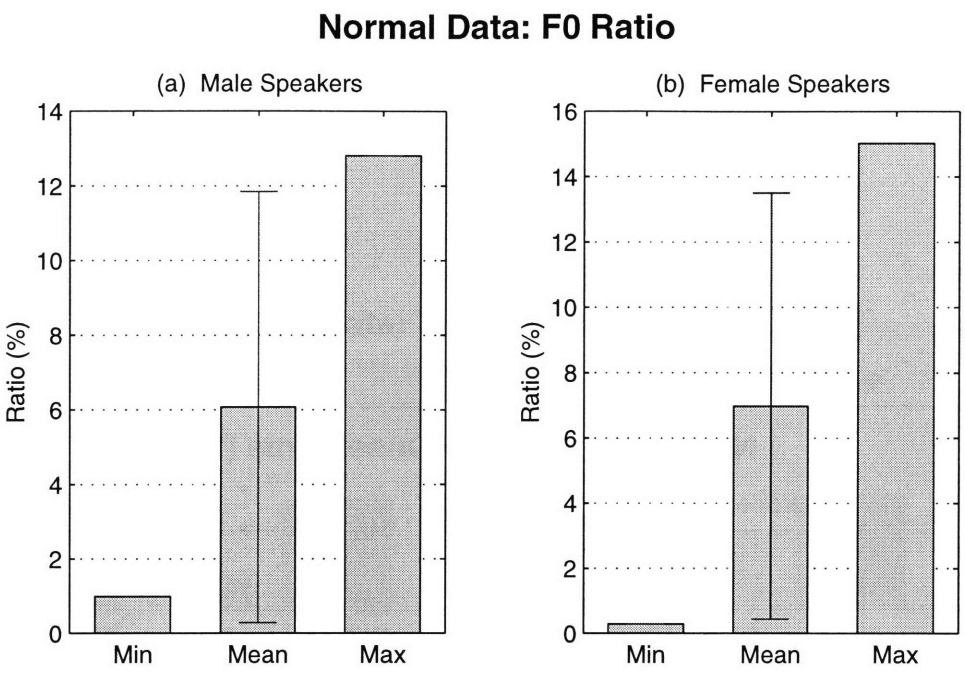

Figure 6-14 : F0 Ratio Acoustic Measure for normal speakers. The F0 Ratio is calculated as $(F0_{vcls} - F0_{vcd})/F0_{vcls}$, expressed as a percentage. For the F0 Ratio mean, $F0_{vcls}$ was averaged across the first four $F0$ values in each repetition (beginning with the VIT-identified glottal pulse at the start of the vowel), 3 repetitions/utterance, the utterances pat and tile and all four (a) male and (b) female speakers. $F0_{vcd}$ was calculated similarly for the utterances bad and dug. The $F0$ Ratio range was calculated by allowing the 6 word repetitions (3 repetitions/utterance $\times$ 2 utterances) to vary for each of the voiced and voiceless utterance subsets, while still averaging across the first four $F0$ values in each repetition. The $F0$ mean, with a one standard deviation error bar, and range extrema are shown for normal (a) male and (b) female speakers.

with Ohde (1982). The $F0$ ratio derived from his data is 16%, which he attributes to coarticulatory interaction of the voiceless frication noise source and vocal-fold vibration of the following vowel. (The magnitude difference between Ohde's $F0$ ratio value and the $F0$ ratio values presented here may potentially be due to differences in how the VIT was defined in the two studies. It is not possible to be certain of this statement, however, since Ohde does not describe the details of how VIT was determined in that study.)

A measure of the air pressure control during labial and alveolar stop production is provided by the $A1_v - A_{high}$ acoustic measure (measured from the burst and vowel average spectra). The results of this measure are shown on the y-axis in Figures 6-5 and 6-6 and replotted in Figure 6-15. Although the amplitude A_{high} predominantly

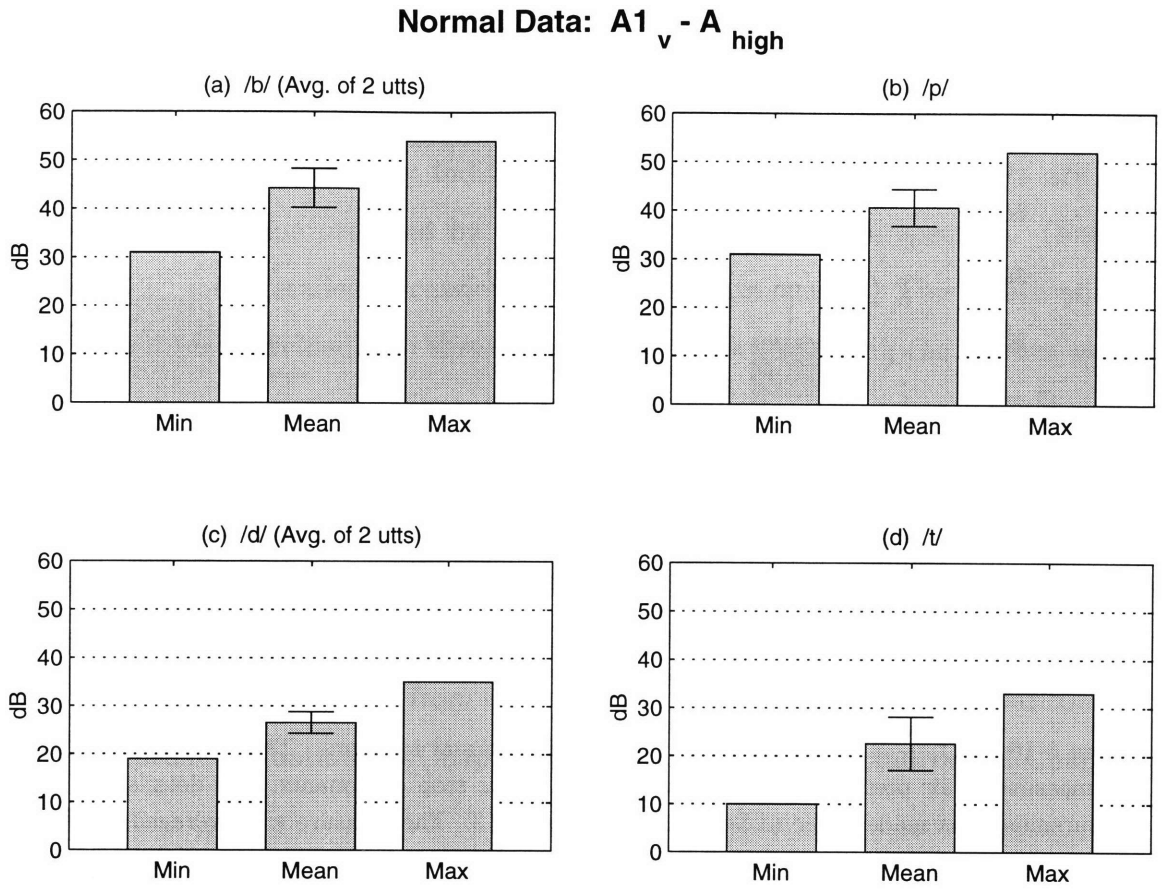

Figure 6-15 : $A1_v - A_{high}$ Acoustic Measure for normal speakers. This amplitude difference is a measure of air pressure control in word-initial labial and alveolar stop consonants. For details of how the measurement was made, refer to Sections 6.1.2 and 6.1.3. The measure was averaged across all 8 normal speakers, 3 repetitions/utterance, and the number of utterances indicated for the labial and alveolar stops in (a) – (d). In each plot, the data shown are characterized by the mean, with a one standard deviation error bar, and the range extrema.

reflects the presence or absence of the 4–5 kHz spectral prominence associated with an alveolar stop or a labial stop, respectively, A_{high} also reflects the intraoral pressure differential between voiced and voiceless stop production at the time of the release. Labial and alveolar voiced stops have average values about 3 dB higher than their voiceless stop counterparts, as seen in Figures 6-5 and 6-15. Stevens et al. (1999) observed /b/ to have a value 4 dB higher than /p/ and /d/ to be 6 dB higher than /t/, on average, using a measure similar to $A1_v - A_{high}$. These findings agree with the theory that the intraoral pressure at the time of the release is lower for voiced stops than voiceless stops, resulting in a lower value of A_{high}, and a larger value for $A1_v - A_{high}$.

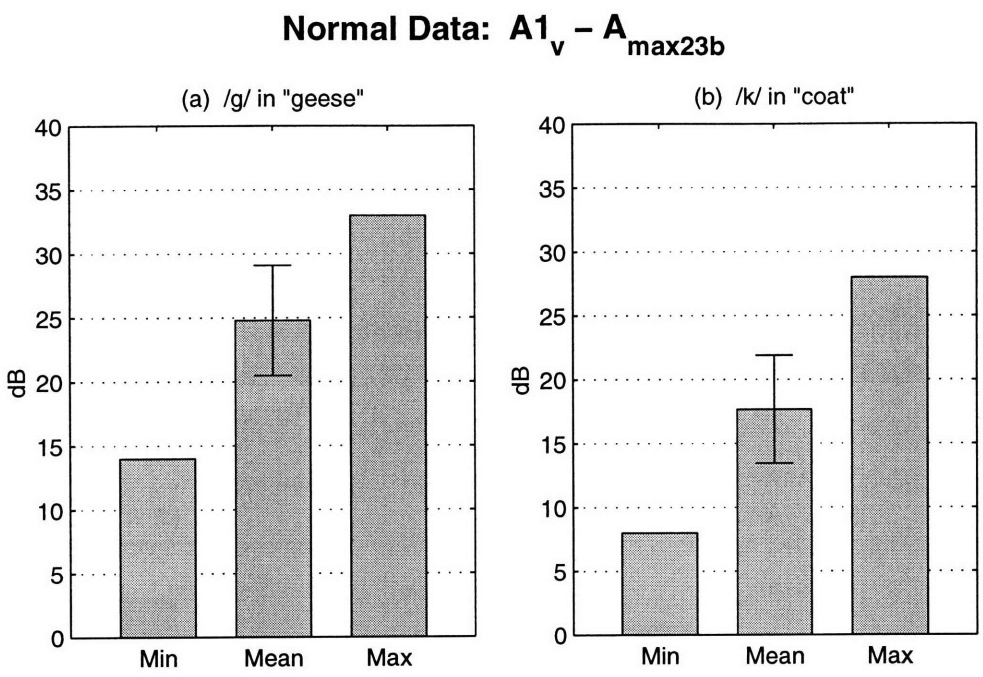

Figure 6-16 : $A1_v - A_{max23b}$ Acoustic Measure for normal speakers. This amplitude difference is a measure of air pressure control in word-initial velar stop consonants. For details of how the measurement was made, refer to Sections 6.1.2 and 6.1.3. The measure was averaged across all 8 normal speakers and 3 repetitions/utterance for the velar stops in the utterances (a) geese and (b) coat. In each plot, the data shown are characterized by the mean, with a one standard deviation error bar, and the range extrema.

A similar measure of air pressure control can be made for velar stops, $A1_v - A_{max23b}$. The results of this measure are shown in Figure 6-16. As seen for labial and alveolar stops, the velar voiced stop has an average value about 7 dB higher than for the voiceless stop, attributed to lower intraoral pressure for the voiced stop at the time of the release.

6.2.2 Dysarthric Speakers

This section contains the results of the acoustic measures performed on word-initial stop consonants produced by individuals with dysarthria. These results are compared to the baseline provided by the results from the normal speakers. (Refer to Section 6.2.1 for a detailed discussion of the normal data results.) This results and discussion section is divided into three subsections below, reflecting the various aspects of the articulatory system. These aspects are the placement and rate of movement of the primary articulator, the laryngeal system, and, to some extent, the respiratory system. In the subsections, the results of the measures are interpreted in terms of the information they reveal about articulator control and coordination for these dysarthric speakers. The dysarthric speakers' data are shown in order of decreasing speaker stop goodness score (from Chapter 4), in order to facilitate comparison with the perceptual evaluations of Chapter 4 and the spectrogram analysis of Chapter 5 as well.

Placement of Primary Articulator

The acoustic measure reflecting primary articulator placement for labial and alveolar stop consonants is $A_{high} - A_{low}$ (a measure of burst tilt, taken from the burst average spectrum). This measure is shown versus $A1_v - A_{high}$ (a measure of air pressure control, taken from the burst and vowel average spectra) in Figure 6-17 for both normal and dysarthric speakers. It can be appreciated from the figure that the circumscribed regions for normal labial and alveolar average stop values are well defined and separated from one another (as originally shown in Fig. 6-6). In contrast, the dysarthric speakers' average values as a whole are not confined to particular regions. In addition, the labial and alveolar average values overlap extensively for these speakers.

In Figures 6-18 and 6-19, the dysarthric speakers' average values for $A_{high} - A_{low}$ and $A1_v - A_{high}$ are shown for the four male and the four female speakers, respectively. It can immediately be appreciated that the dysarthric speakers have very dissimilar results from one another, as well as from normal. The results also differ at times from

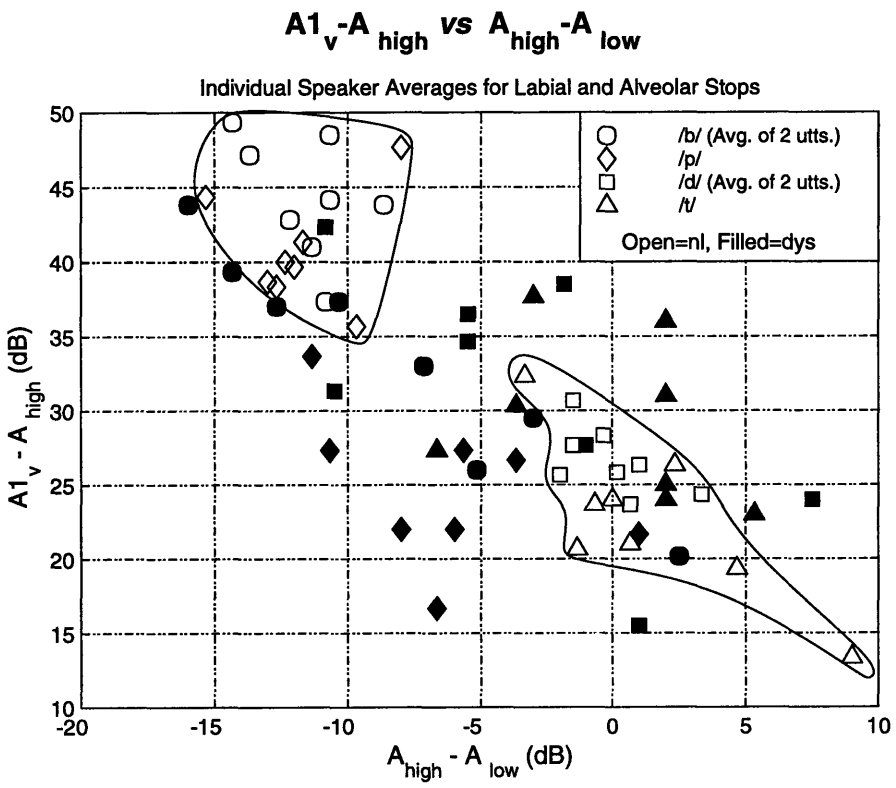

Figure 6-17 : Acoustic Measures $A1_v - A_{high}$ vs. $A_{high} - A_{low}$. Individual speaker averages shown for word-initial labial and alveolar stop consonants spoken by normal and dysarthric speakers. The amplitude difference $A1_v - A_{high}$ is a measure of air pressure control, and the difference $A_{high} - A_{low}$ is a measure of burst tilt. For details of how these measurements were made, refer to Sections 6.1.2 and 6.1.3. Each individual speaker is represented by four data points. A data point is the average of 3 repetitions/utterance across one utterance for the voiceless stops, two utterances for the voiced stops. Lines circumscribe the range of the normal speaker average data for labial and alveolar stop consonants.

the perceptual test results for place of articulation (refer to Question 3 in Fig. 4-8, page 90, and Table 4.2, page 91). For example, speaker DF4 has difficulty with the labial place of articulation but not the alveolar place, according to Figure 6-19(d). Her value of $A1_v - A_{high}$ is too low, and her value of $A_{high} - A_{low}$ is somewhat too high. (Although the value of A_{high} for her labial stops is closer to the typical normal alveolar value, it is important *not* to draw the conclusion that her labial stops are produced as alveolar stops. The amplitude differences $A1_v - A_{high}$ and $A_{high} - A_{low}$ depend upon many factors, not just placement of the articulator, as discussed in the next paragraph.) In contrast to these data, however, the perceptual data for place of articulation (Table 4.2) indicate that, although there was some difficulty detecting labial place of articulation (75–88% detected correctly), listeners had noticeably more difficulty detecting the place for alveolar stops (33–50% detected correctly). Not all results for these two measures differ from perceptual test results, however. For example, for speaker DF3, the acoustic measures in Figure 6-19(c) indicate that she has difficulty with place of articulation for alveolar stops, but not labial stops. Comparing these data to the perceptual data in Table 4.2, listeners also indicated that labial place was correct (100% detected correctly), and alveolar place was incorrect (8–13% detected correctly).

There are several possible reasons why the combination of $A1_v - A_{high}$ and $A_{high} - A_{low}$ may not be a good predictor of place of articulation for some of the dysarthric speakers. The measures are only performed when a stop release is identified. Therefore, some of these average data points may be calculated on as few as one or two repetitions (particularly for speakers DM3 and DF4). If the stop is identified as a glottal stop, the measures are performed, even though the formant excitation probably appears most similar to that of the following vowel. The results of these measures are influenced by other aspects of the speech system, in addition to the placement of the articulator. For example, if there is a high intraoral pressure at the time of the release, such as for an ejective, then the value of A_{high} may be too high. If the velopharyngeal port is faulty, a nasal resonance may appear in the burst average spectrum, typically in the same frequency region as A_{high} is measured, and may oc-

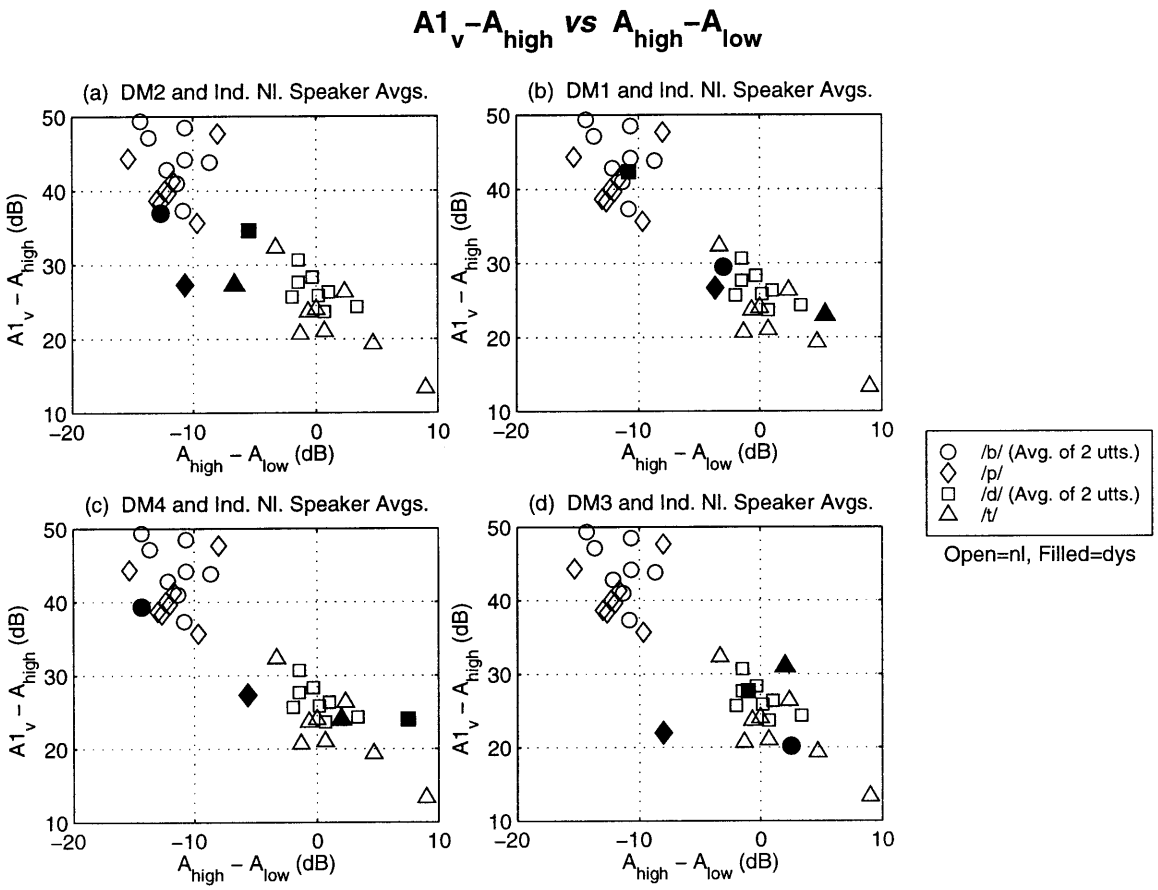

Figure 6-18 : Acoustic Measures $A1_v - A_{high}$ vs. $A_{high} - A_{low}$. The amplitude difference $A1_v - A_{high}$ is a measure of air pressure control, and the difference $A_{high} - A_{low}$ is a measure of burst tilt. For details of how these measurements were made, refer to Sections 6.1.2 and 6.1.3. The four subplots show the results of the measures for word-initial labial and alveolar stop consonants, all 8 individual normal speakers, and the male dysarthric speakers (a) DM2, (b) DM1, (c) DM4, and (d) DM3 (in order of decreasing stop goodness score). In each subplot, each speaker is represented by four data points. A data point is the average of 3 repetitions/utterance across one utterance for the voiceless stops, two utterances for the voiced stops. Repetitions in which the stop is omitted by the speaker are not included in the average.

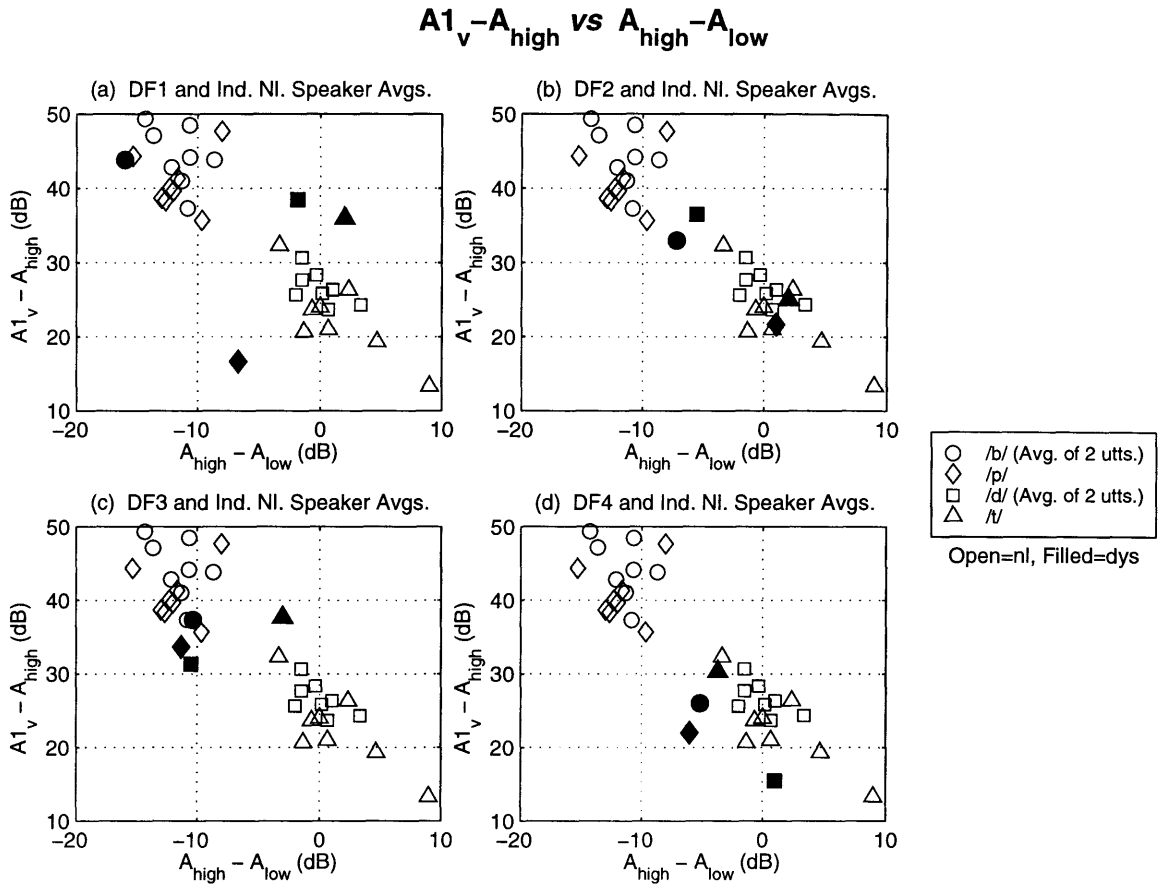

Figure 6-19 : Acoustic Measures $A1_v - A_{high}$ vs. $A_{high} - A_{low}$. The amplitude difference $A1_v - A_{high}$ is a measure of air pressure control, and the difference $A_{high} - A_{low}$ is a measure of burst tilt. For details of how these measurements were made, refer to Sections 6.1.2 and 6.1.3. The four subplots show the results of the measures for word-initial labial and alveolar stop consonants, all 8 individual normal speakers, and the female dysarthric speakers (a) DF1, (b) DF2, (c) DF3, and (d) DF4 (in order of decreasing stop goodness score). In each subplot, each speaker is represented by four data points. A data point is the average of 3 repetitions/utterance across one utterance for the voiceless stops, two utterances for the voiced stops. Repetitions in which the stop is omitted by the speaker are not included in the average.

casionally boost the value of A_{high}. Also, the value of $A1_v$ may be too low, as will be discussed in the upcoming **Laryngeal and Respiratory Systems** subsection.

Although a comparison of the acoustic measure results and perceptual test results reveals inconsistencies in the ability of the acoustic measure to predict place of articulation, the acoustic measure results may still be predictive of the "naturalness" of the stop, as reflected by Question 5 of the perceptual test (refer to Chapter 4). Since each of the dysarthric speaker's acoustic measure results differs from normal in some way, this measure may contribute to an understanding of why the dysarthric speakers differ from normal in their stop goodness scores (as shown in Fig. 4-2, page 82).

The second measure of articulator placement comes from the initial value of $F2$ for the alveolar stops in the formant-frequency transitions for speakers DM2, DM1 and DF1 shown in Figure 6-20, 6-21 and 6-22, respectively. This initial value is an indicator of correct alveolar place of articulation. For DM2 and DF1, the initial mean value of $F2$ for /d/ is within 100 Hz of normal, and the range for the initial value overlaps to a large degree with the normal range (Figs. 6-20 and 6-22, (c) and (d)). These two speakers have most of their /d/ stops identified correctly by the listeners as well (83–96% detected correctly, Table 4.2, page 91). Speaker DM1 has an initial mean $F2$ value that is greater than normal by 300–400 Hz, and the range for that value overlaps the normal range to only a small degree, for the two utterances containing word-initial /d/ (Fig. 6-21 (c) and (d)). One possible explanation for this finding is a longer constriction near the alveolar ridge. If more of the tongue body forms the constriction (not just the tongue tip), then the $F2$ value may be higher. Another possible explanation is that the tongue body is more fronted. Speaker DM1 also has marginally poorer results for the Time Course of F2 Change Qualitative Spectrogram Analysis (SA) attribute (as shown in Fig. 5-14, averaged over all utterances containing voiced stops) compared to speakers DM2, DF1, and the normal speakers. These results do not affect perception of place of articulation for speaker DM1 to a noticeable degree, since the listeners identified 88% of his /d/ stops correctly (Table 4.2). The conclusion is that the location of the constriction near the alveolar ridge for DM1 means the stop is still judged by listeners to have an alveolar rather than velar place

of articulation. The presence of a longer constriction may still contribute to poorer production quality scores for his alveolar stop /d/, however (Fig. 4-2, page 82).

Rate of Primary Articulator Movement

The rate of movement of the primary articulator can be inferred from the rate of movement of the formant frequencies $F1$ and $F2$ shown in Figures 6-20, 6-21, and 6-22 for speakers DM2, DM1, and DF1, respectively. (The formant frequencies were tracked manually for speakers DM2, DM1 and DF1, but not for any of the other speakers because it became too difficult to identify the formants.) These formant-frequency transitions are measured starting at vowel onset (VIT), and consequently they reflect a combination of the primary articulator movement away from the constriction and the tongue movement toward the vowel steady state. From these three figures, it can be seen that all of the formant transitions are in the correct direction. These three speakers also have Time Course of F1 Rise and Time Course of F2 Change SA attributes which are not much different from normal (Figs. 5-13 and 5-14, pages 122 and 123). These findings agree with the perceptual test results that these speakers have stop intelligibilities which are not significantly different from normal (Fig. 4-3, page 83). The subtle differences from normal that do appear in the rates of Figures 6-20, 6-21, and 6-22 and the attribute ratings of Figures 5-13 and 5-14 may contribute to the stop production quality scores (Fig. 4-2, page 82).

The second measure of primary articulator rate following stop release infers the rate from the average number of stop bursts occurring in the release time period for each stop and speaker. This measure may be able to provide some information about the rate of primary articulator movement, particularly for those dysarthric speakers (DF2, DF3, DM4, DM3 and DF4) for whom the formant-frequency transitions were too difficult to track manually. The presence of two or more bursts in a sequence for a given repetition can imply that the primary articulator movement is slower following release. The primary articulator must remain in a superior position to narrow or close the constriction again, which can slow its overall rate of movement downward. However, if the constriction closes again after each burst except the final burst in

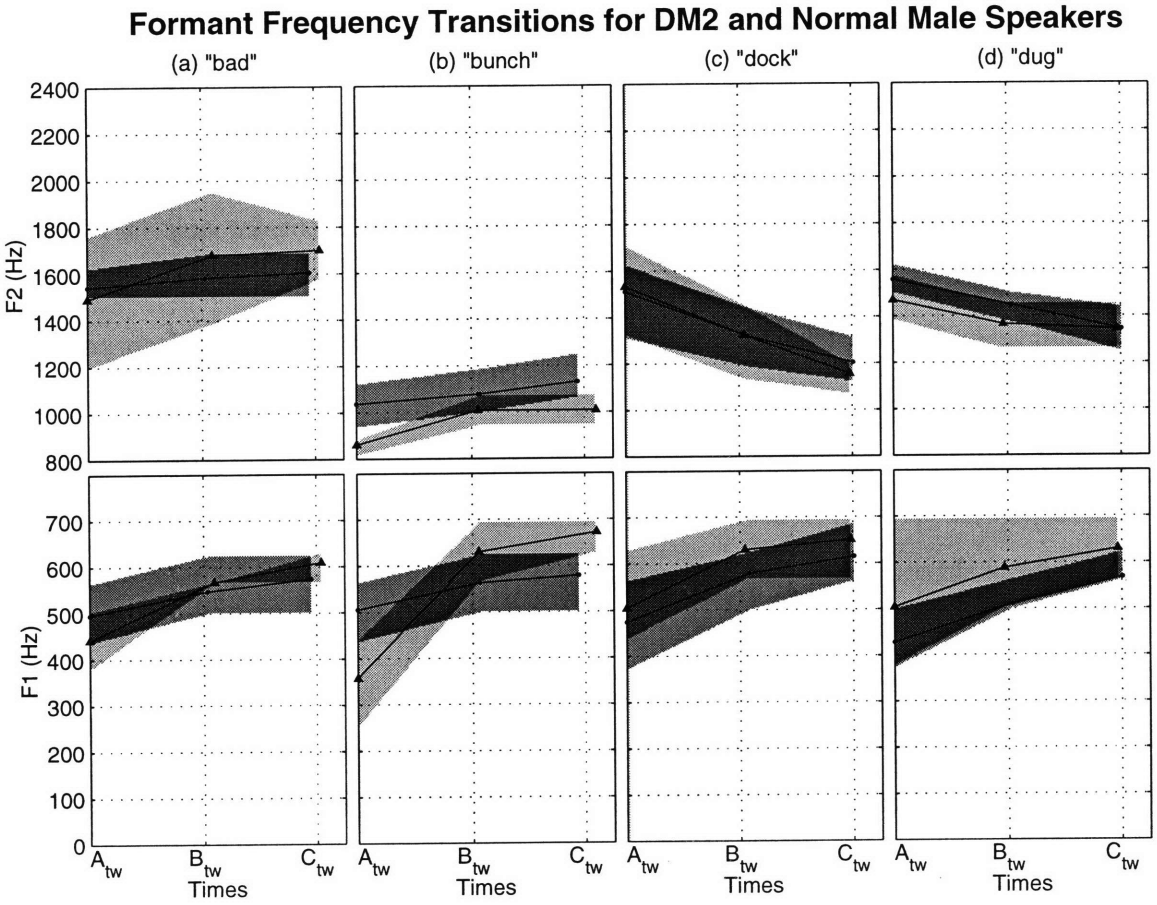

Figure 6-20 : Formant-Frequency Transition Acoustic Measure for DM2 and normal male speakers. Along the x-axis the times have been labeled A_{tw}, B_{tw} and C_{tw}, respectively, as discussed in Section 6.1.2. The subscript tw refers to the time warping that may occur when the formant frequency values are averaged across repetitions. (To a rough approximation, Time A_{tw} can be considered to be at the VIT, Time B_{tw} at VIT + 20 ms and Time C_{tw} at VIT + 40 ms, but, for more accurate times, the reader is referred to the discussion of Section 6.1.2.) This measure is shown averaged across 3 repetitions/utterance for the utterances (a) bad, (b) bunch, (c) dock and (d) dug. The means are shown as solid lines, with the normal mean averaged across all 4 normal male speakers. The light gray shaded region is the range for speaker DM2, the medium gray region is the range extrema across all 4 normal male speakers, and the dark gray region is the region of overlap between normal and dysarthric speakers.

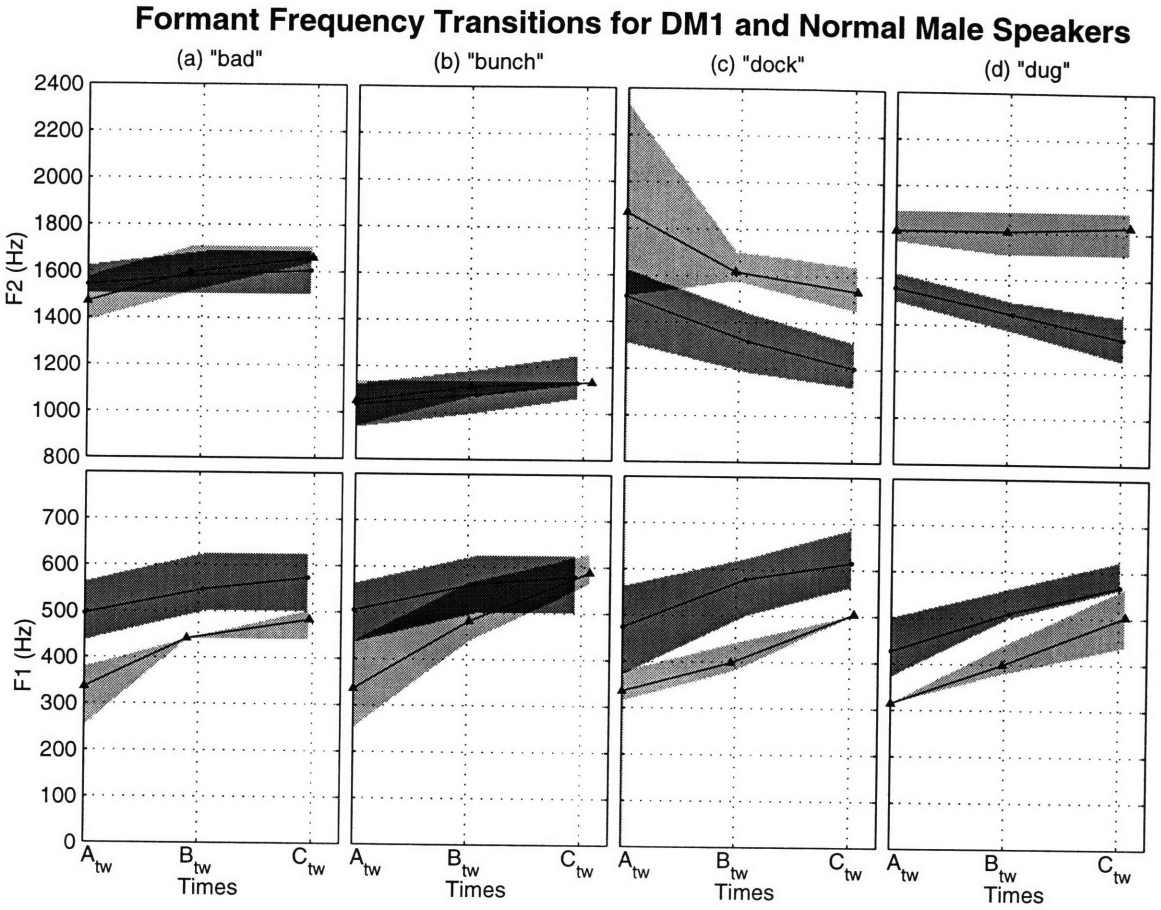

Figure 6-21 : Formant-Frequency Transition Acoustic Measure for DM1 and normal male speakers. Along the x-axis the times have been labeled A_{tw}, B_{tw} and C_{tw}, respectively, as discussed in Section 6.1.2. The subscript tw refers to the time warping that may occur when the formant frequency values are averaged across repetitions. (To a rough approximation, Time A_{tw} can be considered to be at the VIT, Time B_{tw} at VIT + 20 ms and Time C_{tw} at VIT + 40 ms, but, for more accurate times, the reader is referred to the discussion of Section 6.1.2.) This measure is shown averaged across 3 repetitions/utterance for the utterances (a) bad, (b) bunch, (c) dock and (d) dug. The means are shown as solid lines, with the normal mean averaged across all 4 normal male speakers. The light gray shaded region is the range for speaker DM1, the medium gray region is the range extrema across all 4 normal male speakers, and the dark gray region is the region of overlap between normal and dysarthric speakers.

