

An Acoustical Study of Fricatives in the Speech of Dysarthric Speakers

by

Helen Chen

Submitted to the Department of Electrical Engineering and
Computer Science
in partial fulfillment of the requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1998

© Massachusetts Institute of Technology 1998. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part, and to grant others the right to do so.

Author
Department of Electrical and Computer Science
May 8, 1998

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

Accepted by
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUL 14 1998

LIBRARIES

Eng.

An Acoustical Study of Fricatives in the Speech of Dysarthric Speakers

by

Helen Chen

Submitted to the Department of Electrical Engineering and
Computer Science on May 8, 1998,
in partial fulfillment of the requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

The goal of the present study is to better understand dysarthric speakers' speech production through quantitative acoustical analysis of the speech signals. In the past, two acoustical measurements proven to correlate with speech intelligibility in dysarthric speakers are speech rate and slope of the F2 frequency change for certain consonant-vowel syllables. In this thesis, several new spectral parameters from acoustical analysis that reflect neuromotor dysfunctions in production of the fricative /s/ have been devised and are presented. These parameters specify the spectrum shape and relative amplitude of the fricative noise, the presence or absence of voicing, and the time variation of the spectrum preceding, during, and following the fricative. Analysis shows that these parameters are highly correlated with the speakers' overall intelligibility as well as intelligibility of the fricative /s/. The top three parameters that differentiate the intelligibility of speakers are related to the spectral shape (spectral tilt), the amount of variation in spectral energy (time variation within the fricative), and the amount of inadvertent spectral energy present prior to frication (precursor).

The long-term aim of the study is to develop methods for quantifying disordered speech and for determining the deviation from normal speech. In the long run, these methods can be used to objectively measure the effectiveness of speech therapy as well as the amount of speech degeneration in dysarthric speakers.

Thesis Supervisor: Kenneth N. Stevens

Title: Clarence J. Lebel Professor of Electrical Engineering

Acknowledgments

The completion of this thesis would not have been possible without the full support of my thesis advisor, Ken Stevens. His patient guidance, warm encouragement, and insightful comments have left a lasting imprint in my mind. He not only took great care in my research work but also cared for me outside of academics. I feel truly blessed to have the chance to work with Ken and the talented group of researchers at the Speech Communications Group.

My deep gratitude goes to Hwa-Ping Chang. His careful doctoral research has laid a solid foundation for my research to build upon. Heartfelt thanks goes to Corine Bickley for all the assistance with the statistical analysis of the data.

Thanks to AP, JW, LD, KS, ES, KC for taking the time to be my subjects for the study. Thanks to Jeff Kuo for being always available to discuss results, to clarify my thought process, or simply to give me words of encouragement. Thanks to Marilyn Chen for great critique on my thesis work and moral support. Kelly Poort has also given many helpful insights that enabled me to have a better grasp of the data that I was analyzing. Thanks also goes to my wonderful officemate Adrienne Prahler who never failed to bring a smile to my face and keep me to the deadline.

I feel extremely fortunate to have been blessed with two loving roommates who always took great care of me and made me look forward to going home. Thanks also goes to Richard Chung who was always there to help. Thanks to Allen Sun and Mike Sung for helping me to stay awake late at night in order to finish the writing of this thesis.

Special thanks goes to friends I met through Boston Chinese Evangelical Church, especially those at the MIT Chinese Christian Fellowship, the graduate small group, and the China Teaching Team for their faithful prayers, the power of which is a constant source of renewal for me.

Last but not the least, I sincerely thank my parents who made great sacrifices through the years to enable me to have the best education possible. Their unconditional love and support never cease to amaze me.

Soli Deo gloria – To God alone be glory.

CONTENTS

CHAPTER 1 INTRODUCTION.....	9
1.1 THESIS MOTIVATION.....	11
1.2 THESIS OUTLINE	11
CHAPTER 2 BACKGROUND	13
2.1 CHARACTERISTICS OF DYSARTHRIAS	13
2.2 FRICATIVE CONSONANTS - ARTICULATORY AND ACOUSTIC EVENTS ...	15
2.2.1 <i>Acoustics of Unvoiced Alveolar Fricative /s/.....</i>	<i>15</i>
2.3 PAST STUDIES	17
2.4 PRELIMINARY STUDIES	20
CHAPTER 3 METHODS AND PROCEDURES.....	22
3.1 DATA ACQUISITION	22
3.2 SUMMARY OF SUBJECTS	23
3.3 ACOUSTIC ANALYSIS	24
3.3.1 <i>Measurements from Spectra.....</i>	<i>24</i>
3.3.2 <i>Assessments from Spectrograms</i>	<i>26</i>
3.4 PERCEPTUAL RATING OF /s/	29
CHAPTER 4 RESULTS AND DISCUSSION	30
4.1 ACOUSTIC ANALYSIS BASED ON OVERALL INTELLIGIBILITY FROM HWA- PING'S DOCTORAL STUDIES.....	31
4.1.1 <i>Assessment of Formation and Release of Fricative from Spectrograms.....</i>	<i>31</i>
4.1.2 <i>Articulatory and Phonetic Features Based on Measurements of Spectra</i>	<i>33</i>

4.1.3 <i>Total Assessment Average vs. Overall Intelligibility</i>	36
4.1.4 <i>Data Grouping – High and Low Intelligibility Groups</i>	37
4.2 ACOUSTIC ANALYSIS BASED ON /s/ PERCEPTUAL RATING	38
4.2.1 <i>Assessment of Formation and Release of Fricative from Spectrogram</i>	39
4.2.2 <i>Articulatory and Phonetic Features Based on Measurements of Spectra</i>	40
4.2.3 <i>Total Average Assessment vs. New Intelligibility Rating</i>	42
4.3 STATISTICAL ANALYSIS	43
4.3.1 <i>Overall Intelligibility vs. New /s/ Perceptual Rating</i>	43
4.3.2 <i>Pearson Correlation Matrix</i>	44
4.3.3 <i>Discriminant Analysis</i>	45
4.4 DISCUSSION ON INDIVIDUAL SUBJECTS	46
4.4.1 <i>Subject F1</i>	47
4.4.2 <i>Subject F3</i>	49
CHAPTER 5 CONCLUSIONS	52
5.1 SUMMARY	52
5.2 FUTURE WORK	53
CHAPTER 6 REFERENCES	55
APPENDIX A	58

List of Figures

Figure 2.1 a) Midsagittal cross-section of a vocal tract vocal-tract model for an alveolar fricative. b) Vocal tract idealized as a uniform tube with constrictions at the glottis and in the oral region (Wilde, 1995).....	16
Figure 2.2 Histogram of A1v-A1f measures in normal and dysarthric data for /s/ initial utterances.....	21
Figure 3.1 a) & b) Example of time-averaged DFT Spectra of /s/ and /i/ in “see” for normal speakers. Measurements are marked on the plot and the calculations of A1v-A1f (voicing), spectral tilt, and A1v-Ahf are given below the figure.....	26
Figure 3.2 Example of spectrograms of production of “see” by a normal speaker, a dysarthric speaker with highly intelligibility and a dysarthric speaker with low intelligibility	28
Figure 4.1 a)-e) Plot of assessment of formation and release of fricative /s/ in relation to overall intelligibility of individual speakers.....	33
Figure 4.2 Plot of three parameters derived from measurements of spectra of fricative /s/ in relation to overall intelligibility of individual speakers.....	35
Figure 4.3 Plot of spectrogram assessment of voicing characteristic of fricative /s/ in dysarthric speakers in relation to overall intelligibility.....	36
Figure 4.4 Total average assessment of the release and formation of fricative /s/ versus overall intelligibility of the speakers. The data separate into three distinct groups: low intelligibility, high intelligibility, and normal.....	36
Figure 4.5 Spectral assessment separated by low and high intelligibility groups. Each bar represents average data for the speaker group indicated. The thin line on each bar gives the range for one standard deviation above and below the mean value.....	37
Figure 4.6 Measurement of spectra separated by low and high intelligibility groups. Each bar represents average data for the speaker group indicated. The thin line on each bar gives the range for one standard deviation above and below the mean value.....	37

Figure 4.7 a)-e) Plot of assessment of formation and release of fricative /s/ in relation to /s/ perceptual rating of individual speakers.....	40
Figure 4.8 Plot of measurements of spectra of fricative /s/ in relation to /s/ perceptual rating of individual speakers.....	41
Figure 4.9 Total average assessment of the release and formation of fricative /s/ versus /s/ perceptual rating of the speakers.....	42
Figure 4.10 Linear correlation between overall intelligibility determined previously by Hwa-Ping Chang and new /s/ perceptual rating devised in this thesis.....	43
Figure 4.11 Summary of assessment of formation and release of fricative /s/ of three speakers, one normal speaker, one dysarthric speaker with high intelligibility, and one with low intelligibility speaker.....	46
Figure 4.12 Summary of measurements of spectra of fricative /s/ of three speakers as in Figure 4.11.....	47
Figure 4.13 F1' spectrogram of the word "see".....	48
Figure 4.14 F1' time-averaged DFT spectra of /s/, in the 100 ms period preceding the boundary between the fricative and the vowel.....	48
Figure 4.15 F3's spectrogram of the word "see".	50
Figure 4.16 F3's time-averaged DFT spectra of /s/, in the 100 ms period preceding the boundary between the fricative and the vowel.....	51

List of Tables

Table 3.1: Summary pertinent information on dysarthric subjects.....	23
Table 4.1 Summary of spectral parameters and assessments in correlation with intelligibility. The first eight rows are the dysarthric speakers and the bottom two rows are the normals with overall intelligibility of 100%.....	30
Table 4.2 Pearson correlation coefficient matrix for the various spectral measures and intelligibility or perceptual rating scores.....	44

Chapter 1 Introduction

Dysarthrias are types of neuromotor speech disorders that result from damage to the central nervous system. Impairment of the neuromuscular control of speech can affect several or all aspects of speech production: respiration, phonation, resonance, articulation, and prosody (Darley et al., 1975). Extensive clinical research has been performed by Darley, Aronson, and Brown to quantify perceptual impressions based on the psychophysical scaling of multiple dimensions of speech production performance. However, perceptual analysis has limited analytic power, and perceptual judgments are difficult to standardize over time and across different settings. Only clinicians with extensive experience can rely on perceptual assessment alone to evaluate the details of speech production and the consistency of individuals with dysarthria.

In recent years, articulatory and acoustic models for the normal production of various classes of speech sounds have been refined and have led to more quantitative predictions of the properties of the sounds. These models can help to interpret acoustic measurements on speech sounds in terms of the articulatory movements that give rise to the sounds, for both normal and disordered speech production. The broad goal of the present study is to develop improved models of normal as well as dysarthric speakers' speech production and to use these models to interpret acoustic data obtained from the utterances of dysarthric speakers. This study is concerned particularly with the unvoiced alveolar fricative consonant /s/ produced by dysarthric speakers. One reason for selecting

this particular consonant for study is that many dysarthric speakers have been shown to have particular difficulty with the sound (Chang, 1995).

Any articulatory/acoustic model of fricative production should be able to estimate the sound patterns for different fricatives when the subglottal pressure and the time-varying glottal opening and supraglottal movements are known. Such a model must first estimate the flows and pressures, and from these calculate the characteristics of the acoustic sources. Acoustic theory is then used to determine the filtering of these sources by the vocal tract and hence the radiated sound pattern. Generation of the normal acoustic pattern for a fricative consonant requires the coordination of the respiratory system, adjustment of the glottal configuration, closure of the velopharyngeal port, and formation of a constriction by proper placement of the appropriate articulator in the oral cavity. In the case of a voiceless fricative consonant, it can be shown that the cross-sectional area of the constriction formed by the supraglottal articulator should be adjusted to be roughly one-half of the area of the glottis in order to maximize the amplitude of the frication noise source in the vicinity of the supraglottal articulator.

Using this model we can assess the extent to which the above articulatory and aerodynamic requirements for the consonant are being implemented by speakers with motor disorders, based on acoustic measurements of the sound generated by a fricative consonant. We consider here fricatives in word-initial position, as in the word **sat**. Buildup of intraoral pressure before the glottal and supralaryngeal articulators are in place will lead to inadvertent sound that precedes the fricative. If the glottis is not sufficiently abducted, glottal vibration will continue through the fricative, and the noise generated at the tongue-blade constriction will be too weak. If the supraglottal

constriction is not sufficiently narrow, frication noise at the constriction will be weak, and the acoustic source at the glottis will be the dominant one. And, if the tongue blade is improperly placed, the filtering of the frication noise source will not be correct, leading to a deviant spectrum of the radiated sound. At the fricative-vowel boundary, incorrect coordination of changes in the glottal and supraglottal constrictions can lead to premature voicing onset or to weak aspiration noise separating the fricative and the vowel.

1.1 Thesis Motivation

A speaker's speech signals provide a tremendous wealth of knowledge on the quality of the speaker's speech. There have been only limited attempts to characterize dysarthric speech with objective acoustic measurements from speech signals. Interpretations from the source-filter acoustic model of speech production suggest the development of a new set of quantitative acoustic assessments and measurements as analytical tools for dysarthric speech. The long-term motivation for this thesis is to develop methods for characterizing dysarthric speech in order to measure the effectiveness of speech therapy as well as the amount of speech degeneration in dysarthric speakers. In addition, this research could lead to methods for automatic recognition of the speech of these individuals, and ultimately to improve their quality of life.

