Speech Input for Dysarthric Computer Users

by

Hwa-Ping Chang

B.S., National Cheng-Kung University, 1985

M.S., University of Wisconsin at Madison, 1990

Submitted to the Department of Mechanical Engineering in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Mechanical Engineering
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 1995

Author .... Hwa-Ping Chang ....
Department of Mechanical Engineering
May 23, 1995

Certified by .... Michael J R ....
Dr. Michael J. Rosen
Thesis Supervisor

Certified by .... Kenneth N Stevens ....
Professor Kenneth N. Stevens
Thesis Supervisor

Accepted by ....
Ain A. Sonin
Chairman, Department Committee on Graduate Students

Speech Input for Dysarthric Computer Users
by
Hwa-Ping Chang

Submitted to the Department of Mechanical Engineering in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

One practical aim of this research is to determine how best to use speech recognition techniques for augmenting the communication abilities of dysarthric computer users. As a first step toward this goal, we have performed the following kinds of analyses and tests on words produced by eight dysarthric speakers: a closed-set intelligibility test, phonetic transcription, acoustic analysis of selected utterances, and an evaluation of the recognition of words by a commercial speech recognizer. We have examined in detail the data from eight speakers. The analysis and testing have led to several conclusions concerning the control of the articulators for these speakers: production of obstruent consonants was a particular problem (only 56% of syllable-initial obstruents were produced with no error, average from eight speakers), whereas sonorant consonants were less of a problem (72% correct). Of the obstruent errors, most were place errors for alveolar consonants (particularly fricatives) but voicing errors were also high. These obstruent consonants were often produced inconsistently, as inferred from acoustic analysis and from low scores from the recognizer for words with these consonants (only 49% recognition accuracy for the words with syllable-initial obstruents compared with 70% for words without any obstruent consonant). In comparison, vowel errors were less prevalent. The practical use of a speech recognizer for augmenting the communication abilities of the speakers is discussed. Based on the experimental data, word lists that lead to the improvement of recognition accuracy for each speaker have been designed.

Thesis Supervisor: Dr. Michael J. Rosen
Dr. Kenneth N. Stevens

Title: Research Scientist of Mechanical Engineering
Professor of Electrical Engineering and Computer Science

Acknowledgments

I would first like to express my gratitude to Dr. Cheryl Goodenough-Trepagnier who helped me to get started on this project and Professor Michael Rosen who also inspired me to undertake this project. They both gave me invaluable advice during the early and middle stages of my research. I would also like to thank Professor Kenneth Stevens for his help which motivated me to overcome obstacles and complete this project. Because of Kenneth's diligent guidance and encouragement, I was able to make it through the most difficult period in the study of my degree. In addition, I would like to express my gratitude to Professor Woodie Flowers, Dr. Stefanie Shattuck-Hufnagel, Dr. Sharon Manuel, Dr. Corine Bickley, Dr. Joe Perkell, and Ms. Julie Atwood for their stimulating questions, useful suggestions, and special attention to my project.

For funding this project, I am indebted to the partial support from the National Institute on Disability and Rehabilitation Research grant (No. H231A00025) to Tufts University and from the National Institutes of Health grant (No. DC00075) to M.I.T. I would also like to express my gratitude to the subjects who participated in this project. Their involvement was essential to the success of my research.

I also want to thank all of my friends at the Tufts University School of Medicine, Kathy's kindness in lending me the equipment and Harry's enthusiastic advise for the programs used in this project. In addition, I would also appreciate the officemates in the M.I.T. Speech Group. Especially, I would like to thank Anna, Jeff, and Maryline's invaluable assistance in the experiment of the project, Mark's enthusiastic help in analyzing my data, Peter and Jane's useful suggestions, and Wendy's effort in correcting my English Writing at the M.I.T. English Writing Center.

Last, but not least, I am appreciative to my family for their moral and financial support throughout my entire Ph.D. career and their encouragement which helped me survive the most challenging period and finish this project. In addition, I would like to express my appreciation to my friends for their help and support, especially, Wen-Hui Lee, for her comfort and support to finish this project. All of these people were essential in helping me attain my academic goals.

Dedication:

To My Dear Dad and Mom

謹獻給

親愛的：

老爸和老媽

一份最特別的生日礼物.

Speech Input for Dysarthric Computer Users

Figures

Tables:

Appendix

Speech Input for Dysarthric Computer Users

Hwa-Ping Chang

Chapter 1 Introduction

1.1 Dysarthric Population and Characterisitics

The efficient execution of oral communication requires the smooth sequencing and coordination of the following three basic processes (Darley et al., 1975): "(i) the organization of concepts and their symbolic formulation and expression, (ii) the externalization of thought in speech through the concurrent motor functions of respiration, phonation, resonance, articulation, and prosody, and (iii) the programming of these motor skills in the volitional production of individual speech sounds and their combination into sequences to form words. Impairment of any one of these three processes results in a distinctive communication disorder" [p. 1].

Of the 1.5 million children and adults in the U.S. who are speech-impaired (ASHA, 1981 and Eastern and Halpern, 1981), approximately 900,000 are individuals with congenital impairments: over 500,000 have acquired disabilities resulting from severe illness or fever, head trauma or stroke, and 140,000 individuals have progressive disorders of the central nervous system.

The population of individuals with dysarthric speech includes individuals with a variety of acquired neuromotor impairments which can be mild or severe, with or without cognitive deficit. Yorkston (1988) had defined dysarthria as a group of speech disorders resulting from disturbances in the muscular control of the speech mechanism due to neuromuscular disease. Dysarthria results in a weakness, slowness or incoordination of speech and may involve several or all of the basic processes of speech: respiration, phonation, resonance, articulation, and prosody. Dysarthrias can create significant communication problems for the patients who also may have movement limitation.