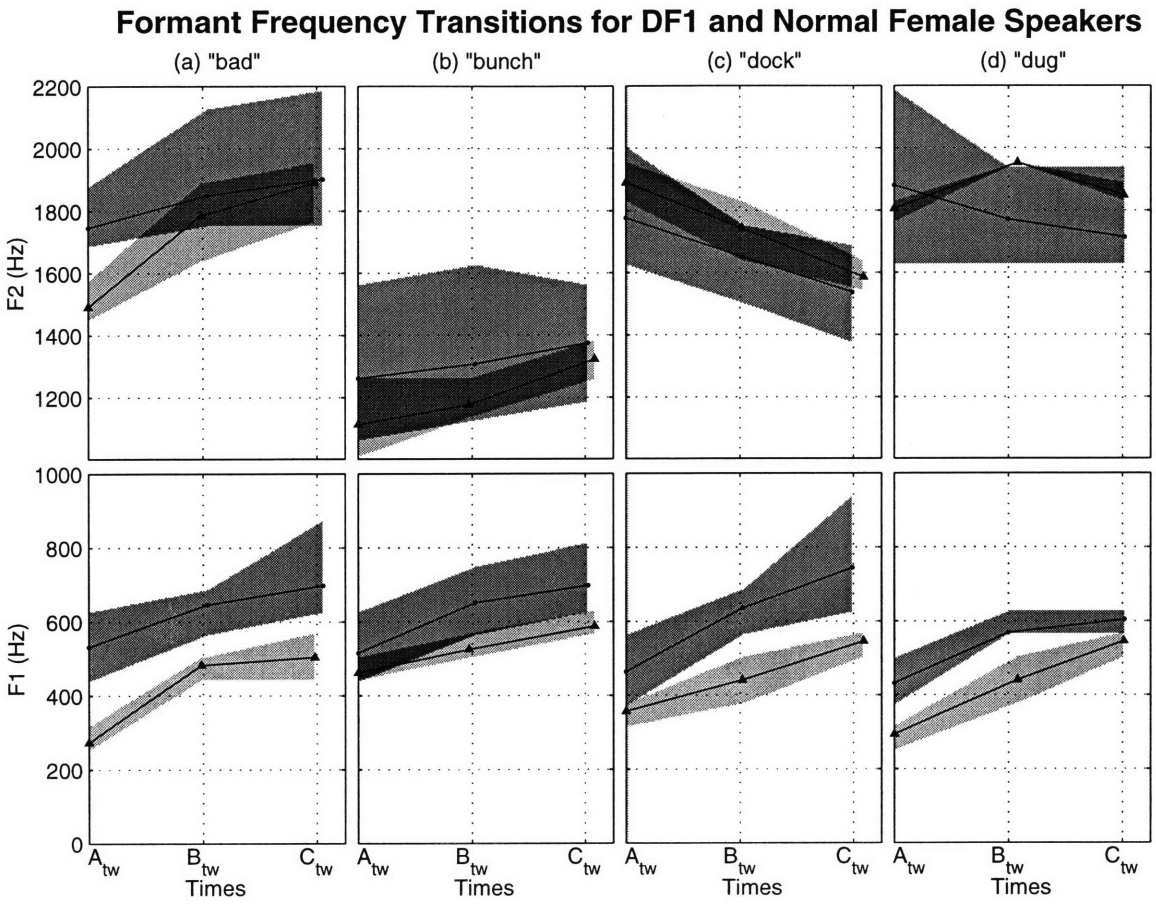

Figure 6-22 : Formant-Frequency Transition Acoustic Measure for DF1 and normal female speakers. Along the x-axis the times have been labeled A_{tw}, B_{tw} and C_{tw}, respectively, as discussed in Section 6.1.2. The subscript *tw* refers to the time warping that may occur when the formant frequency values are averaged across repetitions. (To a rough approximation, Time A_{tw} can be considered to be at the VIT, Time B_{tw} at VIT + 20 ms and Time C_{tw} at VIT + 40 ms, but, for more accurate times, the reader is referred to the discussion of Section 6.1.2.) This measure is shown averaged across 3 repetitions/utterance for the utterances (a) bad, (b) bunch, (c) dock and (d) dug. The means are shown as solid lines, with the normal mean averaged across all 4 normal female speakers. The light gray shaded region is the range for speaker DF1, the medium gray region is the range extrema across all 4 normal female speakers, and the dark gray region is the region of overlap between normal and dysarthric speakers.

a series, then the final burst is considered to be the SRT, and the rate of release after that burst is not predicted by the number of bursts preceding it. Therefore, the number of bursts in a multiple-burst sequence can only *suggest* which speakers may have slower rates of release for certain stop consonants. With this information in mind, the results of this measure are shown in Figure 6-23. From this figure it can be observed that the production of multiple bursts is highly speaker- and stop-dependent. There is a slight trend toward more instances of multiple bursts for velar stops than labial and alveolar stops among the dysarthric speakers, similar to normals. Also, seven of the eight dysarthric speakers have a multiple burst in at least one of their /d/ productions. Otherwise, on a speaker-by-speaker basis, the instances in which the average number of multiple bursts is near two or higher is as follows: DM1 /p/, DF2 /t,g/, DF3 /k/, and DM3 /b,g,k/. This information can be compared to the Abruptness of Release SA attribute (Fig. 5-10, page 118). There is some correspondence between a less abrupt release and more instances in which the average number of multiple bursts is near two or higher, although the correspondence is not one-to-one. (In addition to the presence of multiple bursts for several of the alveolar and velar utterances, the faulty velopharyngeal port opening for DF2 also probably influences her Abruptness of Release attribute rating.)

Laryngeal and Respiratory Systems

The acoustic measure $A1_v - A1_p$ (measured from the precursor and vowel spectra), is designed to assess the duration and amplitude of prevoicing, or voicing present prior to the voiced stop release. The results of this measure are shown in Figure 6-24 for each voiced stop and in Figure 6-25 averaged across all voiced stops. (As discussed in Section 6.2.1, data will not be reported for the voiceless stops.) As the duration and/or amplitude of the prevoicing increases, the value of $A1_p$ increases, and the difference $A1_v - A1_p$ decreases. Strong prevoicing may be generated by a speaker preceding the stop release in a number of ways, such as by building up subglottal pressure too quickly, building up too much subglottal pressure, relaxing or actively expanding the supraglottal cavity walls too much, and/or approximating the vocal

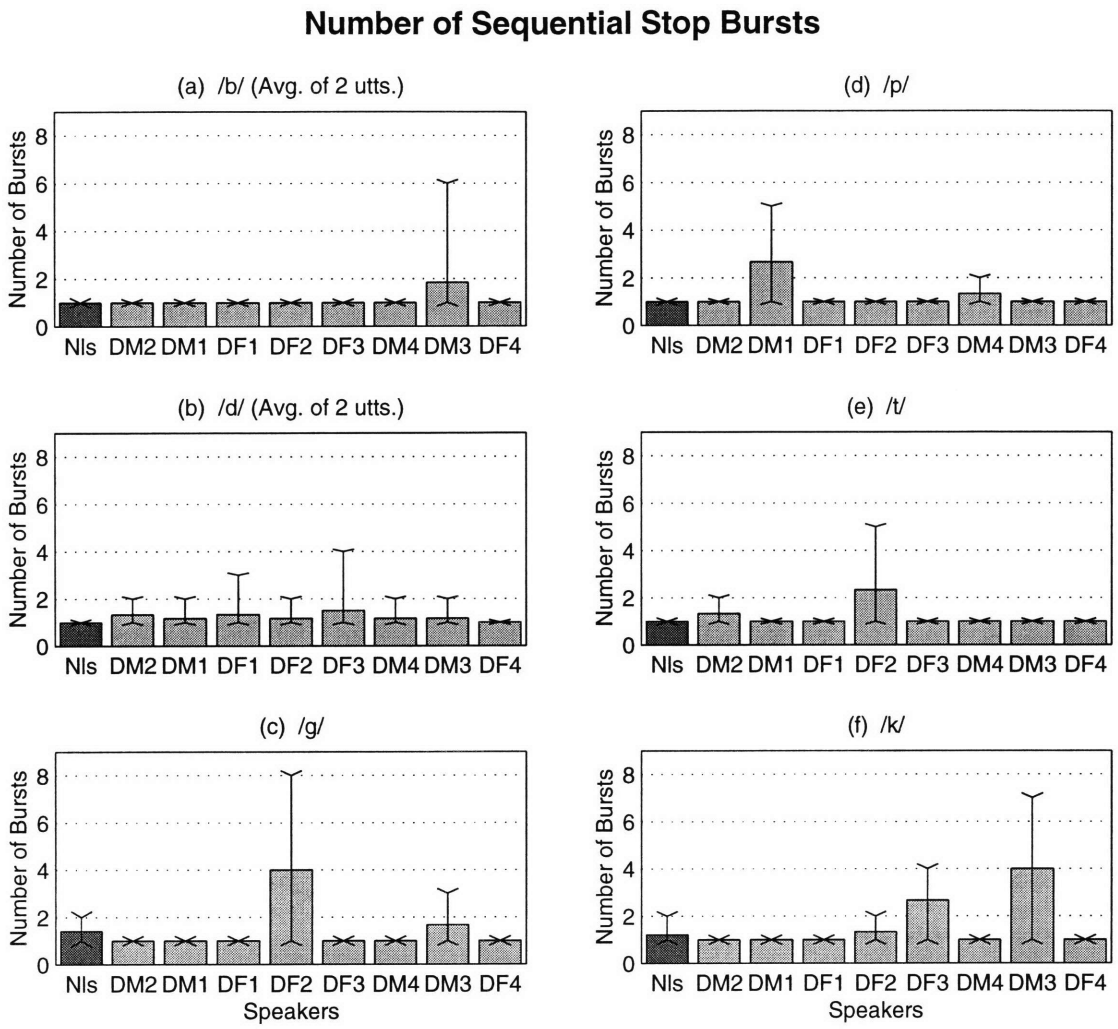

Figure 6-23 : Acoustic Measure of Number of Stop-Consonant Bursts in multiple-burst sequences. For each speaker, number of bursts per word repetition shown averaged across 3 repetitions/utterance, and one utterance for each word-initial stop (with the exception of /b/ and /d/, which each contain two utterances). When a glottal stop was produced, that repetition was considered to have a single burst. When a stop was omitted, that repetition was not included in the measure. For normal speakers, the measure was also averaged across all 8 speakers. The bars represent the mean, and the error bars are the range extrema. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

folds too soon. (Due to variable recording conditions for these dysarthric speakers, it is also possible that $A1_p$ reflects nonspeaker-generated noises (see Section 4.1.3).) Variation that may occur in the value of $A1_v$ will be discussed later in this subsection.

In Figure 6-24, the values of $A1_v - A1_p$ that deviate noticeably from normal are /b/ production for DM2, DM1 and DM4, and /g/ production for DF1 and DM3, in which $A1_v - A1_p$ is too low. Also from this figure it is observed that, for a given dysarthric speaker, the average value of $A1_v - A1_p$ may vary with the place of articulation, unlike for the normal speakers. Comparing the average of $A1_v - A1_p$ across all voiced stops (Fig. 6-25) to the Prevoicing SA attribute shown in Figure 5-9 (page 117), a quite close correspondence is observed between a poorer Prevoicing attribute rating and a lower $A1_v - A1_p$ value (indicating excessive prevoicing). A correspondence can also be shown between listener responses to Question 1 regarding the presence or absence of a precursor prior to the stop release (Fig. 4-6, page 87) and Figure 6-25. This correspondence is somewhat less direct, however, as the definition of precursor for the perceptual evaluations of Chapter 4 included all speaker (subject)-generated sounds prior to the stop release, not just prevoicing.

The second measure of laryngeal function is the voice onset time (VOT), reflecting the duration from the stop release to the onset of the following vowel. The results for this measure are shown in Figure 6-26 for each stop separately and in Figure 6-27 for the average of the voiced stops and the average of the voiceless stops. From Figure 6-26, the variability in VOT values across stops for a given dysarthric speaker can be appreciated. Although the trend for most dysarthric speakers is for VOT to increase as the constriction moves further back in the oral cavity, similar to normals, the average values and the ranges for the dysarthric speakers can vary widely from normals. Variability in the dysarthric speakers' results can also be appreciated in Figure 6-27.

The results shown in Figure 6-27 can be compared to the VOT SA attribute results in Figure 5-12 (page 121) and the perceptual test Question 2 results in Figure 4-7 (page 88). In Figure 5-12(a) the rating is poorer for voiced stops with longer VOT

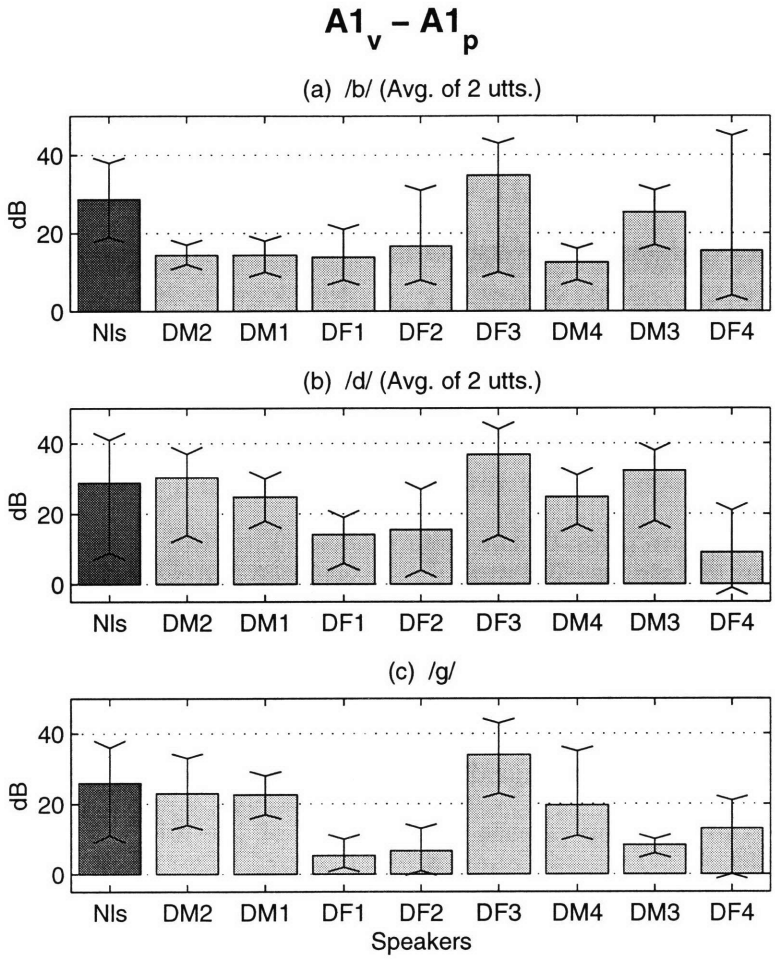

Figure 6-24 : $A1_v - A1_p$ Acoustic Measure by individual word-initial voiced stop. This amplitude difference is a measure of the presence of prevoicing prior to the stop-consonant release. For details of how the measurement was made, refer to Sections 6.1.2 and 6.1.3. This measure was averaged across 3 repetitions/utterance and the number of utterances indicated for the word-initial voiced stops in (a)–(c). For normal speakers, the measure was also averaged across all 8 speakers. The bars represent the mean, and the error bars are the range extrema. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

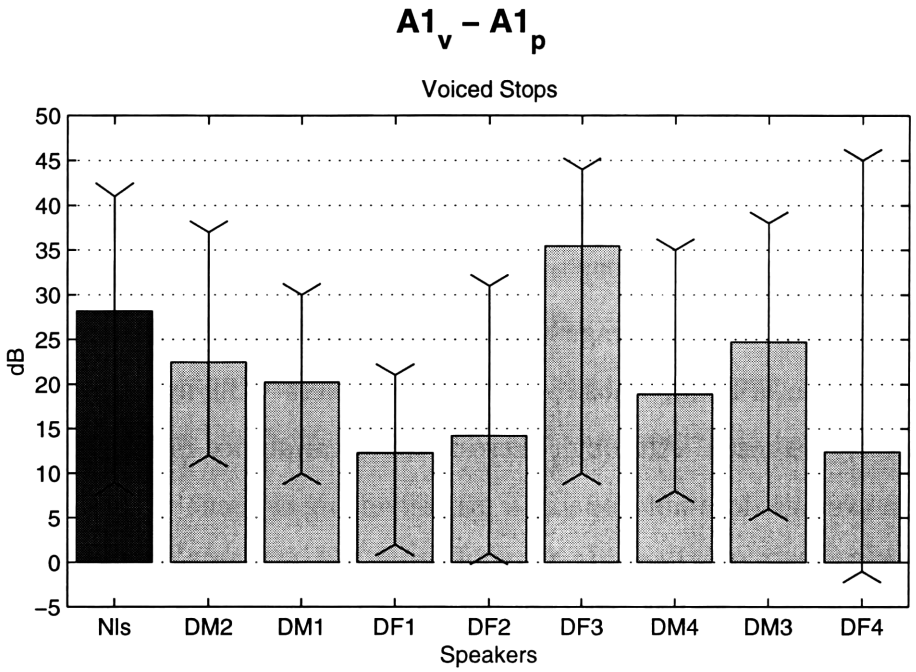

Figure 6-25 : $A1_v - A1_p$ Acoustic Measure across all word-initial voiced stops. This amplitude difference is a measure of the presence of prevoicing prior to the stop-consonant release. For details of how the measurement was made, refer to Sections 6.1.2 and 6.1.3. The measure was averaged across 3 repetitions/utterance and 5 utterances containing word-initial voiced stops. For normal speakers, the measure was also averaged across all 8 speakers. The bars represent the mean, and the error bars are the range extrema. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

values, but remains at a value of 1 for VOT values which are correct or too short. For the speaker with the longest voiced VOT average in Figure 6-27(a), speaker DM3, the attribute rating is poorest, as to be expected. For speaker DF3, however, the acoustic measure shows an average VOT value close to normal while the attribute rating is somewhat poorer than normal. The perceptual test results for voiced stops are in Figure 4-7(a). These results for DM3 and DF3 are similar to the SA results. Based on the acoustic measure alone, average voiced-stop VOT values are too short for DF1 and DM4. When the VOT is too short, it can be an indication of prevoicing, vowel or glottal-stop production. A VOT that is too short for voiced-stop production is not likely to result in misclassification of the type of voicing in the perceptual test. ("Vowel" is considered to be voiced.)

For voiceless stops, the VOT acoustic measure results of Figure 6-27(b) show a general increase in VOT average value with decreasing stop goodness score. The notable exception is DM3, whose voiced and voiceless VOT average values are both approximately 40 msec. Although there is some variability in his VOT values, this speaker on average does not appear to utilize voicing as a cue to distinguish between voiced and voiceless stop consonants. Average voiceless-stop VOT values are too long for DM4 and DF4. A VOT that is too long can be an indication of a slow-moving primary articulator or the prolonged generation of aspiration noise. The acoustic measure results for voiceless stops agree well with the VOT SA attribute results for voiceless stops (shown in Fig. 5-12(b), page 121). Since a poor attribute rating is assigned in voiceless-stop production for VOT values that are either too long *or* too short, the attribute rating of approximately 2.7 for DM3 corresponds well to the observation of a VOT average value for the acoustic measure that is too short (Fig. 6-27(b)). Comparing the acoustic measure and perceptual test results, the expectation would be that a VOT that is too long would still result in the perception of a voiceless stop consonant, whereas a VOT that is too short would result in perception of a voiced stop consonant (for the purposes of determining type of voicing, the "vowel" category in the perceptual test will be considered "voiced"). The results of the perceptual test (Fig. 4-7(b)) correspond nicely to the results for the VOT acoustic measure.

Speakers who have some utterances with VOT values that are too short (the range extends below the normal range), such as DF2, DF3, DM4, DM3 and DF4, each have some voiceless stops judged to be either voiced or vowel. Speaker DM3, the speaker with the shortest VOT average value from the acoustic measure is the only speaker judged to produce more voiced stops or vowels than voiceless stops when attempting to produce voiceless stops.

The F0 ratio represents the third acoustic measure of laryngeal function. The results of this measure are reported in Figure 6-28. The steps involved in calculating the ratio are summarized in the figure caption and are described in more detail in Section 6.1.3. It was hoped that this measure would reflect a difference for voiced and voiceless stop production. The presence of a higher initial F0 value for voiceless stops than voiced stops was anticipated, due to a coarticulation effect attributed to the stiffer vocal-fold position required for the generation of aspiration noise preceding the vowel in the voiceless stop production. However, the results shown in Figure 6-28 are obscured by variability, for both the normal and the dysarthric speakers. Although there is a small positive average percentage difference as expected for normals, the range of variability is large, and the range for each of the dysarthric speakers overlaps the normal range to some degree. It is possible that the F0 ratio might become a better indicator of vocal-fold stiffness if F0 were measured at the very first indication of vocal-fold vibration following the stop release (rather than at the VIT), and if only that initial F0 value was compared across repetitions (rather than an average of the first four F0 values). Since the dysarthric speakers' results do differ from normal, on average, the F0 ratio measured as is may reflect an aspect of stop quality.

There are two measures of air pressure control, $A1_v - A_{high}$ for labial and alveolar stops and $A1_v - A_{max23b}$ for velar stops. These measures are designed to assess the intraoral pressure difference at the time of the release between voiced and voiceless stop consonants with the same place of articulation. The value of $A1_v - A_{high}$ is expected to be larger for voiced than for voiceless stops, within place of articulation, based on a lower intraoral pressure for voiced stops. The results of these measures are shown in Figure 6-29 for labial and alveolar stops and Figure 6-30 for velar stops.

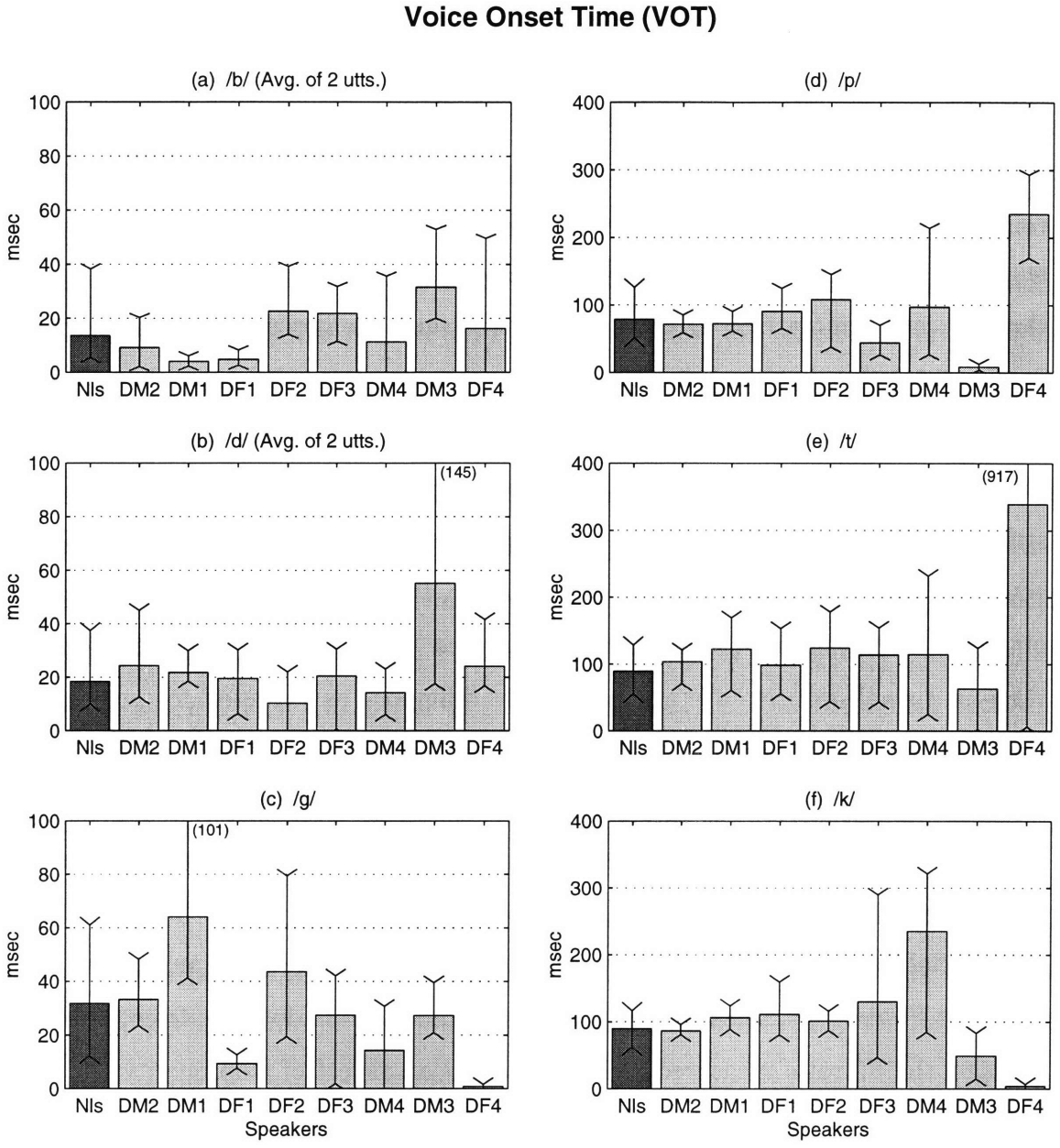

Figure 6-26 : Voice Onset Time (VOT) Acoustic Measure. For each speaker, the measure was averaged across 3 repetitions/utterance and the number of utterances indicated for the individual word-initial stops in (a)–(f). For normal speakers, the measure was also averaged across all 8 speakers. The bars represent the mean, and the error bars are the range extrema. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

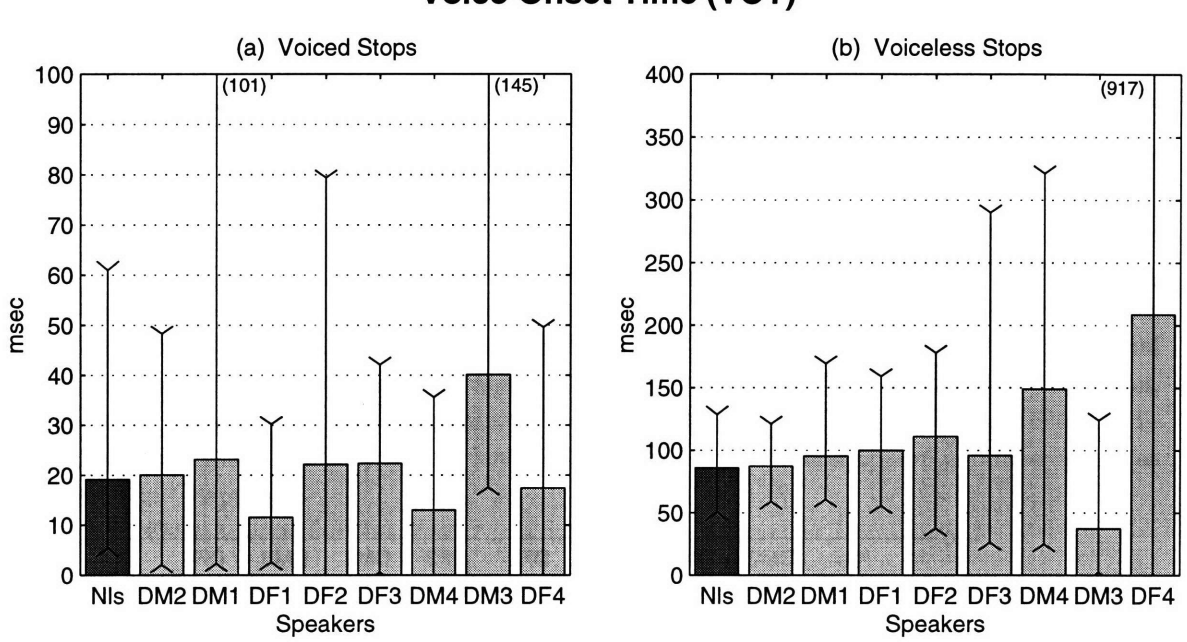

Figure 6-27 : Voice Onset Time (VOT) Acoustic Measure. For each speaker, the measure was averaged across 3 repetitions/utterance and (a) 5 utterances containing word-initial voiced stops or (b) 3 utterances containing word-initial voiceless stops. For normal speakers, the measure was also averaged across all 8 speakers. The bars represent the mean, and the error bars are the range extrema. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

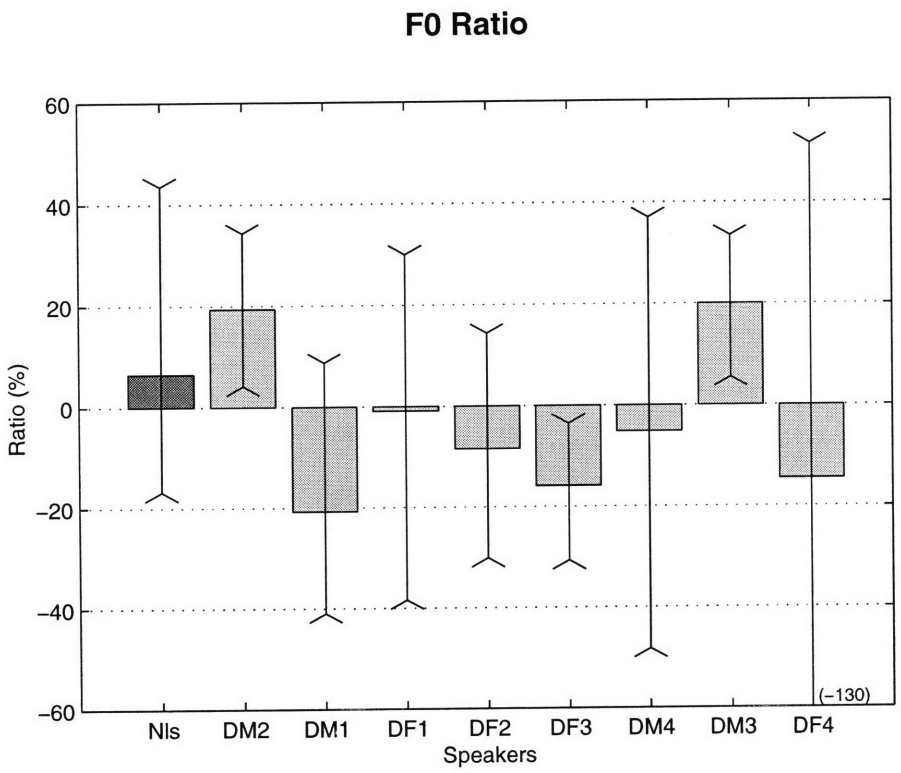

Figure 6-28 : F0 Ratio Acoustic Measure. The F0 Ratio is calculated as $(F0_{vcls} - F0_{vcd})/F0_{vcls}$, expressed as a percentage. For the F0 Ratio mean, $F0_{vcls}$ was averaged across the first four $F0$ values in each repetition (beginning with the VIT-identified glottal pulse at the start of the vowel), 3 repetitions/utterance, the utterances pat and tile, and all 8 normal speakers (Nls) or each individual dysarthric speaker (DF1–DF4, DM1–DM4). $F0_{vcd}$ is calculated similarly for the utterances bad and dug. The $F0$ Ratio range is calculated by allowing the 6 word repetitions (3 repetitions/utterance $\times$ 2 utterances) to vary for each of the voiced and voiceless utterance subsets, while still averaging across the first four $F0$ values in each repetition. For normal speakers, the range also reflects the variation across the 8 speakers. The bars represent the mean, and the error bars are the range extrema. The normal and dysarthric speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

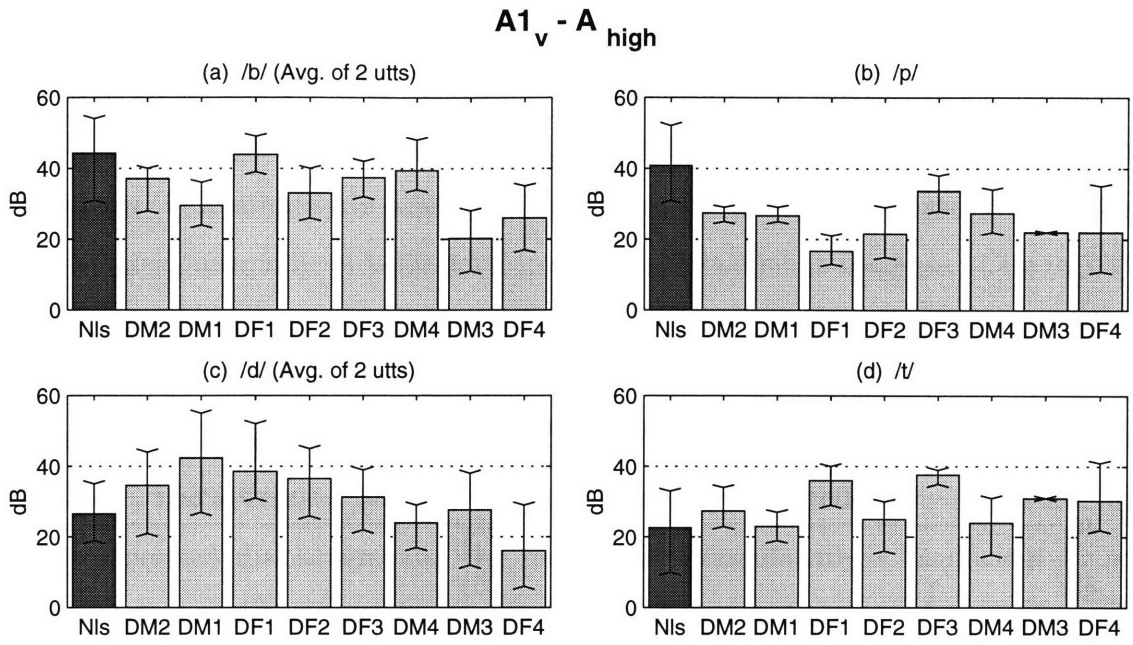

Figure 6-29 : $A1_v - A_{high}$ Acoustic Measure. This amplitude difference is a measure of air pressure control in word-initial labial and alveolar stop consonants. For details of how the measurement was made, refer to Sections 6.1.2 and 6.1.3. The measure was averaged across 3 repetitions/utterance and the number of utterances indicated for the labial and alveolar stops in (a) - (d). For normal speakers, these measures were also averaged across all 8 speakers. The bars represent the mean, and the error bars are the range extrema. The normals (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

The $A1_v - A_{high}$ measure deviates most from normal for /p/ production, in which five of the eight dysarthric speakers have $A1_v - A_{high}$ ranges which are so low that they do not overlap the normal range. These speakers are DM2, DM1, DF1, DF2, and DM3. In Figure 6-30, due to the broad normal range of variability, only speaker DF2 has values of $A1_v - A_{max23b}$ that are outside the normal range (for the utterance coat, Fig. 6-30(b)).

The deviations from normal discussed for the values of $A1_v - A_{high}$ and $A1_v - A_{max23b}$ could be attributed to a value of A_{high} or A_{max23b} that is too large. If A_{high} or A_{max23b} is too high, it can indicate formation of an ejective instead of a pulmonary release. The closed glottis and active contraction of the supraglottal cavity required to form an ejective result in increased intraoral pressure, P_m, compared to normal. At the time of the release, the increased P_m boosts A_{high} or A_{max23b}, and consequently decreases $A1_v - A_{high}$ or $A1_v - A_{max23b}$, respectively.