1.2 Thesis Outline

Background information on the characteristics of various types of dysarthria, an acoustical model of fricative consonants, previous work, and preliminary studies are described in Chapter two. Chapter three discusses the data acquisition methods and procedures used to obtain the new acoustic parameters. Chapter four presents the results

of acoustical analysis and the significance of the findings as well as a few case studies on individual subjects. Finally, Chapter five gives a summary and suggestions for future work.

Chapter 2 Background

Different types of dysarthria have generally been categorized by their unique sets of perceptual characteristics (Darley, Aronson and Brown 1975). These perceptual characteristics are related to the specific site of lesion, and to the disease process involved. This chapter will detail the different types of dysarthria, the acoustic and articulatory events involved in the production of fricative consonants, some past acoustical studies of dysarthria and fricatives, and results of a preliminary acoustic study of the fricative /s/ produced by some dysarthric speakers.

2.1 Characteristics of Dysarthrias

Dysarthria is a term used to describe a group of neuromotor speech disorders resulting from impairment in the muscular control of the speech mechanism. Damage to the neuromuscular control of speech can affect respiration, phonation, resonance, articulation, and prosodic aspects of speech. From early works by Darley, Aronson, and Brown (1975), several identifiable types of dysarthria were linked to neurological conditions and damage to certain levels of the nervous system. The following list summarizes the most common dysarthria classification using five categories (LaPointe, 1994):

- a) Spastic Dysarthria – Caused by damage to the bilateral motor strip in the brain and resulting in poor articulation and strained or strangled voice.

- b) Ataxia Dysarthria – Caused by damage to the cerebellum and resulting in irregular speech and syllable repetition.
- c) Flaccid Dysarthria – Originating from neurological involvement in the brainstem and resulting in poor articulation and hypernasality.
- d) Hyperkinetic Dysarthria – Caused by damage throughout much of the extra-pyramidal motor system and resulting in unsteady rate, pitch and loudness as well as frequent grunting noises.
- e) Hypokinetic Dysarthria – Usually evolving from Parkinson’s disease and resulting in reduced loudness and speech rate.

The different subjects in this study exhibit three types of dysarthria: spastic dysarthria, athetoid (hyperkinetic) dysarthria, and ataxia dysarthria. The first two result from cerebral palsy and the last result from cerebellar ataxia. (Chang, 1995). Dysarthria may also accompany aphasia and apraxia, two other types of speech disorder. To reduce the order of complexity for this study, the subjects did not exhibit symptoms of apraxia and aphasia. In addition, the subjects’ auditory comprehension and reading skills were intact.

Speakers with cerebral palsy are commonly known to have difficulty in “coordinating” the processes of articulation, phonation, and respiration. Therefore dysarthria is a common symptom of this neurological disorder. Two kinds of cerebral palsy are exhibited by the subjects: spastic and athetosis. “Speech of the athetoid presents varying gradations of a pattern in irregular, shallow, and noisy breathing; whispered, hoarse, or ventricular phonation; and articulatory problems varying from the extremes of complete mutism or extreme dysarthria to a single awkwardness in lingual movement. Spastic dysarthrias result in poor articulation and strain/strangled voiced. Some speech symptoms of the spastic dysarthrias are: (i) slow and labored speech, (ii) consistently

imprecise articulation, (iii) low and monotonous pitch, (iv) harsh and often strained or strangled-sounding voice quality, and (v) hypernasality (but usually no audible nasal emission) (Darley et al., 1968)

2.2 Fricative Consonants – Articulatory and Acoustic Events

The source-filter acoustic theory of speech production represents the vocal tract as an acoustic tube with varying cross-sectional area. An acoustic source can form the excitation of this tube either at the glottal end or at points along its length, and the shape of the tube determines how the source is to be filtered.

Fricative consonants are distinguished from other speech sounds by their manner of production. Fricatives are produced by forming a narrow constriction in some region along the length of the vocal tract (Stevens, in press). Air blown through this constriction becomes turbulent in flow, typically near an obstacle in the airstream or at the walls of the vocal tract. The acoustic result of the turbulence is the generation of noise. This noise is then filtered by the vocal tract, with the acoustic cavity in front of the constriction contributing the greatest influence on the filtering. There are eight fricatives in English: labiodentals (f, v), dentals (th, dh), alveolars (s, z), and palato-alveolars (sh, zh).

2.2.1 Acoustics of Unvoiced Alveolar Fricative /s/

The alveolar fricative /s/ can be modeled under the assumption of a concatenated tube with two constrictions: one at the glottis and one at the constriction formed by the supraglottal articulators. Acoustically, the system as shown in Figure 2.1 can be decomposed into a quarter-wave resonator representing the cavity in front of the constriction, which contributes some of the system poles, and a half-wave resonator

representing the portion of the vocal tract between the constriction and the pressure source, which contributes the zeros and the remaining poles. The zeros depend on location of the source. As an example, if the front cavity is 2 cm long with the pressure source located 1 cm in front of the constriction, then the first pole of the system in front of the constriction is the first resonance of the quarter-wave resonator, $c/4l = 4425$ Hz, where c is the speed of sound (354 m/s) and l is 0.02 m. The calculated spectrum of the fricative has a wide bandwidth high frequency peak around 4500 Hz reflecting the filtering of the pressure source by the short cavity in front of the constriction at the alveolar ridge.

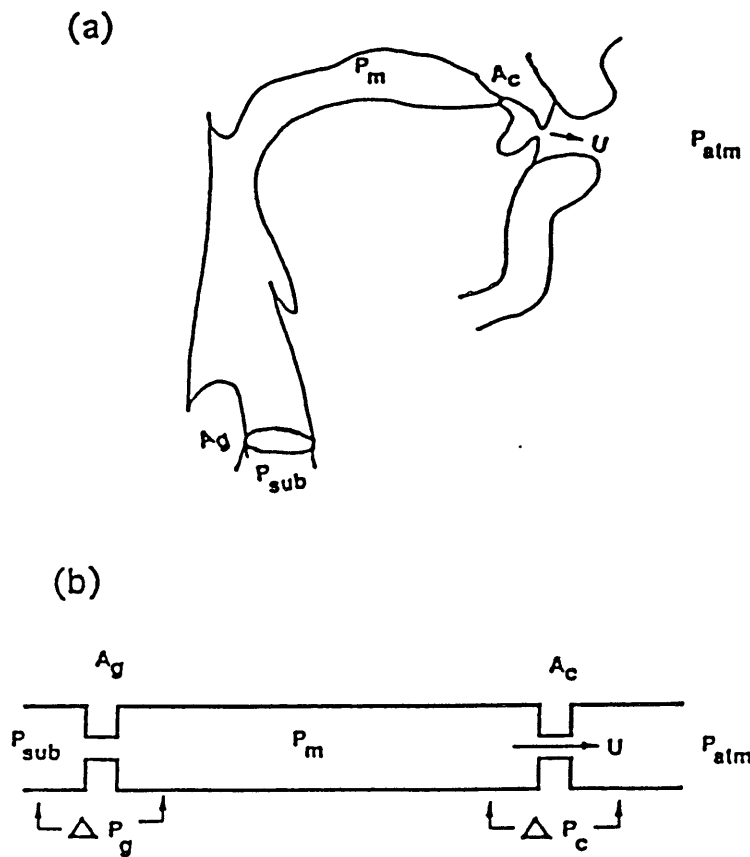


Figure 2.1a) Midsagittal cross-section of a vocal tract vocal-tract model for an alveolar fricative. b) Vocal tract idealized as a uniform tube with constrictions at the glottis and in the oral region (Wilde, 1995).

2.3 Past Studies

In 1975, Kent and Netsell demonstrated that cineradiographic and spectrographic analysis revealed abnormalities in speaking rate, stress patterns, articulatory placements for both vowels and consonants, velocities of articulator movement, and fundamental frequency contours in the speech of ataxic dysarthria patients. Kent et al. (1979) in another acoustical study observed abnormal timing patterns, with prolonged segments and equalized syllable durations.

Fricative production of a Broca's aphasic subject was analyzed spectrographically and compared to that of a normal subject (Code and Ball, 1982). For the aphasic subject, vocal-fold vibration was absent from 'voiced' fricatives, and vowel duration differences preceding fricatives were maintained normally, although friction duration in both voiced and voiceless phonemes was longer than normal speakers.

In a study of a group of 10 women with amyotrophic lateral sclerosis, the mean slope of the second formant F2 for a variety of consonant-vowel syllables was reduced compared with that of a normal geriatric control group (Kent et al., 1992). This F2 slope reduction is interpreted to reflect loss of lingual motorneurons. Acoustic measures of phonatory function for sustained vowel prolongation demonstrated abnormalities in fundamental frequency, perturbations in frequency (jitter) and amplitude (shimmer), and signal-to-noise ratio. Comparison of the results of this study on the group of women to a similar study on a group of men with ALS showed that men were more likely to have impairments of voicing in syllable-initial position (Kent, 1990).

Assessment of the intelligibility of speech of dysarthria patients has been another focus of past studies. Although it may not be necessary for clinicians to quantify a dyarthric individual's intelligibility, it is nonetheless useful in research studies. A small number of item identification tests have been developed specifically for this purpose. These tests include the Frenchay Dysarthria Assessment (Enderby, 1983). In this test, intelligibility is evaluated with three tasks: word task, sentence task, and conversation. Yorkston and Beukelman's (1981) Assessment of Intelligibility in Dysarthric Speakers test involves item identification at both single-word and sentence level. This test eliminates the need for new listeners every time an intelligibility estimate is desired. Procedures developed by Tikofsky and Tikofsky (1964) and by Tikofsky (1970) estimate single-word intelligibility of dyarthric speakers. Platt, Andrews, and their associates at the Spastic Center of New South Wales in Sydney, Australia, determined the intelligibility of cerebral-palsied speakers with a phonetically balanced list of 50 words (Platt, 1980). Most recently, Kent et al. (1989) designed an intelligibility test correlating dysarthric intelligibility impairment and acoustic properties. The test is based on a database consisting of a 70-word list, and is designed to examine 19 acoustic-phonetic contrasts that are likely to be sensitive to dysarthric impairment and contribute significantly to speech intelligibility. Two parameters that have been proven to correlate with speech intelligibility in dysarthric speakers, are speech rate and slope for F2 frequency change (Kent, 1992).

One study that made use of Kent's intelligibility test is Hwa-Ping Chang's Doctoral Thesis (1995) on speech input for dysarthric computer users. He focused on how best to use speech recognition techniques for augmenting the communication

abilities of dysarthric computer users. He examined in detail the data from eight dysarthric speakers based on four types of analysis: a closed-set intelligibility test, phonetic transcription, acoustic analysis of certain utterances, and an evaluation of the recognition of words by a commercial speech recognizer. One important conclusion from the study, which motivated this thesis, is that more of the articulation errors were found in the production of obstruent consonants. More specifically, many of the errors were place errors for alveolar consonants, particularly fricatives.

Another acoustic-phonetic study was performed on four Swedish dysarthric speakers. The study conducted by Hunnicutt, Nord, and Rosengren (1996) was designed to indicate which of the speech processes seem most affected, the type of dysarthria that correlated with the symptom constellation, and the severity of dysarthria. Both qualitative and quantitative data were collected. Four subjects with dysarthria were given a Swedish test (Hartelius and Svensson, 1990) to assess their speech. For the two male speakers, all consonant clusters beginning with /s/ were articulated with an intrusive schwa vowel. Intense energy down to 2 kHz was observed in the dental fricative /s/. In one female speaker, initial consonant clusters beginning with /s/ were never realized as intended. Several other results were described in the report. Due to the limited number of test subjects, however, the results may only be taken as describing examples of the subjects' speech.

Most recently, Chiu, Chen, and Tseng (1996) correlated semiquantitative clinical assessment of cerebellar function to quantitative acoustical analysis of cerebellar speech. Results show that overlapping of voice onset time for voiced and voiceless stop consonants highly correlates with truncal ataxia as well as eye movement abnormality

and hand-eye coordination. The prolongation of vowel duration also correlates significantly with tandem walking, dysdiadochokinesia, and eye movement abnormality. The study concludes that the midline structures-vermis and fastigial nucleus-seem to be the primary focus for the coordination of motor speech in the cerebellum.

These studies showed that acoustic analysis can provide objective measurements of abnormalities in dysarthric speech and might be a reliable means of documenting speech changes due to treatment.

2.4 Preliminary Studies

In a preliminary study, averaged spectra were calculated for utterances (from the 70-word list) that contain English fricatives /s/ and /sh/ in four of the male dysarthric speakers. The speakers ranged in age from 38-61. Three of speakers were diagnosed with cerebral palsy, and one with cerebellar ataxia. The intelligibility of the speakers for words beginning with /s/ or /sh/ ranged from 20 to 90 percent. Characteristics of the noise spectra were analyzed with respect to the adjacent vowel spectra. The spectra for /s/ and /sh/ were taken at the midpoint of the fricative, and the spectrum of the following vowel was sampled 20 ms after the fricative-vowel boundary. In the more severe dysarthric cases, the boundary was less defined and therefore the slices were taken more subjectively. More objective methods have been devised in more in-depth studies to be described later in this thesis. From initial data, the relative amplitude difference between the peak value in the high frequency (> 3 kHz) region of the fricative consonant (A1f) and the first peak value in the vowel region (A1v) was highly correlated with intelligibility scores determined by Hwa-Ping Chang for his Ph.D. thesis (1995). Intelligibility decreased as the amplitude difference between the two peak values

increased. It is hypothesized that the decreased amplitude of the frication noise for the fricatives produced by the dysarthric speakers probably is a consequence of incorrect placement and shaping of the tongue blade when it is raised toward the hard palate. Either the constriction is not sufficiently narrow, or the tongue blade is placed too far forward. For some utterances, a strong low-frequency peak in the fricative spectrum indicates that there is presence of voicing for the fricatives. Large duration variation also exists between the speakers. Both time variation, and presence of voicing along with many other variables present many challenges for this thesis. A histogram of the amplitude difference measurements of A1v and A1f for all normal and dysarthric utterances of /s/ in the data is shown in Figure 2.2. This histogram shows that the high-frequency level of the fricative tends to be weaker for the dysarthric speakers than for the normal speakers. From the preliminary analyses involving examination of several acoustic attributes, it appears that it may be possible to devise combinations of several parameters to understand and model impairments in the placing, manner, and voicing of dysarthric speakers' speech. These preliminary results motivated a more extensive and systematic study with more utterances and speakers.