From clinical studies, cerebral palsy is estimated as having an incidence of at least 900,000 in the U.S. or 4.5 cases per 1000 in the general population. A significant percentage of these people have dysarthria. The average occurrence of verbal communication disorders among cerebral palsied individuals has been estimated at 70-80 percent (Easton and Halpern, 1981). Amyotrophic lateral sclerosis, with a prevalence of approximately 5 cases per 100,000 (Corcoran, 1981), is an adult-onset disease. Many of these people ultimately lose the ability to speak clearly and eventually find writing and typewriter use difficult or impossible. Multiple sclerosis, which occurs in as many as 500,000 people in the U.S., is extremely variable in its symptoms (Kraft, 1981), which often include movement disorders and speech and keyboard-use impairments. Parkinsonism occurs in about 187 cases per 100,000 (Corcoran, 1981) and impairs the ability to use the computer and often impairs intelligibility of speech. As scientific technology keeps progressing, more and more infants with congenital disorders or prenatal impairments survive and people can live longer than ever before. Therefore, the numbers mentioned above keep increasing. Thus, the study of the pathology and the rehabilitation for persons with disabilities has become more important than ever.

1.2 Computer Interfaces for People with Disabilities

The methods for helping persons with disabilities who use commercial personal computers differ according to the source of the motor function impairments. For example, verbal and nonverbal communication in cerebral palsy may not be necessarily equally impaired. When verbal skills are impaired but nonverbal skills are normal, manual or visual language systems may be used and vice versa. Verbal communication requires the capacity for symbolic language plus the complex articulatory skills necessary to pronounce sounds and words with clarity. Nonverbal communication similarly requires the capacity for symbolic language and an alternate means of expression (Easton and Halpern, 1981). Figure 1.1 shows the traditional and commercial input channels, e.g., keyboard, joystick, mouse, and so on, for personal computers. By using these input channels, users can control the output interfaces, e.g., printer, speaker, robot, and so on.

As a consequence of improvements in technology, a new computer input channel, speech recognition, is available for computer users to operate the computer by means of speech. This is an interface which is easier to use than other traditional input interfaces. It is especially useful for people with disabilities since most of them have varying degrees of motor control disorders. Therefore, integration of the available computer interfaces can help persons with disabilities to overcome their difficulties in using computers and communicating with others. Goodenough-Trepagnier and Rosen (1991) have designed an experimental project to identify optimal interface designs for individuals with neuromotor disabilities who are performing specific computer-aided tasks. Since speech is the most natural and familiar communication means for human beings, the speech input and output channels will be specifically focused on in this project. By using speech technology, e.g., speech recognition and speech synthesis, people with disabilities can improve their abilities to communicate with computers and other people.

Several attempts which have been made to assess the effectiveness of speech recognizers for individuals with disabilities have been investigated. Rodman (1985) showed the results of two different training procedures and vocabulary lists for three cerebral palsy adults using speech recognizers. He identified a number of individual and environmental factors that need attention when individuals use speaker-dependent speech recognizers.

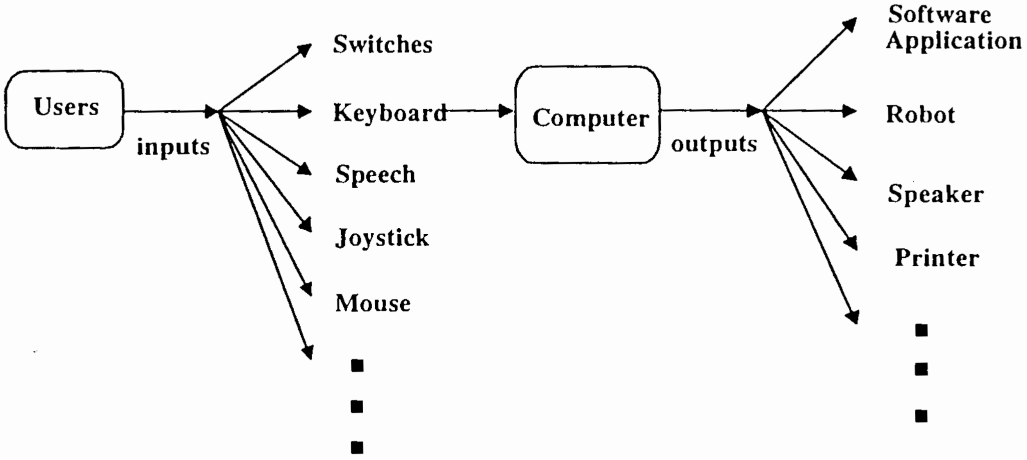


Figure 1.1 Block diagram of modes of computer interfacing. The users can operate the computer by using commercial input channels, e.g., keyboard, speech, joystick, mouse, etc. By operating the computer, the users can control the output interfaces, e.g., printer, speaker, robot, etc. This project will especially focus on the possibility of operating the computer by using the speech input channel.

Unfortunately, the percentage of people with motor impairments and dysarthria is quite high. Lee et al. (1987) have studied the possibility of using a speech recognizer for a dysarthric speaker with cerebral palsy. They used a Keytronics Model 5152 V Keyboard with a built-in Interstate Voice Products' Speech Recognition System. A 28-word list was chosen based on the usefulness in controlling a computer but not considering factors such as the phonetic balance of the word list. A set of training sessions and tests was performed. They found that the overall recognition accuracy still could not be improved beyond 70%, even if they retrained the words whose recognition accuracy was lower than 75%. This conclusion showed that the speakers with speech impairments have difficulty using the speech recognizer as an efficient computer input channel.