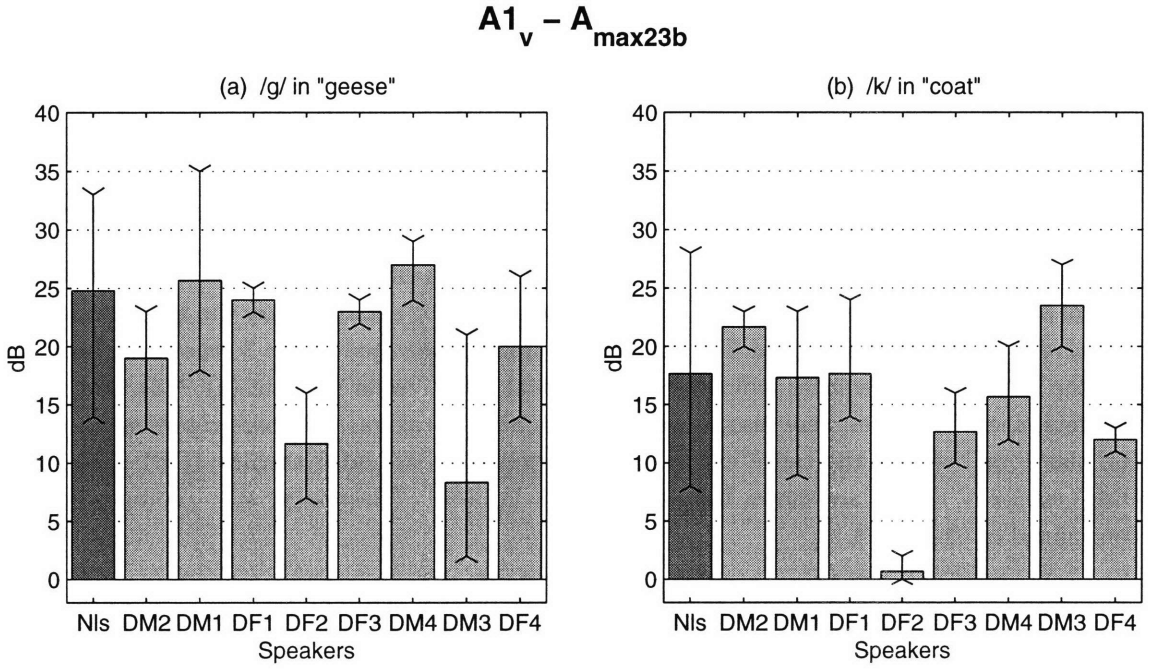

Figure 6-30 : $A1_v - A_{max23b}$ Acoustic Measure. This amplitude difference is a measure of air pressure control in word-initial velar stop consonants. For details of how the measurement was made, refer to Sections 6.1.2 and 6.1.3. The measure was averaged across 3 repetitions/utterance for the velar stops in the utterances (a) geese and (b) coat. For normal speakers, the measure was also averaged across all 8 speakers. The bars represent the mean, and the error bars are the range extrema. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

Each of the measures $A1_v - A1_p$, $A1_v - A_{high}$, and $A1_v - A_{max23b}$ was observed to deviate most from normal by becoming too low. Earlier in this section, the implications of increasing $A1_p$, A_{high} and A_{max23b} were discussed. It is also possible that the decrease in these amplitude differences can be explained by a decrease in $A1_v$. Two different mechanisms have been developed to explain such a decrease. One or both of these mechanisms may occur for a given dysarthric speaker. The first mechanism is inadequate maintenance of the subglottal pressure, P_s, throughout the utterance. There is a high enough P_s at the time of the burst, but by the time the vowel is reached, P_s has decreased because not enough energy is stored in the expanded thorax or depressed diaphragm and the respiratory musculature is insufficient or not adequately recruited to provide the necessary airflow (see Section 3.1.1). This mechanism is more likely for voiceless stops than for voiced stops, due to a longer VOT (or even a prolonged VOT, in the case of some of these dysarthric speakers), during which P_s can decrease. (The definition of VIT results in a longer VOT for voiceless stops than voiced stops as well.) This mechanism is essentially saying that the air pressure in the lungs at the level of the alveoli, P_{alv}, is sufficient at the time of the release, but decreases by the time of vowel onset.

The second mechanism of decreasing $A1_v$ is based on sufficient P_{alv} throughout the utterance, but the airflow is so high following the release that the pressure drop across the lungs results in a decrease in P_s by the time of vowel onset. This mechanism is more likely to occur for voiceless than for voiced stops as well, due to the need to produce more noise and the adducted position of the vocal folds following a voiceless stop release. It may be possible that a higher airflow is seen for voiced stops as well, for example, in the presence of a faulty velopharyngeal port.

6.3 Conclusions

The normal data are in general agreement with published data for normal speakers. Labial and alveolar stops are separated from one another through the combination of $A1_v - A_{high}$ and $A_{high} - A_{low}$ measures. Normal speakers produce only one stop

burst for labial and alveolar stops, but may produce up to two sequential stop bursts for velar stops. The average quantity of prevoicing was not found to depend on the place of articulation. The measure of VOT is somewhat longer than the VOT values reported in other studies, particularly for voiceless stops. This longer VOT duration is attributed to the manner in which VIT is defined, especially with regard to "complete" and "incomplete" glottal pulses. Noise production following voiceless stop release results in more "incomplete" initial glottal pulses for the vowel and a later VIT, lengthening the VOT. The results of the $F0$ ratio indicate a higher $F0$ value for voiceless than voiced stops, consistent with the stiffer vocal-fold position required for the voiceless stop and coarticulatory effects of that vocal-fold position on the following vowel. Measures of air pressure control for labial and alveolar stops ($A1_v - A_{high}$) and velar stops ($A1_v - A_{max23b}$) reflect an intraoral pressure differential between voiced and voiceless stops. The values of $A1_v - A_{high}$ and $A1_v - A_{max23b}$ are higher for voiced stops than voiceless stops, indicating that intraoral pressure is lower for the voiced stops.

A summary of the individual-speaker observations for dysarthric speakers across acoustic measures, combining information from perceptual evaluations (Chapter 4) and spectrogram analysis (Chapter 5) appears in Chapter 7. In that chapter, deviations from normal observed in the placement and rate of movement of the primary articulator, the laryngeal system, and the respiratory system are summarized for each speaker. Additionally, the relationship of some of the findings to the type of dysarthria of the individual will be discussed briefly.

Observations can be made across dysarthric speakers as follows. Table 6.1 includes a list of those acoustic measures found to best correspond to the spectrogram analysis results (Chapter 5) and/or individual questions from the perceptual evaluations (Chapter 4). In addition to these findings, the measure $A1_v - A_{high}$ is lower for production of /p/ for several of the dysarthric speakers (the dysarthric and normal ranges do not overlap). Hypotheses to explain this finding include that A_{high} is too high due to increased intraoral pressure (ejective formation), or $A1_v$ is too low due to either increased airflow immediately following the stop release or lack of sufficient

inspiration and/or respiratory musculature recruitment to maintain adequate breath support at vowel onset. Although some of the acoustic measures, such as $A1_v - A_{high}$, $A1_v - A_{max23b}$, and $A1_v - A1_p$, showed deviations from normal (parameter ranges for some of the dysarthric speakers did not overlap the normal ranges), these measures did not track the stop goodness score well. The only acoustic measure to correspond to a noticeable degree with the stop goodness score is the deviation from normal of the average VOT for voiceless stops.

Acoustic Measures	**Spectrogram Analysis**	**Perceptual Evaluations**
$A1_v - A1_p$ (Voiced)	Prevoicing	Q1
VOT, average deviation from normal (Voiced)	VOT (Voiced)	Q2 (Voiced)
VOT, average deviation from normal (Voiceless)	VOT (Voiceless)	Q2 (Voiceless)
Instances When Avg. No. of stop bursts ≥ 2	Abruptness of Release	—

Table 6.1 : Correspondences observed between acoustic measures, spectrogram attributes (Chapter 5) and individual questions from perceptual evaluations (Chapter 4).

Other measures were thought to be more likely to affect stop production quality judgments (Question 5 of the perceptual test in Chapter 4), such as the acoustic measure $A1_v - A_{high}$ in combination with $A_{high} - A_{low}$; the F0 ratio; and the value of $F2$ in the vowel-onset spectrum as well as the $F1$ and $F2$ formant-frequency transitions (measured only for the three speakers with highest stop goodness scores).

The acoustic measures generated a series of testable hypotheses. One of the measures (deviation from normal of average VOT for voiceless stops) was also observed to correspond to the stop goodness score. Overall, however, it was concluded that none of the acoustic measures, either singly or as a group, was able to capture and quantify most of what the listeners perceived during the perceptual experiment. The acoustic analysis performed in this thesis generally made measurements on short time durations ($\leq$ 100 msec). The amplitude difference measurements also had underly-

ing assumptions about the behavior of the respiratory system, based on observations from normal speech. It was discovered, however, that in the speech of these dysarthric speakers, the respiratory function could also vary, resulting in amplitude measurements that could indicate problems with more than one aspect of the speech system. For example, the measure $A_{high} - A_{low}$ identifies the labial or alveolar place of articulation for normal speakers. For dysarthric speakers, A_{high} can also vary with intraoral pressure (or air leakage through the nasal passageways). As another example, the measure $A1_v - A_{high}$ for normal speakers reveals an intraoral pressure difference between voiced and voiceless stop production. For dysarthric speakers, A_{high} can vary as discussed above, and $A1_v$ can vary with changes in subglottal pressure due to poor inspiration, poor respiratory support or increased airflow. One way of viewing dysarthric stop production is as the "superposition" of several slowly time-varying subsystems, the respiratory system, the laryngeal system, articulatory movements in the oral passageways and the velopharyngeal port leading to the nasal passageways. Since all of these subsystems vary over long time durations ($>$ 100 msec, generally), evaluation of these systems lends itself to a visual inspection and interpretation of the spectrograms. Spectrogram analysis can determine ways in which normal and dysarthric speech differ without relying on certain aspects of the system, such as respiration, to behave normally while other aspects are perturbed. The results of spectrogram analysis were discussed in Chapter 5. The spectrogram analysis of Chapter 5 was performed chronologically *after* the acoustic analysis of the present chapter, for the reasons discussed here.

6.4 Summary

Section 6.1 contains descriptions of the data utilized in the acoustic analysis of this chapter and the development of the acoustic measures used to analyze that data. The data analyzed in this chapter (Sect. 6.1.1) are the same as the data analyzed in the perceptual evaluations of Chapter 4 and the spectrogram analysis (SA) of Chapter 5. In order to perform the acoustic analysis, the data were first processed in the time

and frequency domains (Sect. 6.1.2). In the time domain, the times of the stop-consonant release and vowel onset were identified. In the frequency domain, a series of three average spectra were created prior to the stop release, at the stop release, and after vowel onset. Additionally, three individual spectra were created at and after vowel onset. With the aid of these times and spectra, several acoustic measures were developed, reflecting various aspects of the speech system during stop-consonant production (Sect. 6.1.3). These aspects are the placement and rate of movement of the primary articulator, the laryngeal system and the respiratory system. The acoustic measures include assessments of stop burst tilt and amplitude; formant-frequency transitions; the number of sequential stop bursts in a given repetition; the duration and amplitude of prevoicing, (voicing preceding the stop release); voice onset time (VOT); and fundamental frequency (F0).

Section 6.2 contains the results of performing the acoustic measures on the normal and dysarthric data. In Sect. 6.2.1 the application of the measures to the normal data is described. The mean, standard deviation of the mean, and range of the normal results are provided. These results serve as a baseline for comparison with the dysarthric data results in Section 6.2.2. Section 6.2.2 contains results and discussion of the application of the acoustic measures to the speech of dysarthric individuals. The results of each measure are discussed, and hypotheses are developed to explain some of the differences observed in the speech of normal and dysarthric speakers. Some of the acoustic measures deviated noticeably from normal for some of the dysarthric speakers (parameter range values did not overlap normal ranges), including measures such as $A1_v - A_{high}$, $A1_v - A_{max23b}$, and $A1_v - A1_p$. The acoustic measure results are also compared to the perceptual evaluations (Chapter 4) and the spectrogram attribute results (Chapter 5). Some measures, such as VOT, $A1_v - A1_p$ and the number of consecutive stop bursts, correspond to perceptual and (or spectrographic data observations, while most of the remaining measures may contribute to quality judgments in the perceptual data. Only one of the measures (of voiceless VOT) had a good correspondence to the stop goodness score. Based on observations that several subsystems can vary simultaneously in dysarthric speech (such as the respiratory

system, laryngeal system, and articulatory system) over long durations (> 100 msec), visual inspection of spectrograms was next pursued, in an attempt to further capture and quantify what the listeners heard during the perceptual experiment.

Chapter 7

Dysarthric Speaker Observations

This chapter discusses some of the more prominent findings for each individual dysarthric speaker. Results from perceptual evaluations, spectrogram analysis attribute ratings and acoustic measures are interrelated for each speaker. Several aspects of the speech system are addressed, including the placement of the primary articulator, the rate of primary articulator movement following the stop release, the laryngeal system and, to a lesser extent, the respiratory system. This discussion is not intended to be a comprehensive evaluation of each speaker, but rather includes highlights of some of the more salient observations made from the results of Chapters 4, 5 and 6. Brief mention will be made of the type of speaker dysarthria from Chapter 2. Where possible, a list of dysarthric characteristics which are in agreement with the findings reported in this thesis will be provided.

7.1 Assessment of Individual Dysarthric Speakers

This section consists of one subsection for each dysarthric speaker. The subsections are in order of decreasing stop goodness score for the speakers. Each subsection contains discussion of the prominent findings for that particular speaker. A series of graphs is also included to serve as an overview. These graphs show the noteworthy mean results for the given dysarthric speaker as well as the mean of the eight normal speakers. For the acoustic measures, the error bars indicate the range, either for

the dysarthric individual or across all eight normal speakers, respectively. The stop goodness score from Figure 4-2, is included for every speaker. The remaining graph selection is tailored to the particular dysarthric speaker, with only those results most deviant from normal shown. In the case of the spectrogram analysis (SA), only attributes with ratings > 1.5 are included, with the exception of SA Prevoicing of voiceless stops (included ratings > 1.2) and SA VOT for voiced stops (included ratings > 1.3), for which normal speakers virtually never deviate from 1.0. The SA attribute results are taken from Figures 5-8 to 5-14.

Only some of the acoustic measure results are shown graphically in each subsection. These measures are $A1_v - A1_p$ (Fig. 6-24 and 6-25), VOT (Fig. 6-27), $A1_v - A_{high}$ (Fig. 6-29) and $A1_v - A_{max23b}$ (Fig. 6-30). These measures are primarily associated with the laryngeal and respiratory systems, although $A1_v - A_{high}$ reveals place of articulation information as well. Results are shown for an individual if they deviate notably from normal, typically when the range for the dysarthric speaker does not overlap the normal range, but occasionally when the average value differs notably from normal and/or the range of variation is large for the dysarthric speaker compared to the normal range.

The ability to establish a relationship between a given speaker's perceptual and acoustic findings and their type of dysarthria is confounded by the lack of complete medical histories, including speech-language pathologist evaluations and neurologic assessments. As discussed in Chapter 2, the speech characteristics associated with each type of dysarthria are broad, and the particular characteristics displayed by each dysarthric individual in this study are largely unknown. Furthermore, the severity of the dysarthria, which can influence both the number of sequelae exhibited as well as the degree to which the sequelae affect stop production, is not known for the speakers. To the extent possible, each subsection contains information from the medical history pertinent to the auditory-perceptual and acoustic findings, as well as some of the characteristics of the particular type of dysarthria which are in agreement with the findings.

In summary, each speaker's subsection, findings related to the placement of the

primary articulator, the rate of primary articulator movement following the stop release, the laryngeal system, and the respiratory system are outlined. When pertinent, information about the functioning of the velopharyngeal port is also included. The relevant information from the individual's medical history (Sect. 2.2) appears next. Lastly, information about that particular individual's type of dysarthria appears from Section 2.1.

7.1.1 Subject DM2

- **Placement of primary articulator:** appears normal
- **Rate of primary articulator movement:** appears normal
- **Laryngeal/Respiratory system:** (1) some evidence of abnormal prevoicing based on $A1_v - A1_p$; (2) excessive aspiration and/or air pressure control difficulties based on Spectrogram Analysis (SA) - Time Course of Release attribute and $A1_v - A_{high}$
- **Medical history:** irregularities of loudness
- **Ataxic Dysarthria:** increased variability, inconsistency or instability of intensity

Refer to Figure 7-1 for some of the notable deviations from normal depicted graphically.

7.1.2 Subject DM1

- **Placement of primary articulator:** some evidence that alveolar constrictions are either made slightly longer by utilizing tongue tip and some of tongue body to form constriction or tongue body is more fronted, based on initial value of $F2$ following release (SA - Time Course of $F2$ Change for voiced stops)

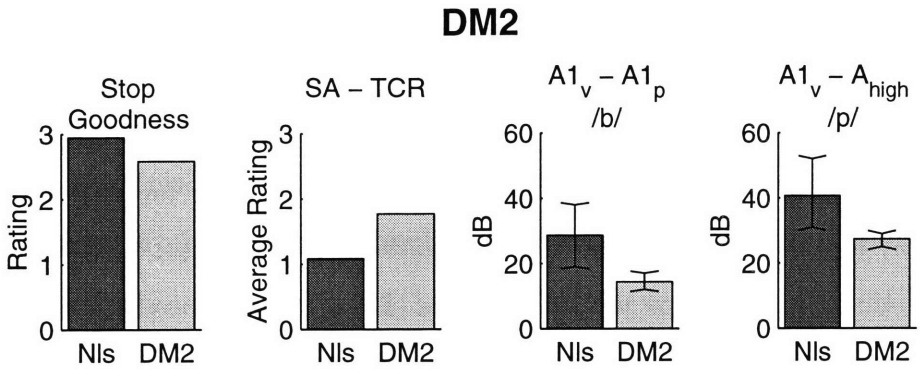

Figure 7-1 : Notable results for dysarthric male speaker DM2, compared to normal (Nls). From left to right, the following results are shown: stop goodness score from perceptual evaluations; spectrogram analysis attribute ratings - time course of release; acoustic measures $A1_v - A1_p$ for /b/, and $A1_v - A_{high}$ for /p/. The bars represent the mean, and the error bars are the range extrema. Refer to the text for references to the figures from which these results were obtained.

- **Rate of primary articulator movement:** some indication of slower rate, based on lower value of $F1$ at vowel onset (SA - Time Course of $F1$ Rise for /b,d/) and presence of multiple stop bursts

- **Laryngeal/Respiratory system:** (1) some evidence of abnormal prevoicing based on $A1_v - A1_p$; (2) excessive aspiration and/or air pressure control difficulties based on Spectrogram Analysis (SA) - Time Course of Release attribute and $A1_v - A_{high}$

- **Medical history:** no information provided

- **Spastic Dysarthria:** shallow breathing, decreased vocal fold abduction during respiration, hyperadduction of true and false vocal folds during speech, reduced speed of tongue movement, reduced acceleration and deceleration of articulators

Refer to Figure 7-1 for some of the notable deviations from normal depicted graphically.

7.1.3 Subject DF1

- **Placement of primary articulator:** appears normal

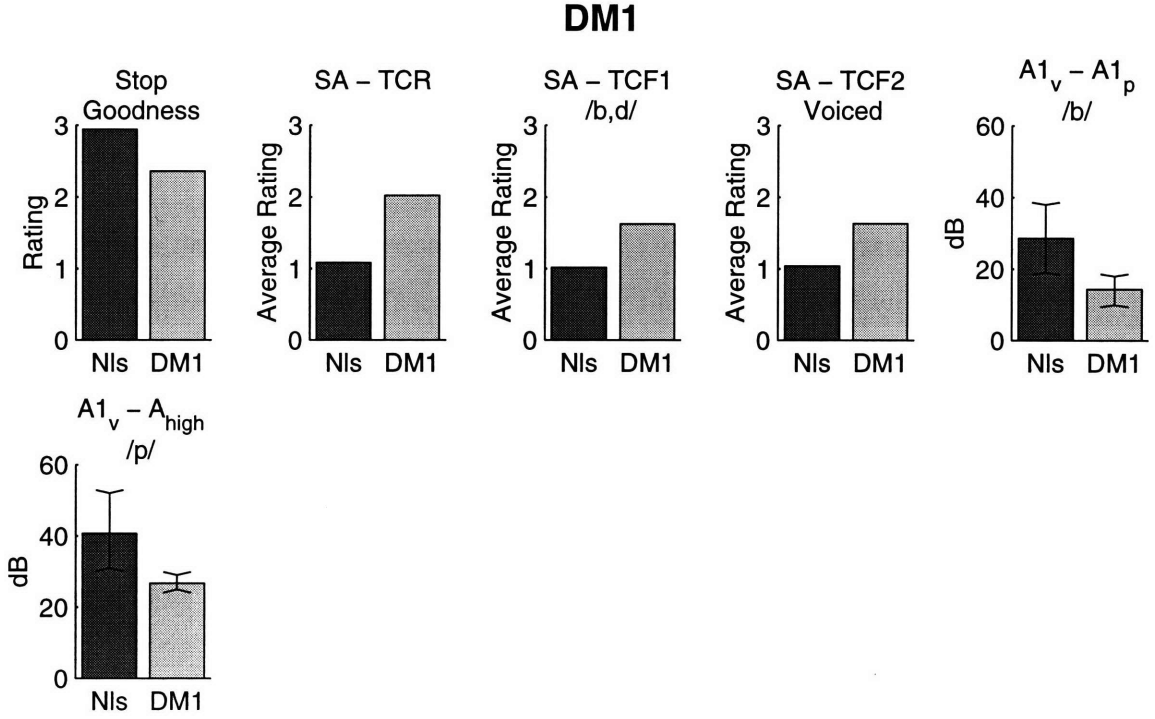

Figure 7-2 : Notable results for dysarthric male speaker DM1, compared to normal (Nls). From left to right, top to bottom, the following results are shown: stop goodness score from perceptual evaluations; spectrogram analysis attribute ratings - time course of release, time course of $F1$ rise for /b,d/, time course of $F2$ change for voiced stops; acoustic measures $A1_v - A1_p$ for /b/, and $A1_v - A_{high}$ for /p/. The bars represent the mean, and the error bars are the range extrema. Refer to the text for references to the figures from which these results were obtained.

- **Rate of primary articulator movement:** some indication of slower rate, based on lower value of $F1$ at vowel onset (SA - Time Course of $F1$ rise for /b,d/)

- **Laryngeal/Respiratory system:** (1) prevoicing too loud and too long preceding voiced stops, based on SA - Prevoicing prior to voiced stops and $A1_v - A1_p$; (2) excessive aspiration and/or air pressure control difficulties based on SA - Time Course of Release attribute and $A1_v - A_{high}$; (3) average VOT too short for voiced stops, based on the VOT acoustic measure for voiced stops

- **Medical history:** airflow and lung vital capacity control problems

- **Spastic Dysarthria:** shallow breathing, decreased vocal fold abduction during respiration, hyperadduction of true and false vocal folds during speech, reduced speed of tongue movement, reduced acceleration and deceleration of articulators, reduced VOT for stops

Refer to Figure 7-1 for some of the notable deviations from normal depicted graphically.

7.1.4 Subject DF2

- **Placement of primary articulator:** appears normal

- **Rate of primary articulator movement:** possible slower rate (based on presence of multiple bursts)

- **Laryngeal/Respiratory system:** (1) prevoicing before voiced and voiceless stops, based on SA - Prevoicing for Voiced and Voiceless stops and $A1_v - A1_p$; (2) possible inadvertent vowel generation prior to release, based on SA - Precursor; (3) possible excessive aspiration and/or air pressure control difficulties, based on SA - Time Course of Release and Voice Onset Time (VOT) for voiceless stops, $A1_v - A_{high}$ and $A1_v - A_{max23b}$

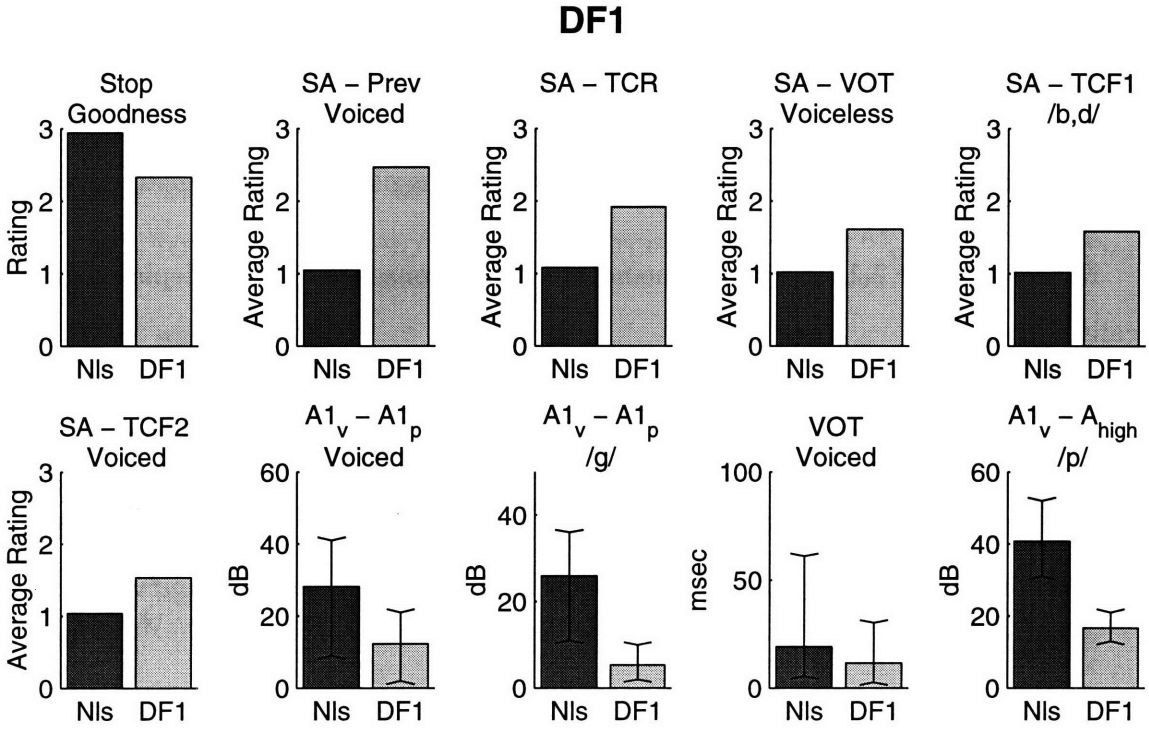

Figure 7-3 : Notable results for dysarthric female speaker DF1, compared to normal (Nls). From left to right, top to bottom, the following results are shown: stop goodness score from perceptual evaluations; spectrogram analysis attribute ratings - prevoicing for voiced stops, time course of release, voice onset time (VOT) for voiceless stops, time course of $F1$ rise for /b,d/, time course of $F2$ change for voiced stops; acoustic measures $A1_v - A1_p$ for voiced stops, $A1_v - A1_p$ for /g/, VOT for voiced stops, and $A1_v - A_{high}$ for /p/. The bars represent the mean, and the error bars are the range extrema. Refer to the text for references to the figures from which these results were obtained.

- **Velopharyngeal port:** incomplete velopharyngeal closure, based on SA - Precursor, Abruptness of Release, Time Course of Release and Time Course of $F2$ Change for Voiced Stops attributes

- **Medical history:** speech is weak sounding, lisping, poor aspiration control, some utterances generated with breathy and explosive noise

- **Spastic Dysarthria:** shallow breathing, decreased vocal fold abduction during respiration, hyperadduction of true and false vocal folds during speech, reduced speed of tongue movement, reduced acceleration and deceleration of articulators, incomplete velopharyngeal closure, slow and sluggish velopharyngeal movement

Refer to Figure 7-1 for some of the notable deviations from normal depicted graphically.

7.1.5 Subject DF3

- **Placement of primary articulator:** alveolar stops produced as velars, based on results for perceptual test Q3 (contributes to stop goodness) and $A1_v - A_{high}$ for /t/

- **Rate of primary articulator movement:** potentially slower rate due to deviant $F1$ rise and multiple bursts (SA - Time Course of $F1$ Rise for /b,d/ stops and Abruptness of Release)

- **Laryngeal/Respiratory system:** (1) lengthens voiced VOT, based on SA - VOT for Voiced Stops, (2) shortens voiceless VOT, based on SA - VOT for voiceless stops

- **Medical history:** involuntary movements of tongue, sudden changes in airflow due to irregular spasmodic contractions of diaphragm and other respiratory muscles, large range of jaw movement, each word is prolonged, speech is weak sounding

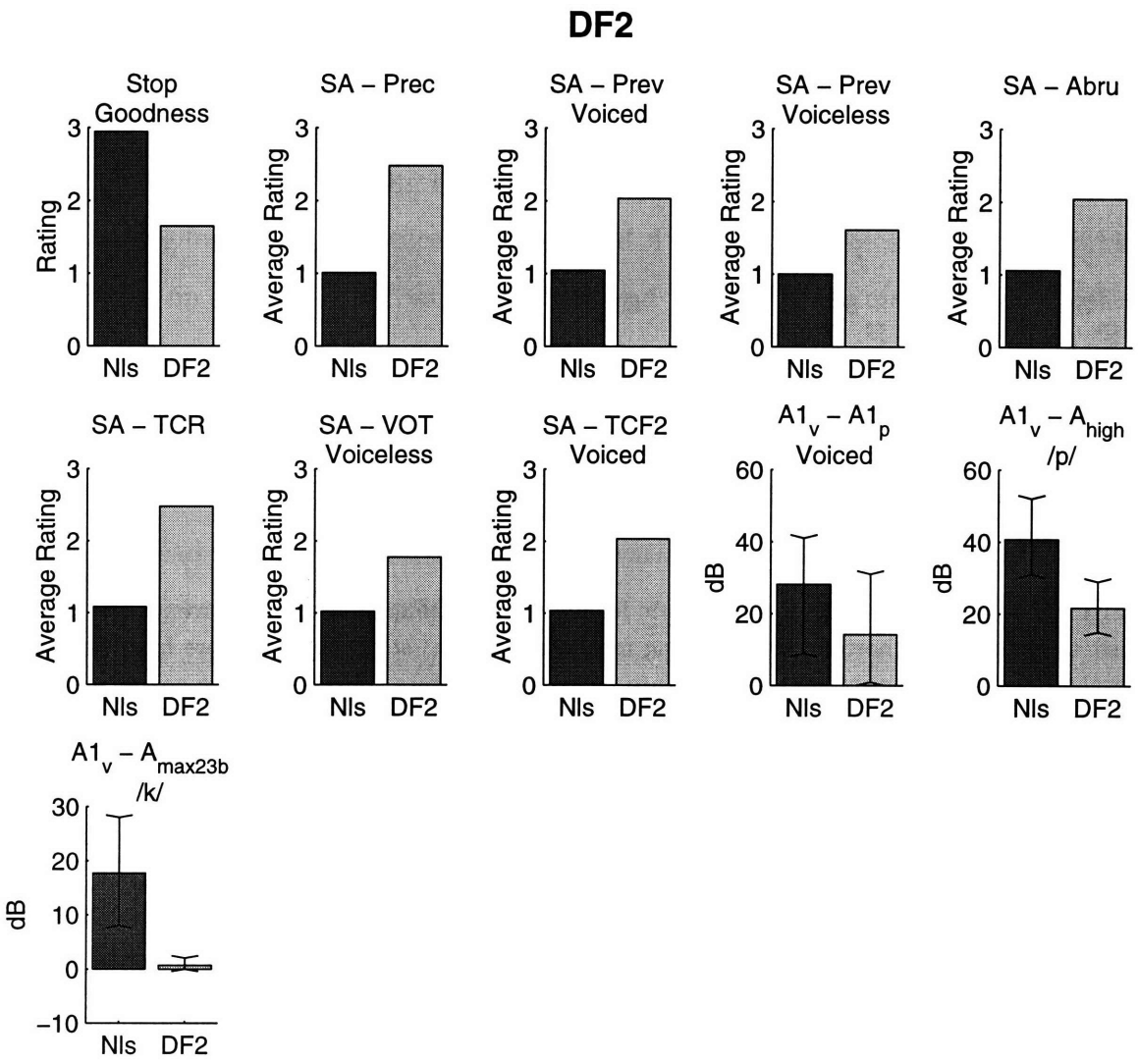

Figure 7-4 : Notable results for dysarthric female speaker DF2, compared to normal (Nls). From left to right, top to bottom, the following results are shown: stop goodness score from perceptual evaluations; spectrogram analysis attribute ratings - precursor, prevoicing for voiced stops, prevoicing for voiceless stops, abruptness of release, time course of release, voice onset time for voiceless stops, time course of $F2$ change for voiced stops; acoustic measures $A1_v - A1_p$ for voiced stops, $A1_v - A_{high}$ for /p/, and $A1_v - A_{max23b}$. The bars represent the mean, and the error bars are the range extrema. Refer to the text for references to the figures from which these results were obtained.

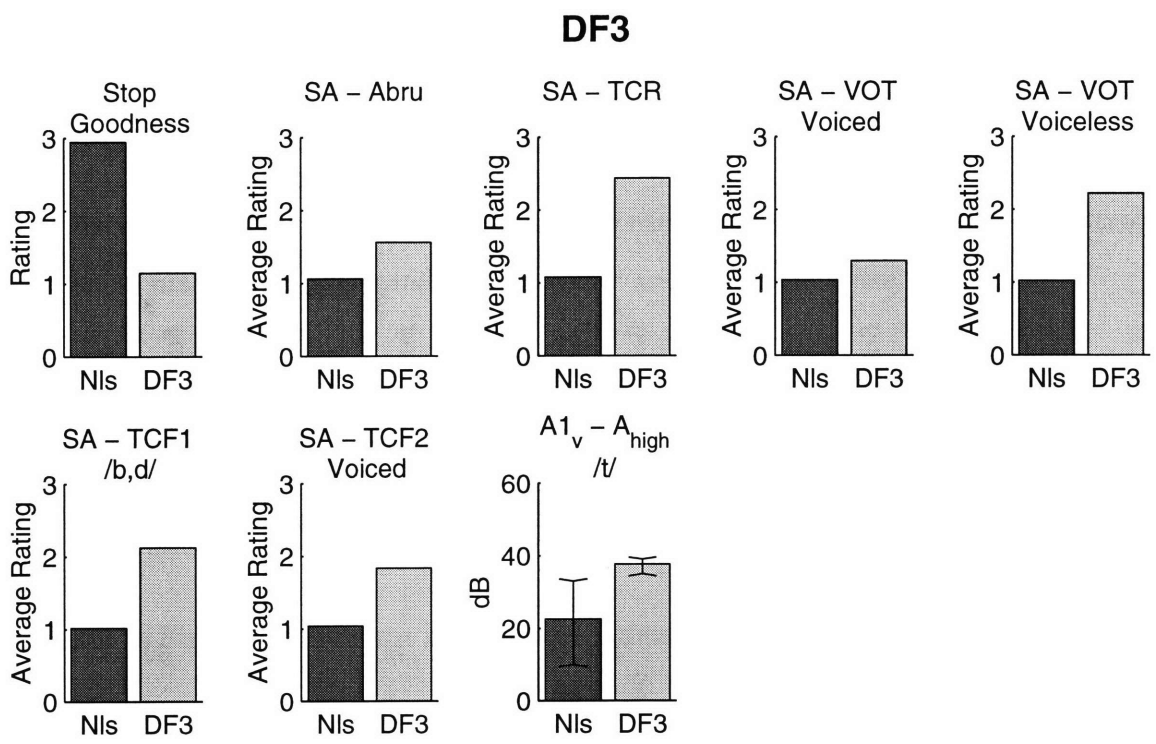

Figure 7-5 : Notable results for dysarthric female speaker DF3, compared to normals (Nls). From left to right, top to bottom, the following results are shown: stop goodness score from perceptual evaluations; spectrogram analysis attribute ratings - abruptness of release, time course of release, voice onset time (VOT) for voiced stops, VOT for voiceless stops, time course of $F1$ rise for /b,d/, time course of $F2$ change for voiced stops; acoustic measure $A1_v - A_{high}$ for /t/. The bars represent the mean, and the error bars are the range extrema. Refer to the text for references to the figures from which these results were obtained.

- **Spastic and Athetoid Dysarthria:** shallow breathing, decreased vocal fold abduction during respiration, hyperadduction of true and false vocal folds during speech, reduced speed of tongue movement, reduced acceleration and deceleration of articulators, increased subglottal air pressure, shortened VOT of voiceless stops, breathy voice quality, prolonged transitions between articulatory movements

Refer to Figure 7-1 for some of the notable deviations from normal depicted graphically.

7.1.6 Subject DM4

- **Placement of primary articulator:** difficulty producing labial and velar

stops, based on perceptual test Q3 (contributes to stop goodness)

- **Rate of primary articulator movement:** $F1$ and $F2$ transitions deviate from normal, based on SA - Time Courses of $F1$ Rise and $F2$ Change

- **Laryngeal/Respiratory system:** (1) noise production precedes voiceless stops, based on SA - Precursor and Q1 of perceptual test (contributes to stop goodness); (2) prevoicing before voiced and voiceless stops, based on SA - Prevoicing for voiced and voiceless stops, and $A1_v - A1_p$; (3) voiceless VOT variable, based on SA - VOT for voiceless stops and VOT acoustic measure for voiceless stops; (4) excessive frication and/or aspiration noise generation, based on SA - Time Course of Release

- **Medical history:** poor respiratory control, forced quality, large range of jaw and head movements, particularly time variant, speech pattern and rate change greatly between utterances

- **Spastic and Athetoid Dysarthria:** shallow breathing, decreased vocal fold abduction during respiration, hyperadduction of true and false vocal folds during speech, reduced speed of tongue movement, reduced acceleration and deceleration of articulators, increased subglottal air pressure, shorten VOT of voiceless stops, breathy voice quality, prolonged transitions between articulatory movements

Refer to Figure 7-1 for some of the notable deviations from normal depicted graphically.