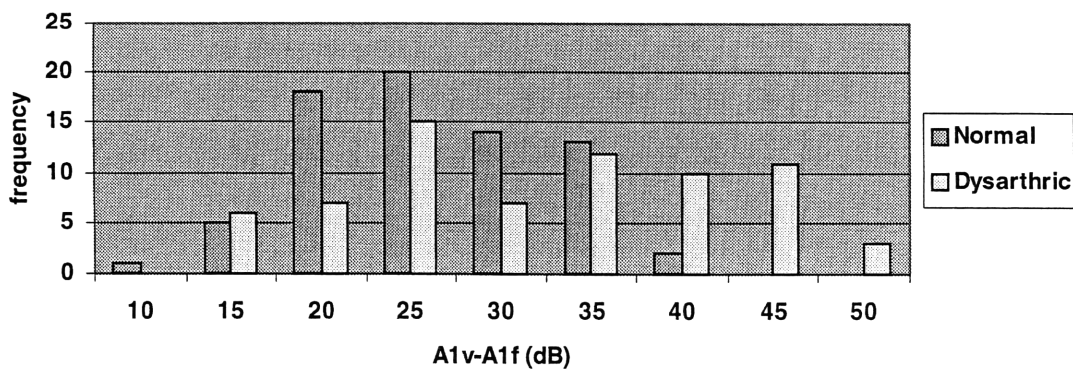


Figure 2.2 Histogram of A1v-A1f measures in normal and dysarthric data for /s/ initial utterances.

Chapter 3 Methods and Procedures

The databases used, the various methods of obtaining the new acoustical spectral measures, and the procedures for analyzing the significance of the measures are described in this chapter.

3.1 Data Acquisition

Utterances of eight dysarthric speakers and two normal speakers were used for this study. In particular, one repetition of nine words with initial /s/ were selected from the larger database. The nine words are: see, seed, sell, sew, side, sigh, sin, sink, and sip.

The database of the speech of eight dysarthric speakers and a normal speaker was available from previous doctoral thesis work performed by Hwa-Ping Chang at MIT. Hwa-Ping Chang obtained several analogue tape recordings of the speakers. Speech recordings were made in a quiet room with an omnidirectional microphone placed 10 cm away from the mouth of the speaker. The recordings consisted of words taken from a 70-word list published by Kent et al. (1989) for intelligibility testing. Both the overall intelligibility and the intelligibility of the /s/ words were also determined previously by Chang.

A new recording of speech from seven normal speakers of the 70-word list was collected and added to the existing database. The subjects recorded the word list, in a random order, five times each. The speech recordings made with digital tapes were recorded in a quiet room with one omnidirectional microphone 10 cm away from the

mouth of the speaker. Since speech signals of normal speakers exhibit little variation, recordings from only two of the normal speakers (one male, one female) were used in this study along with recordings of the eight dysarthric speakers' utterances. More normal speaker data is available for analyzed in future studies.

The recordings were digitized using a VAX computer system in the Speech Communication Laboratory. The sampling rate is 16 kHz with a 16-bit A/D board and a 7.5 kHz low-pass filter.

3.2 Summary of Subjects

Subject information is obtained from Hwa-Ping's Doctoral Thesis. All the special subjects are individuals who have speech disorder due to neurological damage to the muscular control of the speech production mechanism. Out of the nine speakers who served as subjects for his study, eight are speakers with dysarthria (four male and four female) and one is a normal speaker (male) used as a control subject. All subjects are native speakers of English and have no apparent cognitive deficits. Pertinent information on the eight dysarthric speakers (four male and four female) is summarized in the table below:

Subject	Age	O. I. (%)	I. /s/ (%)	Disorder Type
M1	61	95	90	Spastic Cerebral Palsy
M2	38	82	65	Cerebellar Ataxia
M3	48	60	31	Athetosis Cerebral Palsy
M4	45	57	23	Spastic and Athetosis Cerebral Palsy
F1	61	97	98	Spastic Cerebral Palsy
F2	24	89	94	Spastic Cerebral Palsy
F3	62	61	52	Spastic Cerebral Palsy + Surgery damage
F4	22	64	38	Spastic Cerebral Palsy

Table 3.1: Summary pertinent information on dysarthric subjects. O.I. – Overall Intelligibility; I. /s/ - Intelligibility of /s/.

The four males and four female dysarthric subjects range in age from 22 to 62 years. Six of them have cerebral palsy (spastics: 4, athetosis: 1, and spastics plus athetosis: 1), one with spastic cerebral palsy and surgery damage to the brain, and one with cerebellar ataxia. The education level of the subjects range from 5th grade to master's degree.

All seven subjects (four female, three male) for the new recordings of the 70-word list performed for this study are normal native English speakers.

3.3 Acoustic Analysis

Based on preliminary studies, eight measurements were devised to analyze the fricative /s/ in dysarthric speech. Three of them are quantitative measurements obtained from averaged spectra to determine the closeness of the spectral shape to a normative /s/ production model. Five of them are estimates relating to the time course of the formation and release of /s/. These estimates were obtained by making judgements of particular attributes of spectrograms of utterances from the words of each subject in the database.

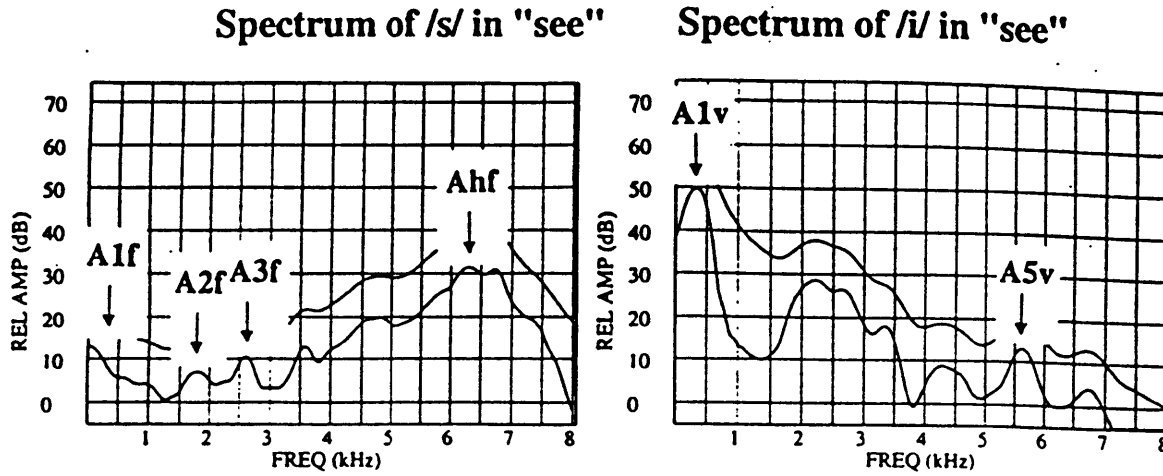
3.3.1 Measurements from Spectra

Time-averaged spectra of the fricative /s/ and the adjacent vowel /i/ in the word “see”, as produced by a normal speaker, are shown in Figure 3.1. Averaged DFT spectra were obtained by stepping a relatively short Hamming window of 6.4 ms every 1 ms over a 100 ms time interval before the fricative-vowel boundary. An average DFT spectrum for the adjacent vowel was obtained similarly in the 100 ms time interval after the fricative-vowel boundary. An average of 100 ms was chosen to minimize fluctuations due to

insufficient sampling. A duration of 100 ms allows enough averaging to incorporate any time variations that may be present in individual utterances. All spectra are calculated without preemphasis.

From the spectra, the following values were measured: a) A_{hf} – peak amplitude above 4 kHz region of the fricative /s/, b) A_{1f} , A_{2f} , A_{3f} – Amplitude of first three formants in the frication noise region, c) A_{1v} – amplitude of the first formant in the adjacent vowel. All these amplitudes are measured in decibels (dB). A set of three acoustic parameters was then calculated from these measurements. For a given utterance, these parameters are all amplitude differences in dB within the utterances. The first parameter calculated is the spectral tilt of the /s/ spectrum ($A_{hf} - \text{average}(A_{2f}+A_{3f})$), a measure of the high frequency prominence relative to the mid frequency spectrum amplitude. A sufficiently large positive value for this parameter would indicate that the constriction formed by the tongue blade is no more than about 2 cm posterior to the lips. A major spectral prominence at a frequency below 4 kHz would suggest a constriction in a more posterior location. In the normative model, the value of this parameter is greater than 10 dB. The difference between A_{1v} and A_{1f} , i.e. $A_{1v}-A_{1f}$, gives a measure of the amount of voicing in the fricative relative to the adjacent vowel. For the voiceless fricative /s/, this difference is expected to be greater than 20 dB. The third parameter is a measure of the amplitude of the first formant in the vowel relative to amplitude of high-frequency prominence in /s/ spectrum. The parameter should indicate whether the amplitude of the fricative is sufficiently high in relation to the vowel. It is an indication of whether the supraglottal constriction is sufficiently narrow or whether there is sufficient

airflow through the constriction. An average value for A1v-Ahf for normal speakers is about 20 dB.



$$\begin{aligned}
 A1v-A1f &= 37 \text{ dB} \\
 Ahf-\text{average}(A2f+A3f) &= 22 \text{ dB} \\
 A1v-Ahf &= 19 \text{ dB}
 \end{aligned}$$

Figure 3.1 a) & b) Example of time-averaged DFT Spectra of /s/ and /i/ in “see” for normal speakers. Measurements are marked on the plot and the calculations of A1v-A1f (voicing), spectral tilt, and A1v-Ahf are given below the figure.

3.3.2 Assessments from Spectrograms

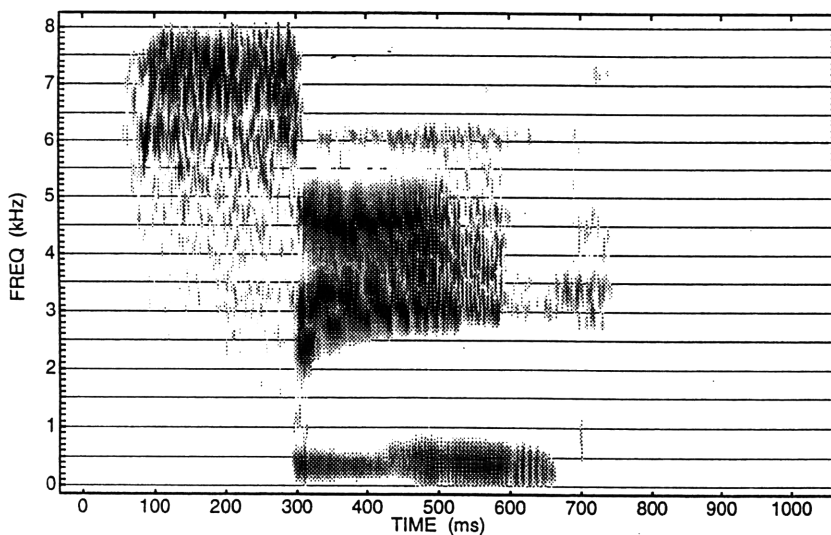
Two experimenters assessed the deviation from normal of the formation and release of the fricatives by examining the spectrogram of each word. Broadband spectrograms (6.4 ms Hamming Window) were made for each of the nine word-initial /s/ utterances by all ten speakers (eight dysarthric, two normal). The assessments made from the spectrograms were on a scale of 1 to 3, with 1 being closest to normal and 3 being most deviant from normal. The experimenters worked together, and arrived at a consensus on the rating

number for each attribute after some discussion. An attempt was made to base the scaling on attributes for which quantitative measures could ultimately be developed. The following five attributes were rated:

1. Precursor – The amount of noise or voicing energy prior to the closure for the /s/
2. Duration – The duration of the /s/ noise
3. Time variation – Variation of noise amplitude or spectrum shape within the /s/
4. Abruptness – Abruptness of the transition from /s/ to its adjacent vowel
5. F2 movement – F2 frequency change in the adjacent vowel in the 50-odd milliseconds following the consonant release

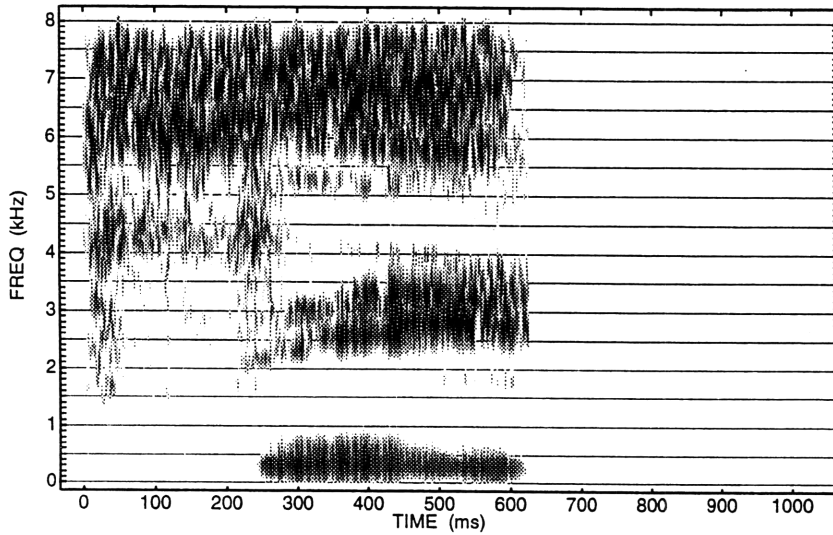
Figure 3.2 gives examples of spectrograms of three productions of “see”: one by a normal speaker, one by a highly intelligible dysarthric speaker, and one by a dysarthric speaker with low intelligibility. Average assessments by the experimenters are shown with the spectrograms.