However, many persons with speech disorders, even severe dysarthrias, still strive to use their voices to communicate. Thus, a new method for improving the accuracy of speech recognition for persons with speech disorders is an important practical goal. Ferrier (1994) said:

> We found that in some cases where the person was too unintelligible for humans to constantly understand, the machine was able to pick up on the speech pattern and make the appropriate word choice. However, the device had more difficulty when a patient had many voice problems or consistently made other nonspeech sounds like glottal catches and saliva control noises [p.13].

The possibility that speech recognition systems may recognize and interpret speech as well as, or better than, an experienced listener opens up the possibility of improved communication for deaf, dysarthric, and other difficult-to-understand speakers. In other words, if the specific speech characteristics of the dysarthrias can be determined, there is the possibility that the speech recognition accuracy can be improved to a level which makes the speech recognizer useful for the dysarthric computer users.

A speech synthesizer can be used as the output channel for a speech recognizer. Then, the computer can pick up the input utterances based on the speech pattern and generate a speech output from the synthesizer to let listeners understand what the specific user with dysarthria is trying to say. This recognition and synthesis can potentially be realized even when the user's speech is very severely impaired. This integration can be used to improve communication for speakers with dysarthria.

1.3 Goals of this Project

This project arose from preliminary work carried out in the Department of Rehabilitation Medicine at Tufts University School of Medicine (Goodenough-Trepagnier and Rosen, 1991 and Goodenough-Trepagnier, Rosen, Hochheiser, and Chang, 1992). These investigators took the approach that the speech recognizer may be exploited as a control mode which can be integrated with other control strategies available to individuals with dysarthria. They used the Beukelman-Yorkston dysarthria test to assess the speech intelligibility of the participants with dysarthria. Participants trained a DragonWriter-1000 by producing ten tokens (instances) of each of 171 words (24 words selected for their phonological properties, 26 words of the military alphabet, 10 digits, and 111 computer operation words). Tools for analysis of the results were developed using the system's software development library. The results of this investigation suggested that instrumented study of the acoustic characteristics of the participants' speech would be useful for designing the speech recognition interface.

The primary goal of this thesis is to derive a method, generalizable to any speaker and any commercial recognizer, for optimal design of a word menu to help a dysarthric speaker use a speech recognizer efficiently as a computer input channel. In order to achieve this goal, objective study of dysarthric speech was undertaken. The complementary purpose of this project has been to improve our understanding of dysarthric speech, in particular those features which limit access to speech recognition technology, and to show how perceptual and instrumented objective methods can be combined with use of commercial recognition technology to achieve this purpose. Therefore, not only the accuracy of speech production but also speech consistency performance are studied. A computer-based speech recognition assessment (R.T.) and acoustic analysis (A.A.) are used to assess dysarthric speech. By combining the traditional perceptual intelligibility test (I.T.) and a phonetic transcription test (T.T.) with these two instrumented assessment techniques (R.T. and A.A.), a detailed understanding of the speech characterisitics of participants with dysarthria can be attained. The I.T. and T.T. are useful for identification of articulatory position-control impairments. Automatic speech recognition studies (R.T.) can yield information regarding the consistency of dysarthric speech. Acoustic analysis (A.A.) allows for more detailed study. A description of these methods and the rationale for their use will be discussed in Section 1.4.

An outcome of this research has been to facilitate communication for dysarthric individuals by increasing the accessibility of speech technology, i.e., speech recognition and speech synthesis. The goals of this thesis project are mutually complementary as follows:

* **Primary Goal:** Determine how best to use speech recognition techniques for augmenting the communication abilities of dysarthric computer users.

* **Complementary Goal 1:** Identify features of dysarthric speech which affect speech recognition performance.

* **Complementary Goal 2:** Supply different viewpoints that can lead to a better understanding of the speech patterns of the dysarthrias.

These concepts will be used as the guidelines leading to achievement of the final goal, and will be discussed again in Chapter 5.

1.4 Analysis Tools

Although a primary goal of this project is to develop new input computer channels - speech recognizer and synthesis - for dysarthric computer users, it also provides information about the speech characteristics of the dysarthrias. This project seeks a better understanding of the speech patterns of the dysarthrias, based on both subjective and objective observation and measurement.

When faced with dysarthric patients, speech pathologists have the challenging task of assessing complex patterns and types of neuromotor disturbances forming the bases for the patient's speech characteristics. The movement disorder is usually complex, involving many muscles, rather than the paralysis of a specific muscle. Many dysarthrias are associated with basal ganglia and/or cerebellar pathologies which not only disturb the muscle strength but also the initiation, temporal organization, and accuracy of movement patterns. Many methods are available for making clinical assessments of these abnormalities.

In recent years, there have been advances in the assessment and diagnosis of dysarthria. An assessment of dysarthric speech must identify and quantify the alterations in speech movement patterns that are due to motor control disturbances. Because the patterns of movement control are disturbed, the integrity of the oral mechanism, vocal tract structure, and muscular components are not predictive of the patient's speech production impairment. There are two types of assessments of speech production impairments that take into account the speech produced by the impaired vocal tract: (i) perceptual and (ii) acoustic (Ludlow and Bassich, 1982). In this project, computer-based speech recognition will be included and used as one of the new types of assessment of impaired speech production.

1.4.1 Perceptual Test:

A perceptually-based assessment involves judgments of the intelligibility of the speech, descriptions such as transcriptions, or judgments using terms like breathy, nasal, etc. In this project, both an intelligibility test and a transcription test are used to quantify dysarthric speech performance. Intelligibility measures have been used to assess dysarthric speech performance for a variety of reasons. Yorkston and Beukelman (1978) summarized these reasons:

> First, reduced intelligibility is a common characteristic of dysarthria and thus intelligibility measures are applicable across a wide variety of types and severity levels of dysarthria. Second, intelligibility provides an overall index of the disorder which takes into account many different neuromuscular factors along with whatever compensatory strategies the dysarthric speaker may have adopted. Third, the quantitative nature of intelligibility measures allows for monitoring of speaker performance during the course of treatment and recovery. Finally, intelligibility measures give an indication of functional communicative performance of dysarthric speakers which can be easily communicated to the speaker, his family, and members of the rehabilitation team [p. 499].