7.1.7 Subject DM3

- **Placement of primary articulator:** inconsistent placement across all stops, voiceless stops typically replaced with glottal stop or vowel, based on perceptual test Q3, SA - Time Course of $F1$ Rise and $F2$ Change

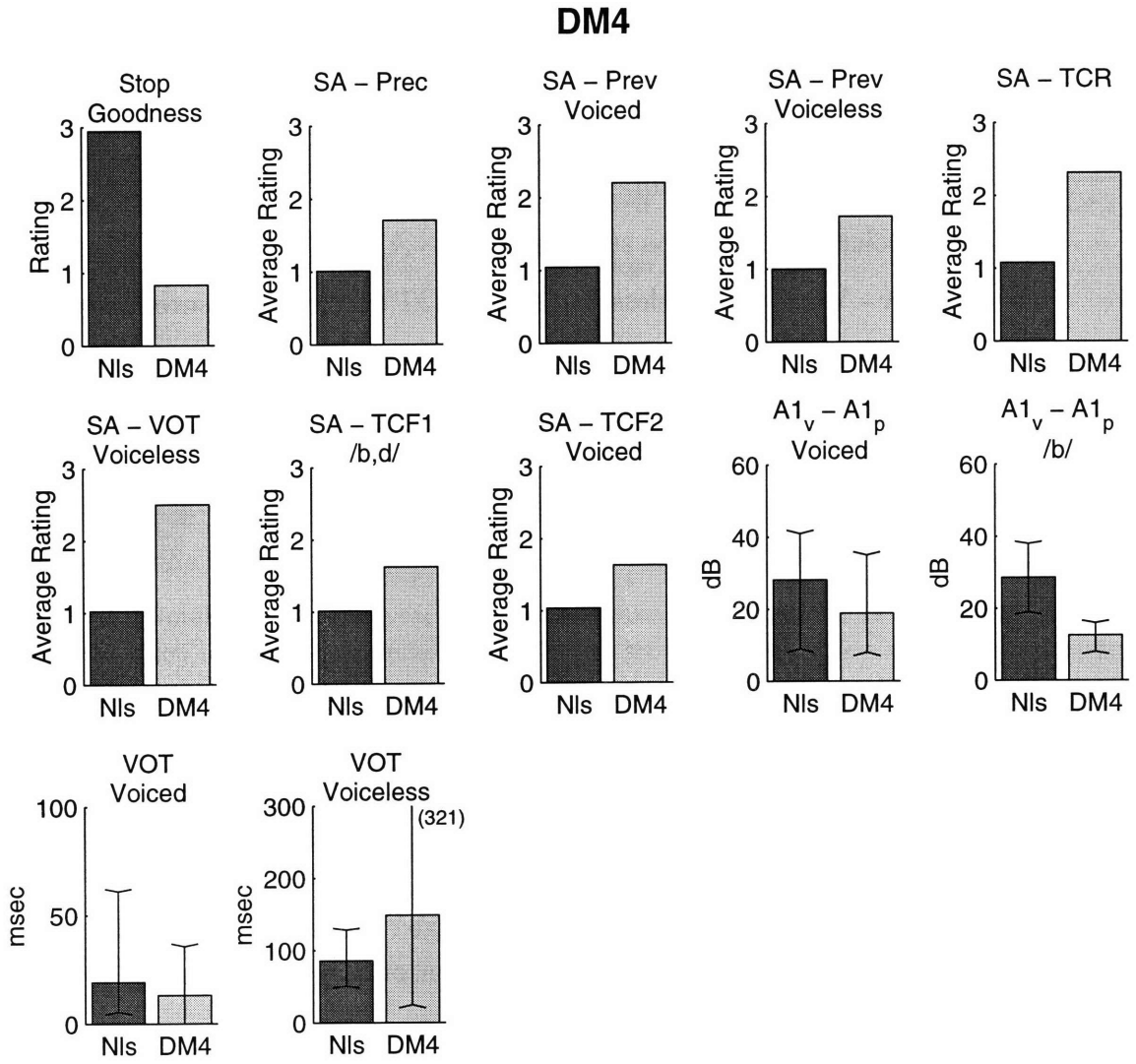

Figure 7-6 : Notable results for dysarthric male speaker DM4, compared to normal (Nls). From left to right, top to bottom, the following results are shown: stop goodness score from perceptual evaluations; spectrogram analysis attribute ratings - precursor, prevoicing for voiced stops, prevoicing for voiceless stops, time course of release, voice onset time (VOT) for voiceless stops, time course of $F1$ rise for /b,d/, time course of $F2$ change for voiced stops; acoustic measures $A1_v - A1_p$ for voiced stops, $A1_v - A1_p$ for /b/, VOT for voiced stops and VOT for voiceless stops. The bars represent the mean, and the error bars are the range extrema. Refer to the text for references to the figures from which these results were obtained.

- **Rate of primary articulator movement:** some evidence for slower rate, based on multiple bursts for 3 of 6 stops and SA - Abruptness of Release

- **Laryngeal/Respiratory system:** (1) shortens voiceless VOT, based on SA - VOT for Voiceless stops and VOT acoustic measure for voiceless stops; (2) lengthens voiced VOT, based on SA - VOT for voiced stops and VOT acoustic measure for voiced stops; (3) evidence of abnormal prevoicing, based on $A1_v - A1_p$; (4) excessive aspiration and/or air pressure control difficulties, based on SA - Time Course of Release and $A1_v - A_{high}$ for /b,p/

- **Medical history:** poor respiratory control, forced quality, large range of jaw movements, severely reduced oral-articulatory abilities, breathiness, whispered and hoarse phonations, intermittent aphonia, throaty noise

- **Athetoid Dysarthria:** increased subglottal pressure, forced, breathy quality, lack of phonation; when phonation does occur, have initial audible glottal attack, inappropriate tongue positioning, inability to finely shape tongue for consonant articulation, prolonged transition times.

Refer to Figure 7-1 for some of the notable deviations from normal depicted graphically.

7.1.8 Subject DF4

- **Placement of primary articulator:** difficulty producing alveolar and velar stops, and difficulty forming complete vocal-tract closure, based on perceptual test Q3 and SA - Abruptness of Release, Time Course of Release attributes

- **Rate of primary articulator movement:** (1) $F1$ and $F2$ transitions deviate from normal, based on SA - Time Courses of $F1$ Rise and $F2$ Change; (2) may have difficulty moving primary articulator rapidly following release, based on SA - Abruptness of Release and Time Course of Release

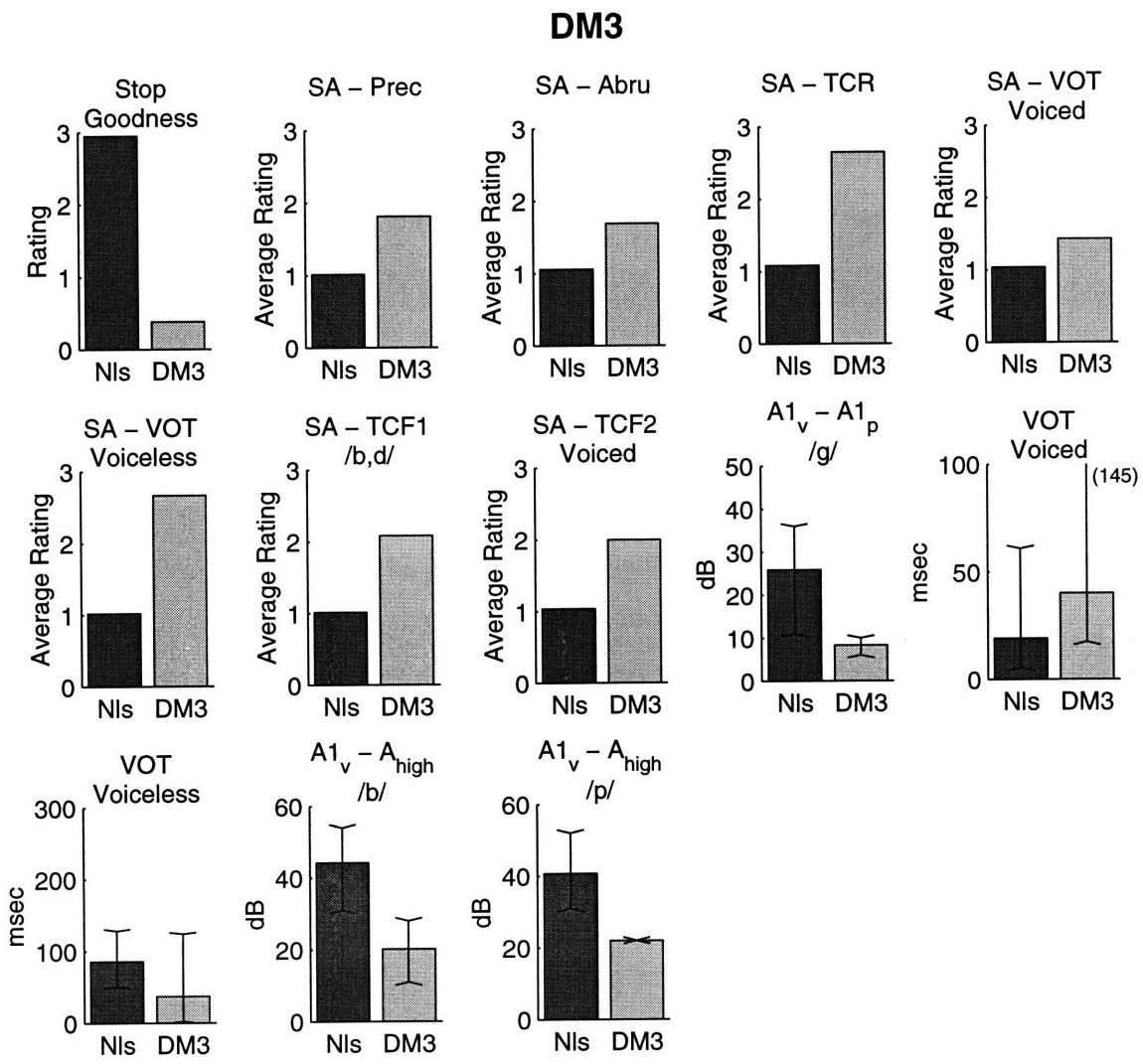

Figure 7-7 : Notable results for dysarthric male speaker DM3, compared to normal (Nls). From left to right, top to bottom, the following results are shown: stop goodness score from perceptual evaluations; spectrogram analysis attribute ratings - precursor, abruptness of release, time course of release, voice onset time (VOT) for voiced stops, VOT for voiceless stops, time course of $F1$ rise for /b,d/, time course of $F2$ change for voiced stops; acoustic measures $A1_v - A1_p$ for /g/, VOT for voiced stops, VOT for voiceless stops, $A1_v - A_{high}$ for /b/ and $A1_v - A_{high}$ for /p/. The bars represent the mean, and the error bars are the range extrema. Refer to the text for references to the figures from which these results were obtained.

- **Laryngeal/Respiratory system:** (1) noise precedes voiceless stops and vocalizations precede voiced stops, based on SA - Precursor; (2) prevoices, based on SA - Prevoicing for voiced and voiceless stops; (3) voiceless VOT too long, based on VOT acoustic measure for voiceless stops

- **Medical history:** paralysis of left side of face, left side of tongue and left vocal fold, poor aspiration control with some breathy and explosive noise, weak sounding

- **Unclassified Dysarthria:** not available.

Refer to Figure 7-1 for some of the notable deviations from normal depicted graphically.

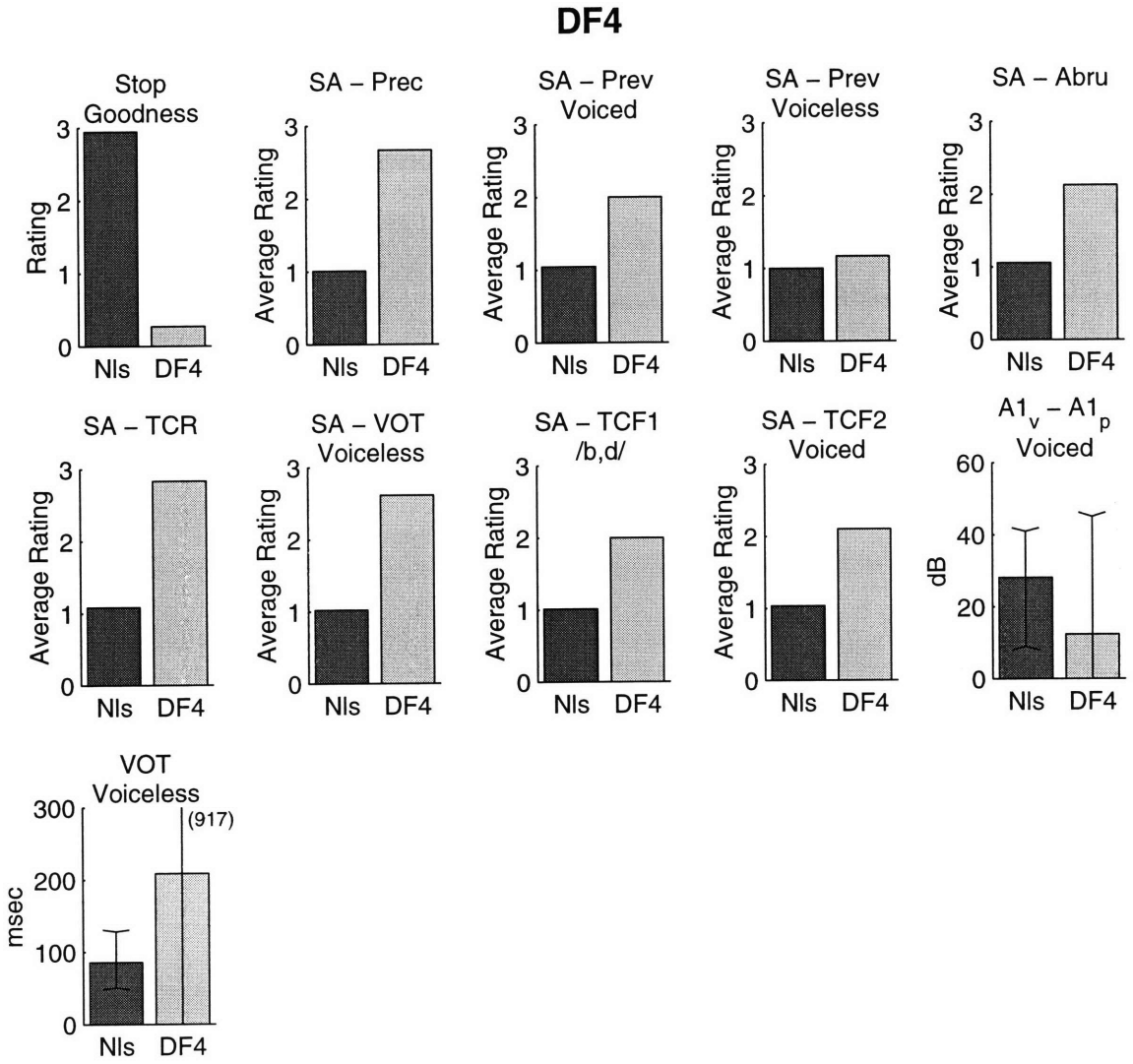

Figure 7-8 : Notable results for dysarthric female speaker DF4, compared to normal (Nls). From left to right, top to bottom, the following results are shown: stop goodness score from perceptual evaluations; spectrogram analysis attribute ratings - precursor, prevoicing for voiced stops, prevoicing for voiceless stops, abruptness of release, time course of release, voice onset time (VOT) for voiceless stops, time course of $F1$ rise for /b,d/, time course of $F2$ change for voiced stops; acoustic measures $A1_v - A1_p$ for voiced stops and VOT for voiceless stops. The bars represent the mean, and the error bars are the range extrema. Refer to the text for references to the figures from which these results were obtained.

7.2 Summary

This chapter consists of assessments of each of the individual dysarthric speakers. Deviations from normal noted in the results of the perceptual evaluations, spectrogram analysis and acoustic analysis from previous chapters are presented in terms of their effects on the placement of the primary articulator, the rate of movement of the primary articulator, the laryngeal system and the respiratory system for each speaker. Where possible, relevant information from the subjects' medical histories is presented. Deviant speech characteristics of the specific type of dysarthria exhibited by the subject are also included in each discussion.

The information presented in this chapter is an initial step toward integrating subjective and objective measures to provide a more complete picture of the way(s) in which a given dysarthric subject's speech deviates from normal. Future applications of assessments of this type include supplementing auditory-perceptual evaluations and establishing a baseline for longitudinal speech comparison.

Chapter 8

Conclusion

8.1 Summary of Results

An auditory-perceptual experiment was designed to evaluate several aspects of stop production, including the presence of a precursor (subject-generated sound prior to the stop release), voicing, place and manner of articulation of the stop; and the quality of the stop production. The primary outcome of this experiment was the development of the stop "goodness" score, a single number for a given dysarthric speaker reflecting listener responses to all the aspects of stop production. The stop goodness score answers the question, "How well is the correctly-identified stop produced?" Values were assigned to the response as follows: Good = 3, Fair = 2, Poor = 1, and, if the stop had been incorrectly produced originally, a value of 0. The average of these values across all word repetitions is the stop goodness score.

The stop goodness score was better able to distinguish all dysarthric speakers from normal than stop intelligibility, which consisted of correctly answering only type of voicing, place and manner of articulation. These results indicate that, at least for some dysarthric speakers, there are aspects of stop production which are still not normal even when the stop consonant itself is identified correctly by the listeners.

Acoustic measures were developed, based on models of normal stop-consonant production. When applied to normal data, the results are as follows. Labial and alveolar stops are separated from one another through the combination of $A1_v - A_{high}$

and $A_{high} - A_{low}$ measures. Normal speakers produce only one stop burst for labial and alveolar stops, but may produce up to two sequential stop bursts for velar stops. The average quantity of prevoicing was not found to depend on the place of articulation. The measure of VOT is somewhat longer than the VOT values reported in other studies, particularly for voiceless stops. This longer VOT duration is attributed to the manner in which VIT is defined, especially with regard to "complete" and "incomplete" glottal pulses. Noise production following voiceless stop release results in more "incomplete" initial glottal pulses for the vowel and a later VIT, lengthening the VOT. The results of the $F0$ ratio indicate a higher $F0$ value for voiceless than voiced stops, consistent with the stiffer vocal-fold position required for the voiceless stop and coarticulatory effects of that vocal-fold position on the following vowel. Measures of air pressure control for labial and alveolar stops ($A1_v - A_{high}$) and velar stops ($A1_v - A_{max23b}$) reflect an intraoral pressure differential between voiced and voiceless stops. The values of $A1_v - A_{high}$ and $A1_v - A_{max23b}$ are higher for voiced stops than voiceless stops, indicating that intraoral pressure is lower for the voiced stops.

When these acoustic measures were applied to the speech of the dysarthric speakers, the following observations were made. The acoustic measure $A1_v - A1_p$ corresponded to the presence of a precursor in the perceptual experiment. (For the purposes of the perceptual experiment, the precursor included abnormal prevoicing.) The VOT corresponded to the type of voicing in the perceptual experiment. The deviation from normal of the average VOT for voiceless stops also corresponded to some degree to the stop goodness score. In addition to these findings, the measure $A1_v - A_{high}$ is lower for production of /p/ for several of the dysarthric speakers (the dysarthric and normal ranges do not overlap). Hypotheses to explain this finding include that A_{high} is too high due to increased intraoral pressure (ejective formation), or $A1_v$ is too low due to either increased airflow immediately following the stop release or lack of sufficient inspiration and/or respiratory musculature recruitment to maintain adequate breath support at vowel onset. It was determined that the acoustic measures reflected not only the aspect of production measured for normal speakers, but

could also reflect other aspects of the speech system as well, such as respiration. For example, the measure $A_{high} - A_{low}$ identifies the labial or alveolar place of articulation for normal speakers. For dysarthric speakers, A_{high} can also vary with intraoral pressure (or air leakage through the nasal passageways). As another example, the measure $A1_v - A_{high}$ for normal speakers reveals an intraoral pressure difference between voiced and voiceless stop production. For dysarthric speakers, A_{high} can vary as discussed above, and $A1_v$ can vary with changes in subglottal pressure due to poor inspiration, poor respiratory support or increased airflow.

Based on the findings from the acoustic measures, a visual-perceptual assessment of spectrograms was conducted. First, seven attributes were designed to span the stop-consonant production time period: Precursor, Prevoicing (voicing preceding the stop release), Time Course of Release, Abruptness of Release, Voice Onset Time (VOT), Time Course of $F1$ Rise, and Time Course of $F2$ Change. Then, judges visually inspected and rated the spectrograms for the production of the seven attributes, Good = 1, Fair = 2 and Poor = 3. The ratings results for all attributes were found to be negatively correlated to the stop goodness score. (In other words, as the stop goodness score decreases, the attribute ratings increase, indicating poorer production.) Across all stops, Time Course of Release was found to be highly correlated with the stop goodness score. For voiceless stops, Time Course of Release and VOT were highly correlated with the goodness score. A high correlation was observed between goodness and Precursor, Abruptness of Release, Time Course of Release and Time Course of $F2$ Change for voiced stops. When velars are no longer under consideration in the voiced stops, the same group of attributes is found to be highly correlated to stop goodness, along with the additional attribute Time Course of $F1$ Rise. The results of the spectrogram analysis reveal that, at least in part, the attribute assessment has been able to capture and quantify what listeners perceived in the speech of the normal and dysarthric speakers.

8.2 Contributions

This thesis makes contributions to the following areas:

- This research represents a first step in the characterization of motor control and coordination difficulties of dysarthric speakers through the development of visual-perceptual and objective acoustic measures which reflect articulatory movements.

- Auditory-perceptual "quality" of production judgments have demonstrated usefulness in distinguishing dysarthric speakers with high word intelligibility from normal speakers.

- Visual-perceptual attributes, developed to assess various aspects of stop production, were able to capture and quantify, at least in part, what the listeners perceived in the auditory-perceptual experiment.

- Objective acoustic measures of the dysarthric speech led to testable hypotheses regarding incorrect articulatory, laryngeal and/or respiratory movements.

- The normal range of variability established for the objective, quantitative acoustic measures has potential applications to speech recognition and synthesis.

8.3 Directions for Future Research

A comparison of the results for the objective acoustic measures (Chap. 6) and the visual-perceptual assessment of spectrogram attributes (Chap. 5) leads to the conclusion that visual inspection of the spectrograms more successfully predicts the stop goodness score than the particular acoustic measures developed in Chapter 6. Although the research was conducted chronologically in the order of acoustic measure development, then spectrogram analysis (i.e., Chap. 6, then Chap. 5), it is recommended that future work occur in chapter order — spectrogram analysis then acoustic measure development. The study of dysarthric speech is more complicated than

simply applying normal measures to it. Several speech subsystems (respiratory, laryngeal, articulatory) can vary simultaneously throughout hundreds of milliseconds during stop production, even for highly-intelligible dysarthric speakers. As stated above, an approach which may lead to improved acoustic measures (improvement compared to the measures in this thesis) would be to perform the data analyses in the order provided by the thesis (perceptual evaluations, spectrogram analysis, then development of acoustic measures). This strategy may better identify those aspects of dysarthric speech which deviate from normal in a manner consistent with the stop goodness score, and facilitate development of acoustic measures reflecting this deviation.

The objective acoustic measures which were performed on the data led to hypotheses which could be tested. For example, respiratory system function could be assessed with the aid of devices that measure air pressure and airflow. Velopharyngeal port closure could be assessed by measuring airflow exiting the nose. In addition to physiologic measures, different acoustic measures could be performed to assess effects of respiration on the waveform. For example, how the amplitude of the waveform envelope for the utterance change with time could be assessed.

A perceptual experiment could be devised to determine the aspects of production that contributed to the listeners' judgments of "quality" of the stop. For example, to determine if the stop production was breathy, nasal, etc.

The study would have benefitted from the use of more phonetic environments for the stops. More repetitions could also have been analyzed. To move toward clinical applications of the research, it is recommended that all speech sounds be analyzed, not just stop consonants. Additionally, in order to aid diagnosis of the type of dysarthria, assessment of severity, and identification of the lesion location, groups of dysarthric speakers with the same type of dysarthria and similar degrees of severity should be studied.

Appendix A

Corpus

This corpus is composed of the 70 words spoken by the eight dysarthric speakers and the eight normal speakers studied in this thesis. The word list was designed by Kent et al. (1989) in the context of developing a word intelligibiility test for use in the clinical evaluation of dysarthric speakers.

ache	ease	leak	sheet
air	feed	leak	ship
at	feet	lip	shoot
ate	fill	meat	side
bad	fork	much	sigh
beat	geese	nice	sin
bill	had	pat	sink
blend	hail	pit	sip
blow	hall	rake	slip
bunch	hand	read	spit
cake	harm	reap	steak
cash	hash	rise	sticks
chair	hat	rock	tile
cheer	heat	row	wax
chop	him	see	witch
coat	hold	seed	write
dock	knew	sell	
dug	knot	sew	

Table A.1: Corpus (leak is the only word to appear twice on the list)

Appendix B

Instructions for Digitizing Speech from an Audio Cassette Tape

This appendix describes the procedure for converting analog speech data stored on an audio cassette tape to digital speech data stored on a computer. The data is converted with the aid of a VAX computer, then transferred to a DEC Alpha computer running the UNIX operating system. This set of instructions was developed with the aid of Hale Ozsoy in the spring of 1998.

The instructions are as follows.

B.1 VAX and Hardware Setup

1. Use VAX called "Nasal" in the Kassel (Computer Rm. 36-553).

2. Turn on the Shure Professional Microphone Mixer. Make sure that knobs 1, 3 and 4 are all set to zero. The knob "Master" is a coarse adjustment of the gain (sound amplitude) and knob 2 is a fine adjustment of the gain. Each of these two knobs should be set somewhere in the range 5 - 7 as an initial setting. Above the knobs, the switches labeled "Lo Cut" and "Limiter" should always be slid to the "out" position. The switch below and between knobs 1 and 2 should be at the "Osc 1" position. As an aside, it is important to note

that the quantization of the time waveform amplitude is always 16 bits (16-bit A/D board is located in the VAX computer hardware), with no opportunity to alter the number of bits while utilizing the laboratory software and hardware described in this Appendix.

3. Turn on the Realistic SA-102 Integrated Stereo Amplifier. Make sure the settings are as follows:

   - Selector on Tape
   - Tone on Hi (fully clockwise)
   - Balance at middle
   - Volume: About "8 o'clock"
   - Mono: In
   - Speakers: In - if you desire sound from speaker; Out - if you plan to use headphones. (The headphones should then be connected to the phone jack on the Realistic SA-102.)

4. On the gray metal panel, from left to right, make sure the settings are as follows:

   - "Cassette Playback Out" L port connected to "Line Input".
   - Play/Record: Initially set on "Record" to listen while digitizing. Later, will set on "Play" to playback utterances while still in record mode, i.e., to verify you've digitized the correct utterance.
   - The two filter switches should be set as follows: Dysarthric speech is routinely sampled at 16 kHz (due to significant high-frequency content), necessitating a filter cutoff frequency of 7.5 kHz. Therefore, switch A (on the left) is UP and switch B (on the right) is DOWN. Be certain you have set the filter switches correctly, as this setting is a common source of error! If you are digitizing normal speech (speech produced by an individual with no known speech or hearing difficulties), you may wish to select a different sampling rate, and corresponding filter cutoff frequency, as follows:

- To filter at 4.8 kHz (corresponding to a sampling rate of 10 kHz), switch A is UP and switch B is UP.
- To filter at 6.2 kHz (corresponding to a sampling rate of 13 kHz), switch A is DOWN and the position of switch B does not matter.

- The top port of "Line Out" connected to L port of "Cassette Record In". The bottom port of "Line Out" connected to R port of "Cassette Record In".

5. Turn on Marantz Stereo Double Cassette Deck PMD500. Insert tape into one of the tape decks. Verify that row of knobs has the following settings:

Knob	Setting
Timer/Sync Rec	off
Dolby NR	off
Balance	middle of range
Rec Level	does not matter (since you are not recording to tape)

Table B.1: Cassette Deck Knob Settings

6. Use cassette player, along with fast forward button, rewind button and word list, to locate desired position on the tape.

TROUBLESHOOTING:

If you do not hear any sound, check the following:

- Connections may have come loose, especially "Cassette/Playback Out" L to "Line Input" on gray panel.
- Play/Record switch may not be in the right position. It must be set to "Record" in order to hear sound while digitizing and set to "Play" in order to hear sound during computer playback or when playing tape just to listen to it.
- Knob settings on Shure Professional Mike Mixer. See (2).

- "Speakers" button on Realistic SA-102. See (3).

B.2 VAX Digitizing Procedure

7. Login to DISORDER account on Nasal (VAX). Use the command 'cd' to switch to the subdirectory Dysarthria and to switch to the correct subdirectory within Dysarthria in which to store the data (i.e., a particular subject's name). (Note that the VAX command `'set def d$users1:[disorder.subdirectory]'` has been mapped to the UNIX command 'cd'.) To verify you are in the correct subdirectory, type 'whoami' at any time.

   - Type 'record -s16000' (Replace 16000 with 10000 or 13000 if you desire 10 kHz or 13 kHz sampling rates, respectively, instead of a 16 kHz rate.)
   - Within record, enter a gain of 1 (changing the default gain from 4 to 1).
   - The default recording duration is 15 seconds, which (as you will quickly realize) is often too short. To increase the recording duration from 15 to 60 seconds (the longest duration available) do the following:

     i. Press return to start a recording/digitizing session. (Since you are only changing the duration, the tape should not be running at this time.)

     ii. Press any key to stop your recording session.

     iii. Press space bar to (momentarily) accept your (bogus) recording.

     iv. Press 'r' to rerecord, and you will be prompted to give a new duration for the new recording session. Type 60 at this prompt. (You will next be asked for the gain again, so you may reenter 1 or just press return, as 1 should now be the new default gain anyway.)

   - To start the digitizing session, press return on keyboard and press Play on cassette player.

8. Next, the procedure for locating the appropriate knob gain settings will be described. You must first decide whether you would like to be able to compare

sound amplitudes across utterances for your subject, or whether you would like to normalize the amplitudes across all utterances. The procedure described here effectively normalizes the gain (amplitude) across all utterances by setting the gain on an utterance-by-utterance basis. The decision was made to normalize this dysarthric speech data set because the recording environment was not well controlled between recording sessions of the same subject (i.e., distance between microphone and subject could vary) and subjects also exhibited large volume changes due to poor respiratory control. If you do not wish to normalize the data, then the following procedure can easily be adapted to set the amplitude gain only once and leave it at that value (those knob settings) for the entire duration of the digitization.

The process of finding a desirable amplitude gain is a little bit tedious. You do not want to end up with data that is clipped (too loud) or too soft. The major step in determining appropriate knob settings for digitizing a specific utterance is to watch the numbers scrolling by in the righthand column of the window on the screen (the window in which you typed 'record' earlier) while simultaneously watching the analog VU needle on the Shure Professional Microphone Mixer. Make sure the following always holds true:

- The maximum number appearing in the righthand column of the screen stays in the range -2.8 to -5.0 for the utterance. (Be careful that you are not measuring the amplitude of any sounds preceding or following the utterance, as the number in the column reflects the peak amplitude value encountered in the section of tape you are playing, and you want to find the maximum only within the utterance.)
- Keep an eye and an ear on the VU needle and make sure that the needle does not go too far into the red region. The needle makes an audible click when it hits the right side of the window.
- If the maximum number appearing on the screen is larger than -2.8 (closer to zero), then the utterance is too loud to be digitized with the current

choice of amplitude gain. You must decrease the knob settings on knob 2 and/or knob "Master". (Recall that "Master" is the coarse adjustment, and knob 2 is the fine adjustment.) If the maximum number appearing on the screen is smaller than -5.0 (more negative), then the utterance is too soft to be digitized with the current choice of amplitude gain. You must increase the knob settings. As you adjust the knob settings, you will notice that there (unfortunately) seem to be nonlinear regions within the knob positions, so you may need to adjust the knob settings (and consequently, rewind and replay the tape several times) until the maximum gain falls within the desired -2.8 to -5.0 range.

- There are rare instances when the gain is in the region -2.8 to -5.0 and you may still observe the VU needle hitting the right side of the window (on the Shure Prof. Mike Mixer). In these instances, you will need to check the peak values on the dual window display, as explained shortly, to verify that you have not clipped the waveform.
- To rerecord, which you will need to do until the knob settings are appropriate, rewind the tape, press 'r' and follow the instructions.
- When the numbers scrolling by on the screen are within the range given above, press space bar to accept the recording. A dual window display of the waveforms will appear. The top window will contain the entire recording session (i.e., all 60 sec) and the bottom window will contain a magnified section of the waveform near the cursor. In each window the amplitude is autoscaled and the time axis scale is controlled by the "up" and "down" arrow keys on the keyboard.

  As a double-check on the knob gain settings (or for the unusual case described above when the VU needle hits the right side of the window even though the amplitude falls between -2.8 and -5.0), verify that 6000 **does** appear on the y axis in the top window, but 8000 **does not** appear.

  a. If 6000 does not appear (that is, if 4000 is the largest number), then

you need to increase the values on one or both of the knobs Master and 2, and rerecord by rewinding the tape and pressing 'r'.

b. If 8000 appears, then you need to decrease the values on one or both knobs, and rerecord by rewinding the tape and pressing 'r'.

c. To check the specific peak value (rather than approximate it with these 4000, 6000, and 8000 y-axis value estimates), first identify the location of the peak by eye in the top window, then place the cursor in the top window as near that peak as possible, make the bottom window active (click on bar at top) and use the "up" arrow to magnify the waveform around the cursor position. Position the cursor directly over the selected peak and press the right mouse button. The value of the peak appears in the original call window (not one of the two windows in the dual window display). If the VU needle hits the right side of the window even though the value of the gain is acceptable, check the peak value in this manner to be certain that it does not exceed 7500. If it does exceed that value, adjust knob settings and rerecord.

- **Caution:** **NEVER** change the knob settings during the recording session itself. If the gain is changed while an utterance is being spoken, later you will be unable to distinguish whether the volume changed as a result of something you did or something the subject did. Always change the knob settings first, then rewind and rerecord, in a serial fashion.

9. To save an utterance to a .wav file, use left mouse button to place a mark at start of utterance and press 's'. Use left mouse button to place a mark at end of utterance and press 'e'. Press 'p' on keyboard or use middle mouse button to play the utterance to verify it is the one you want and that you have included all of the utterance between cursor markers. (Do not forget Play/Record switch on gray panel has to be set on "Play" to hear the utterance.) Type 'W' (uppercase is important), the name of the file, and press return. Do not append ".wav" as the extension will be appended for you.

At this point, you can save other utterances from this same recording, or move on to recording new utterances by typing 'r' for rerecord and pressing Play on the cassette player. When you are finished digitizing the speech from the cassette, press 'q' to quit the recording/digitizing session.

To see a directory listing of the files saved, type 'dir'. If desired, you can type 'klspec93 filename' to look at/listen to a waveform you have previously saved. (Do not append ".wav" to the filename.)

If you have accidentally created multiple files with the same name and different version numbers (the number on the far right in the filename), you must rename the files. If you do not, when they are transferred to the UNIX system (see below) they will not be recognized as separate versions, and files with the same name will overwrite one another. Also, you may find a need to rename files that were accidentally saved to an undesirable filename due to mistyping. To rename files, type the following at the Nasal prompt:

```
rename oldfilename [disorder.dysarthria.subdirname]newfilename
```

For example, to change the version number from 2 to 1 on dock03 in Mike's subdirectory, type:

```
rename dock03.wav;2 [disorder.dysarthria.mike]dock03.wav;1
```

To delete unwanted files, use the command 'del'.

B.3 Copying Data to UNIX System

10. To copy files to the UNIX system, go to "palate", a DEC Alpha located in Rm. 36-568. It is important that you physically go to "palate" to perform the data transfer. Do **not** remotely login to palate (i.e., telnet, etc.) from another machine in the lab as this approach can (intermittently) add noise to the data in the transfer process. (It is uncertain why noise is occasionally added to the signal, but it is thought to perhaps result from poor ethernet connections

between machines in the office area, particularly between PCs running Linux and the DEC Alphas, including "palate".)