Normal Speaker



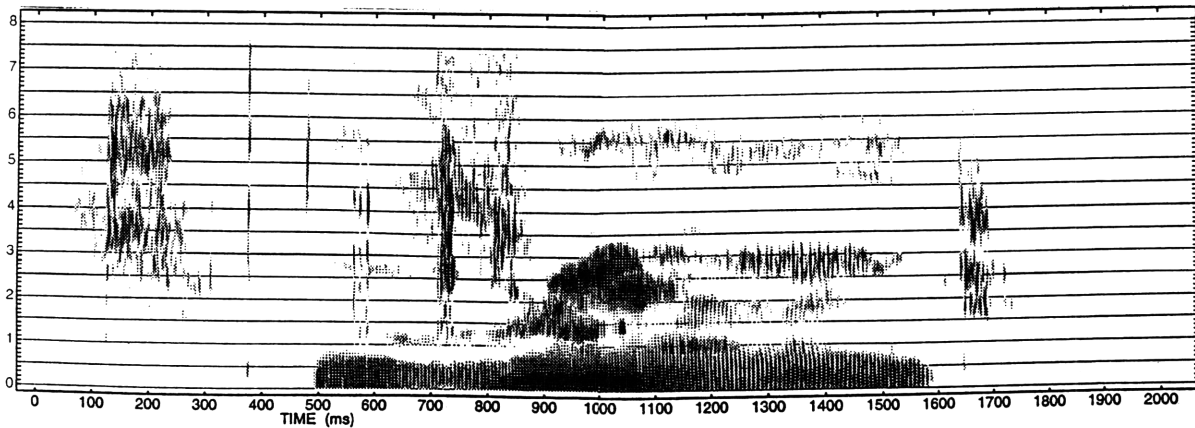
Precursor (1)
Duration (1)
Time Variation (1)
Abruptness (1)
F2 Movement (1)

High Intelligibility



Precursor (1)
Duration (2)
Time Variation (2)
Abruptness (2)
F2 Movement (1)

Low Intelligibility



Precursor (1)
Duration (1)
Time Variation (1)
Abruptness (1)
F2 Movement (1)

Figure 3.2 Example of spectrograms of production of “see” by a normal speaker, a dysarthric speaker with highly intelligibility and a dysarthric speaker with low intelligibility. Average assessments (on a scale of 1 to 3) of the different attributes by two experimenters are also given.

3.4 Perceptual Rating of /s/

Available from Hwa-Ping Chang's research were data on the overall intelligibility for each dysarthric speaker and the intelligibility for words that begins with /s/. The latter intelligibility was determined based on the entire /s/ word rather than the /s/ characteristic alone. In the intelligibility test, listeners had to select one out of four possible choices, one of these was the correct word, and the others differed from the target word in one or more features in the initial or final consonant, or in the vowel. Consequently, an error in a response did not necessarily indicate that /s/ was unintelligible.

A new perceptual rating of based on /s/ words in the 70-word list was proposed in order to obtain a more accurate perceptual correlate of the spectral acoustic parameters for the /s/ in each word. The new test was given to four trained speech pathologists or scientists. Each experimenter was asked to listen to an analogue tape of 90 utterances and select the description that best describes the naturalness of the /s/: Good /s/; some characteristics of /s/ but not a good /s/; fricative but not /s/; not a fricative. The 90 utterances were composed of nine /s/-initial words from the 70-word list utterance produced by each of the eight dysarthric subjects and two normal subjects (one male, one female). An overall numerical rating was calculated for each utterance. This rating was determined by assigning 4, 3, 2, and 1 respectively to the four categories: "good /s/", "some characteristics of /s/ but not a good /s/", "fricative but not /s/", and "not a fricative". The experimenters were also asked to judge the voicing characteristic of each utterance.

Chapter 4 Results and Discussion

The new spectral parameters devised in this thesis were examined in relation to intelligibility scores determined previously by Hwa-Ping Chang as well as the new /s/ perceptual rating scores. This chapter summarizes and interprets the results in detail and compares the effectiveness of the new parameters in determining the intelligibility of the fricative /s/ in dysarthric speech. Table 4.1 summarizes the spectral parameter values in relation to the individual intelligibility scores. (See Appendix A for detailed individual data.) The rating scores for the five spectrogram attributes and the three spectral measures were averaged over the nine utterances containing /s/ in initial position.

Subject	/s/ Perceptual Rating	I. of /s/ (HP)	O.I.	Abruptness	Duration	Time Variation	Precursor	F2 Movement	Total Average	A1v-A1f (dB)	Spectral Tilt (dB)	A1v-Ahf (dB)
M4	1.4	23%	57%	2.2	1.6	2.7	2.9	2.0	2.3	4	-3.3	31.9
F3	1.4	52%	61%	2.4	2.2	2.2	2.9	1.7	2.3	2.4	-1.8	17.4
F4	1.9	38%	64%	1.7	1.3	2.3	1.2	2.3	1.8	32.8	-11.8	43.8
M3	1.9	31%	60%	2.0	1.8	2.1	1.6	1.6	1.8	9.4	-1.4	31
M2	2.1	65%	82%	1.3	1.1	2.0	1.2	1.4	1.4	3.4	-3.8	36.9
F1	2.8	98%	97%	1.2	2.0	1.1	1.0	1.0	1.3	28.4	-0.4	31.6
F2	2.9	94%	89%	1.8	1.3	1.3	1.0	1.0	1.3	21.8	5.9	21.3
M1	3.2	90%	95%	1.2	1.8	1.2	1.0	1.1	1.3	31.5	5.7	28.7
NF	3.4	100%	100%	1.1	1.0	1.0	1.0	1.0	1.0	34.9	15.8	28.9
NM	4.0	100%	100%	1.0	1.0	1.3	1.1	1.1	1.1	35.3	13.2	24.6

Table 4.1 Summary of spectral parameters and assessments in correlation with intelligibility. The first eight rows are the dysarthric speakers and the bottom two rows are the normals with overall intelligibility of 100%. I. /s/ - Intelligibility of /s/; O.I. – Overall Intelligibility.

4.1 Acoustic Analysis Based on Overall Intelligibility from Hwa-Ping's Doctoral Studies

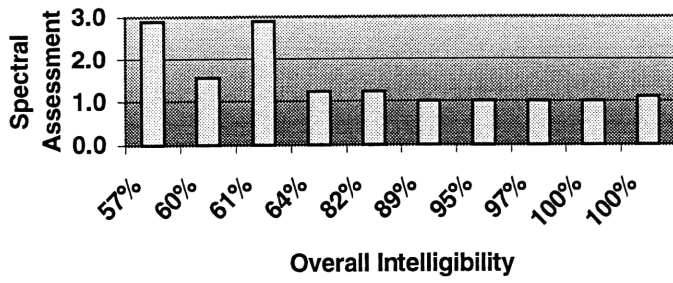
Correlations between the new acoustic parameters and overall intelligibility are presented in this section. The first group of parameters spectrally assesses the proper formation and release of fricative from broadband spectrograms. The second group of parameters describes the acoustic and phonetic deviations from normal based on measurements of spectra.

4.1.1 Assessment of Formation and Release of Fricative from Spectrograms

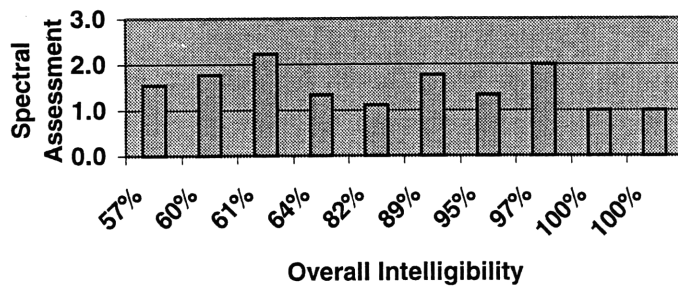
The various parts of Figure 4.1 summarizes the results of spectral assessments of formation and release of fricative /s/ in relation to overall intelligibility. Speakers are labeled by their overall intelligibility, from least to most intelligible, on the horizontal axis. Assessments have values between 1 and 3 as labeled on the vertical axis. Each bar represents assessment averages of the nine /s/ initial words produced by the specified speaker, and judged by two experimenters.

In general, speakers with low intelligibility have a tendency to have poor spectral ratings. Speakers with similar intelligibility can, however, have large individual differences in spectral characteristics. Depending on the speaker, different factors appear to contribute to the decrease in intelligibility.

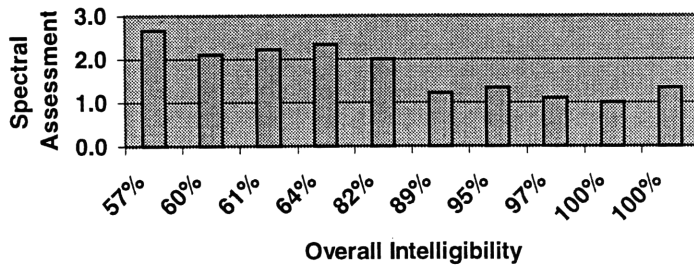
4.1a) Precursor



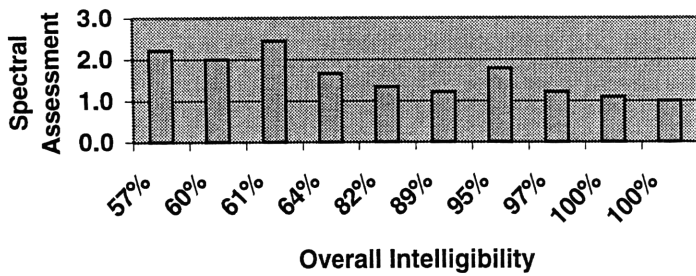
4.1b) Duration



4.1c) Time Variation



4.1d) Abruptness



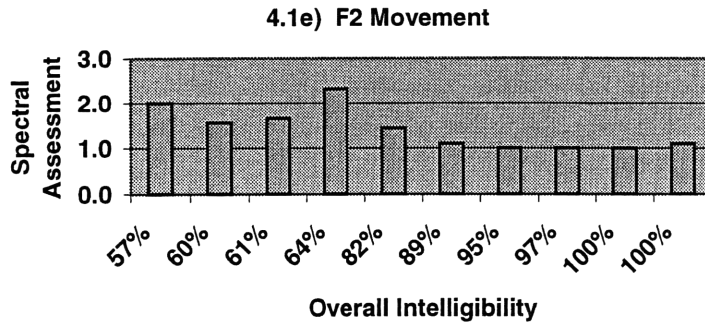


Figure 4.1 a)-e) Plot of assessment of formation and release of fricative /s/ in relation to overall intelligibility of individual speakers.

4.1.2 Articulatory and Phonetic Features Based on Measurements of Spectra

Figure 4.2 summarizes the results for the three parameters derived from spectral measurements of the fricatives and their adjacent vowel spectra. The spectral parameters give an indication of the proper adjustment of the articulators in the production of /s/. Speakers are labeled by their overall intelligibility, from least to most intelligible, on the horizontal axis. Vertical axes are labeled with the respective parameters. Each bar represents assessment averages of the nine /s/ initial words produced by the specified speaker.

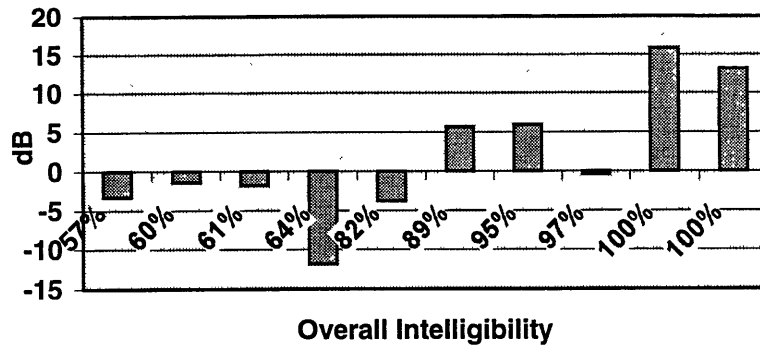
From the plot of spectral tilt, six out of the eight dysarthric speakers showed more low frequency energy than high frequency energy in the noise spectrum. Four of these speakers have relatively low intelligibility. As discussed earlier in Chapter 2, a normal speaker with a 2-cm front cavity has a front-cavity natural frequency at about 4500 Hz. The energy at the natural resonant frequency is usually significantly greater (>10dB) than energy in the lower frequency region corresponding to the second and third formants

which, for /s/, are normally the natural frequencies of the vocal tract behind the constriction.