Darley et al. (1969a; 1969b) have contributed a perceptually-based system for a differential diagnosis classifying various types of dysarthria. Yorkston and Beukelman (1981a; 1981b) have provided clinicians with a reliable perceptually-based assessment of dysarthric speech intelligibility.

However, perceptually-based assessments have limited analytic power for determining which aspects of the speech motor patterns are affected. Perceptual judgments are difficult to standardize over time and across different settings. Perceptual rating systems are often subjective and difficult to replicate. Further, perceptual assessment can not easily be used to evaluate the speech consistency of particular individuals. Flanagan (1972) pointed out that an intelligibility score is not an absolute quantity but rather is "a function of such parameters as test material, personnel, training, and test procedures" [p. 311].

1.4.2 Recognition Test:

This test provides an objective and quantitative method to assess the ability of a speaker to produce utterances in a consistent manner. In the perceptual tests, the accuracy of pronunciation and the pattern of errors are explored, but not the consistency with which a particular word is pronounced. To recognize a speaker's words, a speaker-dependent speech recognition machine analyzes samples of that person's speech. These samples capture the normal variations of that speaker. Large variations can cause the recognizer to give an increased rejection rate or misrecognition of an utterance. A speech recognition test, therefore, supplies information about a subject's speech consistency. There is both anecdotal and research evidence to indicate that some consistency exists in the speech production of speakers with dysarthria (Neilson and O'Dwyer, 1984). Coleman and Meyers (1991) mentioned, "If speakers with dysarthria are consistent and if there is enough variation in their speech so that they are capable of articulating a number of sounds that can be differentiated, then recent technology in speech recognition may be useful to them". [p. 35]. The speech recognizer test measures the variability in a particularly relevant way, i.e., by determining how often an utterance falls within the "boundary" defined by a language model derived from five earlier utterances of the same word by the same speaker. However, although the speech recognition test can give an objective method to understand the dysarthric speech, it can not detect the articulatory position accuracy and the emotional expression, e.g., tired and excited voice. In addition, this test can not describe the dysarthric speech in terms of a set of acoustic parameters, e.g., formant frequency, fundamental frequency, etc.

1.4.3 Acoustic Analysis:

Some of the problems mentioned above can be avoided by acoustic analysis. An acoustic assessment involves quantitative analysis of the speech signal. Analyses of speech spectrograms or other types of acoustic analysis help to standardize the speech assessment process. However, there has been some concern that objective acoustic measures may not assess those aspects of speech production that are important for the patient to communicate. That is, the intelligibility of the patient's speech is subjective, and an acoustic analysis may not arrive at an appropriate measure of speech intelligibility, at least based on current methods of analysis. Sataloff (1992) noted that acoustic analysis can evaluate a variety of attributes: "the formant structure and strength of the voice, the fundamental frequency, breathiness, the harmonics-to-noise ratio (or clarity of the voice) and perturbations of the cycle-to-cycle amplitude ("shimmer") and of the cycle-to-cycle frequency ("jitter")" [p. 114]. He also mentioned, "Subtle characteristics, however, still cannot be detected. For instance, in studies of voice fatigue in trained singers, the difference between a rested and tired voice is usually obvious to the ear, but significant changes cannot be detected consistently even with sophisticated equipment" [p.114]. However, clinicians view instrumentation as the sole realm of the speech scientist, expounding it to be too time-consuming and complicated. In fact, acoustic analysis takes less time than scoring and analyzing many speech tests for the purpose of assessment (Simmons, 1982). Acoustic analysis can be one of the least expensive and readily available methods for speech assessment. It can graphically represent multiple acoustic parameters which can be used in extracting the dependent variables of speech intelligibility. Acoustic analysis can provide data that are very difficult to obtain by perceptual analysis. It also serves as an excellent guide to improving the clinician's understanding of what the patient is doing to produce the impaired speech (Simmons, 1982). Kent et al. (1979) demonstrated that sonographic recordings can provide an objective measure of the acoustic properties of ataxic speech. That study suggested that acoustic analysis might be a reliable means of documenting speech changes due to treatment. Kent et al. (1989) have designed an intelligibility test which correlates dysarthric intelligibility impairment and acoustic properties. This test is designed to examine 19 acoustic-phonetic contrasts that are likely to (a) be sensitive to dysarthric impairment and (b) contribute significantly to speech intelligibility.

Despite the development of a variety of techniques to estimate or objectively quantify dysarthric speech, the measurement methods mentioned above have their own advantages and disadvantages. The comparability of these methods has yet to be evaluated and the methods have not yet been combined together to make an assessment of the dysarthric speech. The purpose of this dissertation is first to explore the relationship that exists between these assessment tests obtained from human listeners and from machines. Then, the assessment findings are applied to improve the speech recognition accuracy. The intelligibility and transcription tests are used to test the articulatory accuracy and pattern of phonetic errors. The recognition test is used to explore the speech consistency issue for each subject. Then, the acoustic analysis supplies a quantitative analysis in the form of various acoustic attributes.