Login to DISORDER account, 'cd' to Dysarthria, and then 'cd' to the temp subdirectory. (If there is no temp subdirectory, use 'mkdir' to create one.) Do not 'cd' into a subject's subdirectory at this point in time, as existing files can be overwritten during the ftp process (even if the files have write protection). To prevent overwriting files, the files will be copied to the UNIX system in a two-step process. To verify you are in the correct subdirectory, type 'pwd' at any time. To copy the files, do the following:

- Type 'ftp -i speech.mit.edu'
- You will be prompted to login to disorder on the VAX system. At the ftp prompt, type **binary**, and 'cd' into desired subdirectory on VAX from which you would like to copy the files.
- Type 'mget *.*' to copy all the files from that subdirectory on the VAX to the temp subdirectory on "palate".
- Type 'quit' to quit the ftp process.
- Type 'ls' to verify all the files were copied.
- Compare the contents of the temp subdirectory with the contents of the subject's subdirectory you plan to move the files into, to make sure that there are no files with the same name. Verify that the files in the subject's subdirectory have write protection by typing 'ls -l' and making certain that "w" does not appear in any of the ten columns on the left. If "w" does appear, type 'chmod 444 *.wav' to provide write protection to the data. Now the data cannot be overwritten when files are transferred from the temp subdirectory to the subject's subdirectory unless you respond 'y' to a prompt.
- While in the temp subdirectory, use the following command to copy the files to the subject's subdirectory:

`'cp -i filename.wav ~/Dysarthria/subjectsubdir/.'`

You may replace 'filename.wav' with '*.*' to copy multiple files. Then, go to the subject's subdirectory and use the 'chmod' command described above to write protect the new files as well.

- You should use xkl within the subject's subdirectory to run "spot checks" on the data, verifying that it looks and sounds good.
- Do not forget to return to the temp subdirectory to delete all the files there, using 'rm *.*'. You may wish to use the '-f' flag to speed up removal of the files. Make absolutely certain that you are within the **temp** subdirectory before you issue this command!
- As the account on the VAX gets full (you can check your usage and your quota by typing 'show quota' at the Nasal prompt on the VAX), you will need to delete the waveforms you have already copied over to the UNIX system. To delete these files, go to the Nasal VAX machine in the Kassel, 'cd' into the desired subdirectory and type 'del *.*;*vernum*', where '*vernum*' is replaced by the appropriate version number, such as 1.

11. Backup the entire DISORDER account on the UNIX system onto a backup DAT tape weekly, using the set of instructions available in the lab, and the tape drive affiliated with the DEC Alpha "palate". Please backup the data regularly... once you have put this much time into digitizing the data, you will not want to have to digitize it again!

Appendix C

Guidelines for Composition of the Word List and Subject Instructions

This appendix presents some guidelines for how to properly construct a word list for an experiment and how to instruct the subject who reads the list. The guidelines are general and should be adapted to fit the needs of each specific experiment at the time it is conducted. For best results, the experiments should be conducted in the Eastham Sound Room, Rm. 36-530. This set of guidelines was developed with the aid of Adrienne Prahler in the spring of 1998.

C.1 Considerations in Word List Composition

1. Compose list keeping within-word coarticulatory effects in mind. (Coarticulatory effects are the effects of production of surrounding sounds on production of the sound in which you are interested.) For example, when you are examining the production of a particular sound, such as the vowel /i/, ask yourself what sounds are on either side of that vowel in each utterance on your list, and how will those sounds affect the production of the vowel.

2. Randomize list. (Randomizes confounding coarticulatory effects that occur across word boundaries.)

3. Place additional utterances at the beginning and end of your list ("utterance padding") to give subjects some utterances to practice on (at the beginning) and to avoid F0 changes as they reach the end of your list. If your list is divided into several short lists or several pages, consider utterance padding for each short list/page.

4. Arrange actual word list in a format that is easy to read and legible. Make list in a large enough font that the subject can read it from over 1 foot away. Do not use flash/index cards for your word list as they cause too much rustling noise, which can interfere with the clean recording of your utterances.

5. Test run on yourself. Enables you to work the kinks out of your choice of utterances (as well as the recording protocol) and allows you to determine how best to instruct the subject for your particular experiment.

6. Go over list with subject to see if they have any questions regarding pronunciation, etc.

7. Make two copies of the lists, so you can follow along as the subject says the utterances. Utterances that are mispronounced or skipped can be identified for repetition by the subject.

8. Some additional things to consider/discuss with research supervisor: What vowels and/or consonants should be placed in each word of each utterance? Should your utterances be placed between two other words (i.e., in a carrier phrase) or spoken in isolation? How many repetitions of each utterance are needed?

C.2 Instructions to Give the Subject

1. Keep in mind there are many things that can generate noise in a room, such as shifting in chairs, coughing/sneezing, moving around, placing things on table, etc. Ask subject (and anybody else in the room, including yourself!) to be as quiet as possible.

2. Instruct subject to stand approximately 1 foot from the microphone. Adjust microphone until it is level with the subject's mouth. The length of a sheet of paper (11 inches) is a good guide for establishing the proper distance. If the subject stands too close, the omnidirectional microphone can pick up puffs of air emitted during the production of certain sounds, distorting the recording, and if the subject stands too far away the recording level may be too soft.

3. Have the subject hold the word list BEHIND the microphone (on the other side of the microphone than the subject), so that the paper doesn't block the microphone, resulting in a poor recording.

4. Ask subject to speak slowly, separating utterances with some silence/pauses. Demonstrate rate to your subject, if necessary. Have subject practice by saying a few words from the list and verify that the words are being spoken in the desired manner, with the desired separation.

5. Tell the subject to feel free to request break/water as needed. You may even want to take a glass of water into the sound room for the subject, but please do not take other beverages or foods into the room.

6. If the subject has a cold or any atypical hoarseness, you may wish to reschedule your recording session.

Appendix D

Instructions for Recording Speech with a DAT Player

This appendix describes the procedure for recording speech using a DAT (digital audio tape) player in the Eastham Sound Room (Rm. 36-530). This set of instructions and guidelines was developed with the aid of Adrienne Prahler in the spring of 1998.

The instructions are as follows.

1. Plug DAT tape player into wall power strip in the Eastham Sound Room, using AC adapter. (If you do not find the DAT tape player within the Sound Room, it is likely to be located outside the room on the table near the MacIntosh computer called "Perceive".) Insert tape after pressing and sliding the open button on side of DAT tape player (side with volume controls).

2. Use cable located with DAT tape player to connect Line In on the tape player to Headphones output on back of Shure Professional Microphone Mixer. The cable is black with a large plug on one end and a small plug on the other end.

3. Be sure the settings on the side of the DAT player are as follows. Settings:

   - Rec Mode: **Manual**
   - SP: **48 kHz** (Or, you could choose 44.1 kHz, although it will take more computer computational power later to downsample from this noninteger

value, since you must first upsample by a large integer value prior to down-sampling. Do not choose 32 kHz, because the DAT player is constructed by the manufacturer to accept this sampling rate only when connected to other devices, and not from a subject speaking into a microphone.)

4. Turn on Shure Professional Microphone Mixer.

5. Have subject stand in center of room approximately 1 foot from omnidirectional microphone. Adjust microphone height until it is level with subject's mouth.

6. Set recording level. Select Pause, then Record, on the DAT player. (This step is similar to pressing pause and record simultaneously on the cassette tape players with which you may be more familiar.) The tape player is now receiving input from the microphone but the tape is not advancing. Have the subject practice saying some of the utterance list while you adjust the recording level using Rec level knob on tape player. Observe dB values indicated by bars appearing on screen of tape player. Your objective is to have approximately **-6 dB** when subject is speaking. Be very careful **not to max out** (be at or very near 0 dB) at any time, but also do not have the recording level too soft (near -24 dB). You are attempting to find a recording level that does not clip the data but also does not result in a poor SNR. If you are having difficulty achieving the desired recording level, be sure to also check the knob settings on the Shure Microphone Mixer. In particular, check the Master knob and the knob associated with the microphone input (currently the microphone is connected to Mic 1 input). These knob settings should typically be in the range 6–7. Setting the recording level can be tricky. If you have any difficulties, please do not hesitate to ask for help from other lab members.

   Caution: **NEVER** change the recording level during the recording session itself. If the recording level is changed while an utterance is being spoken, later you will be unable to distinguish whether the volume changed as a result of something you did or something the subject did. If you need to change the

recording level at a later point in the experiment, stop the tape first and repeat Step 6, then start the recording session again.

7. To record, press Pause a second time to "release" the Pause button. (The tape player will not allow you to record by simply pressing the Record button!) It is advisable to listen to the subject as they say each word on your utterance list in order to keep track of any utterances that are mispronounced or skipped. Simultaneously, you will need to continue to watch the recording level on the tape player, to make certain the subject has not changed speaking volume to the extent that the utterances become clipped or too soft. At the end of each section of words on your list, or at the end of your recording session, make any adjustments needed and ask your subject to repeat the poorly-recorded utterances.

8. Press Stop button at the end of recording. Since the DAT player (for unknown reasons) occasionally rewinds the tape a little at the end of a recording session (i.e., any time it is allowed to sit for a while following a recording), it is recommended that you advance the tape for a short distance beyond the end of your recording session before you remove it from the tape player. Then the tape will be in the proper position for your next recording session. Take tape out before turning off power to DAT player. (You can not remove the tape without power!)

9. Disconnect and put away equipment. Please leave the room as you found it.

Appendix E

Instructions for Copying Data from a DAT Tape to a Computer File, Incorporating Downsampling of the Data

This appendix describes the procedure for transferring digital speech data sampled at a high sampling rate (44.1 kHz or 48 kHz) and stored on a DAT (digital audio tape) to digital speech data sampled at a low sampling rate (10 kHz, 13 kHz, or 16 kHz) and stored in a Klatt .wav file on a DEC Alpha computer running the UNIX operating system. Klatt .wav files are the type of file utilized by the xkl software on the UNIX system. This set of instructions was developed with the aid of Adrienne Prahler and Mengkiat Goh in the spring and summer of 1998.

The instructions are as follows.

E.1 Required Hardware and Software

1. The required hardware and software are listed below.

    - MacIntosh computer called "Perceive", located on table outside Eastham

Sound Room (Rm. 36-530), with digital I/O sound card and Digidesign software installed

- Sony DAT player (portable Walkman unit) (This tape player is usually found in the Eastham Sound Room.)
- Sony RMRD3 - affectionately referred to as the "black box"
- DEC Alpha computer running the UNIX operating system (usu. "palate")
- PC running Linux operating system

E.2 Connecting MacIntosh and Hardware

2. DAT player

   - AC adapter - plug into DAT player and power strip (Make sure power strip is turned on!)*
   - Remote digital I/O (on side of DAT player) connected to built-in timing cord of RMRD3*
   - Set sampling rate switch to the sampling rate of your (previously-recorded) tape (either 44.1 or 48 kHz)

3. RMRD3 (Black Box)

   - Coaxial input and output should already be connected to sound card on the back of "Perceive" (the MacIntosh computer). Connection uses cable with black and red plugs at each end.*
   - Built-in timing cord connected to DAT digital I/O (as stated above)*
   - Plug into power strip*
   - Digital input is set on **coaxial**
   - Input select is set on **digital**
   - Timer is set to **off**

- Power is **on (DO LAST!)**

4. Data Transfer Switch Located to left of the computer on the table.

   - Select **A** for monitor outside the sound room (where you should be working!). The position of this switch should be set prior to turning on the computer.

5. Turn on computer. (Press arrow key at top of keyboard.)

6. Before initiating the data transfer protocol, which utilizes the Digidesign Sound Designer II MacIntosh software (Section E.3), verify the following:

   a. The computer is set on **General Settings** (not Psyscope Settings). To check settings, select the Apple icon menu, Control Panels, Extensions Manager and verify that Selected Set indicates General Settings. If it does not, choose General Settings from the pop-up settings menu, then restart computer for new settings to be active.

   b. The Sound Out setting is **DigiDesign** (not Built-In speakers). To check (or change) select Apple icon menu, Control Panel, Sound, Sound Out, then Digidesign and Quit to exit window. (The volume can be adjusted by selecting Volumes instead of Sound Out.)

   * These steps should already have been done for you.

E.3 Procedure for Copying Data from DAT Tape to MacIntosh

7. Select Apple icon menu, Applications, Sound Designer II.

8. In Sound Designer II, select New from File menu. Click on the panel marked Sound Designer II to select file type **PC WAV (.wav)**. The PC .wav file type shown here is **different** from the Klatt .wav file type that is used in xkl on

the UNIX platform. You will be given instructions about how to convert these PC .wav files to the Klatt .wav file format later in this handout. Next, be sure you are storing the data in your folder within the USERS folder on the Mac. You will likely need to choose Perceive from the box at the top of the window, then go to your folder within the USERS folder. Then, type in the appropriate filename in which to store the data. Do not append .wav to your filename! You will want to avoid confusing these files with the Klatt .wav files created later. Instead, you should append .mswav (for Microsoft .wav file format).

9. Choose desired bit quantization and mono/stereo setting (usually 16 bit, mono) then Save to close window. For the normal data in this thesis, 16 bits and mono were the selected settings, respectively.

10. Verify that Hardware Setup of Sound Designer II under Setup Menu has these settings:

```
Card Type: Audiomedia
Cards to Use: Card 1: Slot 13
Track Mapping
DSP: Slot 13
Plays:  Stereo Mix (L)
        Stereo Mix (R)
Peripheral: No Peripheral
Sample Rate: <sampling rate should be same as DAT tape
             and tape player setting> (44.1 or 48 kHz)
Synch Mode: Digital (a very important setting)
Ch 1,2 input: Digital (a very important setting)
```

When changes are complete, click on Recalibrate Inputs and select OK.

11. Select Record button on screen (looks like a tape reel). Position Input slider bar at about 4. Select Monitor (i.e., be certain box is checked). The indication of Mono or Stereo should reflect the choice made at time of opening a new

file (refer to Step 9). Sampling rate is preselected by the program to be the same as the input device (i.e., DAT tape) being used (either 44.1 kHz or 48 kHz). Select Pre-Allocate (check the box) and change the Disk Buffer Size from 4 to 12. By choosing Pre-Allocate, the data will be stored in a contiguous block and will not be broken up and placed in different spots on the hard drive, whenever possible. A contiguous file is less susceptible to the disk access and general playback problems that can occur when the hard drive data become fragmented. The Disk Buffer Size determines how much memory is allocated as a record buffer. Increasing the buffer size to 12 will help compensate for a slow or fragmented hard drive. (The settings for Pre-Allocate and Disk Buffer Size are made to attempt to prevent the occasional corruption of data files attributed in the past to the MacIntosh "Perceive's" hard drive being small and very full.)

12. Use headphones or speakers with DAT player to determine the appropriate location on the tape to begin copying data to the computer (recording data onto the computer).

13. To start recording, press Play on DAT player and select Record button (REC) on screen. During the recording time period, watch for clipping by observing the green bars on the screen, making sure that the "clip and hold" feature has not been activated (if clipping does occur, the tops of the bars will remain green). If it has been activated, you must reposition the Input slider bar to a smaller value (see Step 11) and rerecord the data to the hard drive. You should avoid recording more than a minute's worth of data from the DAT tape into any given file. This limitation is because the performance of xkl (which you will be using on the UNIX system to read in and examine these files), is extremely slow (and is prone to crash) for files that are longer than one minute. (The primary limitation actually centers around the quantity of RAM available on the PC machines on which you will run xkl, and is not inherent to xkl itself.) To stop recording, select the Stop button on the screen and on the DAT player. Notice that the Sound Designer II software has an interface similar to a tape

recorder, with additional buttons (such as rewind and fast forward) which you may find useful to manipulate the data once it is stored in the computer. To exit recording mode, select Done.

14. Save file after recording process is finished by selecting File Menu, Save.

15. After all files have been recorded and saved, close program by selecting File Menu, Quit.

E.4 Copying and Converting Data from MacIntosh Computer (PC .wav Format) to UNIX System (Klatt .wav Format)

E.4.1 Copying PC .wav Files from MacIntosh to UNIX System

16. At the MacIntosh, select Apple icon menu, Internet Apps, Fetch program.

17. In File Menu, open New Connection window, log into the specific UNIX machine to which files are being copied (usually "palate") with appropriate username and password and Select OK.

18. Choose Binary file, then Put File, and select appropriate file to transfer (i.e., choose Perceive from box at top of window, then go to your folder in USERS to find appropriate file), then select Open. In new window that opens, leave format as default of Raw Data then select OK again.

19. Select Quit from File menu.

E.4.2 Converting Data from PC .wav Format to Klatt .wav Format and Downsampling the Data

20. This series of steps must currently be performed on a PC running Linux (any PC in the laboratory should work, such as "septum" in the Kassell, Rm. 36-553, or "brogino" in the Library, Rm. 36-515). You will need to be on a PC rather than a DEC Alpha or an SGI because the code has currently been compiled only for use with machines running the Linux version of UNIX. (Perhaps in the future the code will become available for use on other machines as well.) The PC needs to be running Linux; therefore, if it is running Microsoft Windows when you first sit down at the terminal, restart the computer and be sure to press the "Shift" key when you see the "LILO" prompt appear on the screen. Then, type "Linux" at the LILO prompt and press return to boot the computer using the Linux platform. Once the computer has booted, login and type "startx" at the prompt to start the X windows emulator. Within the emulator you can open a window, etc., similar to the UNIX machines.

21. Use the program "ms2klmod" to convert the PC .wav files to Klatt .wav files. This program has been placed in the bin/Linux subdirectory of the DISORDER account on palate, and you may copy it to your account for your use. (To copy the file, type 'cd /usr/palate1/disorder', use the 'cd' command to move to the correct subdirectory, then utilize the 'cp' command. When finished copying the program, type 'cd' to return to the top level directory of your own account.) The usage of the program is as follows: 'ms2klmod file1.mswav file2.mswav ...'. This program is a modification of the program "ms2kl", which is available as a satellite utility program affiliated with xkl. The program uses Sox (version 12.14) to convert from one file type to another and is able to handle multiple files using a wildcard. When the file type conversion is complete, you will have both *.wav and *.mswav files in your account. After verifying that the *.wav files (Klatt .wav files) can be opened using xkl and that they look and sound fine, you should delete the *.mswav (Microsoft .wav files) unless you wish to

save them to tape for backup purposes first.

22. Open each file within xkl and separate the file into its individual utterances. For example, to isolate a desired utterance, place the cursor prior to the beginning of that particular utterance, and type 'w'. Place the cursor after the end of the utterance, and type 'e'. Use the middle mouse button or press 'p' to listen to the utterance and make sure it is your desired utterance, as well as to verify that you have included the entire utterance between the 'w' and 'e' cursor positions. (Utilizing your word list will assist you in identifying the utterances in each file.) When you are satisfied with the cursor placement, type 'o' to save that utterance to a file. You will be prompted for a filename and asked whether you would like to view the utterance. Continue in this fashion for each utterance in the file.

23. The next step is to downsample the data. Since the frequency range of interest in speech is limited to lower frequencies, the data is downsampled in order to reduce the amount of hard drive space each file requires. Downsampling will be used to convert data sampled at a high sampling rate (44.1 kHz or 48 kHz) to data sampled at a low sampling rate (10 kHz, 13 kHz, or 16 kHz). The actual downsampling process involves first upsampling, then filtering, then downsampling.

    - The downsampling is performed by a Matlab script called "downsample.m", located in the matlab subdirectory of the DISORDER account on palate, and you may copy it to your account for your use. You will also need to copy the programs "resamplemod.m" and "mat2klmod.m", which have been modified for use with the downsampling script. Other programs called by downsample.m, including "kl2mat.m" and "raw2kl*" are satellite utility programs affiliated with xkl, so they should be available on any machine on which xkl has already been installed.
    - Prior to running the script on your data, you may need to use the emacs editor to edit the script, depending on your desired final sampling rate.

If your final sampling rate is 16 kHz, then you will not need to edit the script. However, if your final sampling rate is 10 kHz or 13 kHz, then you will need to replace all occurrences of 16000 (there are three occurrences) with the final sampling rate of your choice, 10000 or 13000, respectively.

- To run the downsample script, first 'cd' into the subdirectory in which your files are located, then type 'matlab' at the prompt. Once matlab is started, type 'downsample' at the matlab prompt. The script will downsample all the *.wav files in the present directory. The script will rename all the original files as *ORIG.wav, and will save the downsampled files as *.wav. After you have utilized xkl to verify that the downsampling was done to your satisfaction, you may delete all the *ORIG.wav files (unless you wish to save them to tape for backup purposes first).

TROUBLESHOOTING ON THE PC:

This section is for troubleshooting the operation of the programs utilized on a PC running Linux.

- If the version of xkl available online (accessed by typing 'xkl' at the UNIX prompt) crashes when you load in or manipulate waveform files with higher sampling rates (44.1 kHz or 48 kHz), then try copying to your account and using the compiled version of xkl called "xkl-linuxbeta2.4" located in the bin/Linux subdirectory in the DISORDER account on palate.
- In order to view spectra for the waveform files possessing higher sampling rates (44.1 kHz or 48 kHz), you will need to change the window duration to a value under 10 ms. Be sure to change both the 'd' and 's' window duration parameters. This requirement is because of a memory problem in xkl when computing the DFT for longer window durations.
- When you run "downsample.m", if the programs "kl2mat.m" and "raw2kl" are not able to be found on your machine, then copy "kl2mat.m" from the

matlab subdirectory and copy "raw2kl" from the bin/Linux subdirectory of the DISORDER account on palate.

- If you would like to view the filter used in the downsampling routine, you may edit "resamplemod.m", uncommenting the final few lines of code. Then, when "resamplemod.m" is called within "downsample.m" in matlab, the filter will be displayed on the screen.

24. **Please don't forget this step** (be a considerate lab member!): Return to MacIntosh computer "Perceive" and delete from the hard drive the files that are now no longer needed. These files can be huge (several hundred MBs in size apiece) and take up a great deal of hard drive space, consequently it is very important that you delete your files at the end of each transferring session. (Of course, do not delete them until you have verified that they were transferred correctly, but by the time you have reached this step, you will have verified that the files are OK.)

25. Turn off MacIntosh by selecting Shutdown in Special Menu.

26. Return DAT player to sound room.

27. Put RMRD3 back in Standby power mode.

Appendix F

Additional Perceptual Test Results and Experiment Data

F.1 13-Utterance Results

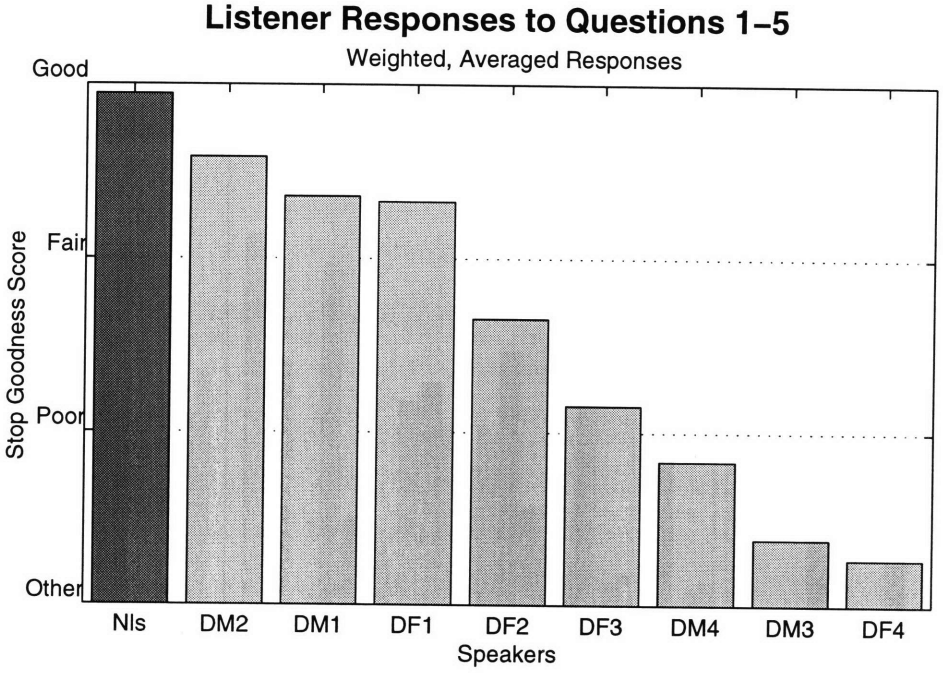

Figure F-1 : Combined, weighted listener responses to Q1–Q5 provide a measure of "stop goodness". Word repetitions in which the listener correctly identified the presence of a consonant (with or without precursor), the type of voicing, and the place and manner of articulation for the consonant were quantified according to the response to Q5: Good = 3, Fair = 2 and Poor = 1. Repetitions in which the initial sound was identified to be a vowel, or the initial consonant was incorrectly identified with regard to voicing, place or manner of articulation, were given a value of 0 (Incorrect). Scores were then averaged across all 13 utterances, 3 repetitions/utterance and 4 listeners to generate one value reflecting stop goodness for a given speaker. In the case of normal speakers (Nls), the scores were also averaged across all 8 speakers. The normal and dysarthric (DF1–DF4, DM1–DM4) speakers are organized from left to right in order of decreasing stop goodness score.

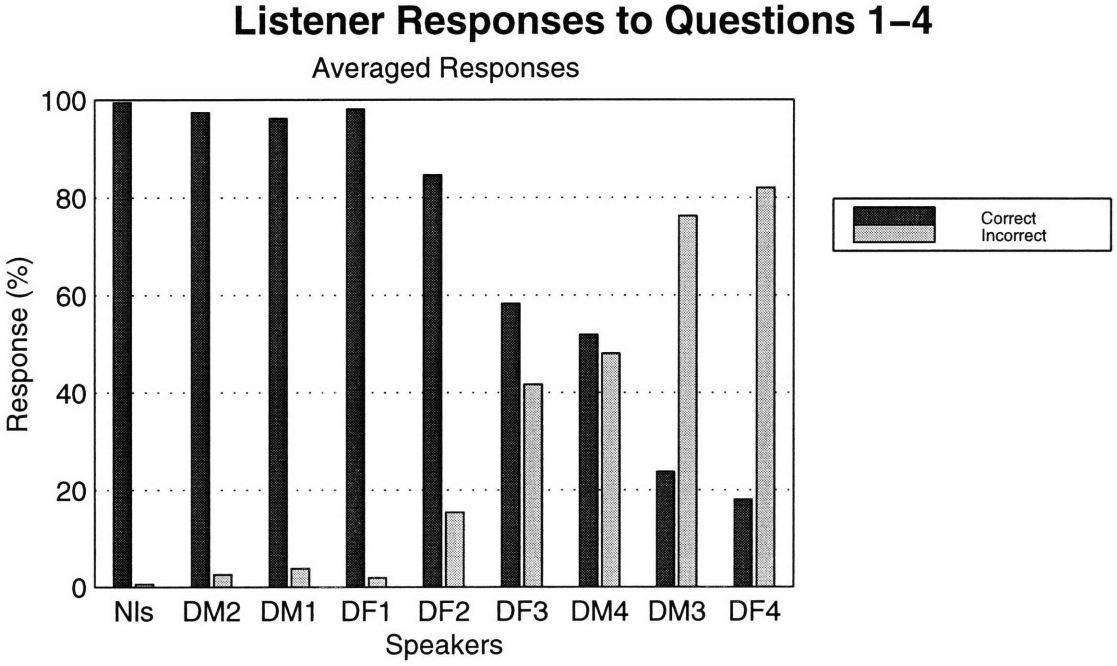

Figure F-2 : Combined listener responses (%) to Q1–Q4. The category "Correct" contains all word repetitions in which the listener correctly identified the presence of a consonant (with or without precursor), the type of voicing, and the place and manner of articulation of the consonant. The category "Incorrect" contains all remaining word repetitions. For each speaker, responses shown averaged across all 13 utterances, 3 repetitions/utterance and 4 listeners. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness, as determined in Figure 4-2.

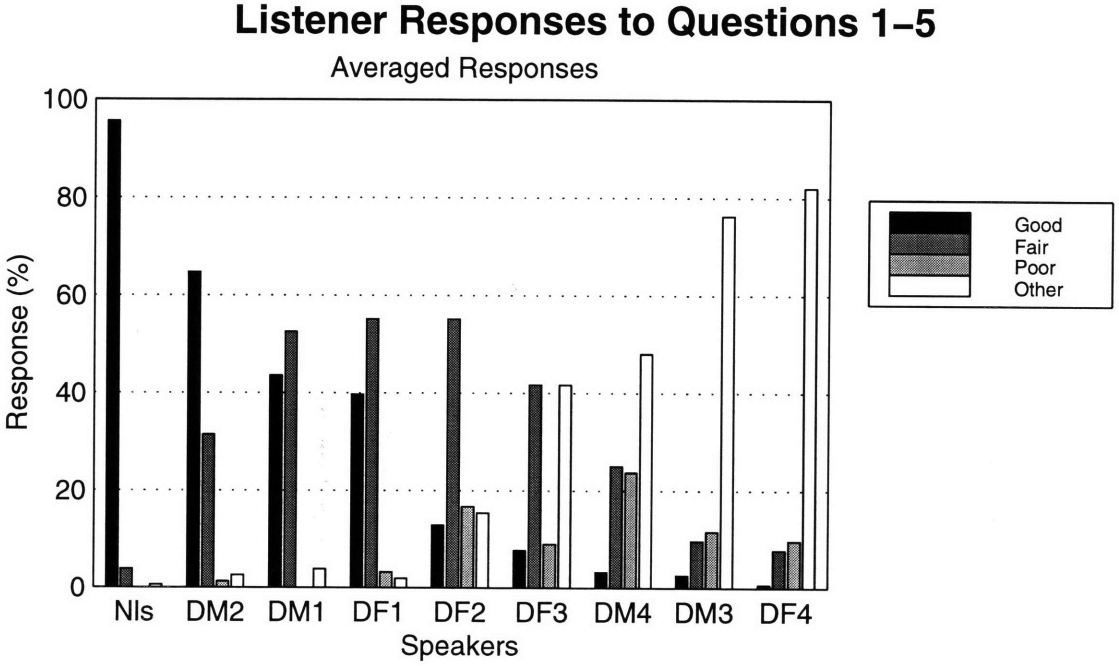

Figure F-3 : Combined listener responses (%) to Q1–Q5. Word repetitions in which the listener correctly identified the presence of a consonant (with or without precursor), the type of voicing, and the place and manner of articulation of the consonant are divided into Good, Fair and Poor ratings according to the responses to Q5. The category "Incorrect" contains all remaining word repetitions. For each speaker, responses shown averaged across all 13 utterances, 3 repetitions/utterance and 4 listeners. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

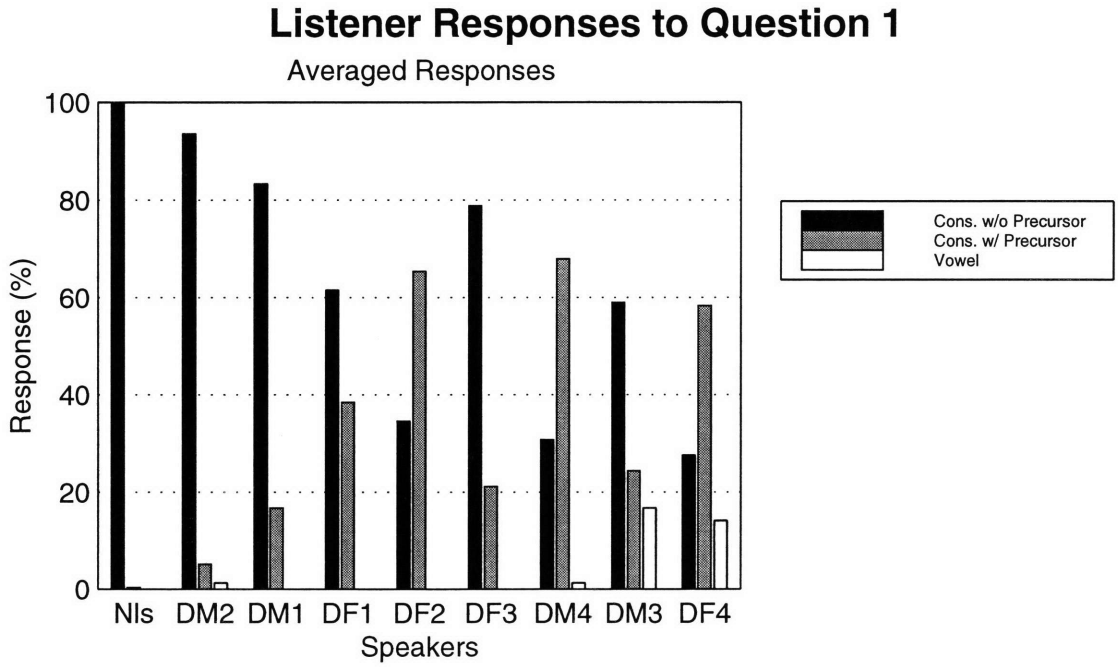

Figure F-4 : Listener responses (%) to Q1, identifying the initial sound in the utterance as a vowel, a consonant with a precursor or a consonant without a precursor. Responses shown averaged across all 13 utterances, 3 repetitions/utterance and 4 listeners for each speaker. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

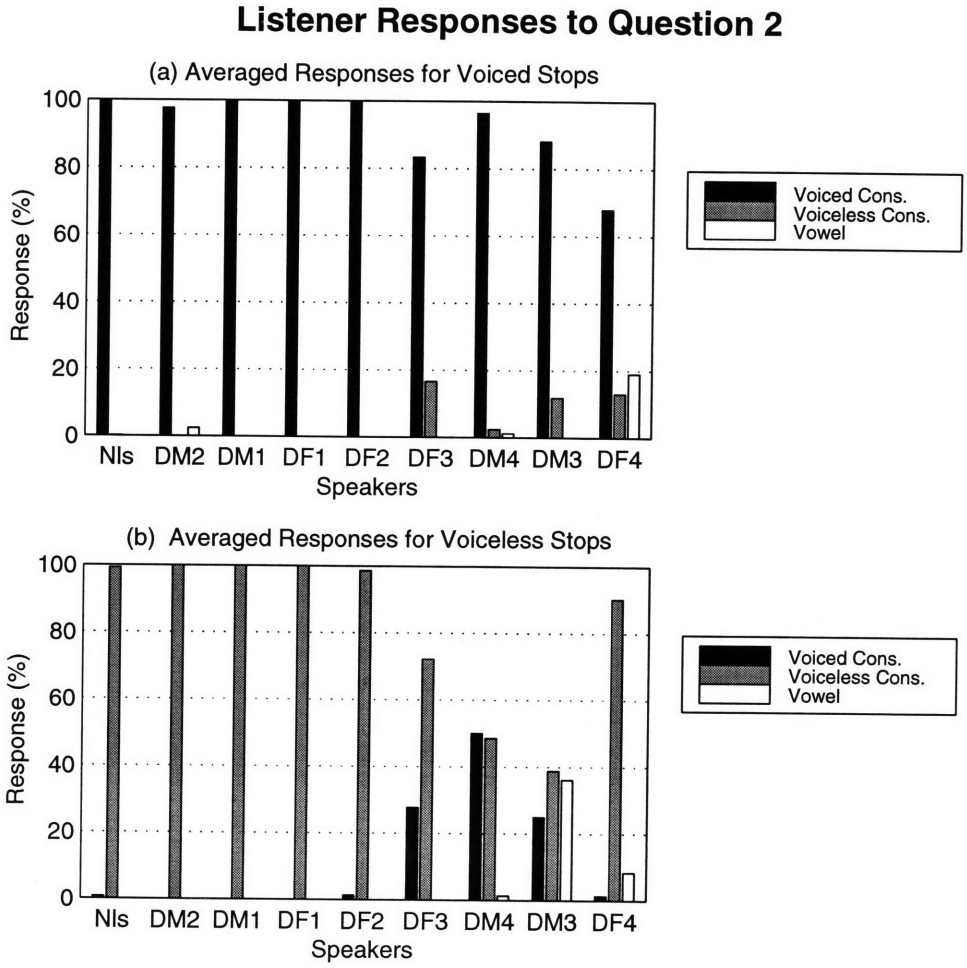

Figure F-5 : Listener responses (%) to Q2, identifying the type of voicing (voiced or voiceless) of the consonant. Instances in which the initial sound was identified as a vowel are also indicated. For each speaker, responses shown averaged across 4 listeners, 3 repetitions/utterance, and (a) 7 utterances containing intended word-initial voiced stops or (b) 6 utterances containing intended word-initial voiceless stops. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

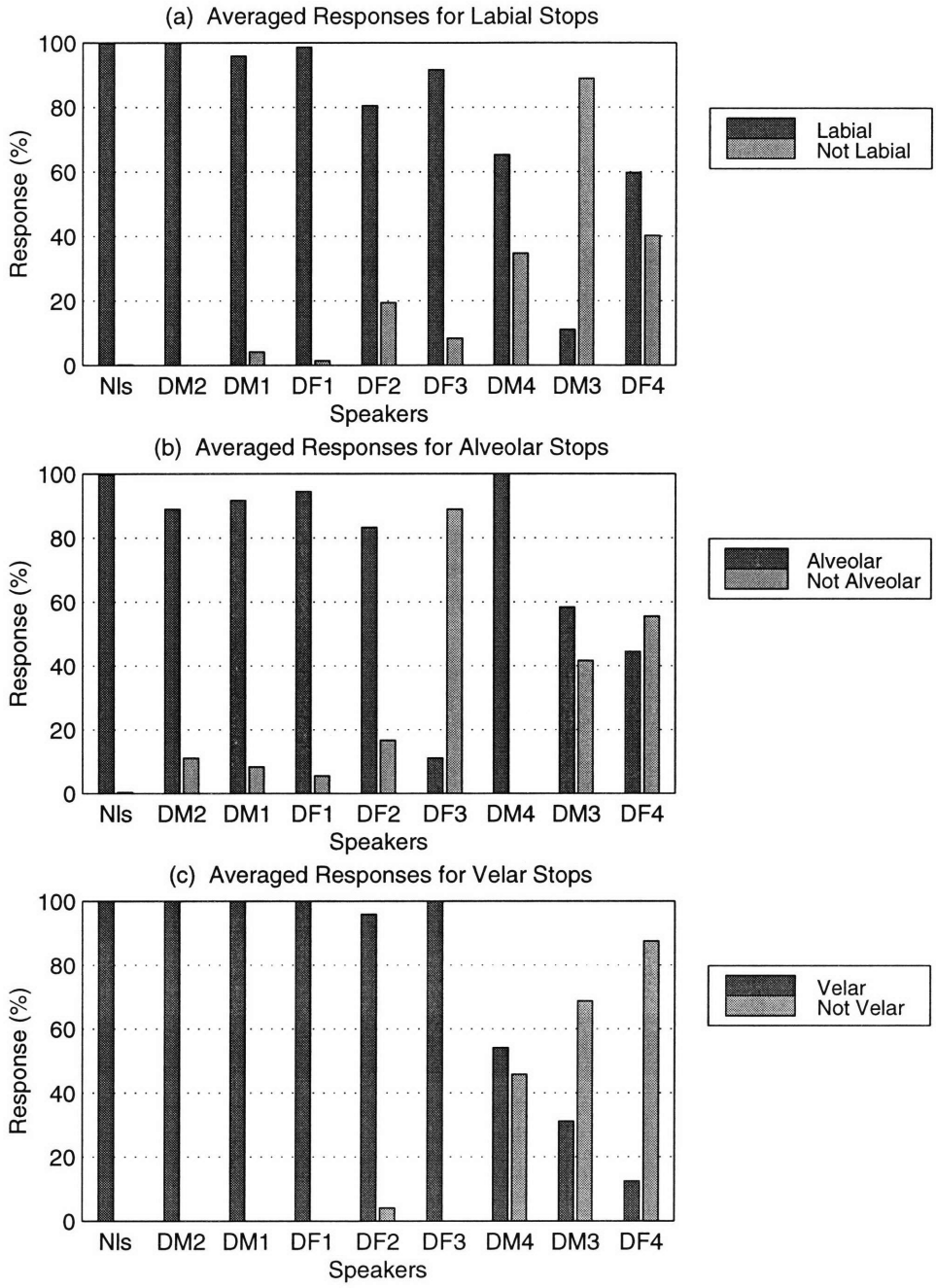

Figure F-6 : Listener responses (%) to Q3, identifying the place of articulation of the consonant. Instances in which the initial sound was identified as a vowel are included in the category "Not [place of articulation]" in each subplot. For each speaker, responses shown averaged across 4 listeners, 3 repetitions/utterance and (a) 6 utterances containing intended word-initial labial stops or (b) 3 utterances containing intended word-initial alveolar stops or (c) 4 utterances containing intended word-initial velar stops. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

Word-Initial Stop	Labial	Alveolar	Velar	GS or V	Other
Normals					
/b/ (Avg. of 4 utts.)	100.0	0.0	0.0	0.0	0.0
/p/ (Avg. of 2 utts.)	99.5	0.5	0.0	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	99.5	0.0	0.0	0.5
/t/	0.0	100.0	0.0	0.0	0.0
/g/	0.0	0.0	100.0	0.0	0.0
/k/ (Avg. of 3 utts.)	0.0	0.0	100.0	0.0	0.0
DM2					
/b/ (Avg. of 4 utts.)	100.0	0.0	0.0	0.0	0.0
/p/ (Avg. of 2 utts.)	100.0	0.0	0.0	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	83.3	0.0	8.3	8.3
/t/	0.0	100.0	0.0	0.0	0.0
/g/	0.0	0.0	100.0	0.0	0.0
/k/ (Avg. of 3 utts.)	0.0	0.0	100.0	0.0	0.0
DM1					
/b/ (Avg. of 4 utts.)	100.0	0.0	0.0	0.0	0.0
/p/ (Avg. of 2 utts.)	87.5	0.0	12.5	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	87.5	4.2	0.0	8.3
/t/	0.0	100.0	0.0	0.0	0.0
/g/	0.0	0.0	100.0	0.0	0.0
/k/ (Avg. of 3 utts.)	0.0	0.0	100.0	0.0	0.0
DF1					
/b/ (Avg. of 4 utts.)	97.9	2.1	0.0	0.0	0.0
/p/ (Avg. of 2 utts.)	100.0	0.0	0.0	0.0	0.0
/d/ (Avg. of 2 utts.)	0.0	95.8	0.0	0.0	4.2
/t/	0.0	91.7	0.0	0.0	8.3
/g/	0.0	0.0	100.0	0.0	0.0
/k/ (Avg. of 3 utts.)	0.0	0.0	100.0	0.0	0.0

Table F.1 : Confusion matrices containing listener responses (%) to Q3, identifying the place of articulation for each stop. The rows represent the intended word-initial stop and the columns the listeners' responses, where GS or V = Glottal Stop or Vowel and Other = Labiodental, Dental or Palatal. For each speaker, responses shown averaged across 3 repetitions, 4 listeners and the number of utterances indicated. For normal speakers, responses also averaged across all 8 speakers. The confusion matrices are in order of decreasing stop goodness for the normal and dysarthric (DF1–DF4, DM1–DM4) speakers. Confusion matrices for speakers DF2, DF3, DM4, DM3, and DF4 are continued on next page.