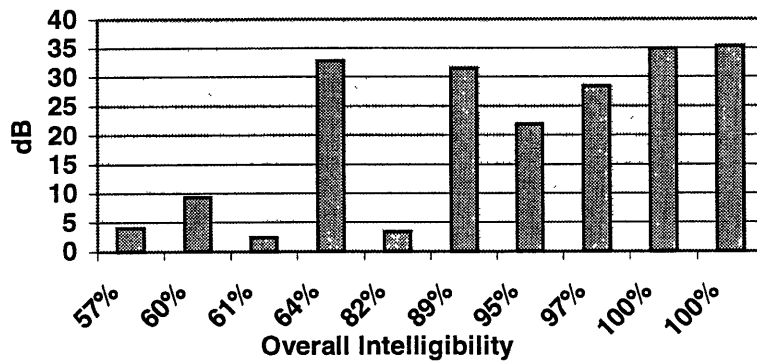
The plot of A1v-A1f shows the amount of voicing that is present in the attempts of individual speakers to produce /s/. Four out of the eight dysarthric speakers voiced their /s/. Normally the amplitude of the first formant (F1) for the fricative at the onset of glottal vibration in the following vowel is much higher (>30 dB) than the amplitude of in the F1 range for the fricative. However, in the case where speaker voices the fricative, F1 is excited in the fricative and the amplitude difference become much smaller. As expected, categorization of the /s/ produced by dysarthric speakers into voiced and unvoiced using the A1v-A1f measurement is consistent with observations made from spectrograms. Figure 4.3 gives the experimenters' assessment based on examination of spectrograms, and shows agreement with the measurements in Figure 4.2b.

Although preliminary data showed that the relative amplitude difference between high frequency region of fricative and its adjacent vowel F1 measurement was well correlated with intelligibility, results in Figure 4.2c show that, with a larger set of utterances and speakers, there is very little correlation between the two factors. The lack of correlation may be due in part to improper production of the vowel, leading to an abnormally high or low spectrum amplitude of the vowel in the F1 region.

4.2 a) Spectral Tilt



4.2 b) Voicing (A1v-A1f)



4.2 c) A1v-Ahf

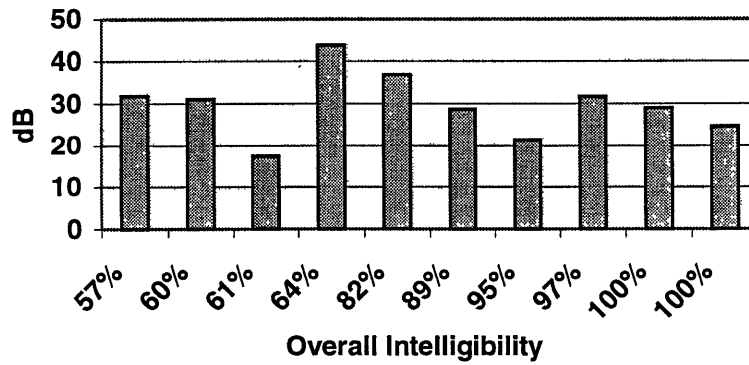


Figure 4.2 Plot of three parameters derived from measurements of spectra of fricative /s/ in relation to overall intelligibility of individual speakers.

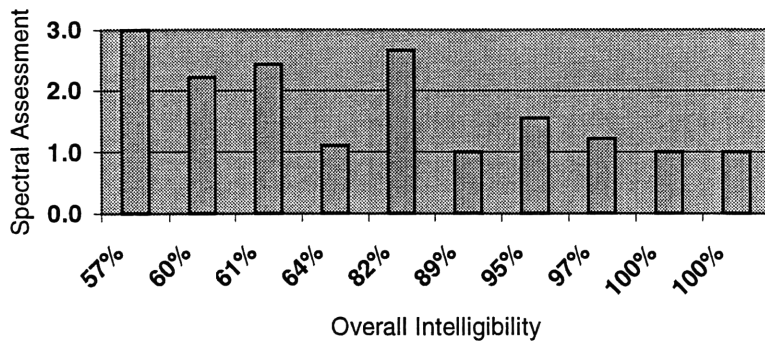


Figure 4.3 Plot of spectrogram assessment of voicing characteristic of fricative /s/ in dysarthric speakers in relation to overall intelligibility.

4.1.3 Total Assessment Average vs. Overall Intelligibility

Figure 4.4 shows a plot of the overall average rating for all five assessments (on a scale of 1 to 3) based on spectrograms of /s/ versus overall intelligibility. The data separate into three groups: normal, high intelligibility, and low intelligibility as marked. For the low intelligibility group, the average assessment on a three-point scale is about 2.0.

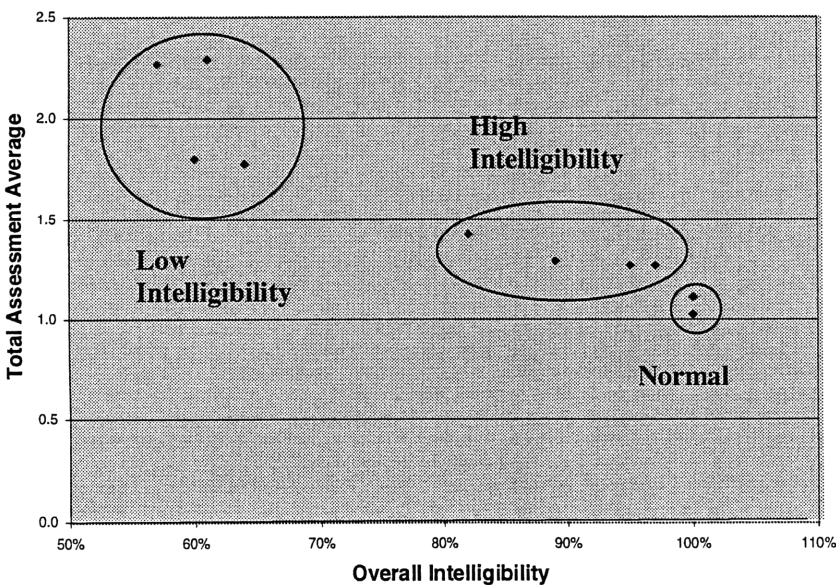


Figure 4.4 Total average assessment of the release and formation of fricative /s/ versus overall intelligibility of the speakers. The data separate into three distinct groups: low intelligibility, high intelligibility, and normal.

4.1.4 Data Grouping – High and Low Intelligibility Groups

Based on the groupings, spectral assessment and measurement data are summarized in Figures 4.5 and 4.6.

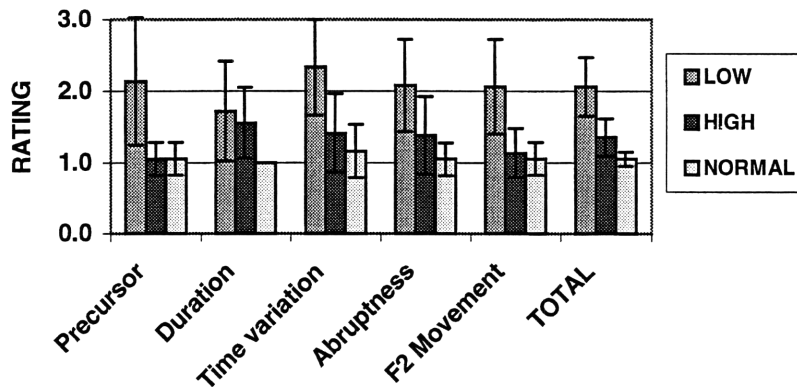


Figure 4.5 Spectral assessment separated by low and high intelligibility groups. Each bar represents average data for the speaker group indicated. The thin line on each bar gives the range for one standard deviation above and below the mean value.

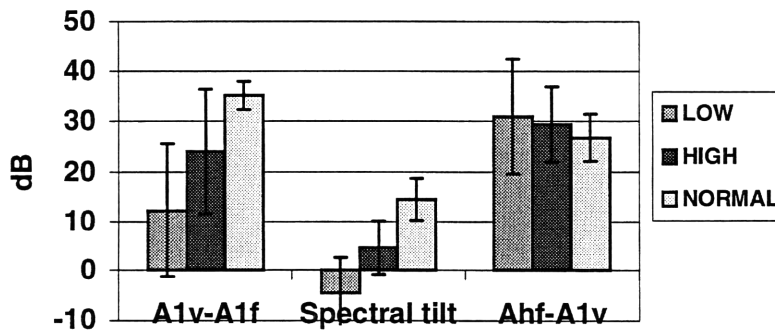


Figure 4.6 Measurement of spectra separated by low and high intelligibility groups. Each bar represents average data for the speaker group indicated. The thin line on each bar gives the range for one standard deviation above and below the mean value.

The low intelligibility group deviates from the normal more than the high intelligibility group. The assessments on precursor and F2 movement do not differentiate the high intelligibility group from normals. Speakers who are highly intelligible are able to make a

narrow enough constriction with their tongue blade in generating turbulence to produce the /s/. In these speakers, high energy noise or voicing is not observed prior to frication, i.e. the precursor rating is close to 1.0. With good tongue movement control, speakers with high intelligibility are also able to produce the correct transition of F2 into the adjacent vowel. The duration of /s/ does not seem to differentiate the low from the high intelligibility group of speakers very well. Ratings on time variation and abruptness differentiate all three groups well. The two spectral measures related to voicing and spectral tilt show the expected trends, whereas high frequency peak amplitude in the fricative relative to low frequency peak amplitude in the vowel does not. In general, variability increases with decreased intelligibility.

4.2 Acoustic Analysis Based on /s/ Perceptual Rating

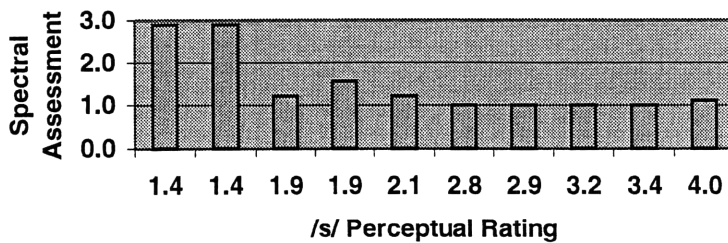
Correlation between the new acoustic parameters and new perceptual /s/ perceptual rating are presented in this section. The first group of parameters spectrally assesses the proper formation and release of the fricative from observation of broadband spectrograms. The second group of parameters describes the acoustic and phonetic deviations from normal based on measurements of spectra. The new /s/ perceptual rating gives ratings solely on the adequacy of production of the word initial /s/ in dysarthric speaker. We expect the /s/ perceptual ratings have higher correlation with the new acoustic parameters.

4.2.1 Assessment of Formation and Release of Fricative from Spectrogram

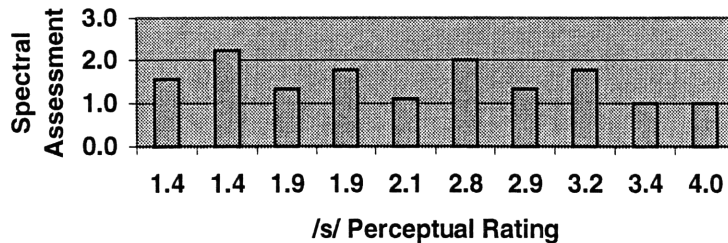
Figure 4.7 summarizes the results of spectral assessments of formation and release of fricative /s/ in relation to /s/ perceptual rating. Speakers are labeled by their /s/ perceptual rating, from least to most intelligible, on the horizontal axis. Assessments are on scale from 1 to 3 as labeled on the vertical axis. Each bar represents assessment averages of the nine /s/ initial words produced by the specified speaker.

In general, speakers with low intelligibility tend to have worse spectral ratings. The trend is more evident when correlated with the new /s/ perceptual rating, especially in assessment of precursor and time variation. Again, speakers with similar intelligibility have large individual differences in some spectral characteristics. Depending on the speaker, different factors contribute to the decrease in intelligibility.

4.7 a) Precursor



4.7 b) Duration



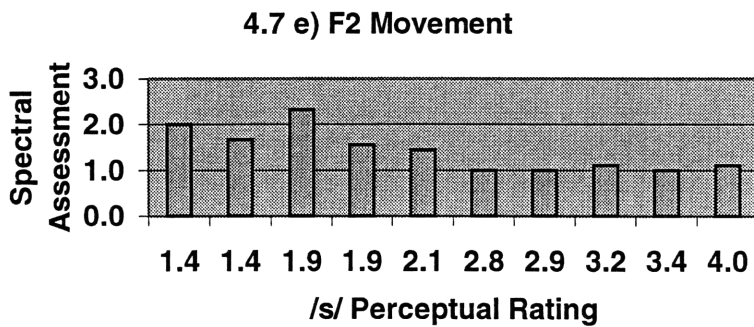
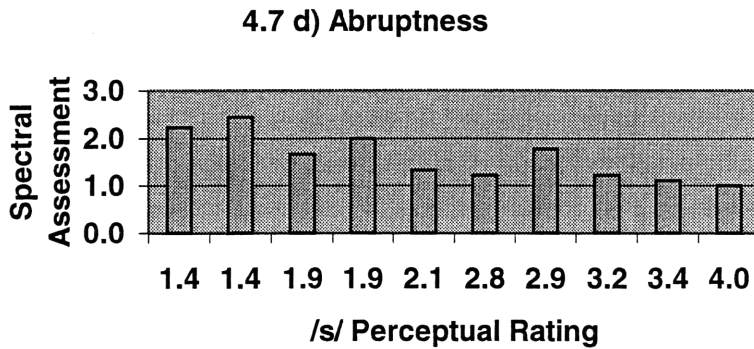
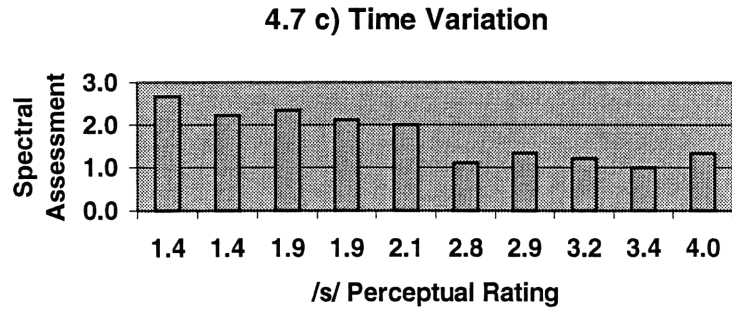


Figure 4.7 a)-e) Plot of assessment of formation and release of fricative /s/ in relation to /s/ perceptual rating of individual speakers.