1.5 Subject Information and Dissertation Outline

This dissertation presents an assessment method that combines four approaches (including intelligibility tests, transcriptions, recognition tests, and acoustic analyses) for eight speakers with speech impairment to obtain their specific speech characterisitics. Eight speakers with dysarthria, JF, MW, JS, DW, ES, JR, CH, and GG, and one normal speaker (MJ) as a control served as subjects for this project. Table 1.1 shows the subjects' information: sex, age, type of disability, education, and usual typing method. Four males and four females participated in the project. The ages range from 22 to 62 years. There are six subjects with cerebral palsy (spastics: 4, athetosis: 1, and spastics plus athetosis: 1), one with spastic cerebral palsy plus surgery damage, and one with cerebellar ataxia. The education level ranged from 5th grade of the primary school to master's degree. The original typing methods for the subjects included: one finger, fingers, one thumb, pencil, or nose. One of the subjects, GG, lost contact since she moved to another place. However, her data are still helpful as a reference.

A detailed discussion of the neurology and pathology for the speech disorders and the integration of speech recognition and speech synthesis is introduced in Chapter 2. The main functions for each component of the articulatory system are presented in Chapter 2, e.g., upper motor neuron system, lower motor neuron system, nervous system, and the stages of human speech processing. A review of the etiology of speakers with dysarthria and the classification of the dysarthrias is presented later in the same chapter. Three types of dysarthrias (cerebral palsy - spastic, cerebral palsy - athetosis, and cerebral ataxia) are discussed. At the end of chapter, the configuration of combining speech recognition and speech synthesis will be discussed. In particular, speech recognition is the main tool served in this project. Its main functions and operating algorithm are introduced in brief.

In Chapter 3, the flow chart and detailed steps that are followed in this project are described. The theoretical basis for interpretation of each test is explained and compared with the relevant literature. In addition, the test formats, figures, and tables, are shown for one example, JF. The word pairs or lists for all of the three tests - intelligibility test, transcription test, and recognition test - are presented in appendices.

All of the detailed information and discussion of test data for each subject are described in Chapter 4. Interpretation of results from the intelligibility test is presented in the form of figures. The transcription test results are shown in the form of tables. The recognition test and the intelligibility test accuracy are presented in the form of accuracy tables. Acoustic analysis for a limited set of utterances has also been applied to study the actual production of individual utterances and to attempt to explain any discrepancies between the three tests mentioned above. The speech characteristics for each subject are summarized at the end of the corresponding subject section. Based on the information about the speech characterisitics for each subject, a special draft word list is designed for each dysarthric speaker. The recognition accuracy percentage improvement between the 70-word list and the draft word list is shown. An application test, which compares the speed of the traditional typing methods and speech dictation, was given to each subject. These data will demonstrate the possibility of using the speech recognition system as the input channel for operating the computer. The final list for each subject is presented in the appendix. Some of the simple rules used to design the draft word list across all of these eight subjects are summarized at the end of Chapter 4.

Finally, the overall discussion for all of these subjects is presented in Chapter 5. In this chapter, the findings emerging from the project will be discussed for each of the components which are mentioned in Section 1.3 and the results will be compared with past work reported in the literature. The relationship between the human listening test and the speech recognition test is noted and used to interpret the data. The influence of utterance duration and articulatory deficits on the speech recognition accuracy is discussed separately. Future investigations and alternatives of this project along with the main contributions are also described at the end of the chapter.

Table 1.1 Subject information and status. Eight speakers with dysarthria, JF, MW, JS, DW, ES, JR, CH, and GG, and one normal speaker (MJ) as a control served as subjects for this project. The columns show the subjects' information: sex, age, type of disability, education, and usual typing method.

Subj. \ Status	**M / F**	**Age**	**Type**	**Education**	**Typing**
JF	M	61	CP	High	1 Finger
MW	M	38	Ataxia	B.S.	Fingers
JS	M	48	CP	Under.	Nose
DW	M	45	CP	M.S.	1 Finger
ES	F	61	CP	B.S.	1 Finger
JR	F	22	CP	Under.	Thumb
CH	F	62	CP+OP	5th Grade	1 Finger
(GG)	F	24	CP	Under.	Pencil
MJ	M	27	Normal	Ph.D.	Fingers

Chapter 2 Background

2.1 Introduction

The development of normal speech and language skills from infancy depends upon the following: (i) an intact hearing mechanism, (ii) an intact brain, (iii) an environment of interactions between peers and adults, and (iv) sufficient motor skills to communicate (Easton and Halpern, 1981).

Language has evolved as a communication tool for interfacing human beings' thinking. The procedures of producing a sentence include: (i) generating the linguistic rules (i.e., organizing the ideas into sentences; converting the sentences into grammatical structure; and finally choosing the right words that express the exact meaning) and (ii) making practical acts of speech production for the right words chosen in step (i). Netsell (1986) said, "Speech is the motor-acoustic expression of language and speech motor control is defined as the motor-afferent mechanisms that direct and regulate speech movements. Mechanisms refer to the muscle and neural anatomy and physiology that have evolved for the purpose of producing speech movements" [p. 33].

The production of speech or song, or even just a vocal sound, entails a complex orchestration of mental and physical actions. Figure 2.1 shows the procedures for the production of speech. The "idea" for a sound originates in the cerebral cortex, e.g., the speech area in the temporal cerebral cortex and travels to the motor nuclei in the brain stem and spinal cord. The speech motor area in the primary motor cortex (in the precentral gyrus) sends out complicated messages for coordinating the activities of the larynx, the thoracic and abdominal musculature, and the vocal-tract articulators. For instance, the movement of the larynx is controlled from the speech motor area in the primary motor cortex and is transmitted to the larynx by various nerves, e.g., the superior laryngeal nerve and the recurrent laryngeal nerve. As a result, if the aerodynamic conditions are right, the vocal folds will vibrate and generate a buzzing sound. That buzzing sound is filtered throughout the area of the vocal tract above the glottis, including the pharynx, tongue, palatal, oral cavity and nose, which produces the sound perceived by listeners (Sataloff, 1992). Different vocal tract shapes result in different sounds even if the sound source from larynx does not change.