Word-Initial Stop	Labial	Alveolar	Velar	GS or V	Other
DF2					
/b/ (Avg. of 4 utts.)	83.3	10.4	0.0	0.0	6.2
/p/ (Avg. of 2 utts.)	75.0	12.5	4.2	0.0	8.3
/d/ (Avg. of 2 utts.)	4.2	87.5	0.0	0.0	8.3
/t/	0.0	75.0	8.3	8.3	8.3
/g/	0.0	0.0	100.0	0.0	0.0
/k/ (Avg. of 3 utts.)	0.0	0.0	94.4	0.0	5.6
DF3					
/b/ (Avg. of 4 utts.)	91.7	6.2	2.1	0.0	0.0
/p/ (Avg. of 2 utts.)	91.7	4.2	0.0	0.0	4.2
/d/ (Avg. of 2 utts.)	0.0	12.5	87.5	0.0	0.0
/t/	0.0	8.3	83.3	0.0	8.3
/g/	0.0	0.0	100.0	0.0	0.0
/k/ (Avg. of 3 utts.)	0.0	0.0	100.0	0.0	0.0
DM4					
/b/ (Avg. of 4 utts.)	81.2	8.3	8.3	2.1	0.0
/p/ (Avg. of 2 utts.)	33.3	16.7	33.3	12.5	4.2
/d/ (Avg. of 2 utts.)	0.0	100.0	0.0	0.0	0.0
/t/	0.0	100.0	0.0	0.0	0.0
/g/	0.0	50.0	33.3	8.3	8.3
/k/ (Avg. of 3 utts.)	0.0	19.4	61.1	11.1	8.3
DM3					
/b/ (Avg. of 4 utts.)	14.6	60.4	10.4	0.0	14.6
/p/ (Avg. of 2 utts.)	4.2	8.3	0.0	87.5	0.0
/d/ (Avg. of 2 utts.)	0.0	70.8	16.7	0.0	12.5
/t/	0.0	33.3	0.0	66.7	0.0
/g/	25.0	0.0	50.0	0.0	25.0
/k/ (Avg. of 3 utts.)	0.0	11.1	25.0	58.3	5.6
DF4					
/b/ (Avg. of 4 utts.)	62.5	4.2	0.0	33.3	0.0
/p/ (Avg. of 2 utts.)	54.2	4.2	0.0	37.5	4.2
/d/ (Avg. of 2 utts.)	8.3	50.0	0.0	8.3	33.3
/t/	0.0	33.3	8.3	58.3	0.0
/g/	0.0	8.3	8.3	83.3	0.0
/k/ (Avg. of 3 utts.)	0.0	2.8	13.9	80.6	2.8

Table F.1 : (continued)Confusion matrices containing listener responses (%) to Q3, identifying the place of articulation for each stop. The rows represent the intended word-initial stop and the columns the listeners' responses, where GS or V = Glottal Stop or Vowel and Other = Labiodental, Dental or Palatal. For each speaker, responses shown averaged across 3 repetitions, 4 listeners and the number of utterances indicated. For normal speakers, responses also averaged across all 8 speakers. The confusion matrices are in order of decreasing stop goodness for the normal and dysarthric (DF1–DF4, DM1–DM4) speakers.

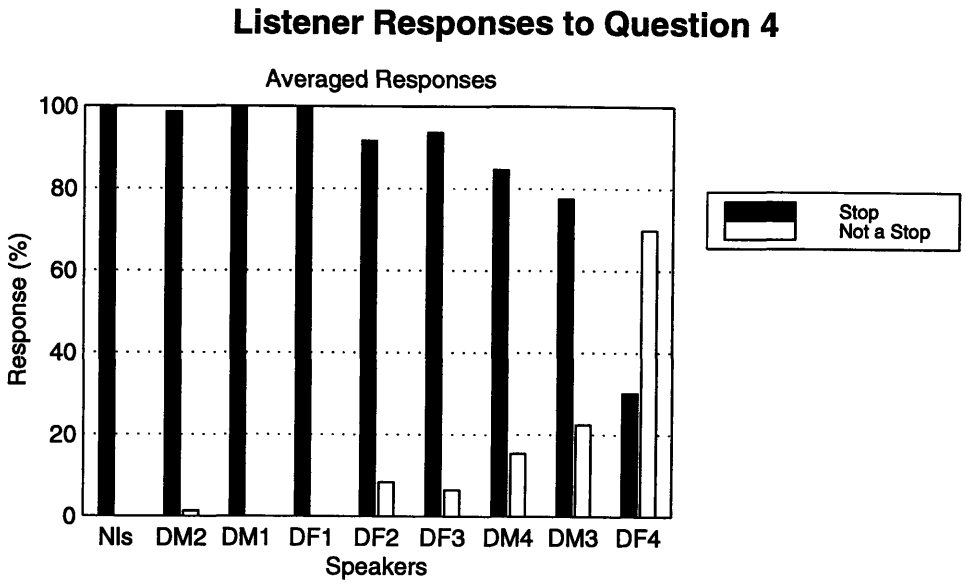

Figure F-7 : Listener responses (%) to Q4, identifying the initial sound in the utterance as a stop consonant or not. Instances in which the initial sound was identified as a vowel are included in the category "Not a Stop". Responses shown averaged across all 13 utterances, 3 repetitions/utterance and 4 listeners for each speaker. For normal speakers, responses also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers are shown from left to right in order of decreasing stop goodness score.

	Stop	Other Obstruent	Sonorant	Vowel
Normals				
Voiced	100.0	0.0	0.0	0.0
Voiceless	100.0	0.0	0.0	0.0
DM2				
Voiced	97.6	0.0	0.0	2.4
Voiceless	100.0	0.0	0.0	0.0
DM1				
Voiced	100.0	0.0	0.0	0.0
Voiceless	100.0	0.0	0.0	0.0
DF1				
Voiced	100.0	0.0	0.0	0.0
Voiceless	100.0	0.0	0.0	0.0
DF2				
Voiced	94.0	2.4	3.6	0.0
Voiceless	88.9	9.7	1.4	0.0
DF3				
Voiced	92.9	3.6	3.6	0.0
Voiceless	94.4	5.6	0.0	0.0
DM4				
Voiced	85.7	2.4	10.7	1.2
Voiceless	83.3	8.3	6.9	1.4
DM3				
Voiced	92.9	7.1	0.0	0.0
Voiceless	59.7	0.0	4.2	36.1
DF4				
Voiced	20.2	4.8	56.0	19.0
Voiceless	41.7	50.0	0.0	8.3

Table F.2 : Confusion matrices containing listener responses (%) to Q4, identifying the manner of articulation of the stop consonants. The rows indicate the intended type of voicing, and the columns are the listeners' responses. For each speaker, responses shown averaged across 3 repetitions, 4 listeners and 7 utterances containing intended word-initial voiced stops (first row) or 6 utterances containing intended word-initial voiceless stops (second row). For normal speakers, responses also averaged across all 8 speakers. The confusion matrices are shown in order of decreasing stop goodness for the normal and dysarthric (DF1–DF4, DM1–DM4) speakers.

F.2 Raw Data

This section details the listener responses recorded during the auditory-perceptual testing. Refer to Chapter 4 for a detailed description of the experiment. The perceptual test data listed below are organized in seven columns as follows: Speaker ID, utterance, repetition number and responses for each of the four listeners.

F.2.1 Question 1

The listener answers to Question 1 are abbreviated as follows: Vowel = `vow`, Consonant with Precursor = `pre`, and Consonant without Precursor = `nop`.

DF1	bad	2	pre	nop	pre	pre
		3	nop	nop	nop	pre
		4	nop	nop	pre	pre
	beat	2	pre	pre	pre	pre
		3	pre	pre	pre	pre
		4	nop	nop	pre	pre
	bill	2	pre	pre	pre	pre
		3	pre	nop	nop	pre
		4	pre	pre	pre	pre
	bunch	2	pre	pre	pre	pre
		3	pre	nop	nop	pre
		4	nop	pre	pre	pre
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	pre	pre	pre	pre
		3	nop	nop	nop	pre
		4	nop	pre	nop	pre
	dug	2	pre	nop	nop	pre
		3	pre	nop	pre	pre
		4	nop	nop	pre	pre
	geese	2	pre	pre	pre	pre
		3	pre	nop	nop	pre
		4	nop	nop	pre	pre
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	pre	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	pre	pre
		4	nop	nop	nop	nop
DF2	bad	2	pre	nop	pre	pre
		3	pre	nop	pre	nop
		4	pre	pre	pre	pre
	beat	2	pre	pre	pre	pre
		3	pre	pre	pre	pre
		4	pre	pre	nop	pre
	bill	2	pre	pre	pre	pre
		3	pre	pre	nop	pre
		4	pre	nop	pre	pre
	bunch	2	pre	pre	pre	pre
		3	nop	nop	nop	pre
		4	nop	nop	nop	pre
	cake	2	nop	nop	nop	pre
		3	nop	nop	nop	pre
		4	nop	pre	pre	pre
	coat	2	pre	pre	pre	pre
		3	nop	pre	nop	pre
		4	pre	pre	pre	pre
	cash	2	pre	pre	pre	pre
		3	nop	pre	nop	pre
		4	nop	pre	nop	pre
	dock	2	pre	nop	pre	pre
		3	pre	nop	pre	pre
		4	nop	nop	pre	pre
	dug	2	pre	pre	pre	pre
		3	pre	pre	pre	pre
		4	nop	nop	nop	pre
	geese	2	pre	nop	pre	pre
		3	pre	nop	nop	pre
		4	pre	nop	pre	pre
	pat	2	pre	pre	nop	pre
		3	nop	nop	nop	pre
		4	nop	pre	nop	pre
	pit	2	nop	pre	nop	pre
		3	pre	pre	pre	pre
		4	nop	nop	nop	nop
DF4	bad	2	nop	nop	pre	nop
		3	pre	pre	pre	pre
		4	nop	nop	nop	nop
	beat	2	pre	pre	pre	pre
		3	vwl	nop	nop	vwl
		4	pre	vwl	vwl	vwl
	bill	2	vwl	nop	vwl	vwl
		3	pre	pre	pre	pre
		4	nop	nop	nop	pre
	bunch	2	nop	nop	nop	pre
		3	pre	pre	nop	pre
		4	pre	vwl	nop	nop
	cake	2	pre	pre	pre	pre
		3	vwl	pre	pre	pre
		4	vwl	nop	pre	pre
	coat	2	pre	nop	pre	pre
		3	nop	nop	nop	nop
		4	vwl	pre	vwl	vwl
	cash	2	pre	nop	pre	pre
		3	pre	pre	pre	vwl
		4	nop	pre	pre	pre
	dock	2	pre	pre	pre	pre
		3	pre	pre	pre	pre
		4	pre	pre	pre	pre
	dug	2	pre	pre	pre	pre
		3	pre	pre	pre	pre
		4	nop	pre	vwl	nop
	geese	2	vwl	nop	vwl	vwl
		3	vwl	pre	pre	vwl
		4	vwl	pre	nop	pre
	pat	2	pre	pre	pre	pre
		3	pre	nop	nop	pre
		4	nop	nop	nop	nop
	pit	2	pre	pre	pre	pre
		3	pre	pre	pre	pre
		4	nop	nop	nop	nop
	tile	2	pre	pre	pre	nop
		3	pre	pre	pre	pre
		4	pre	nop	pre	pre
DF3	bad	2	pre	pre	pre	pre
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	nop
		3	pre	pre	nop	pre
		4	nop	nop	nop	nop
	bill	2	nop	pre	nop	pre
		3	nop	pre	nop	pre
		4	nop	nop	nop	nop
	bunch	2	nop	nop	nop	nop
		3	nop	nop	pre	pre
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	pre
		3	nop	nop	nop	pre
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	pre
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	pre	pre	nop	pre
		4	nop	nop	nop	pre
	dug	2	nop	nop	nop	nop
		3	pre	nop	nop	pre
		4	pre	pre	nop	pre
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	pre
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	pre	nop	pre
		3	nop	nop	nop	nop
		4	pre	pre	nop	nop
	tile	2	pre	nop	nop	pre
		3	nop	nop	nop	nop
		4	nop	nop	nop	pre
DM1	bad	2	nop	nop	pre	pre
		3	nop	nop	pre	pre
		4	nop	nop	nop	nop
	beat	2	nop	nop	pre	pre
		3	pre	nop	pre	pre
		4	nop	nop	nop	pre
	bill	2	nop	nop	nop	pre
		3	nop	nop	nop	nop
		4	nop	nop	nop	pre
	bunch	2	nop	nop	nop	nop
		3	nop	nop	nop	pre
		4	nop	nop	nop	pre
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop

	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	pre	pre
		4	nop	nop	nop	pre
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	pre	pre
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	pre	pre	pre	pre
		3	nop	nop	nop	pre
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	pre	pre
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
DM2	bad	2	nop	nop	nop	nop
		3	nop	nop	pre	pre
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bill	2	nop	nop	pre	nop
		3	nop	nop	nop	pre
		4	nop	nop	nop	pre
	bunch	2	nop	nop	nop	nop
		3	nop	nop	nop	pre
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	vwl	nop	vwl
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	pre
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	nop	pre
		4	nop	nop	nop	nop
DM3	bad	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	pre	nop	nop	pre
	beat	2	nop	nop	nop	pre
		3	pre	nop	pre	pre
		4	pre	nop	nop	pre
	bill	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bunch	2	pre	nop	pre	pre
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cake	2	vwl	pre	nop	pre
		3	pre	nop	nop	pre
		4	nop	nop	nop	pre
	coat	2	vwl	nop	nop	vwl
		3	pre	pre	pre	pre
		4	vwl	nop	nop	vwl
	cash	2	nop	nop	nop	vwl
		3	vwl	vwl	pre	vwl
		4	vwl	nop	nop	vwl
	dock	2	nop	pre	pre	pre
		3	nop	nop	nop	nop
		4	pre	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	nop	nop	pre	pre
		3	pre	pre	pre	pre
		4	nop	nop	pre	pre
	pat	2	vwl	nop	vwl	vwl
		3	vwl	nop	nop	vwl
		4	vwl	nop	nop	vwl
	pit	2	pre	pre	vwl	nop
		3	vwl	nop	nop	pre
		4	vwl	nop	nop	vwl
	tile	2	nop	nop	nop	nop
		3	vwl	nop	nop	vwl
		4	vwl	pre	pre	vwl
DM4	bad	2	pre	pre	pre	pre
		3	nop	nop	nop	pre
		4	nop	nop	nop	pre
	beat	2	pre	pre	pre	pre
		3	nop	nop	pre	pre
		4	nop	nop	nop	nop
	bill	2	pre	pre	nop	pre
		3	pre	nop	pre	pre
		4	pre	pre	pre	pre
	bunch	2	pre	pre	pre	pre
		3	pre	pre	pre	pre
		4	pre	pre	pre	pre
	cake	2	nop	nop	nop	pre
		3	nop	nop	nop	nop
		4	pre	pre	pre	pre
	coat	2	pre	nop	nop	pre
		3	pre	pre	nop	pre
		4	pre	pre	pre	pre
	cash	2	pre	nop	nop	pre
		3	nop	nop	pre	pre
		4	pre	pre	pre	pre
	dock	2	pre	pre	pre	pre
		3	pre	nop	nop	pre
		4	pre	pre	pre	pre
	dug	2	nop	nop	pre	pre
		3	pre	pre	pre	pre
		4	pre	nop	pre	pre
	geese	2	pre	pre	pre	pre
		3	pre	pre	vwl	pre
		4	pre	pre	vwl	pre
	pat	2	nop	nop	nop	pre
		3	pre	nop	nop	pre
		4	nop	nop	nop	nop
	pit	2	nop	pre	vwl	pre
		3	nop	nop	nop	nop
		4	pre	pre	pre	pre
	tile	2	pre	pre	pre	pre
		3	pre	pre	pre	pre
		4	pre	nop	pre	pre
NF1	bad	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bill	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bunch	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
NF2	bad	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bill	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bunch	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
NF4	bad	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bill	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bunch	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop

	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
NM1	bad	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bill	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bunch	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
NM2	bad	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bill	2	nop	nop	pre	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bunch	2	nop	nop	nop	nop
		3	nop	nop	pre	nop
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
NM3	bad	2	nop	nop	nop	nop
		3	nop	nop	nop	pre
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	pre
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bill	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bunch	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
NM4	bad	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	beat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bill	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	bunch	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cake	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	coat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	cash	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dock	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	dug	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	geese	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pat	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	pit	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop
	tile	2	nop	nop	nop	nop
		3	nop	nop	nop	nop
		4	nop	nop	nop	nop

F.2.2 Question 2

The listener answers to Question 2 are abbreviated as follows: Voiced = `vcd`, Voiceless = `vcls`. A '-' appears if the listener answered Vowel in Question 1 for that particular token.

DF1	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd

	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
DF2	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
DF3	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcls	vcd
		4	vcd	vcd	vcls	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcls	vcls	vcls	vcls
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcls	vcd
		4	vcd	vcd	vcls	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcls	vcd	vcd
		4	vcd	vcd	vcd	vcls
	dug	2	vcls	vcls	vcls	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcd	vcls	vcls
	pit	2	vcd	vcd	vcd	vcd
		3	vcd	vcls	vcd	vcd
		4	vcd	vcd	vcd	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
DM1	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	-	vcd	-
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
DM2	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	-	vcd	-
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	-	vcd	vcd
	pat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcls	vcls	vcls	vcls
DM3	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcls	vcls	vcls	vcls
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcls	vcls	vcls
		4	vcd	vcd	vcd	vcd
	cake	2	-	vcd	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcd	vcd	vcd	vcd
	coat	2	-	vcd	vcls	-
		3	vcls	vcd	vcls	vcls
		4	-	vcls	vcls	-
	cash	2	vcd	vcls	vcd	-
		3	-	-	vcls	-
		4	-	vcd	vcls	-
	dock	2	vcd	vcd	vcd	vcd
		3	vcls	vcls	vcls	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	-	vcls	-	-
		3	-	vcd	vcls	-
		4	-	vcd	vcls	-
	pit	2	vcls	vcd	-	vcls
		3	-	vcd	vcls	vcd
		4	-	vcd	vcls	-
	tile	2	vcls	vcls	vcls	vcls
		3	-	vcd	vcls	-
		4	-	vcd	vcls	-
DM4	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcls
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcls	vcd	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcd	vcd	vcd	vcd
		4	vcls	vcd	vcls	vcls
	cash	2	vcd	vcd	vcd	vcd
		3	vcls	vcls	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dock	2	vcd	vcd	vcd	vcd
		3	vcls	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	-	vcd
	pat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcls	vcls	vcls	vcls

	pit	2	vcd	vcd	-	vcd
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcd	vcd	vcd	vcd
NF1	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
NF2	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcd
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
NF3	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
NF4	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
NM1	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcls	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
NM2	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcd	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
NM3	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcd	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls

	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
NM4	bad	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	beat	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bill	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	bunch	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	cake	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	coat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	cash	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	dock	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	dug	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	geese	2	vcd	vcd	vcd	vcd
		3	vcd	vcd	vcd	vcd
		4	vcd	vcd	vcd	vcd
	pat	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	pit	2	vcls	vcls	vcls	vcls
		3	vcls	vcls	vcls	vcls
		4	vcls	vcls	vcls	vcls
	tile	2	vcls	vcls	vcls	vcls
		3	vcls	vcd	vcls	vcls
		4	vcls	vcls	vcls	vcls

F.2.3 Question 3

The listener answers to Question 3 are abbreviated as follows: Labial = `lab`, Alveolar = `alv`, and Velar = `vel`. A '-' appears if the listener answered Vowel in Question 1 for that particular token.

DF1	bad	2	lab	lab	lab	lab
		3	lab	alv	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	dent	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	dent	alv	alv
		4	alv	alv	alv	alv
DF2	bad	2	lab	alv	lab	alv
		3	lab	lab	lab	lab
		4	lab	lab	lab	labd
	beat	2	lab	alv	lab	alv
		3	lab	dent	lab	lab
		4	lab	alv	lab	lab
	bill	2	lab	lab	lab	labd
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	pal	vel	pal
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	lab	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	dent	alv	alv
	dug	2	lab	alv	lab	alv
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	geese	2	vel	glot	glot	-
		3	lab	-	-	-
		4	vel	vel	vel	vel
	pat	2	lab	-	-	-
		3	glot	glot	glot	glot
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	-	lab	lab
	tile	2	vel	glot	glot	glot
		3	-	glot	glot	glot
		4	-	dent	glot	alv
DF4	bad	2	lab	lab	lab	lab
		3	lab	alv	lab	alv
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	-	glot	glot	-
		4	lab	-	-	-
	bill	2	-	glot	-	-
		3	glot	glot	glot	glot
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	-	lab	lab
	cake	2	vel	glot	glot	glot
		3	-	glot	glot	glot
		4	-	dent	glot	alv
	coat	2	glot	glot	glot	glot
		3	glot	glot	glot	glot
		4	-	glot	-	-
	cash	2	vel	glot	glot	glot
		3	vel	glot	vel	-
		4	vel	glot	glot	glot
	dock	2	alv	dent	alv	alv
		3	alv	pal	pal	alv
		4	alv	pal	pal	alv
	dug	2	alv	dent	alv	alv
		3	alv	pal	pal	alv
		4	lab	glot	-	lab
	geese	2	-	glot	-	-
		3	-	glot	glot	-
		4	-	vel	glot	alv
	pat	2	lab	alv	glot	lab
		3	lab	lab	lab	lab
		4	lab	dent	lab	lab
	pit	2	glot	glot	glot	glot
		3	glot	glot	glot	glot
		4	lab	lab	lab	lab
	tile	2	alv	vel	glot	alv
		3	glot	glot	glot	glot
		4	alv	alv	glot	glot
DF3	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	alv	lab	vel	alv
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	alv
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dug	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	alv	alv	vel	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	pal	lab	lab
		4	lab	lab	lab	alv
	tile	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	alv	pal	vel	vel
DM1	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab

	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	dent	alv	alv
		4	alv	alv	alv	vel
	dug	2	alv	dent	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	vel	lab	vel
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	vel	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
DM2	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	pal	alv	alv
		3	alv	alv	alv	alv
		4	alv	pal	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	-	alv	-
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
DM3	bad	2	lab	dent	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	vel	alv
	beat	2	alv	alv	alv	alv
		3	vel	vel	pal	alv
		4	vel	pal	alv	alv
	bill	2	lab	dent	alv	alv
		3	alv	pal	pal	alv
		4	alv	alv	alv	pal
	bunch	2	lab	lab	vel	lab
		3	lab	lab	alv	alv
		4	alv	alv	alv	alv
	cake	2	-	glot	glot	vel
		3	vel	alv	alv	alv
		4	vel	vel	vel	vel
	coat	2	-	glot	glot	-
		3	vel	glot	vel	vel
		4	-	glot	glot	-
	cash	2	alv	pal	pal	-
		3	-	-	glot	-
		4	-	glot	glot	-
	dock	2	alv	pal	alv	alv
		3	vel	alv	alv	alv
		4	vel	vel	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	pal	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	vel	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
NF1	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	pal	alv
		3	alv	alv	alv	alv
		4	vel	vel	-	alv
	pat	2	lab	lab	alv	lab
		3	lab	labd	vel	alv
		4	lab	vel	alv	lab
	pit	2	lab	glot	-	lab
		3	vel	vel	vel	vel
		4	vel	glot	alv	vel
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
DM4	bad	2	vel	vel	alv	vel
		3	lab	lab	alv	lab
		4	lab	alv	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	vel	alv	glot
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	pal	vel
		3	alv	dent	vel	alv
		4	vel	alv	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	alv	alv	vel
		3	alv	alv	glot	glot
		4	vel	glot	pal	glot
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	pal	alv
		3	alv	alv	alv	alv
		4	vel	vel	-	alv
	pat	2	lab	lab	alv	lab
		3	lab	labd	vel	alv
		4	lab	vel	alv	lab
	pit	2	lab	glot	-	lab
		3	vel	vel	vel	vel
		4	vel	glot	alv	vel
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
NF2	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
NF3	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
NF4	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab

	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
NM1	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
NM2	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
NM3	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	dent	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
NM4	bad	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	beat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bill	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	bunch	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	cake	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	coat	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	cash	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	dock	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	dug	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv
	geese	2	vel	vel	vel	vel
		3	vel	vel	vel	vel
		4	vel	vel	vel	vel
	pat	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	pit	2	lab	lab	lab	lab
		3	lab	lab	lab	lab
		4	lab	lab	lab	lab
	tile	2	alv	alv	alv	alv
		3	alv	alv	alv	alv
		4	alv	alv	alv	alv

F.2.4 Question 4

The listener answers to Question 4 are abbreviated as follows: Stop = `stop`, Fricative = `fric`, Glide = `glid`, Nasal = `nasl`, Liquid /l/ = `liql`, Liquid /r/ = `liqr` and Affricate = `aff`. A '-' appears if the listener answered Vowel in Question 1 for that particular token.

DF1	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop

	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
DF2	bad	2	stop	nasl	stop	nasl
		3	stop	stop	stop	stop
		4	stop	stop	stop	fric
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	fric
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	fric	stop	fric
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	nasl	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	fric
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	fric	fric	fric
		4	stop	stop	stop	stop
	tile	2	stop	stop	nasl	fric
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
DF4	bad	2	nasl	nasl	nasl	nasl
		3	nasl	nasl	nasl	nasl
		4	nasl	nasl	nasl	nasl
	beat	2	nasl	nasl	nasl	nasl
		3	-	stop	stop	-
		4	stop	-	-	-
	bill	2	-	stop	-	-
		3	fric	fric	fric	fric
		4	nasl	nasl	nasl	nasl
	bunch	2	stop	stop	stop	stop
		3	stop	nasl	nasl	nasl
		4	stop	-	nasl	nasl
	cake	2	stop	fric	fric	fric
		3	-	stop	stop	stop
		4	-	fric	stop	fric
	coat	2	fric	fric	fric	fric
		3	fric	fric	fric	fric
		4	-	stop	-	-
	cash	2	stop	fric	fric	fric
		3	stop	fric	stop	-
		4	stop	stop	stop	fric
	dock	2	nasl	nasl	nasl	nasl
		3	nasl	glid	nasl	nasl
		4	nasl	glid	glid	nasl
	dug	2	nasl	nasl	nasl	nasl
		3	nasl	glid	glid	nasl
		4	stop	stop	-	stop
	geese	2	-	stop	-	-
		3	-	stop	stop	-

		4	-	nasl	stop	nasl
	pat	2	stop	stop	fric	stop
		3	stop	stop	stop	stop
		4	stop	fric	stop	stop
	pit	2	fric	fric	fric	fric
		3	fric	fric	fric	fric
		4	stop	stop	stop	stop
	tile	2	stop	fric	fric	stop
		3	fric	fric	fric	fric
		4	stop	stop	fric	fric
DF3	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	fric	stop	stop	fric
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	fric
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	fric	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	nasl	nasl	stop	nasl
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	fric
	tile	2	stop	fric	stop	stop
		3	stop	stop	stop	stop
		4	stop	fric	stop	stop
DM1	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	nasl	nasl	stop	nasl
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
DM2	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop

		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	-	stop	-
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
DM3	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	aff	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	aff	stop
		4	stop	stop	stop	aff
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	-	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	-	stop	stop	-
		3	stop	stop	stop	stop
		4	-	stop	stop	-
	cash	2	glid	glid	glid	-
		3	-	-	stop	-
		4	-	stop	stop	-
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	fric
		3	stop	stop	stop	aff
		4	stop	stop	stop	aff
	pat	2	-	stop	-	-
		3	-	stop	stop	-
		4	-	stop	stop	-
	pit	2	stop	stop	-	stop
		3	-	stop	stop	stop
		4	-	stop	stop	-
	tile	2	stop	stop	stop	stop
		3	-	stop	stop	-
		4	-	stop	stop	-
DM4	bad	2	stop	stop	stop	stop
		3	nasl	nasl	nasl	nasl
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	nasl	nasl	fric
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	glid	stop
		3	stop	stop	stop	stop
		4	stop	nasl	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	glid	stop	stop
		4	stop	fric	stop	stop

	cash	2	stop	stop	stop	stop
		3	stop	stop	fric	fric
		4	stop	glid	glid	fric
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	glid	glid	glid
		3	stop	stop	stop	stop
		4	stop	fric	-	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	fric	-	stop
		3	stop	stop	stop	stop
		4	stop	fric	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
NF1	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
NF2	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
NF3	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
NM1	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
NM2	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
NM3	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop

	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
NM4	bad	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	beat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bill	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	bunch	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cake	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	coat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	cash	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dock	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	dug	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	geese	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pat	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	pit	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop
	tile	2	stop	stop	stop	stop
		3	stop	stop	stop	stop
		4	stop	stop	stop	stop

F.2.5 Question 5

A '-' appears if the listener answered either Vowel in Question 1 or anything other than Stop in Question 4 for that particular token.