4.2.2 Articulatory and Phonetic Features Based on Measurements of Spectra

Figure 4.8 summarizes the results of spectral measurement of the fricative and its adjacent vowel spectrum in relation to /s/ perceptual rating. The calculations give an indication as to the proper adjustment of articulators and vocal tract in the production of

/s/. Speakers are labeled by their overall intelligibility, from least to most intelligible, on the horizontal axis. Vertical axes are labeled with the respective parameters. Each bar represents assessment averages of the nine /s/ initial words produced by the specified speaker.

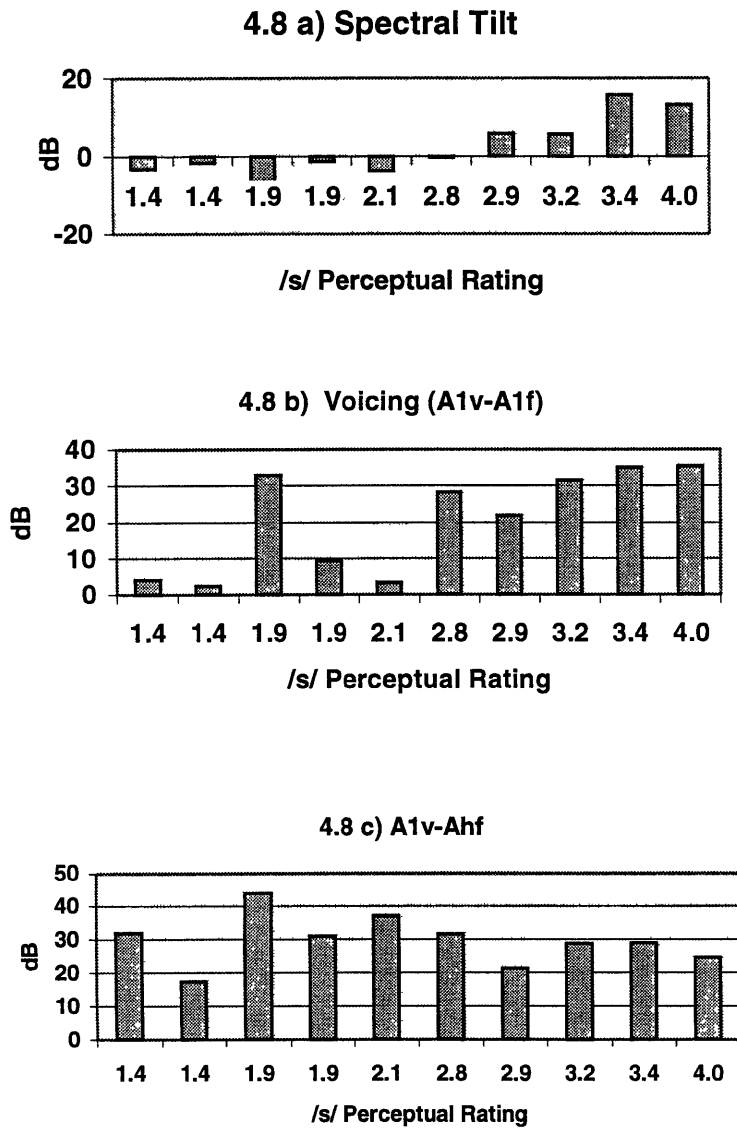


Figure 4.8 Plot of measurements of spectra of fricative /s/ in relation to /s/ perceptual rating of individual speakers.

Trends similar to the analysis based on overall intelligibility are seen in the new analysis. As expected, spectral tilt values are better correlated with /s/ perceptual rating. The six speakers that have higher low frequency energy than high frequency energy noise have the lowest perceptual ratings. A combination of factors leads to a decrease in the rating. An example is the speaker with the third lowest perceptual rating. Although this speaker did not exhibit a significant amount of voicing, the spectral shape was more deviant from the normal, and this factor apparently gave rise to lower rating.

4.2.3 Total Average Assessment vs. New Perceptual Rating

Figure 4.9 shows a plot of the total average rating for assessments of spectrograms of /s/ versus /s/ perceptual rating. The data shows a relative linear correlation between intelligibility and the average combination of all spectral assessments.

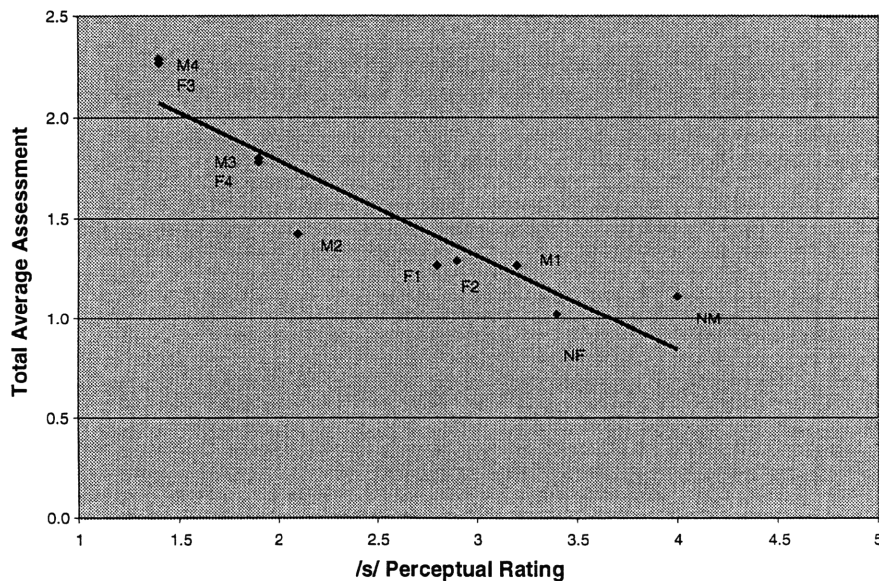


Figure 4.9 Total average assessment of the release and formation of fricative /s/ versus /s/ perceptual rating of the speakers.

4.3 Statistical Analysis

4.3.1 Overall Intelligibility vs. New /s/ Perceptual Rating

Figure 4.10 shows a plot of the overall intelligibility determined previously by Hwa-Ping Chang, and the new /s/ perceptual rating for the eight dysarthric speakers and the two normals. The two sets of data are highly correlated, although there are some deviations. The two data sets have a Pearson correlation coefficient of 0.926. Since the two tests were performed separately by different experienced listeners, the high correlation gives validity to both sets of data. From results presented in the previous sections, the /s/ perceptual rating is indeed better correlated with the new acoustical parameters that assess the proper production of /s/ quantitatively. The two data have a person correlation coefficient of 0.926.

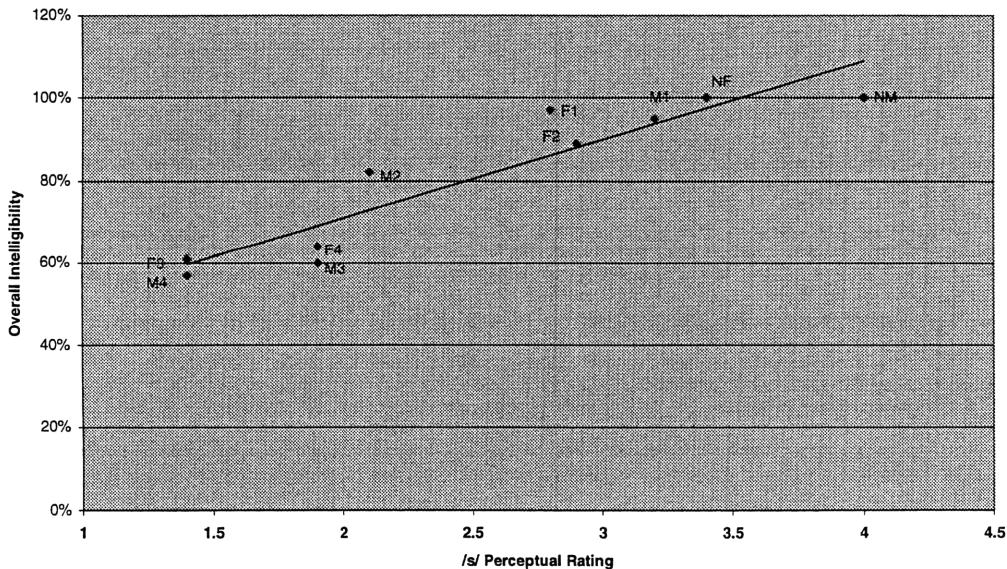


Figure 4.10 Linear correlation between overall intelligibility determined previously by Hwa-Ping Chang and new /s/ perceptual rating devised in this thesis.

4.3.2 Pearson Correlation Matrix

The Pearson product moment correlation coefficient, r , is a dimensionless index that ranges from -1.0 to 1.0 inclusive, and reflects the extent to which there is a linear relationship between two data sets. If the correlation is negative, then large values of one data set associate with small values of the other. If the correlation is positive, then large values of one data set are associated with large values of the other. Results are summarized in Table 4.2.

	Overall Intelligibility	/s/ perceptual rating	Precursor	Duration	Time Var.	Abrupt ness	F2 Move ment	Spectral Tilt	Voicing
Overall Intelligibility	1.000								
/s/ Perceptual Rating	0.926	1.000							
Precursor	-0.767 (4)	-0.750 (6)	1.000						
Duration	-0.403 (7)	-0.496 (7)	0.478	1.000					
Time Variation	-0.952 (1)	-0.887 (1)	0.750	0.213	1.000				
Abruptness	-0.861 (3)	-0.859 (2)	0.845	0.557	0.772	1.000			
F2 Movement	-0.871 (2)	-0.775 (5)	0.778	0.125	0.912	0.603	1.000		
Spectral Tilt	0.753 (5)	0.828 (3)	-0.374	-0.398	-0.770	-0.546	-0.803	1.000	
Voicing	0.695 (6)	0.801 (4)	-0.735	-0.375	-0.706	-0.750	-0.383	0.486	1.000
A1v-Ahf	-0.220 (8)	-0.216 (8)	-0.266	-0.294	0.305	-0.229	0.520	0.132	0.132

Table 4.2 Pearson correlation coefficient matrix for the various spectral measures and intelligibility or perceptual rating scores.

Rankings of correlation between the parameter and intelligibility are listed in parentheses. Since the correlation coefficients are extremely close to one another, the rankings should only be taken as a guide and one should not place too much importance on the absolute significance of each of the parameters. The top three parameters that correlate with the /s/ perceptual rating are: time variation, abruptness, and spectral tilt, compared with time variation, F2 movement, and abruptness when correlating with overall intelligibility. Of some interest, perhaps, is the high correlation between the perceptual rating for /s/ and the overall intelligibility for all 70 words in the test of Kent et al. (1989). The implication is that if a speaker has difficulty in producing /s/, then he or she might also have difficulty with other sounds. Another point of interest is that duration does not seem to correlate well with any of the other parameters and has the lowest correlation with overall intelligibility and /s/ perceptual rating among all the parameters.

4.3.3 Discriminant Analysis

Given the high correlation coefficients, we proceeded to perform a discriminant analysis. The discriminant analysis was performed to determine the factors that contribute to the classification of the speakers into three intelligibility groups: low intelligibility, high intelligibility, and normal. Speakers (M3, M4, F3, F4) were all classified into the low intelligibility group based on the overall intelligibility score as well as the perceptual ratings. The other four speakers (M1, M2, F1, F2) were classified into the high intelligibility group. The contributions of all eight of the measures to this classification were statistically significant. Correlations were calculated between each of the measures and both factors in the discriminant analysis. The three measures that showed the highest

correlation with the first factor are ranked as following: 1) spectral tilt, 2) time variation within /s/, and 3) precursor.

4.4 Discussion on Individual Subjects

Eight speakers with varying degrees of dysarthria and etiologies were the subjects for this study. In this section, significant findings for two of the speakers will be discussed in detail. One of the subjects, F1, has the highest overall intelligibility score, but ranked third on the /s/ perceptual rating scale. The other subject, F3, is one of the least intelligible speakers. The ratings based on spectrograms for each of these subjects together with those of a normal speaker are summarized in Figures 4.11 and 4.12.