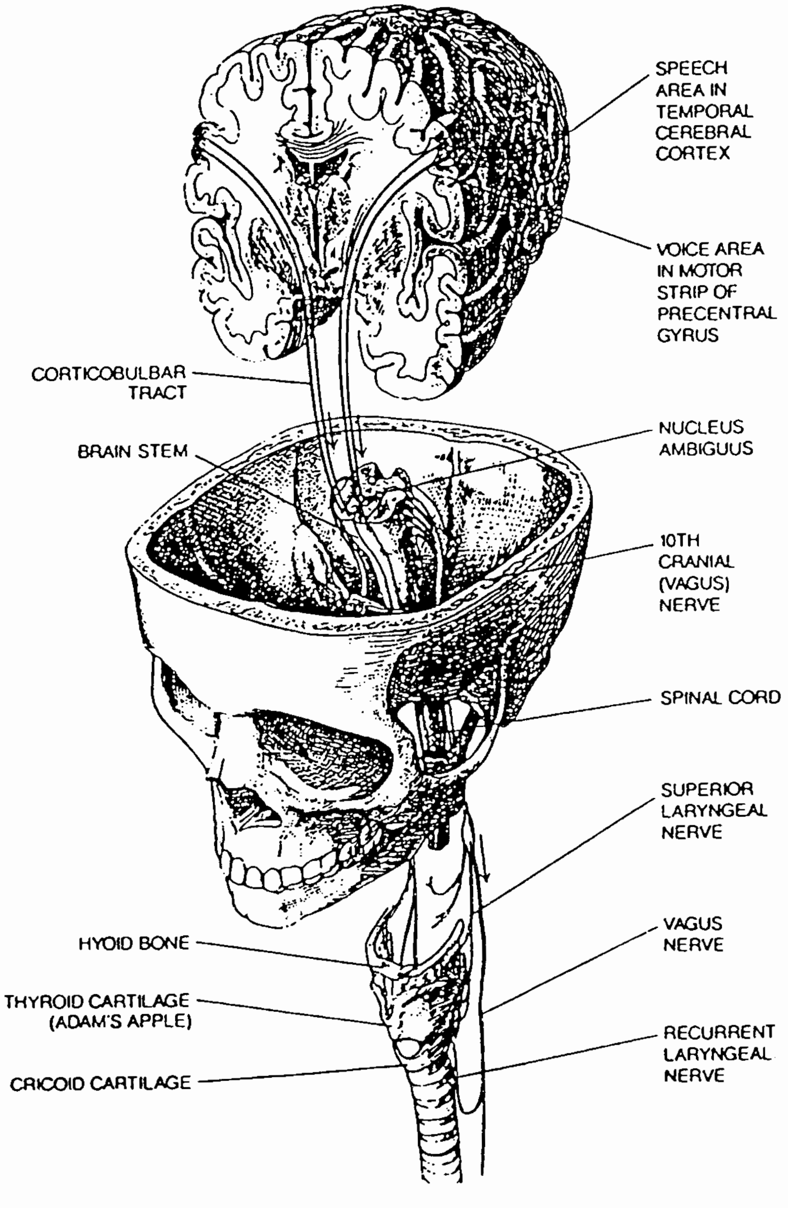

SPEECH AREA IN TEMPORAL CEREBRAL CORTEX
VOICE AREA IN MOTOR STRIP OF PRECENTRAL GYRUS
CORTICOBULBAR TRACT
NUCLEUS AMBIGUUS
BRAIN STEM
10TH CRANIAL (VAGUS) NERVE
SPINAL CORD
SUPERIOR LARYNGEAL NERVE
HYOID BONE
VAGUS NERVE
THYROID CARTILAGE (ADAM'S APPLE)
RECURRENT LARYNGEAL NERVE
CRICOID CARTILAGE

Figure 2.1 The anatomical structure of the speech production, from Sataloff (1992). The idea for making a sound originates in the cerebral cortex of the brain and is transmitted to the larynx and the relative parts of vocal tract by various nerves. As a result, the vocal folds vibrate, generating a buzzing sound and resonating in the vocal tract.

Moreover, nerves also provide feedback to the brain about speech production. Auditory feedback, which is transmitted from the ear through the brain stem to the cerebral cortex, allows the speakers to compare the sound spoken with the sound intended. Thus, auditory feedback and tactile feedback from throat and muscles enable the speakers to achieve fine-tuning of the vocal output (Sataloff, 1992)

When particular neural pathways or regions are damaged, the motor control of speech becomes abnormal. Some of these abnormality are too severe to perform intelligible speech (Netsell, 1986). For example, the nerves that control the muscles of the vocal tract are potential sources of voice problems. The two recurrent laryngeal nerves which control most of the intrinsic muscles in the larynx, shown in Figure 2.1, are easily injured by trauma or surgery of the neck and chest since they run through the neck, down into the chest, and then back up to the larynx (Sataloff, 1992). Furthermore, if the impaired region is in the area of language processing, e.g., Wernicke's area, it will possibly cause communication disorders. Communication disorders of the speech processes include aphasia (a kind of language disorder with damage of the cerebrum which has the primary function of language processing) and apraxia (a kind of articulatory disorder with damage of certain brain areas which are devoted to the programming of articulatory movement). Speech disorders of motor control are referred as dysarthrias.

Speech is one of the most essential tools for human communication. Due to advances in speech technology, a speech input signal can be used as a new interface to communicate with the computers. The tool applied for this purpose is called as automatic speech recognition system (ASR). It has recently become as an important tool to input or control computers. In particular, persons with disabilities have difficulties using the keyboard and mouse. Speech recognition technology supports a new kind of input channel to help them control the computer by using their speech.