DF1	bad	2	fair	good	fair	poor
		3	good	fair	good	fair
		4	fair	fair	good	poor
	beat	2	fair	fair	fair	fair
		3	fair	fair	fair	fair
		4	good	fair	fair	fair
	bill	2	fair	fair	good	poor
		3	fair	fair	fair	fair
		4	fair	fair	fair	fair
	bunch	2	fair	fair	fair	fair
		3	fair	fair	fair	good
		4	good	fair	fair	fair
	cake	2	good	good	good	fair
		3	good	good	fair	good
		4	good	fair	good	good
	coat	2	good	good	good	good
		3	fair	poor	good	good
		4	good	good	good	good
	cash	2	good	fair	good	good
		3	good	fair	good	good
		4	good	fair	good	good
	dock	2	fair	fair	fair	fair
		3	fair	fair	fair	fair
		4	fair	fair	fair	fair
	dug	2	fair	fair	fair	poor
		3	fair	fair	fair	fair
		4	good	fair	fair	fair
	geese	2	fair	fair	fair	good
		3	fair	fair	good	fair
		4	fair	good	fair	fair
	pat	2	good	fair	fair	good
		3	good	fair	fair	good
		4	good	fair	good	fair
	pit	2	good	fair	fair	good
		3	fair	good	good	good
		4	good	fair	fair	fair
	tile	2	good	fair	good	good
		3	good	fair	good	good
		4	good	good	good	good
DF2	bad	2	poor	-	fair	-
		3	fair	fair	fair	fair
		4	fair	poor	fair	-
	beat	2	fair	fair	fair	fair
		3	fair	poor	fair	poor
		4	fair	poor	good	poor
	bill	2	fair	fair	fair	-
		3	poor	poor	fair	poor
		4	fair	fair	good	fair
	bunch	2	fair	poor	fair	poor
		3	good	good	good	fair
		4	fair	fair	fair	fair
	cake	2	fair	-	fair	-
		3	good	fair	good	fair
		4	fair	fair	good	fair
	coat	2	fair	poor	fair	fair
		3	fair	fair	fair	fair
		4	fair	poor	fair	poor
	cash	2	fair	poor	fair	poor
		3	fair	-	-	-
		4	good	poor	fair	poor
	dock	2	poor	poor	-	-
		3	good	fair	fair	fair
		4	good	fair	fair	fair
DF4	bad	2	-	-	-	-
		3	-	-	-	-
		4	-	-	-	-
	beat	2	-	-	-	-
		3	-	poor	fair	-
		4	poor	-	-	-
	bill	2	-	fair	-	-
		3	-	-	-	-
		4	-	-	-	-
	bunch	2	fair	poor	fair	poor
		3	fair	-	-	-
		4	poor	-	-	-
	cake	2	poor	-	-	-
		3	-	-	-	-
		4	-	poor	-	-
	cash	2	poor	-	-	-
		3	poor	-	poor	-
		4	fair	fair	fair	-
	dock	2	-	-	-	-
		3	-	-	-	-
		4	-	-	-	-
	dug	2	-	-	-	-
		3	-	-	-	-
		4	poor	poor	-	poor
	geese	2	-	fair	-	-
		3	-	poor	fair	-
		4	-	-	fair	-
	pat	2	poor	poor	-	poor
		3	poor	poor	fair	poor
		4	fair	-	fair	fair
	pit	2	-	-	-	-
		3	-	-	-	-
		4	fair	fair	good	fair
	tile	2	poor	-	-	fair
		3	-	-	-	-
		4	poor	poor	-	-
DF3	bad	2	poor	fair	fair	fair
		3	fair	fair	fair	fair
		4	fair	fair	fair	fair
	beat	2	fair	poor	poor	poor
		3	-	fair	fair	-
		4	good	poor	fair	fair
	bill	2	fair	poor	fair	poor
		3	fair	poor	fair	-
		4	good	fair	good	good
	bunch	2	fair	fair	fair	good
		3	fair	fair	fair	poor
		4	fair	fair	fair	good
	cake	2	poor	-	fair	poor
		3	good	fair	fair	fair
		4	good	good	good	fair
	coat	2	fair	fair	fair	fair
		3	fair	fair	fair	fair
		4	fair	poor	fair	fair
	cash	2	fair	fair	good	fair
		3	fair	poor	fair	fair
		4	fair	fair	good	fair
	dock	2	poor	good	good	fair
		3	poor	fair	good	fair
		4	fair	fair	good	fair
	dug	2	poor	fair	fair	fair
		3	poor	fair	good	poor
		4	-	-	good	-
	geese	2	fair	fair	good	fair
		3	poor	poor	fair	poor
		4	fair	fair	fair	fair
	pat	2	fair	fair	fair	fair
		3	fair	fair	fair	fair
		4	fair	fair	fair	fair
	pit	2	fair	poor	fair	poor
		3	fair	poor	good	fair
		4	fair	poor	fair	-
	tile	2	poor	-	poor	poor
		3	fair	poor	fair	fair
		4	fair	-	poor	poor
DM1	bad	2	good	fair	fair	fair
		3	good	good	fair	fair
		4	good	fair	good	fair

	beat	2	good	fair	good	fair
		3	fair	fair	fair	fair
		4	good	fair	fair	fair
	bill	2	good	good	fair	fair
		3	good	fair	good	good
		4	good	fair	good	fair
	bunch	2	good	fair	good	good
		3	good	fair	good	fair
		4	fair	fair	good	fair
	cake	2	good	fair	good	good
		3	good	fair	good	fair
		4	good	good	fair	good
	coat	2	good	fair	good	good
		3	fair	fair	good	fair
		4	good	fair	good	good
	cash	2	fair	good	good	fair
		3	fair	good	good	fair
		4	good	good	good	fair
	dock	2	fair	fair	good	fair
		3	good	fair	good	good
		4	fair	fair	fair	fair
	dug	2	good	fair	fair	fair
		3	good	fair	good	good
		4	good	fair	fair	fair
	geese	2	fair	fair	fair	fair
		3	fair	fair	fair	fair
		4	good	fair	fair	good
	pat	2	fair	fair	fair	fair
		3	good	fair	fair	good
		4	good	fair	fair	fair
	pit	2	fair	fair	good	good
		3	good	fair	fair	fair
		4	good	fair	fair	fair
	tile	2	good	fair	fair	fair
		3	good	good	good	good
		4	good	fair	fair	good
DM2	bad	2	good	good	good	good
		3	good	good	fair	fair
		4	good	good	good	good
	beat	2	fair	good	fair	good
		3	good	fair	good	fair
		4	good	fair	fair	good
	bill	2	good	fair	good	good
		3	good	good	good	fair
		4	good	fair	fair	fair
	bunch	2	good	good	good	good
		3	good	good	good	fair
		4	good	good	good	good
	cake	2	good	good	good	good
		3	good	good	fair	fair
		4	good	fair	fair	good
	coat	2	good	good	good	fair
		3	good	fair	fair	fair
		4	good	fair	good	good
	cash	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dock	2	fair	good	good	good
		3	good	fair	good	fair
		4	fair	poor	fair	poor
	dug	2	good	fair	good	fair
		3	good	fair	fair	good
		4	fair	-	fair	-
	geese	2	good	good	fair	fair
		3	good	fair	good	fair
		4	good	good	good	good
	pat	2	good	good	good	good
		3	good	fair	good	good
		4	good	good	good	good
	pit	2	good	good	good	good
		3	good	good	good	good
		4	good	good	fair	good
	tile	2	good	fair	fair	fair
		3	good	fair	fair	poor
		4	fair	fair	fair	fair
DM3	bad	2	good	poor	good	fair
		3	fair	fair	good	fair
		4	poor	fair	good	poor
	beat	2	poor	poor	poor	poor
		3	poor	poor	poor	poor
		4	poor	-	good	poor
	bill	2	fair	fair	good	poor
		3	fair	fair	-	fair
		4	fair	fair	fair	-
	bunch	2	poor	poor	fair	poor
		3	fair	fair	fair	fair
		4	fair	fair	good	fair
	cake	2	-	fair	fair	poor
		3	fair	poor	fair	fair
		4	poor	fair	fair	fair
	coat	2	-	poor	good	-
		3	poor	fair	poor	poor
		4	-	good	fair	-
	cash	2	-	-	-	-
	dock	2	poor	poor	fair	poor
		3	poor	poor	fair	poor
		4	poor	fair	fair	poor
	dug	2	fair	poor	good	fair
		3	good	fair	good	fair
		4	fair	fair	fair	poor
	geese	2	poor	poor	fair	-
		3	poor	poor	fair	-
		4	poor	poor	fair	-
	pat	2	-	fair	-	-
		3	-	poor	fair	-
		4	-	poor	good	-
	pit	2	fair	poor	-	fair
		3	-	poor	fair	poor
		4	-	fair	fair	-
	tile	2	fair	poor	fair	poor
		3	-	poor	good	-
		4	-	poor	fair	-
DM4	bad	2	poor	fair	fair	poor
		3	-	-	-	-
		4	fair	fair	fair	poor
	beat	2	poor	poor	fair	fair
		3	fair	fair	fair	fair
		4	good	fair	good	fair
	bill	2	poor	poor	fair	poor
		3	poor	poor	poor	fair
		4	poor	-	-	-
	bunch	2	poor	poor	fair	poor
		3	poor	fair	fair	fair
		4	poor	poor	fair	fair
	cake	2	poor	fair	-	poor
		3	good	fair	good	fair
		4	poor	-	fair	poor
	coat	2	poor	poor	fair	fair
		3	poor	-	fair	poor
		4	poor	-	poor	poor
	cash	2	poor	poor	poor	poor
		3	poor	poor	-	-
		4	poor	-	-	-
	dock	2	poor	fair	poor	poor
		3	poor	fair	good	fair
		4	poor	poor	fair	poor
	dug	2	good	fair	fair	good
		3	poor	fair	fair	fair
		4	poor	fair	fair	poor
	geese	2	poor	-	-	-
		3	poor	poor	fair	poor
		4	poor	-	-	poor
	pat	2	fair	poor	good	poor
		3	poor	poor	good	poor
		4	fair	poor	fair	fair
	pit	2	poor	-	-	poor
		3	fair	fair	fair	fair
		4	poor	-	poor	poor
	tile	2	fair	poor	fair	fair
		3	poor	fair	poor	fair
		4	fair	fair	fair	fair
NF1	bad	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	beat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	bill	2	good	good	fair	good
		3	good	good	good	good
		4	good	good	good	good
	bunch	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cake	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	coat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cash	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dock	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dug	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	geese	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pat	2	good	good	fair	good
		3	good	good	good	good
		4	good	good	good	good
	pit	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	tile	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
NF2	bad	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	beat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	fair	good
	bill	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	bunch	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cake	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	coat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	fair	good
	cash	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dock	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dug	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	geese	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pat	2	good	good	good	good
		3	good	good	fair	good
		4	good	good	good	good
	pit	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	tile	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
NF3	bad	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	beat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	bill	2	good	good	good	good
		3	good	fair	good	good
		4	good	good	good	good
	bunch	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cake	2	good	good	fair	good
		3	good	good	good	good
		4	good	good	fair	good
	coat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cash	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dock	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dug	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	geese	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pit	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	tile	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
NF4	bad	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	beat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good

	bill	2	good	fair	good	good
		3	good	fair	good	good
		4	good	fair	good	good
	bunch	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	fair
	cake	2	good	good	fair	good
		3	good	good	fair	good
		4	good	good	fair	good
	coat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cash	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dock	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dug	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	geese	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pit	2	good	good	good	good
		3	good	fair	good	good
		4	good	good	good	good
	tile	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
NM1	bad	2	good	good	good	good
		3	good	fair	good	good
		4	good	good	good	good
	beat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	bill	2	good	good	good	good
		3	good	fair	good	good
		4	good	good	fair	good
	bunch	2	good	good	good	good
		3	good	good	fair	good
		4	good	fair	fair	fair
	cake	2	good	fair	good	good
		3	good	good	good	good
		4	good	good	good	good
	coat	2	good	good	good	good
		3	good	good	fair	good
		4	good	good	good	good
	cash	2	good	good	fair	good
		3	good	good	good	good
		4	good	good	good	good
	dock	2	good	good	good	fair
		3	good	good	good	good
		4	good	good	good	good
	dug	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	geese	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good

	pat	2	good	fair	good	good
		3	good	good	good	good
		4	good	good	good	good
	pit	2	good	fair	good	good
		3	good	fair	good	good
		4	good	good	good	good
	tile	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
NM2	bad	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	beat	2	good	fair	fair	good
		3	good	good	good	good
		4	good	good	good	good
	bill	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	bunch	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cake	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	coat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cash	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dock	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dug	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	geese	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pit	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	tile	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
NM3	bad	2	good	good	good	good
		3	good	good	good	fair
		4	good	good	good	good
	beat	2	good	good	good	fair
		3	good	good	good	good
		4	good	good	good	good
	bill	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	bunch	2	good	good	good	good
		3	good	good	good	good
		4	good	fair	good	good
	cake	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good

	coat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cash	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dock	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dug	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	geese	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	pit	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	tile	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
NM4	bad	2	good	good	good	good
		3	good	fair	good	good
		4	good	good	good	good
	beat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	bill	2	good	fair	good	good
		3	good	good	good	good
		4	good	good	good	good
	bunch	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cake	2	good	good	fair	good
		3	good	good	good	good
		4	good	good	good	good
	coat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	cash	2	good	good	fair	good
		3	good	good	fair	good
		4	good	good	good	good
	dock	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	dug	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good
	geese	2	good	good	good	good
		3	good	good	good	good
		4	good	good	fair	good
	pat	2	good	good	good	good
		3	good	good	good	good
		4	good	good	fair	good
	pit	2	good	fair	good	good
		3	good	fair	good	good
		4	good	good	good	good
	tile	2	good	good	good	good
		3	good	good	good	good
		4	good	good	good	good

Appendix G

Additional Spectrogram Analysis Results and Experiment Data

G.1 Additional Results

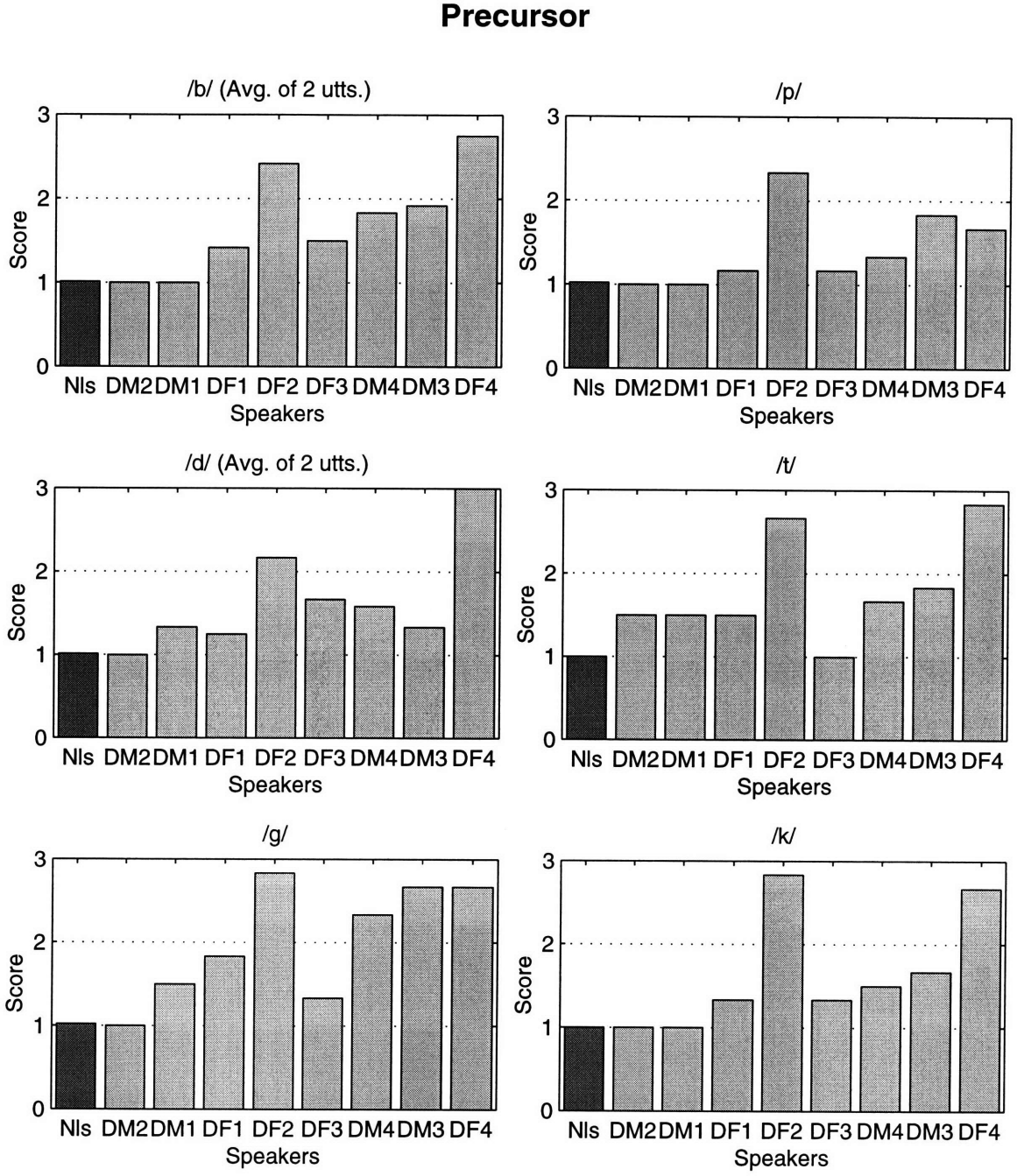

Figure G-1 : Precursor Spectrogram Analysis Attribute results. Ratings averaged across 2 judges, 3 repetitions/utterance and the number of utterances shown, for each speaker. The normal speakers' ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

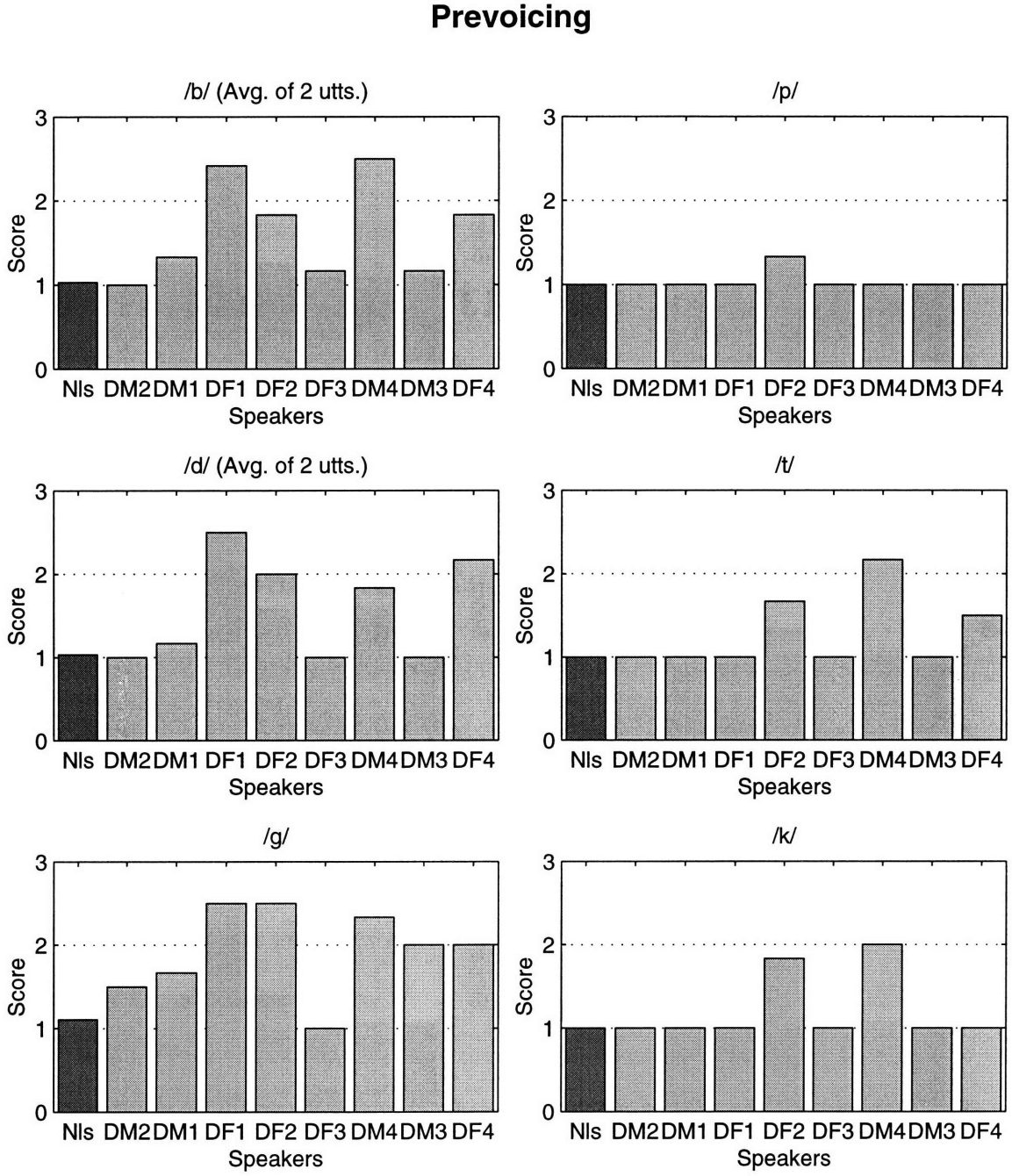

Figure G-2 : Prevoicing Spectrogram Analysis Attribute results. Ratings averaged across 2 judges, 3 repetitions/utterance and the number of utterances shown, for each speaker. The normal speakers' ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

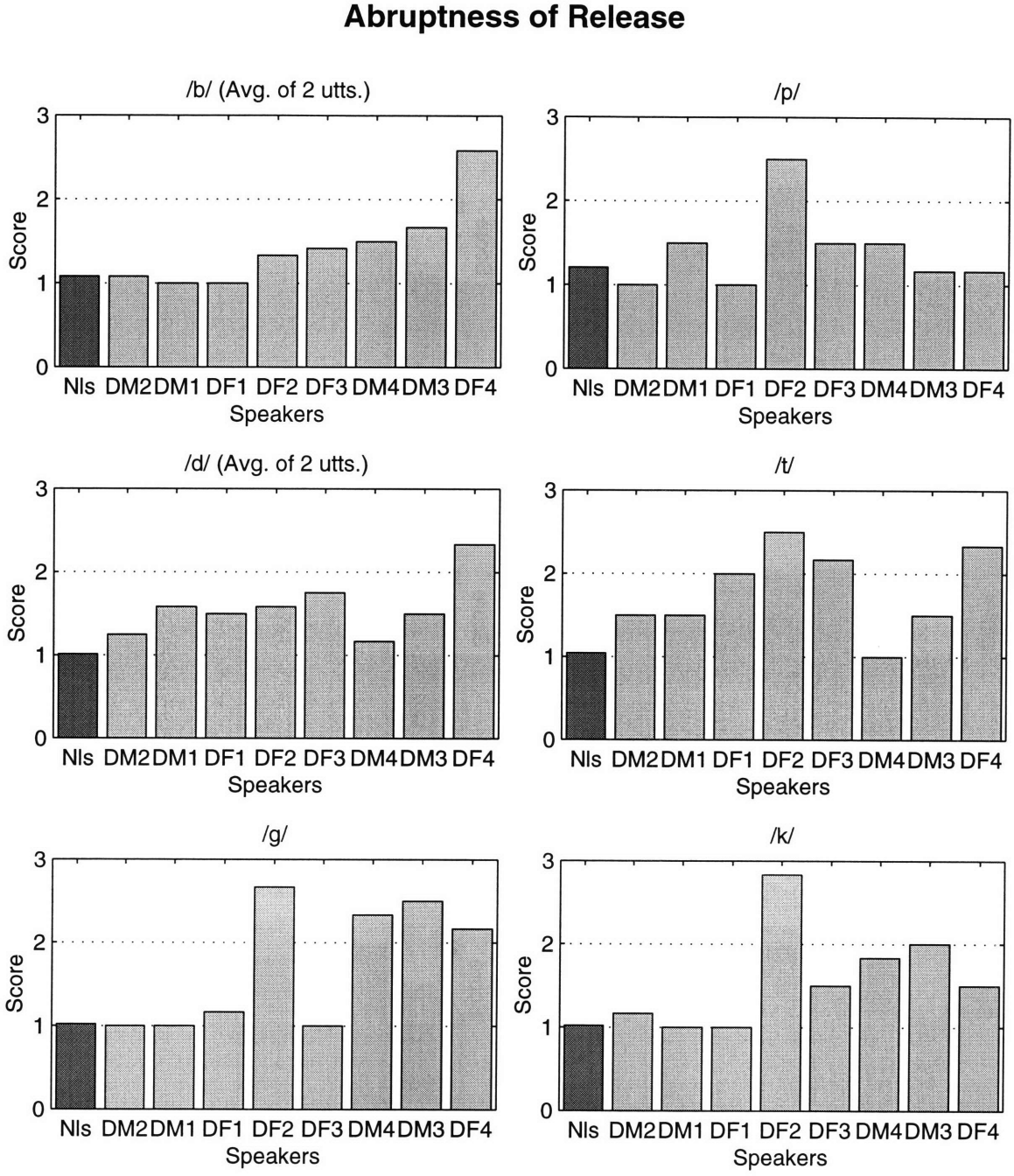

Figure G-3 : Abruptness of Release Spectrogram Analysis Attribute results. Ratings averaged across 2 judges, 3 repetitions/utterance and the number of utterances shown, for each speaker. The normal speakers' ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

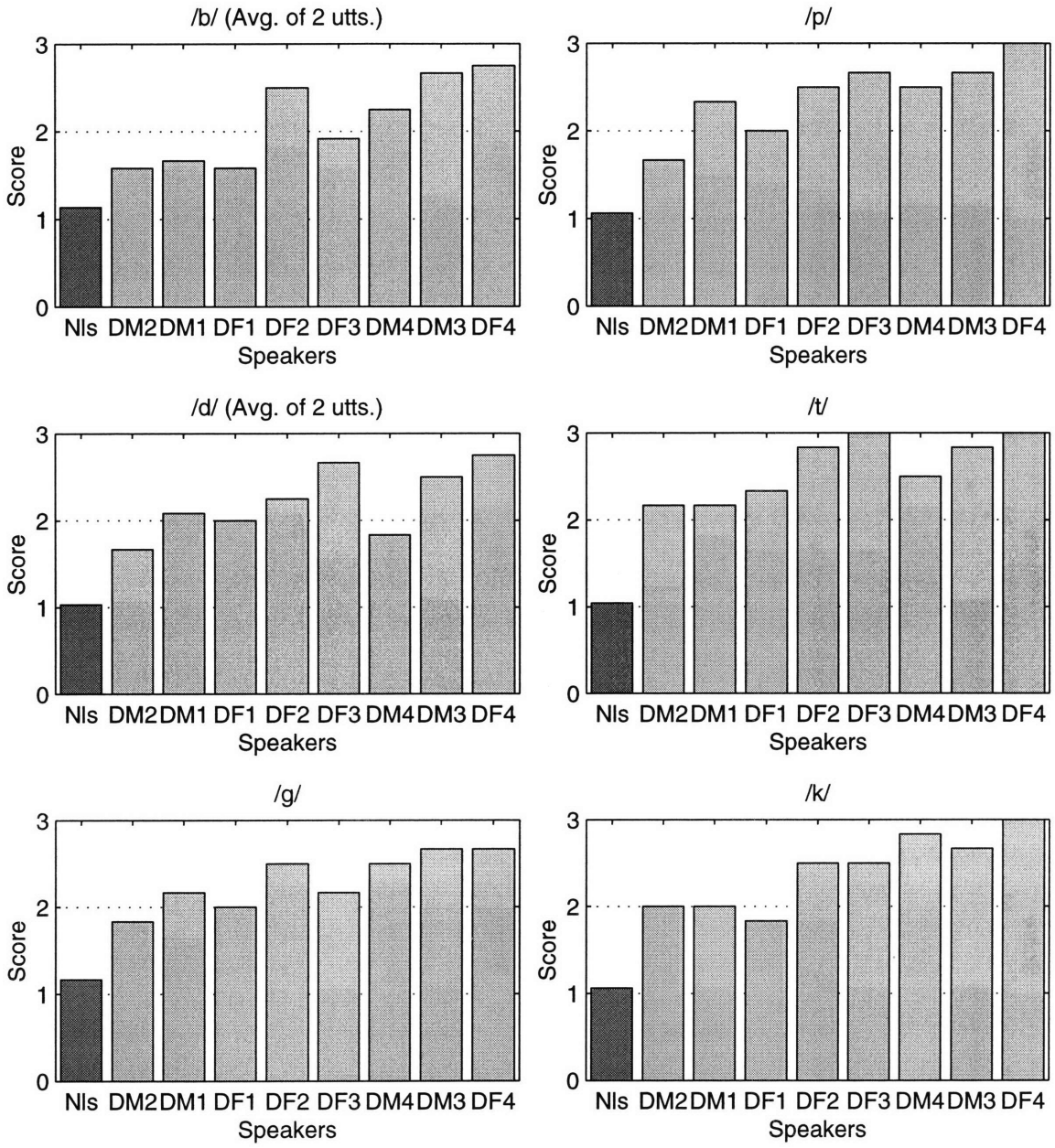

Figure G-4 : Time Course of Release Spectrogram Analysis Attribute results. Ratings averaged across 2 judges, 3 repetitions/utterance and the number of utterances shown, for each speaker. The normal speakers' ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

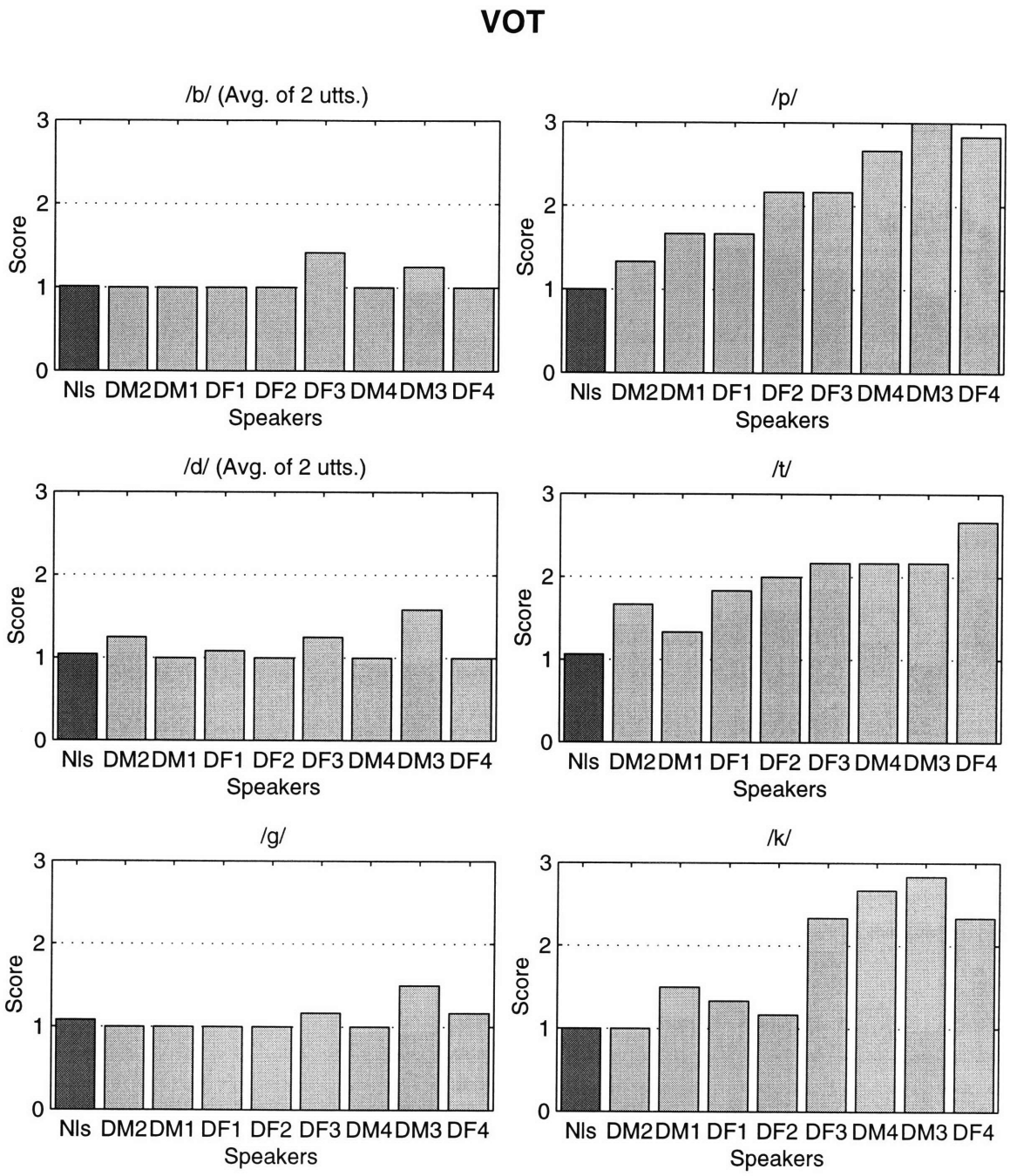

Figure G-5 : Voice Onset Time (VOT) of Release Spectrogram Analysis Attribute results. Ratings averaged across 2 judges, 3 repetitions/utterance and the number of utterances shown, for each speaker. The normal speakers' ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

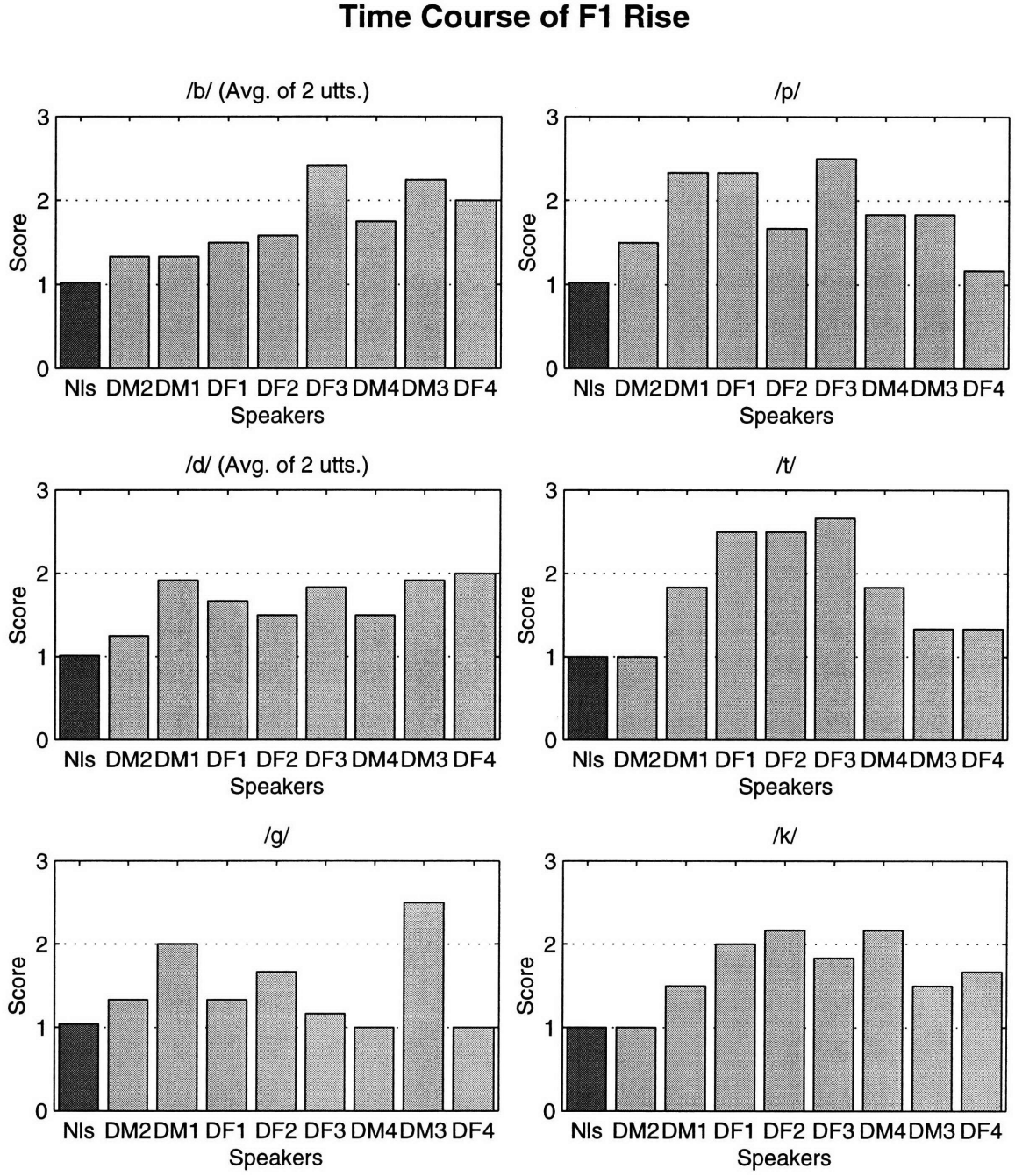

Figure G-6 : Time Course of $F1$ Rise Spectrogram Analysis Attribute results. Ratings averaged across 2 judges, 3 repetitions/utterance and the number of utterances shown, for each speaker. The normal speakers' ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

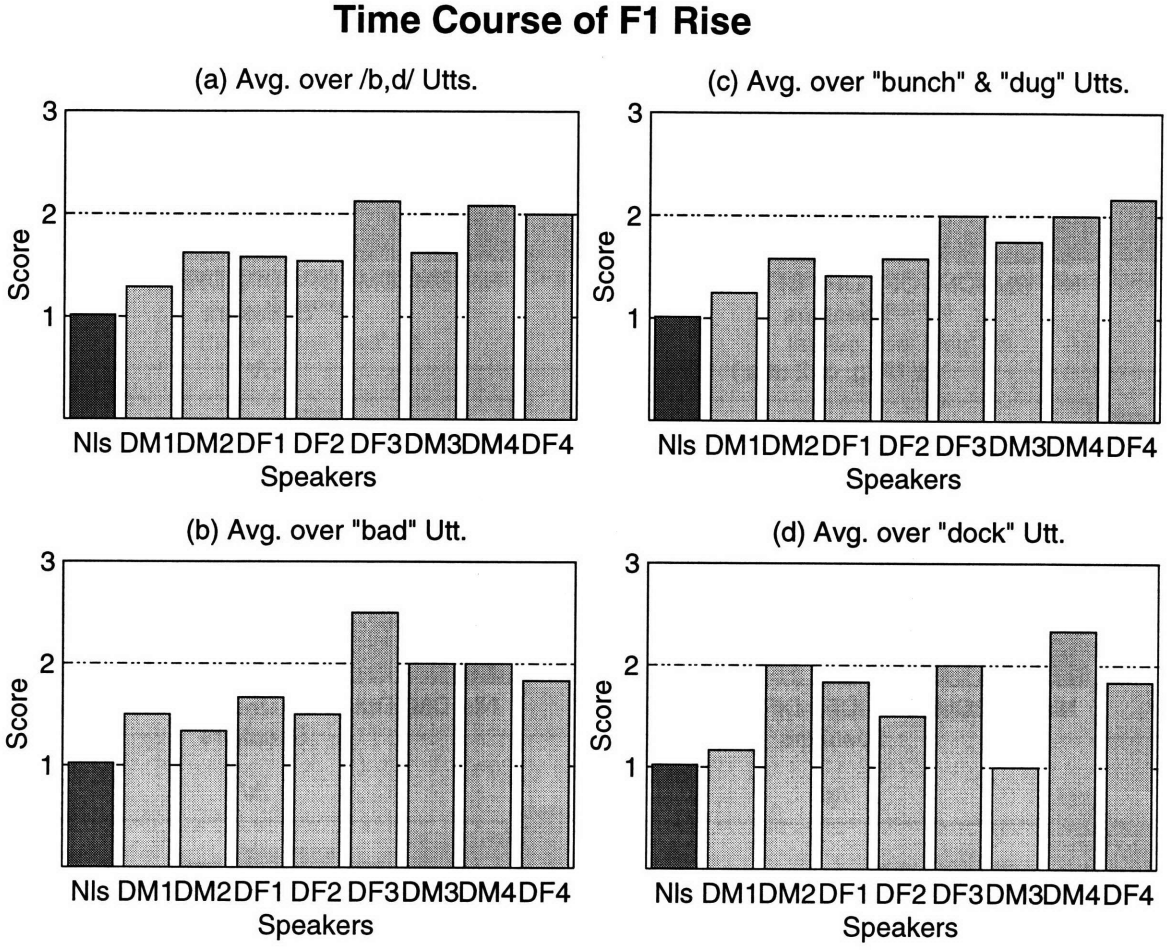

Figure G-7 : Time Course of $F1$ Rise attribute results from Spectrogram Analysis. For each speaker, ratings averaged across 2 judges, 3 repetitions/utterance, and (a) 4 utterances containing either intended word-initial /b/ or /d/, or (b) the utterance bad, or (c) the utterances bunch and dug, or (d) the utterance dock. For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