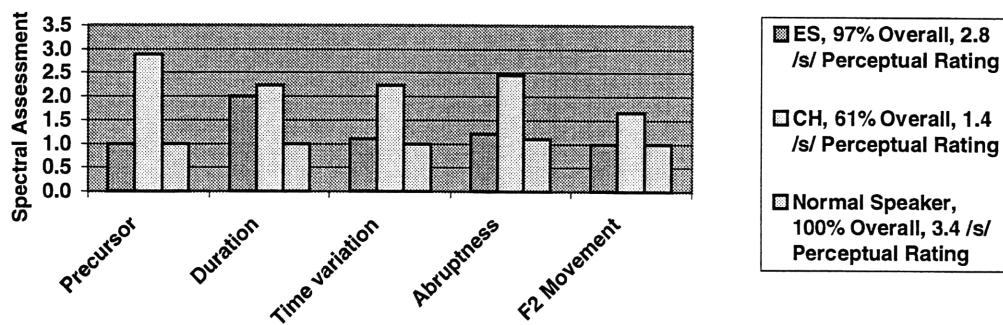


Figure 4.11 Summary of assessment of formation and release of fricative /s/ of three speakers, one normal speaker, one dysarthric speaker with high intelligibility, and one with low intelligibility speaker.

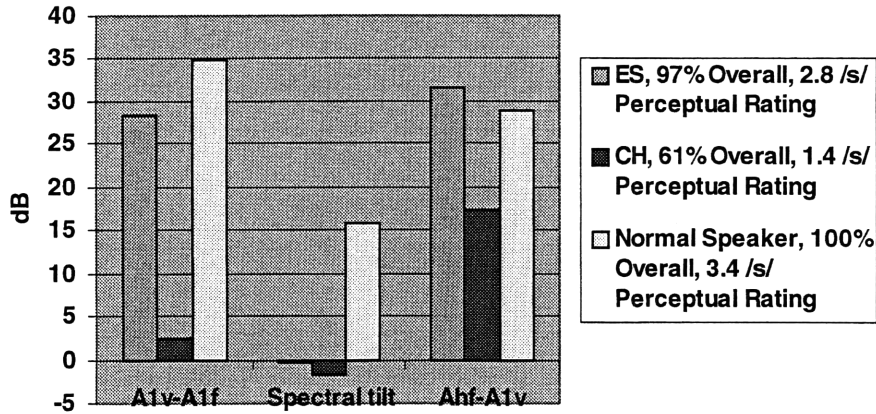


Figure 4.12 Summary of measurements of spectra of fricative /s/ of three speakers as in Figure 4.11.

4.4.1 Subject F1

F1 is 61 years old with a bachelor’s degree. She has had spastic cerebral palsy since birth. Her muscle movements are very stiff and she is only able to move through a limited range. Her cognitive and linguistic abilities are normal, and her speech is the most intelligible of all eight subjects. She has an air flow and vital capacity problem and her speech becomes weak after some time. This is reflected as a decrease in amplitude of her speech spectrum. In separate utterances she is very clear and intelligible, but not in continuous speech. F1’s overall intelligibility is 97%, and her /s/ perceptual rating is 2.8 on a scale of 1 to 4, with 1 being a good /s/ and 4 being not a fricative.

A combination of several slight deviations from normal acoustical patterns as illustrated in F1’s spectrogram of the word “see” in Figure 4.13, results in a small decrease in intelligibility. Intense frication noise energy down to 2 kHz is observed, leading to a relatively flat (approximately 0 dB) spectral tilt. The spectrum of the /s/ for the utterance in Figure 4.13 is shown in Figure 4.14. This may be due to improperly

placed tongue resulting in incorrect filtering of the frication noise source. F1's /s/ perceptual rating is ranked third best rather than the most intelligible of all speakers. The spectrogram also shows that the transition between the frication noise and the adjacent vowel is not as abrupt as that of a normal /s/.

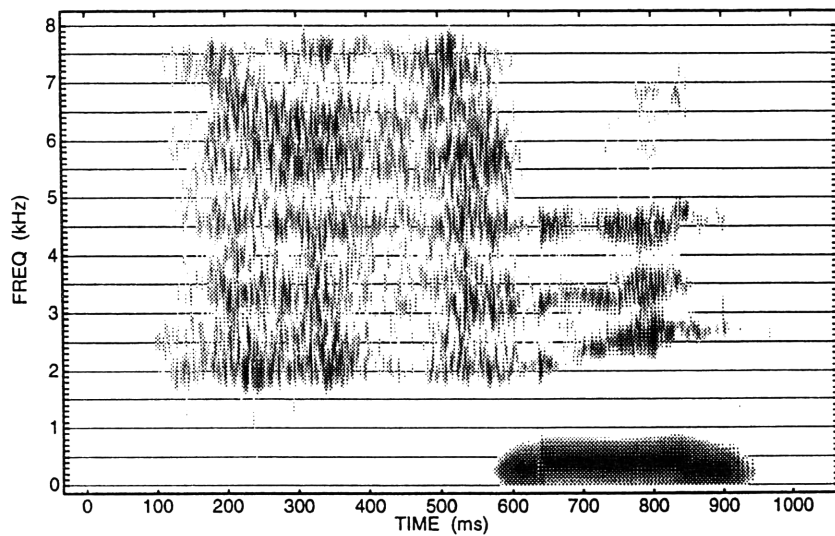


Figure 4.13 F1's spectrogram of the word "see".

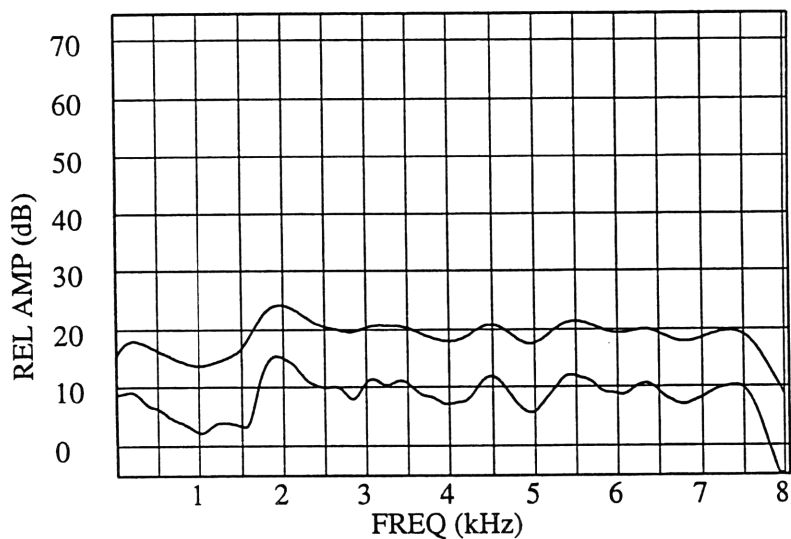


Figure 4.14 F1's time-averaged DFT spectra of /s/, in the 100 ms period preceding the boundary between the fricative and the vowel.

4.4.2 Subject F3

F3 is 62 years old with a 5th grade education. She has also had spastic cerebral palsy since birth. Her muscles are generally still and the movements are awkward. Her speech was intact originally until ten years ago when she had surgery to remove an acoustic neuroma. After the operation, the left side of her face, the left side of her tongue, her left ear, and her vocal folds were paralyzed (Chang, 1995). Her vocal fold and vocal tract nerves were damaged and her speech became abnormal and lisping. Her speech is especially breathy. Her cognitive and linguistic abilities are intact.

The spectrogram of “see” in Figure 4.15 shows a long pre-voicing without a consonantal constriction before the frication of /s/. This precursor to the fricative was given a rating of 3 on a scale from 1 (normal) to 3 (strongly deviant from normal). Apparently the speaker has poor control of glottal opening, which causes her to not be able to build up enough intraoral pressure before the glottal and supralaryngeal articulators are in place. For all of the /s/ initial utterances, the /s/ sounded like /h/, which is not a fricative at all. The /h/ sound can be directly inferred from the time-averaged noise spectra of /s/ as shown in Figure 4.16. There is a lack of front cavity resonance at high frequencies and high noise energy in the low frequency region, giving rise to a negative spectral tilt. Also, F3’s frication noise varies in time more and is much weaker relative to the adjacent vowel than that of a normal speaker. Moreover, the small difference in A1v and A1f amplitudes indicates voicing during frication of the voiceless /s/. Theoretically, if the glottis is not sufficiently abducted, glottal vibration should continue through the fricative. Most likely, paralysis of the speaker’s vocal folds contributed to these deviant acoustic patterns. Transition from frication noise to

vocalization of the adjacent vowel /i/ is not very abrupt again due to poor coordination of changes in the glottal and supraglottal constrictions. Interestingly, there is little elongation in duration of the fricative noise relative to the normal despite the vast difference in intelligibility.

More variability was exhibited across F3's utterances of /s/ initial words than F1's utterances. Overall, F3 tends to have more difficulties with underlying neuromotor functions such as respiration, phonation and gross motor control of the tongue. In order to make direct and accurate inferences from the acoustical features to the physiological events, one needs to have much better understanding of the neuromotor functions that are specifically associated with the production of particular sounds.

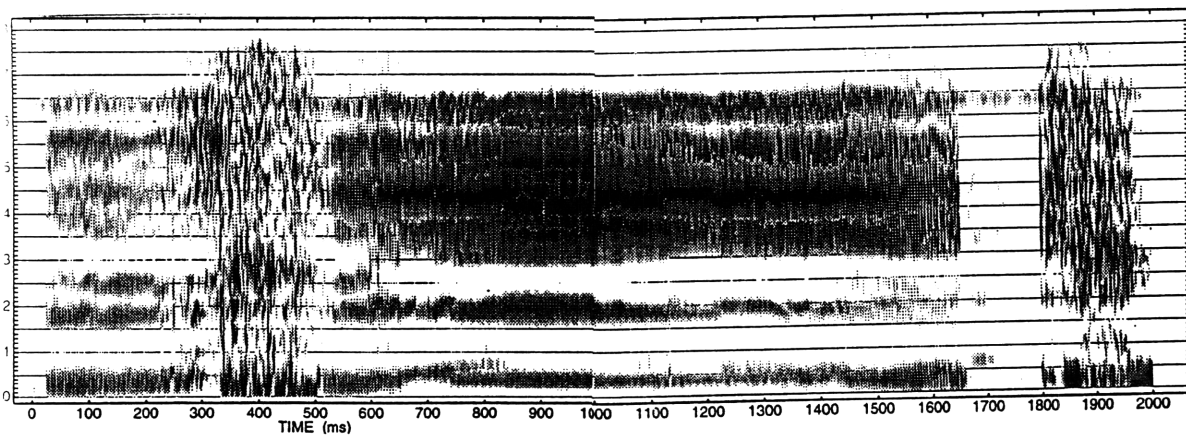


Figure 4.15 F3's spectrogram of the word "see".

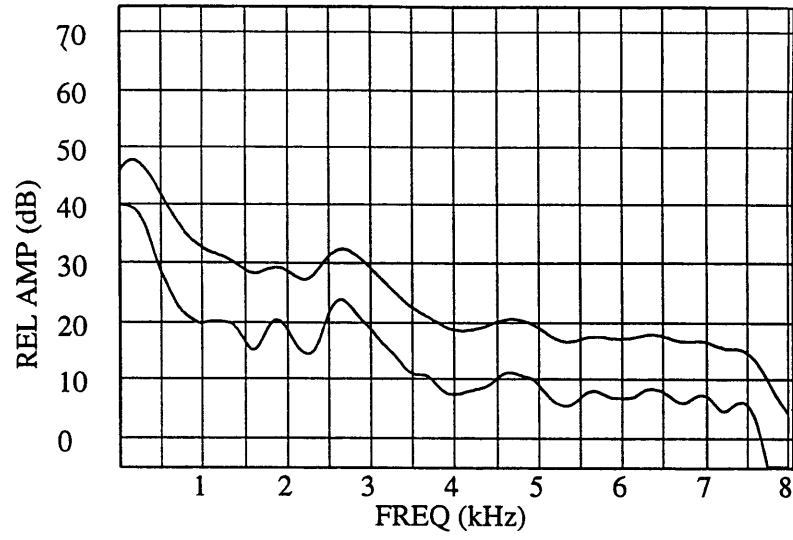


Figure 4.16 F3's time-averaged DFT spectra of /s/, in the 100 ms period preceding the boundary between the fricative and the vowel.

Chapter 5 Conclusions

5.1 Summary

We have obtained a variety of acoustically-based measures of word-initial /s/ produced by several dysarthric and normal speakers, using spectral analysis methods as well as judgements based on spectrographic observations. These measures have provided preliminary quantitative assessments that reflect the neuromotor dysfunctions of individual speakers. The study has examined measures related to the initiation of the sound, maintenance of the proper tongue blade configuration for /s/, and release of the fricative into a following vowel. Although the eight dysarthric subjects have different symptoms and suffered damage to various parts of the brain leading to varying degrees of speech disorder, all except for one of the spectral measurements and ratings show a good correlation with the intelligibility of the speakers.

Extreme variability in the speech of people with various types of dysarthria makes it difficult to draw conclusions that span across repetitions or speakers. More severely disordered speakers tend to have more difficulties with underlying neuromotor functions such as respiration, phonation and gross motor control of the tongue. These difficulties may not be adequately characterized with the use of acoustic measures traditionally used to study the speech of normal speakers. A combination of new parameters such as the ones presented in this thesis can be used to quantify dysarthric speech. Determining such sets of parameters requires better understanding of the underlying processes that contribute to deviation in the speech of these speakers. Although considerable difficulties

are involved in making inferences from acoustical features to the physiological events that produced them, acoustic measure of dysarthric speech can provide important information that associate the different disordered articulatory events to acoustic events and thereby determine the intelligibility of the speech.