In the following sections, some background on the neurology of speech and on the control of speech production will be introduced. Then, one particular speech disorder, dysarthria, will be described and classified. Furthermore, three types of dysarthrias - spastic cerebral palsy, athetoid cerebral palsy, and cerebellar ataxia - will be discussed in detail. At the end of this chapter, a basic introduction to the speech recognition algorithm will be described since the speech recognizer is the main equipment used in this research project.

2.2 Functional Anatomy of Speech and Language

Perkins and Kent (1986), Kolb and Whishaw (1990), and Darley et al. (1975) have discussed in depth the main functions of the neurology of speech, and the present discussion is adapted from these reviews (in particular from Perkins and Kent (1986)). Before describing the stages of speech processing, the main functions for each speech-relevant motor control area and the relevant aspects of the nervous system are summarized first. This information is necessary to understand the speech disorders discussed in the next section.

Nervous System:

From an anatomical view, the nervous system is composed of two subsystems: (i) the central nervous system and (ii) peripheral nervous system, as shown in Figure 2.2. The central nervous system can be divided into the brain and spinal cord, which are protected by the skull or vertebral column. The peripheral nervous system includes cranial nerves and spinal nerves, which are not protected by bony housing. The central nervous system is the main part of the nervous system, whereas the peripheral nervous system is peripheral to the central nervous system. These two sub-nervous systems consist largely of transmission lines, like optical fibers, to translate the signal from the sense organs and to the muscles. By using these transmission lines, the signal from the brain can be sent even to the extremities in a short period of time.

Left Hemisphere:

Most right handers are dominated by the left hemisphere for speech and language. Figure 2.3-a shows some of the areas that influence the speech and language processes: Broca's area, Wernicke's area, supplementary motor speech area, supramarginal gyrus, angular gyrus, fissure of Sylvius, precentral gyrus, and postcentral gyrus. The two primary speech centers in the dominant hemisphere are Broca's area and Wernicke's area, which were identified about one hundred years ago. In more recent years, neurologists have found that angular gyrus, supramarginal gyrus, supplementary motor cortex of the dominant hemisphere, the nondominant hemisphere (which is mainly responsible for emotional and intonational components of speech), and subcortical structures (which are important in mapping cognition with language) are also important.

Broca's area, located in the lower frontal lobe of the dominant hemisphere, is associated with planning and executing spoken language. If this area is damaged, the

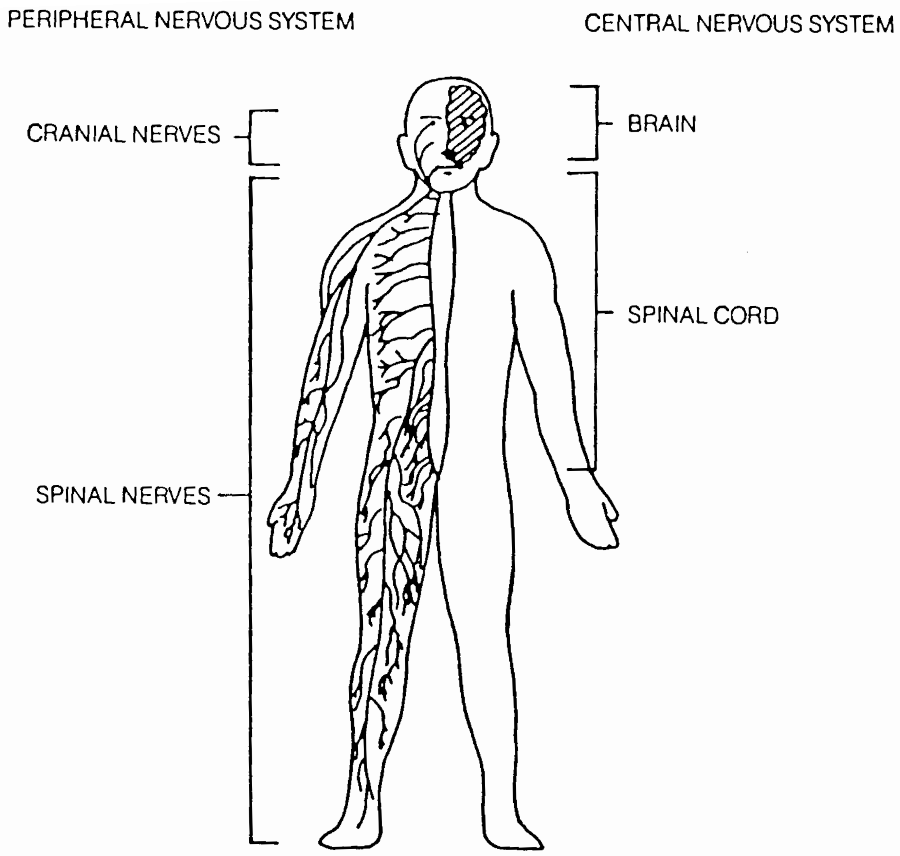


Figure 2.2 The nervous system, from Perkins and Kent (1986). The nervous system includes two subsystems: the central and the peripheral nervous systems. The central nervous system contains the brain and the spinal cord. The peripheral nervous system has cranial nerves and spinal nerves.