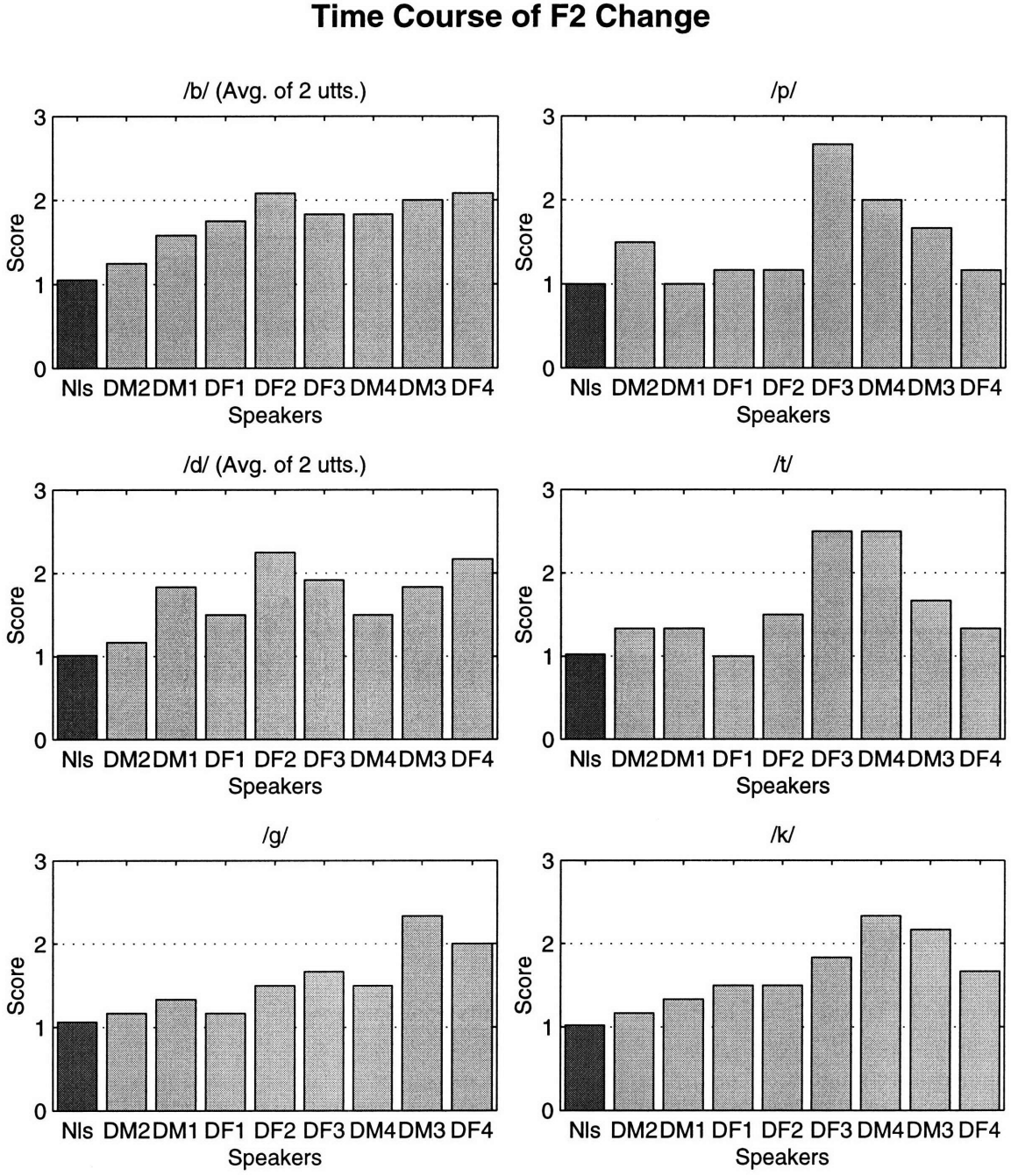

Figure G-8 : Time Course of $F2$ Change Spectrogram Analysis Attribute results. Ratings averaged across 2 judges, 3 repetitions/utterance and the number of utterances shown, for each speaker. The normal speakers' ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

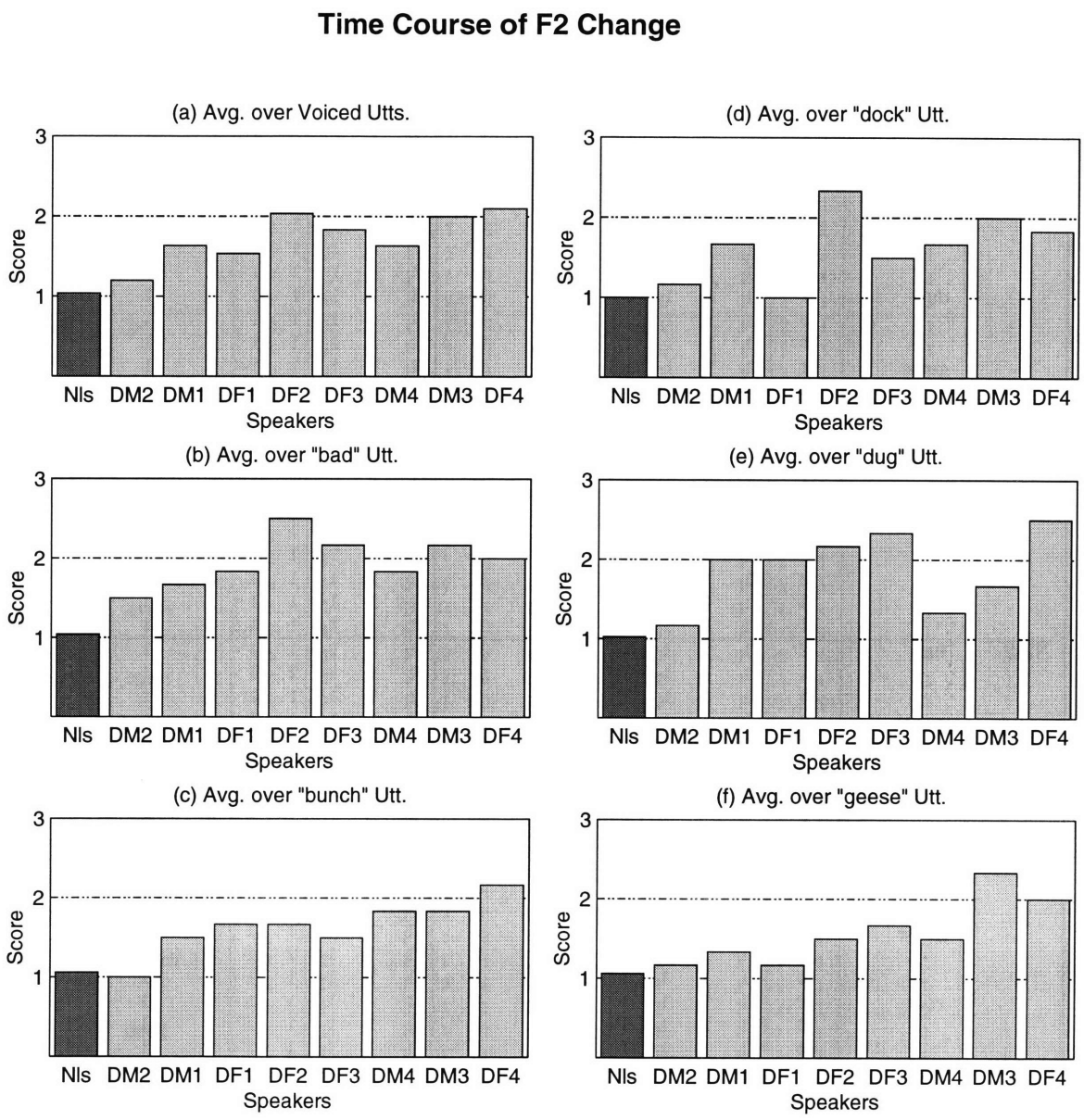

Figure G-9 : Time Course of F2 Change attribute results from Spectrogram Analysis. For each speaker, ratings averaged across 2 judges, 3 repetitions/utterance, and (a) 5 utterances containing either intended word-initial voiced stops, or (b) the utterance bad, or (c) the utterance bunch, or (d) the utterance dock, or (e) the utterance dug, or (f) the utterance geese. For normal speakers, ratings were also averaged across all 8 speakers. The normal (Nls) and dysarthric (DF1–DF4, DM1–DM4) speakers' results are shown from left to right in order of decreasing stop goodness score, as determined in Chapter 4.

G.2 Raw Data

			Judge 1							Judge 2						
Speaker	Utterance	Rep.	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2
NM1	bad	2	1	1	1	1	1	1	1	1	2	2	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	2	1	1	1	1
	bunch	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	2	1	1	1	1	1	2	1	1	1	1
	dock	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dug	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	2	1	1	1	1	1	1	1	1	1
	geese	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	2	1	1	1	1
		4	1	1	1	1	2	1	1	1	1	1	1	1	1	1
	coat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	pat	2	1	1	2	1	1	1	1	1	1	2	1	1	1	1
		3	1	1	2	1	1	1	1	1	1	2	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	tile	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NM2	bad	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	2	1
	bunch	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dock	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	2	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dug	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	geese	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	2	1	1	1	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	coat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	pat	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	2	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	tile	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NM3	bad	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	2	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	bunch	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	2	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	2	1	1	1	1	1	1	1	1	1	1	2
	dock	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	2	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dug	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1

			Judge 1							Judge 2						
Speaker	Utterance	Rep.	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	2	1	1	1	1	1
	geese	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	2	1	1	2
	coat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	pat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	tile	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NM4	bad	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	bunch	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	2	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dock	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	2	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dug	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	geese	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	2	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	2	1	1	1	1	1	1	1	1
	coat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	2
	pat	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	2	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	tile	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NF1	bad	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	2	1	1	1	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	2	1	1	2
	bunch	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	2	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dock	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dug	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	2	1	1	1	1	1	1	1	1	1
	geese	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	2	1	1	1	1	1	1	1	1	1
	coat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	pat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	2	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	tile	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1

			Judge 1							Judge 2						
Speaker	Utterance	Rep.	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NF2	bad	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	2	1	1	1	1	1	1	2	1	1	2
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	bunch	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	2	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dock	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	2	1	1	1	1	1	1
	dug	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	2	1	1	1	1	1	1	1	1	1
	geese	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	2	1	1	1	1	1	1	1	1	2
	coat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	2	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	pat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	2	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	tile	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	2	1	1	1	1
NF3	bad	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	2	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	bunch	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	2	1	1	1	1	2
	dock	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dug	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	geese	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	2	1	1	1	1	1
		4	1	1	1	1	1	1	1	2	2	1	2	1	1	1
	coat	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	2	1	1	1	1	1	1	1	1	1	1
	pat	2	1	1	2	1	1	1	1	1	1	2	1	1	1	1
		3	1	1	1	1	1	1	1	2	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	tile	2	1	1	1	1	1	1	1	1	1	2	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NF3	bad	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	1	2	1	2	1	1	1	1
	bunch	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2
		3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	dock	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		4	1	2	1	1	1	1	1	1	1	1	2	1	1	1
	dug	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1

			Judge 1							Judge 2						
Speaker	Utterance	Rep.	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2
		3	1	1	1	1	1	1	1	1	1	1	1	1	1	2
		4	1	1	1	1	1	1	1	1	1	1	2	1	1	1
	geese	2	1	1	1	2	1	1	1	1	1	1	2	1	1	1
		3	1	1	1	2	1	1	1	1	1	1	2	1	1	1
		4	1	2	1	1	1	1	1	1	1	1	1	1	1	2
	coat	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	pat	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		3	1	1	1	1	1	1	1	1	1	2	1	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	tile	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	1	1	1	2
DM2	bad	2	1	1	1	1	1	1	1	1	1	1	2	1	2	1
		3	1	1	1	1	1	1	2	1	1	2	2	1	2	2
		4	1	1	1	2	1	2	2	1	1	1	2	1	1	1
	bunch	2	1	1	1	1	1	2	1	1	1	1	2	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	1	1	1	1	2	1	1	1
	dock	2	1	1	2	1	1	2	1	1	1	2	2	1	1	1
		3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		4	1	1	2	2	1	1	2	1	1	1	2	2	1	1
	dug	2	1	1	1	2	1	2	1	1	1	1	2	1	1	1
		3	1	1	1	1	1	2	1	1	1	1	2	1	1	1
		4	1	1	1	2	2	1	2	1	1	1	1	2	1	1
	geese	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1
		3	1	1	1	2	1	2	1	1	2	1	2	1	2	2
		4	1	2	1	2	1	1	1	1	2	1	2	1	1	1
	coat	2	1	1	1	2	1	1	1	1	1	1	2	1	1	1
		3	1	1	1	2	1	1	1	1	1	2	2	1	1	2
		4	1	1	1	2	1	1	1	1	1	1	2	1	1	1
	pat	2	1	1	1	2	1	2	2	1	1	1	2	3	3	1
		3	1	1	1	1	1	1	2	1	1	1	2	1	1	1
		4	1	1	1	1	1	1	2	1	1	1	2	1	1	1
	tile	2	1	1	1	2	1	1	1	1	1	2	2	2	1	1
		3	2	1	1	2	1	1	1	2	1	1	2	2	1	2
		4	1	1	1	3	2	1	2	2	1	3	2	2	1	1
DM1	bad	2	1	1	1	1	1	1	2	1	2	1	2	1	1	1
		3	1	1	1	2	1	1	2	1	1	1	2	1	2	1
		4	1	1	1	1	1	1	2	1	1	1	2	1	2	2
	bunch	2	1	1	1	1	1	1	1	1	2	1	2	1	2	2
		3	1	1	1	2	1	1	1	1	2	1	2	1	1	2
		4	1	1	1	1	1	1	1	1	2	1	2	1	2	2
	dock	2	1	1	1	3	1	1	2	1	1	2	3	1	3	2
		3	1	1	1	1	1	1	2	1	1	1	2	1	3	1
		4	1	1	2	2	1	1	2	1	1	2	2	1	3	1
	dug	2	2	1	1	2	1	2	2	2	2	1	2	1	2	2
		3	1	1	2	2	1	1	2	1	1	2	2	1	2	2
		4	2	1	2	2	1	2	2	2	2	2	2	1	2	2
	geese	2	2	1	1	2	1	2	1	2	3	1	3	1	2	2
		3	1	1	1	2	1	2	1	1	1	1	2	1	2	2
		4	1	1	1	2	1	2	1	2	3	1	2	1	2	1
	coat	2	1	1	1	2	1	1	1	1	1	1	2	1	2	1
		3	1	1	1	2	2	1	1	1	1	1	2	3	3	2
		4	1	1	1	2	1	1	1	1	1	1	2	1	1	2
	pat	2	1	1	2	2	1	2	1	1	1	2	3	1	3	1
		3	1	1	1	2	2	2	1	1	1	1	2	3	3	1
		4	1	1	1	2	2	2	1	1	1	2	3	1	2	1
	tile	2	1	1	2	2	1	2	2	2	1	2	3	1	3	2

			Judge 1							Judge 2						
Speaker	Utterance	Rep.	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2
		3	2	1	1	2	1	1	1	2	1	1	2	1	1	1
		4	1	1	1	2	2	2	1	1	1	2	2	2	2	1
DF1	bad	2	2	2	1	1	1	1	1	2	2	1	2	1	2	2
		3	1	2	1	2	1	1	2	1	3	1	2	1	2	2
		4	2	2	1	2	1	2	2	2	3	1	2	1	2	2
	bunch	2	1	2	1	1	1	1	2	2	3	1	2	1	2	2
		3	1	2	1	2	1	1	1	1	3	1	1	1	2	2
		4	1	2	1	1	1	1	1	1	3	1	1	1	1	2
	dock	2	2	2	1	2	1	2	1	1	3	1	2	1	2	1
		3	1	2	1	3	2	2	1	1	3	3	2	1	2	1
		4	1	2	1	2	1	1	1	1	3	1	2	1	2	1
	dug	2	2	2	1	2	1	1	2	1	3	3	2	1	1	2
		3	1	2	1	1	1	2	2	1	3	2	2	1	2	2
		4	1	2	1	2	1	2	2	2	3	2	2	1	1	2
	geese	2	1	2	1	2	1	1	1	1	3	1	2	1	1	1
		3	2	2	2	2	1	1	2	2	3	1	2	1	2	1
		4	2	2	1	2	1	1	1	3	3	1	2	1	2	1
	coat	2	1	1	1	2	1	2	2	1	1	1	1	1	2	1
		3	2	1	1	2	2	2	1	2	1	1	2	2	2	1
		4	1	1	1	2	1	2	2	1	1	1	2	1	2	2
	pat	2	1	1	1	2	1	2	1	1	1	1	2	3	3	1
		3	1	1	1	2	1	2	2	1	1	1	2	2	3	1
		4	1	1	1	2	2	2	1	2	1	1	2	1	2	1
	tile	2	1	1	2	2	1	2	1	1	1	2	2	1	3	1
		3	2	1	3	3	1	2	1	3	1	3	3	3	3	1
		4	1	1	1	2	2	2	1	1	1	1	2	3	3	1
DF2	bad	2	2	2	1	1	1	1	2	3	2	1	3	1	1	3
		3	2	2	1	2	1	1	2	3	2	2	3	1	2	3
		4	2	2	2	3	1	1	2	3	2	1	3	1	3	3
	bunch	2	2	2	2	2	1	1	2	3	3	1	3	1	3	2
		3	2	1	2	2	1	1	1	2	1	1	3	1	2	2
		4	2	1	1	2	1	1	1	3	2	1	3	1	2	2
	dock	2	2	2	2	3	1	1	2	3	3	3	3	1	1	3
		3	3	1	1	2	1	1	2	3	2	1	2	1	3	3
		4	1	1	1	2	1	2	2	1	2	1	2	1	1	2
	dug	2	2	2	1	2	1	1	2	2	3	1	2	1	1	3
		3	2	2	2	2	1	2	2	2	2	2	2	1	2	2
		4	2	2	1	2	1	2	2	3	2	3	3	1	1	2
	geese	2	3	2	3	2	1	2	1	3	3	3	3	1	2	2
		3	2	1	2	2	1	1	2	3	3	3	3	1	3	2
		4	3	3	2	2	1	1	1	3	3	3	3	1	1	1
	coat	2	3	2	3	2	1	2	2	3	3	3	3	1	2	1
		3	2	1	2	2	2	2	2	3	1	3	3	1	3	1
		4	3	2	3	2	1	2	2	3	2	3	3	1	2	1
	pat	2	3	2	3	2	2	1	1	3	2	3	3	2	2	2
		3	2	1	2	2	1	1	1	1	1	2	3	3	2	1
		4	2	1	2	2	2	2	1	3	1	3	3	3	2	1
	tile	2	3	3	2	3	2	2	1	3	3	3	3	3	3	2
		3	2	1	3	3	2	2	1	3	1	3	3	1	3	2
		4	3	1	2	2	1	2	1	2	1	2	3	3	3	2
DF3	bad	2	1	2	1	2	1	2	2	1	2	1	2	1	3	3
		3	1	1	1	2	2	2	2	1	1	2	2	1	3	2
		4	1	1	1	2	2	2	2	1	1	2	2	1	3	2
	bunch	2	2	1	1	1	2	2	2	2	1	1	2	1	3	2
		3	2	1	2	2	2	2	1	2	1	3	2	1	3	1
		4	2	1	1	2	2	1	1	2	1	1	2	1	3	2
	dock	2	1	1	1	3	1	2	2	1	1	1	2	1	2	2
		3	1	1	2	3	1	1	2	1	1	3	2	1	3	1
		4	2	1	1	3	2	1	1	2	1	2	3	1	3	1
	dug	2	2	1	2	3	2	2	2	2	1	3	2	1	1	3

			Judge 1							Judge 2						
Speaker	Utterance	Rep.	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2
		3	2	1	2	3	1	2	2	2	1	2	3	1	1	2
		4	2	1	1	2	2	2	2	2	1	1	3	1	2	3
	geese	2	1	1	1	2	1	1	2	1	1	1	2	1	1	1
		3	2	1	1	3	2	1	2	2	1	1	2	1	1	2
		4	1	1	1	2	1	2	1	1	1	1	2	1	1	2
	coat	2	1	1	1	2	2	2	2	1	1	1	2	2	1	2
		3	1	1	1	3	2	2	2	1	1	1	2	2	3	2
		4	2	1	2	3	3	2	1	2	1	3	3	3	1	2
	pat	2	1	1	2	3	2	3	2	1	1	1	3	1	2	3
		3	1	1	2	3	3	3	3	2	1	1	3	2	2	3
		4	1	1	2	2	3	2	2	1	1	1	2	2	3	3
	tile	2	1	1	3	3	2	3	2	1	1	2	3	2	2	2
		3	1	1	1	3	3	3	3	1	1	1	3	3	3	3
		4	1	1	3	3	1	3	3	1	1	3	3	2	2	2
DM4	bad	2	1	2	2	2	1	2	2	1	2	3	3	1	3	3
		3	3	2	2	2	1	2	1	2	3	2	3	1	2	2
		4	1	2	1	2	1	2	1	1	3	1	2	1	1	2
	bunch	2	2	3	1	3	1	2	2	3	3	2	3	1	1	3
		3	3	2	1	2	1	1	2	2	3	1	2	1	2	1
		4	2	2	1	2	1	1	2	1	3	1	1	1	2	1
	dock	2	2	1	1	2	1	1	2	1	3	2	2	1	1	1
		3	1	1	1	1	1	1	2	1	2	1	2	1	1	2
		4	2	2	1	2	1	1	2	2	3	2	2	1	1	1
	dug	2	2	1	1	2	1	2	2	1	2	1	2	1	2	1
		3	2	1	1	1	1	2	2	1	3	1	2	1	2	1
		4	2	1	1	2	1	2	1	2	2	1	2	1	2	1
	geese	2	3	3	3	3	1	1	1	3	3	3	3	1	1	1
		3	1	1	1	2	1	1	1	1	1	1	2	1	1	2
		4	3	3	3	2	1	1	2	3	3	3	3	1	1	2
	coat	2	1	1	1	3	3	1	1	1	1	3	2	3	1	2
		3	1	2	1	3	3	3	3	1	3	1	3	3	3	3
		4	3	2	3	3	2	2	3	2	3	2	3	2	3	2
	pat	2	1	1	1	2	3	3	1	2	1	1	3	3	2	3
		3	1	1	1	2	2	2	2	1	1	1	2	2	1	2
		4	1	1	2	3	3	1	1	2	1	3	3	3	2	3
	tile	2	2	2	1	3	3	2	3	1	3	1	3	3	1	3
		3	2	3	1	2	1	2	2	2	3	1	3	1	2	3
		4	1	1	1	1	2	2	2	2	1	1	3	3	2	2
DM3	bad	2	1	1	2	3	1	2	1	1	1	3	3	2	3	2
		3	1	1	1	2	1	2	2	2	1	1	3	1	1	3
		4	2	1	2	2	1	2	2	3	2	1	3	1	2	3
	bunch	2	1	1	1	3	2	2	2	2	2	1	3	1	3	3
		3	2	1	2	2	1	2	1	3	1	2	3	2	3	2
		4	2	1	2	2	1	2	1	3	1	2	3	1	3	2
	dock	2	1	1	2	2	2	2	2	1	1	3	2	2	3	3
		3	2	1	1	3	3	1	1	2	1	1	3	2	3	2
		4	2	1	1	3	2	2	2	2	1	2	3	2	3	2
	dug	2	1	1	1	2	1	1	2	1	1	1	3	1	2	3
		3	1	1	2	2	1	2	1	1	1	2	2	1	1	1
		4	1	1	1	2	1	2	1	1	1	1	3	1	1	2
	geese	2	3	1	2	3	3	2	2	3	2	3	3	1	3	2
		3	2	2	2	2	2	2	2	3	3	3	3	1	3	3
		4	2	1	2	2	1	2	2	3	3	3	3	1	3	3
	coat	2	1	1	1	2	3	1	2	1	1	1	3	3	1	3
		3	2	1	3	3	3	1	2	2	1	2	3	2	1	1
		4	2	1	3	3	3	3	3	2	1	2	2	3	2	2
	pat	2	2	1	1	3	3	3	2	2	1	1	3	3	1	1
		3	2	1	1	3	3	3	2	2	1	2	3	3	1	2
		4	1	1	1	2	3	2	2	2	1	1	2	3	1	1
	tile	2	1	1	1	3	1	1	2	2	1	2	3	2	2	1

			Judge 1							Judge 2						
Speaker	Utterance	Rep.	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2	Prec	Prev	Abr	TCR	VOT	TCF1	TCF2
		3	2	1	1	3	3	1	2	2	1	1	3	3	2	1
		4	2	1	2	3	2	1	3	2	1	2	2	2	1	1
DF4	bad	2	3	1	3	3	1	2	1	2	1	3	3	1	3	2
		3	3	1	2	2	1	1	2	3	3	3	3	1	2	3
		4	3	1	2	2	1	1	2	3	2	3	3	1	2	2
	bunch	2	2	2	2	2	1	2	2	3	2	3	3	1	2	1
		3	3	2	3	3	1	2	2	3	3	2	3	1	2	2
		4	3	3	3	3	1	3	3	2	1	2	3	1	2	3
	dock	2	3	1	2	2	1	2	2	3	3	2	3	1	2	2
		3	3	1	2	2	1	1	2	3	3	2	3	1	2	2
		4	3	1	3	2	1	1	1	3	3	3	3	1	3	2
	dug	2	3	1	3	3	1	2	3	3	3	2	3	1	3	2
		3	3	2	2	3	1	3	3	3	3	1	3	1	1	3
		4	3	2	3	3	1	3	2	3	3	3	3	1	1	2
	geese	2	1	2	3	3	1	1	2	3	1	2	3	2	1	2
		3	3	2	1	1	1	1	2	3	2	1	3	1	1	1
		4	3	2	3	3	1	1	2	3	3	3	3	1	1	3
	coat	2	3	1	2	3	1	2	2	3	1	2	3	3	2	2
		3	3	1	1	3	1	2	2	3	1	1	3	3	2	1
		4	1	1	2	3	3	1	1	3	1	1	3	3	1	2
	pat	2	3	1	1	3	2	1	1	3	1	1	3	3	1	2
		3	1	1	1	3	3	1	1	1	1	2	3	3	1	1
		4	1	1	1	3	3	1	1	1	1	1	3	3	2	1
	tile	2	3	1	3	3	3	1	1	2	1	1	3	3	2	2
		3	3	3	3	3	3	1	1	3	1	2	3	1	2	1
		4	3	2	3	3	3	1	1	3	1	2	3	3	1	2

Bibliography

Ansel, B. M. and Kent, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *J. Speech and Hear. Res.*, 35:296–308.

Berry, W. and Goshorn, E. (1983). Immediate visual feedback in the treatment of ataxic dysarthria: A case study. In Berry, W., editor, *Clinical Dysarthria*, pages 253–266. College Hill Press, San Diego, CA.

Chang, H.-P. (1995). *Speech Input for Dysarthric Computer Users.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Charcot, J. (1877). *Lectures on the Diseases of the Nervous System*, volume 1. The New Sydenham Society, London.

Coelho, C. A., Gracco, V. L., Fourakis, M., Rossetti, M., and Oshima, K. (1994). Application of instrumental techniques in the assessment of dysarthria: A case study. In Till, J. A., Yorkston, K. M., and Beukelman, D. R., editors, *Motor Speech Disorders: Advances in Assessment and Treatment*, chapter 8, pages 103–117. Paul H. Brookes Publishing Co. Clinical Dysarthric Conference Proceedings.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *J. Acoust. Soc. Am.*, 24:597–606.

Darley, F. L., Aronson, A. E., and Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *J. Speech and Hear. Res.*, 12:462–496.

Darley, F. L., Aronson, A. E., and Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *J. Speech and Hear. Res.*, 12:246–269.

Darley, F. L., Aronson, A. E., and Brown, J. R. (1975). *Motor Speech Disorders.* W. B. Saunders Co., Philadelphia, PA.

Duffy, J. R. (1995). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management.* Mayo-Foundation, Mosby-Year Book, Inc., St. Louis, MO.

Enderby, P. (1983). *Frenchay dysarthria assessment.* College-Hill, San Diego.

Fant, G. (1960). *Acoustic Theory of Speech Production.* Mouton, The Hague.

Fant, G. (1972). Vocal tract wall effects, losses, and resonance bandwidths. Technical report, Speech Transmission Laboratory Quarterly Progress and Status Report 2-3, Royal Institute of Technology, Stockholm, Sweden.

Fant, G. (1973). Stops in CV-syllables. In *Speech Sounds and Features*, chapter 7, pages 110–139. MIT Press, Cambridge, MA.

Gerratt, B. R., Till, J. A., Rosenbek, J. C., Wertz, R. T., and Boysen, A. E. (1991). Use and perceived value of perceptual and instrumental measures in dysarthria management. In Moore, C. A., Yorkston, K. M., and Beukelman, D. R., editors, *Dysarthria and Apraxia of Speech: Perspectives on Management*, chapter 7, pages 77–93. Paul H. Brookes Publishing Co. Clinical Dysarthric Conference Proceedings.

Glass, J. (1986). Electrical Engineering and Computer Science Departmental Area Exam. Massachusetts Institute of Technology, Cambridge, MA.

Hertegård, S. (1994). *Vocal Fold Vibrations as Studied with Flow Inverse Filtering.* PhD thesis, Department of Logopedics and Phoniatrics, Karolinska Institute, Huddinge University Hospital, Stockholm, Sweden.

Hodge, M. M. and Hall, S. D. (1994). Effects of syllable characteristics and training on speaking rate in a child with dysarthria secondary to near-drowning. In Till, J. A., Yorkston, K. M., and Beukelman, D. R., editors, *Motor Speech Disorders: Advances in Assessment and Treatment*, chapter 17, pages 229–250. Paul H. Brookes Publishing Co. Clinical Dysarthric Conference Proceedings.

Ishizaka, K., French, J. C., and Flanagan, J. L. (1975). Direct determination of vocal tract wall impedance. In *IEEE Trans. on Acoust., Speech and Signal Processing,* ASSP-23, pages 370–373.

Isshiki, N. and Ringel, R. (1964). Air flow during the production of selected consonants. *J. Speech and Hear. Res.*, 7:233–244.

Kent, R., Netsell, R., and Abbs, J. (1979). Acoustic characteristics of dysarthria associated with cerebellar disease. *J. Speech and Hear. Res.*, 22:627–648.

Kent, R. D., Kent, J. F., Duffy, J., and Weismer, G. (1998). The dysarthrias: Speech-voice profiles, related dysfunctions, and neuropathology. *J. Med. Speech-Language Path.*, 6(4):165–211.

Kent, R. D., Weismer, G., Kent, J. F., and Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *J. Speech and Hear. Disorders*, 54:482–499.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *J. Speech and Hear. Res.*, 18(4):686–706.

Klatt, D. H. and Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.*, 87(2):820–857.

Ladefoged, P. (1963). Some physiological parameters in speech. *Language and Speech*, 6:109–119.

Lehiste, I. (1965). Some acoustic characteristics of dysarthric speech. *Bibliotheca Phonetica*, Fasc. 2.

Lord, J. (1984). Cerebral palsy: A clinical approach. *Archives of Physical Medicine and Rehab.*, 65:542–548.

Love, R. (1992). *Childhood Motor Speech Disability.* Macmillan Publishing Co., New York, NY.

Ludlow, C. and Bassich, C. (1984). Relationships between perceptual ratings and acoustic measures of hypokinetic speech. In McNeil, M., Rosenbek, J., and Aronson, A., editors, *The Dysarthrias: Physiology, Acoustics, Perception, Management.* San Diego: College Hill Press.

Massey, N. S. (1994). Transients at stop-consonant releases. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

McNeil, M. R. (1986). A critical appraisal of instrumentation methods in the evaluation and management of dysarthria. Paper presented at the Clinical Dysarthria Conference, Tucson, AZ.

Müller, E. M. and Brown, Jr., W. S. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretations. In Lass, N. J., editor, *Speech and Language: Advances in Basic Research and Practice*, volume 4, pages 317–389. Academic Press, New York.

Ohde, R. N. (1982). Coarticulatory effects of stop voicing on F0 from voicing onset to vowel target. *J. Acoust. Soc. Am.* paper presented at 103rd Mtg. of ASA, Chicago, IL.

Ohde, R. N. and Sharf, D. J. (1992). *Phonetic Analysis of Normal and Abnormal Speech.* Macmillan Publishing Company, New York.

Pastel, L. (1987). Turbulent noise sources in vocal tract models. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Poort, K. L. (1995). Stop consonant production: An articulation and acoustic study. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Ramig, L. A., Scherer, R. C., Titze, I. R., and Ringel, S. P. (1988). Acoustic analysis of voices of patients with neurologic disease: Rationale and preliminary data. *Ann. Otol. Rhinol. Laryngol.*, 97:164–172.

Rosenbek, J. C. and LaPointe, L. L. (1985). The dysarthria: Description, diagnosis and treatment. In Johns, D. F., editor, *Clinical Management of Neurogenic Communication Disorders.* Boston: Little, Brown.

Rothenberg, M. (1968). The breath stream dynamics of simple-released-plosive production. Bibliotheca Phonetica No. 6, S. Karger, Basel.

Shadle, C. (1985). The acoustics of fricative consonants. RLE Technical Report 506, Massachusetts Institute of Technology, Cambridge, MA.

Stevens, K., Manuel, S., and Matthies, S. (1999). Revisiting place of articulation measures for stop consonants: Implications for models of consonant production. In *ICPHS*, volume 2, pages 1117–1120. ICPHS.

Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants. *J. Acoust. Soc. Am.*, 50:1180–1192.

Stevens, K. N. (1993). Models for the production and acoustics of stop consonants. *Speech Comm.*, 13:367–375.

Stevens, K. N. (1998). *Acoustic Phonetics.* MIT Press, Cambridge, MA.

Strand, E. A. and Yorkston, K. M. (1994). Description and classification of individuals with dysarthria: A 10-year review. In Till, J. A., Yorkston, K. M., and Beukelman, D. R., editors, *Motor Speech Disorders: Advances in Assessment and Treatment*, chapter 3, pages 37–54. Paul H. Brookes Publishing Co. Clinical Dysarthric Conference Proceedings.

Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., and Wilhelms-Tricarico, R. (1997). Tongue surface displacement during bilabial stops. *J. Acoust. Soc. Am.*, 102(1):562–571.

Weismer, G. (1984). Acoustic descriptions of dysarthric speech: Perceptual correlates and physiological inferences. *Seminars in Speech and Language*, 5:293–314.

Weismer, G. and Liss, J. M. (1991). Reductionism is a dead-end in speech research: Perspectives on a new direction. In Moore, C. A., Yorkston, K. M., and Beukelman, D. R., editors, *Dysarthria and Apraxia of Speech: Perspectives on Management*, chapter 2, pages 15–27. Paul H. Brookes Publishing Co. Clinical Dysarthric Conference Proceedings.

Westbury, J. R. (1979). Aspects of the temporal control of voicing in consonant clusters in English. Texas Linguistic Forum 14, Department of Linguistics, University of Texas.

Yorkston, K. M. and Beukelman, D. R. (1981). *Assessment of Intelligibility of dysarthric speech.* CC Publications, Tigard, Oregon.

Yorkston, K. M., Beukelman, D. R., and Bell, K. R. (1988). *Clinical Management of Dysarthric Speakers.* Pro-Ed, Inc., Austin, TX.

Yumoto, E., Sasaki, Y., and Okamura, H. (1984). Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *J. Speech and Hear. Res.*, 27:2–5.

Zeplin, J. and Kent, R. D. (1996). Reliability of auditory-perceptual scaling of dysarthria. In Robin, D. A., Yorkston, K. M., and Beukelman, D. R., editors, *Disorders of Motor Speech: Assessment, Treatment, and Clinical Characterization*, chapter 8, pages 145–154. Paul H. Brookes Publishing Co. Clinical Dysarthric Conference Proceedings.

Zue, V. W. (1976). Acoustic characteristics of stop consonants: A controlled study. Technical report, Indiana University Linguistics Club, Indiana University, 310 Lindley Hall, Bloomington, Indiana 47405. (first reproduced in 1980).