5.2 Future Work

One task of future studies is to automate the measurements that are based on the assessment methods used in this research and to apply the measures to a larger population of speakers and various types of utterances. A goal is to assemble a set of parameters that, in combination, can predict the intelligibility of dysarthric speech signal. The work done in this thesis is merely the beginning.

Future work should also focus on a variety of speech sound classes. In this study, we examined what may be one of the most difficult sounds to produce by speakers with neuromotor disorders. The sound /s/ involves fine control of the shape and position of the tongue blade. Do other tongue-blade consonants have similar problems? Are labials and velars produced with greater accuracy and consistency? Acoustic analysis should help to answer these questions.

The objective measures devised should also be related to assessments of clinicians along several different dimensions. The dysarthric subjects studied all had similar disorder types, cerebral palsy and cerebellar ataxia. It would be interesting to use the methods proposed here to study speech disorders of other origin. The ultimate aim is to specify an inventory of quantitative acoustic measures that can be used to accurately



Room 14-0551
77 Massachusetts Avenue
Cambridge, MA 02139
Ph: 617.253.5668 Fax: 617.253.1690
Email: docs@mit.edu
<http://libraries.mit.edu/docs>

DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Page 54 is missing from original

Chapter 6 References

Caruso, A.J., and E.K. Burton (1987). Temporal acoustic measures of dysarthria associated with amyotrophic lateral sclerosis, *J. Speech and Hearing Research* 24, pp. 80-87.

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

Chiu, M.J., R.C. Chen, C. Tseng (1996). Clinical correlates of quantitative acoustic analysis in ataxic dysarthria, *Eur. Neurol.* 36, pp. 310-314.

Code, C. and M. Ball (1982). Fricative production in Broca's aphasia: a spectrographic analysis, *J. of Speech and Hearing Disorders* 40, pp. 325-331.

Darley, F., A. Aronson, and J. Brown (1968) *Motor Speech Signs in Neurologic Disease*, *Medical Clinics of North America*, 52, pp. 835-844.

Darley, F., A. Aronson, and J. Brown (1975). *Motor Speech Disorders*, Philadelphia: W.B. Saunders.

Dworkin, J.P., A.E. Aronson, and D.W. Mulder (1980). Tongue force in normals and in dysarthric patients with amyotrophic lateral sclerosis, *J. of Speech and Hearing Research* 23, pp 828-837.

Enderby, P.M. (1983). *Frenchay Dysarthria Assessment*, San Diego: College-Hill Press.

Gentil, M., P. Pollak, and J. Perret (1995). La dysarthrie parkinsonienne, *Rev. Neurol (Paris)* 151, pp. 105-112.

Hartelius, L. L. Nord, and E.H. Buder (1995). Acoustic analysis of dysarthria associated with multiple sclerosis. *Clinical Linguistics & Phonetics*, v.9, no.2, pp 95-120.

Hartelius, L. and P. Svensson. (1990). *Dysarthritest*, Hagersten, Sweden: Psykologiforlaget AB.

Hunnicut, S., L. Nord, and E. Rosengren. (1996). An acoustic-phonetic study of several Swedish dysarthric speakers, *Proceedings from Phonetik 96*, Swedish Phonetics Conference, Nasslingen.

Kent, J.F., R.D. Kent, J.C. Rosenbek, G. Weismer, R. Martin, R. Sufit, and B.R. Brooks (1992). Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis, *J. of Speech and Hearing Research* 35, pp 723-733.

Kent, R.D., and R. Netsell (1975). A case study of an ataxic dysarthric: Cinneradiographic and spectrographic observations, *J. of Speech and Hearing Research* 40, pp 115-134.

Kent, R.D., R. Netsell, and J.H. Abbs (1979). Acoustic characteristics of dysarthria associated with cerebellar disease, *J. Speech and Hearing Research* 22, pp. 627-648.

Kent, R.D., G. Weismer, J.F. Kent, and J.C. Rosenbek (1989). Toward Phonetic Intelligibility Testing in Dysarthria, *J. of Speech and Hearing Disorders* 54, pp. 482-499.

Kent, R.D., J.F. Kent, G. Weismer, R. Sufit, J.C. Rosenbek, R. Martin, and B.R. Brooks (1990). Impairment of Speech Intelligibility in men with amyotrophic lateral sclerosis, *J. of Speech and Hearing Disorders* 55, pp 721-728.

LaPointe, L.L. (1994). Diagnosis: Dysarthria. In Robert Trace (Assistant Editor), *Advance for Speech-Language Pathologists and Audiologists*, December 5th, pp. 12-13.

Minifie, F.D. (1994). Introduction to Communication Sciences and Disorders, San Diego: Singular Publishing Group.

Peterson, G.E., and H.L. Barney (1952). Control methods used in a study of the vowels, *J. Acoust. Soc. Am.* 24, pp. 175-182.

Platt, L.J., G. Andrews, and P.M. Howie (1980). Dysarthria of adult cerebral palsy: II. Phonemic analysis of articulation errors, *Journal of Speech and Hearing Research* 23, pp. 41-55.

Stevens, K.N. (1971). Airflow and turbulence for fricative and stop consonants: Static considerations, *Journal of Acoustic Society of America* 70 , pp. 1180-1192.

Stevens, K.N. (in press). *Acoustic Phonetics*. Cambridge, MIT Press.

Tikofsky, R.S. (1970). A revised list for the estimation of dysarthric single word intelligibility, *Journal of Speech and Hearing Research* 13, pp. 59-64.

Tikofsky, R.S. and R.P. Tikofsky (1964). Intelligibility measures of dysarthric speech, *Journal of Speech and Hearing Research* 7, pp. 325-333.

Turner, G.S. and G. Weismer (1993). Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis, *J. Speech and Hearing Research* 36, pp. 1134-1144.

Wilde, Lorin (1995). *Analysis and Synthesis of Fricative Consonants*, Massachusetts Institute of Technology, PhD Thesis.

Yorkston, K.M., D.R. Beukelman, and K.R. Bell (1981). Assessment of intelligibility of dysarthric speech, Tigard, OR: C.C. Publications.

Appendix A

Individual data giving the spectral measures and the ratings from spectrograms (on a scale between 1 and 3) are summarized here. The individual intelligibility scores from the test of Kent et al. (1989) give the overall intelligibility and the intelligibility of the subset of words with initial /s/. Also given is the perceptual rating of /s/ on a scale of 1 to 4, as described in the text.

**Average DFT-Spectrogram Analysis for
Dysarthric Data**

Date: April 7,
1998

Name	Overall Intelligi bility	Intelligi bility of /s/	New Perceptual Rating of /s/	Word	A1v- A1f	Spectral tilt	Ahf-A1v	Abrupt ness	Duration	Time variation	Precursor	F2 Movement	Total				
M4	57%	23%	1.4	see1	1	-1.5	23	3	2	3	3	3	14				
				seed1	-5	5	25	2	2	3	3	2	12				
				sell1	8	-10.5	41	2	3	3	2	2	12				
				sew1	4	-9	38	2	1	3	3	3	12				
				side1	1	-6	23	3	1	3	3	2	12				
				sigh1	2	5.5	22	2	1	2	3	2	10				
				sin1	9	-3.5	43	3	1	2	3	1	10				
				sink1	10	0	35	2	1	3	3	2	11				
				sip1	6	-10	37	1	2	2	3	1	9				
				AVER AGE	4	-3.3	31.9	20	14	24	26	18	102				
				M3	60%	31%	1.9	see1	20	4	23	2	1	2	1	2	8
								seed1	6	3	22	3	2	2	2	2	11
								sell1	6	3	25	2	1	2	3	2	10
sew1	7	-8.5	38					2	3	2	1	2	10				
side1	12	17	18					2	2	1	2	1	8				
sigh1	13	-11.5	40					2	2	2	1	1	8				
sin1	5	-6.5	35					2	2	3	1	2	10				
sink1	2	-5	33					2	2	3	1	1	9				
sip1	14	-8	45					1	1	2	2	1	7				
AVER AGE	9.4	-1.4	31					18	16	19	14	14	81				
M1	95%	90%	3.2	see1	26	9.5	32	2	1	1	1	1	6				
				seed1	23	0	33	1	2	1	1	1	6				

				sell1	33	3.5	27	1	2	1	1	1	6
				sew1	34	-5.5	39	1	2	2	1	1	7
				side1	30	10	21	1	2	1	1	1	6
				sigh1	32	8	21	1	1	1	1	1	5
				sin1	40	5.5	27	1	2	1	1	1	6
				sink1	31	6.5	33	1	2	1	1	1	6
				sip1	35	14	25	2	2	2	1	2	9
				AVER	31.6	5.7	28.7	11	16	11	9	10	57
				AGE									
M2	82%	65%	2.1	see1	10	-1.5	42	1	2	2	1	2	8
				seed1	-1	-5	35	1	1	2	1	2	7
				sell1	0	-5	41	1	1	2	1	1	6
				sew1	13	-7.5	38	2	1	2	2	1	8
				side1	5	-5	37	1	1	2	1	1	6
				sigh1	12	-6	39	1	1	2	1	2	7
				sin1	-2	-2.5	36	1	1	1	1	1	5
				sink1	-12	1	25	3	1	3	2	2	11
				sip1	6	-2.5	39	1	1	2	1	1	6
				AVER	3.4	-3.8	36.9	12	10	18	11	13	64
				AGE									
F3	61%	52%	1.4	see1	-9	-8.5	21	1	1	1	3	1	7
				seed1	-9	-1.5	12	2	2	1	3	3	11
				sell1	14	1.5	21	3	2	3	3	1	12
				sew1	7	-2.5	12	2	3	3	3	2	13
				side1	10	1	18	3	3	3	3	2	14
				sigh1	-6	-1	14	3	2	2	3	2	12
				sin1	10	0	26	3	2	2	3	2	12
				sink1	1	-7	22	2	2	2	2	1	9
				sip1	4	1.5	11	3	3	3	3	1	13
				AVER	2.4	-1.8	17.4	22	20	20	26	15	103
				AGE									
F1	97%	98%	2.8	see1	35	-2	38	2	2	1	1	1	7
				seed1	30	-2.5	31	1	2	1	1	1	6
				sell1	28	1.5	26	1	2	1	1	1	6

				sew1	34	2	33	1	2	2	1	1	7
				side1	28	3.5	30	1	2	1	1	1	6
				sigh1	25	-1.5	29	1	2	1	1	1	6
				sin1	33	0	31	1	2	1	1	1	6
				sink1	13	0	36	2	2	1	1	1	7
				sip1	30	-5	30	1	2	1	1	1	6
				AVER	28.4	-0.4	31.6	11	18	10	9	9	57
				AGE									
F2	94%	89%	2.9	see1	29	7	26	2	1	1	1	1	6
				seed1	25	10.5	13	2	1	1	1	1	6
				sell1	21	4	19	2	1	1	1	1	6
				sew1	11	3	21	2	1	1	1	1	6
				side1	24	1.5	24	1	1	2	1	1	6
				sigh1	20	6	25	2	2	2	1	1	8
				sin1	18	9.5	11	1	2	1	1	1	6
				sink1	28	6.5	26	2	2	2	1	1	8
				sip1	20	5	27	2	1	1	1	1	6
				AVER	21.8	5.9	21.3	16	12	12	9	9	58
				AGE									
F4	64%	38%	1.9	see1	27	-11	41	1	2	3	2	2	10
				seed1	31	-17	43	1	2	2	1	2	8
				sell1	35	-13	49	2	2	3	2	2	11
				sew1	30	-12.5	45	2	1	3	1	2	9
				side1	35	-12	45	1	1	3	1	2	8
				sigh1	37	-18	44	2	1	1	1	3	8
				sin1	26	-9	41	2	1	2	1	3	9
				sink1	36	-6	43	2	1	2	1	3	9
				sip1	38	-8	43	2	1	2	1	2	8
				AVER	32.8	-11.8	43.8	15	12	21	11	21	80
				AGE									
NF	100%	100%	3.4	see1	27	18	24	1	1	1	1	1	5
				seed1	35	14	30	1	1	1	1	1	5
				sell1	39	12.5	33	1	1	1	1	1	5
				sew1	36	15.5	31	1	1	1	1	1	5

				side1	37	10.5	35	1	1	1	1	1	5
				sigh1	36	17	31	2	1	1	1	1	6
				sin1	33	14.5	25	1	1	1	1	1	5
				sink1	38	19	31	1	1	1	1	1	5
				sip1	33	21.5	20	1	1	1	1	1	5
				AVER	34.9	15.8	28.9	10	9	9	9	9	46
				AGE									
NM	100%	100%	4	see1	37	13.5	21	1	1	1	1	1	5
				seed1	36	13	27	1	1	1	1	1	5
				sell1	36	17.5	23	1	1	1	1	1	5
				sew1	35	3	29	1	1	2	1	2	7
				side1	36	16	22	1	1	1	2	1	6
				sigh1	36	13	28	1	1	2	1	1	6
				sin1	37	11.5	28	1	1	2	1	1	6
				sink1	30	10.5	25	1	1	1	1	1	5
				sip1	35	20.5	18	1	1	1	1	1	5
				AVER	35.3	13.2	24.6	9	9	12	10	10	50
				AGE									