articulation becomes labored and difficult, and the ability to generate a grammatical sentence is impaired but the comprehension ability may not be greatly affected. Wernicke's area, located in the posterior left temporal lobe, does not play a role in motor organization of articulation but is responsible for language planning. This area is essential to meaningful speech, reading, and writing. If this area is damaged, comprehension ability will be affected. Speech production could still be phonetically intact and fluent but because of a lack of grammar, it can be meaningless and full of babble. The supplementary motor speech area, located in the upper left frontal lobe, is involved in both speech and language functions. This area plays a role in control of rhythm, phonation, and articulation. It has functions similar to those of the dominant side of thalamus, which is closely connected with the supplementary motor speech area and is involved with language function. The supramarginal gyrus, located behind the postcentral gyrus, is also thought to play a role in planning volitional motor speech. The angular gyrus is the area that connects the spoken language systems with visual language mechanisms. If this area is damaged, the visual cortex will disconnect with Wernicke's area. The area bordering the fissure of Sylvius has been found to be mainly concerned with articulatory processing. The precentral gyrus, immediately anterior to the fissure of Rolando, has a preponderance of pyramidal motor cells, which belong to the upper motor neurons and connect with the lower motor neurons. The lower precentral gyrus area has the main motor function for speech mechanisms, and sends out complex signals for coordinating the articulatory mechanisms. The postcentral gyrus, immediately posterior to the fissure of Rolando, is the primary somatic sensory cortex center. It receives a feedback signal from the lower motor neurons. The lower postcentral gyrus is responsible for the sensory feedback of the articulation system. Overall, certain regions in the left hemisphere are relatively frequently involved in the speech and language process, but none of them are exclusively involved for a specific function.

Right Hemisphere:

There is no distinct area in the right hemisphere for speech production comparable to Broca's or Wernicke's areas, as shown in Figure 2.3-b. Currently, the right hemisphere is being investigated as a mechanism for processing some components of speech. What these components are is still unclear. The main functions of the right hemisphere for speech production are thought to be control of the prosodic expressions of attitude and emotion since these prosodic and emotional expressions are less structured. The right hemisphere is also thought to organize motor mechanisms of emotional expression, but is better at managing negative emotional states; most of the positive emotions still seem to

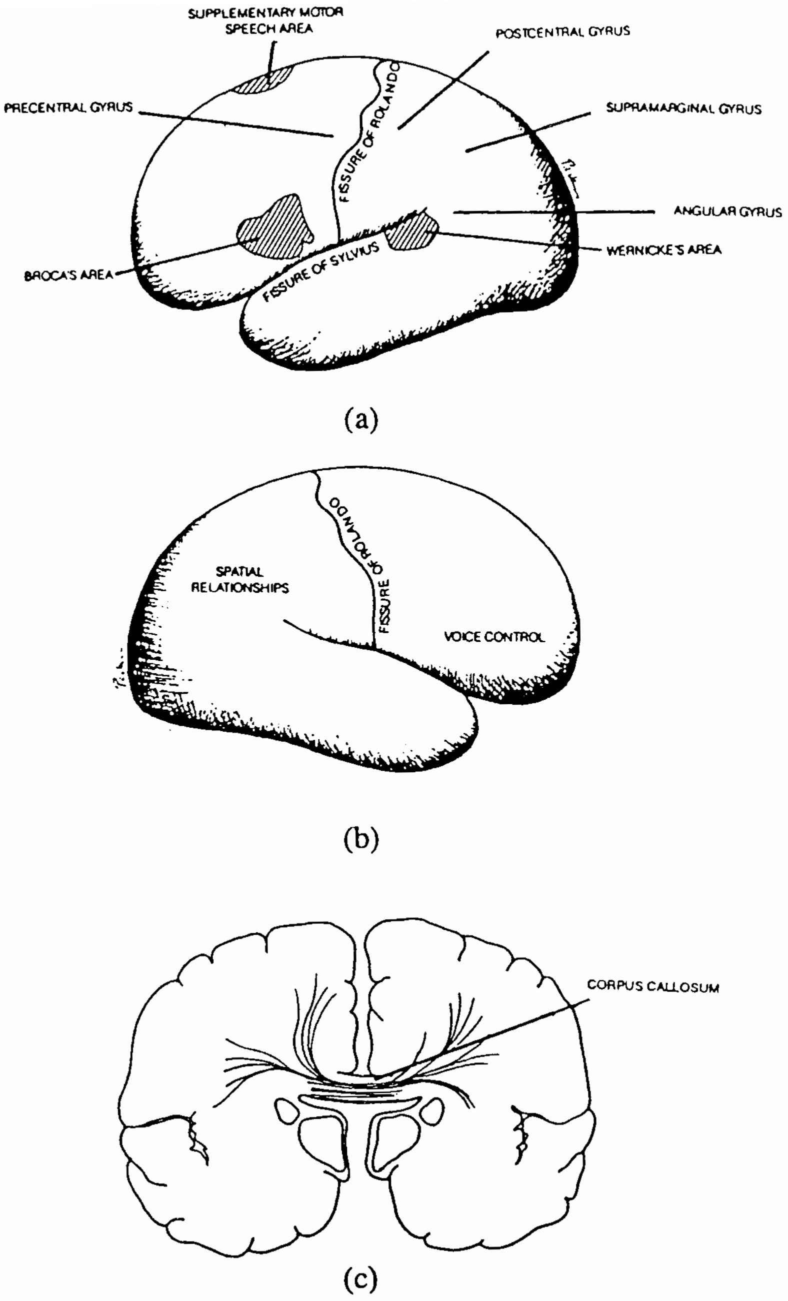


Figure 2.3 Major left and right hemisphere speech centers and the commissural pathways, adapted from Perkins and Kent (1986). (a) Left hemisphere speech center contains Broca's area, Wernicke's area, supplementary motor speech area, angular gyrus, precentral gyrus, postcentral gyrus, fissure of sylvius, and supramarginal gyrus. (b) Right hemisphere has no apparent distinct areas for speech production compared to Broca's and Wernicke's areas, but it has more concern for emotional expression. (c) Along with four small commissural pathways interconnecting the hemispheres is the great central commissure, the corpus callosum.

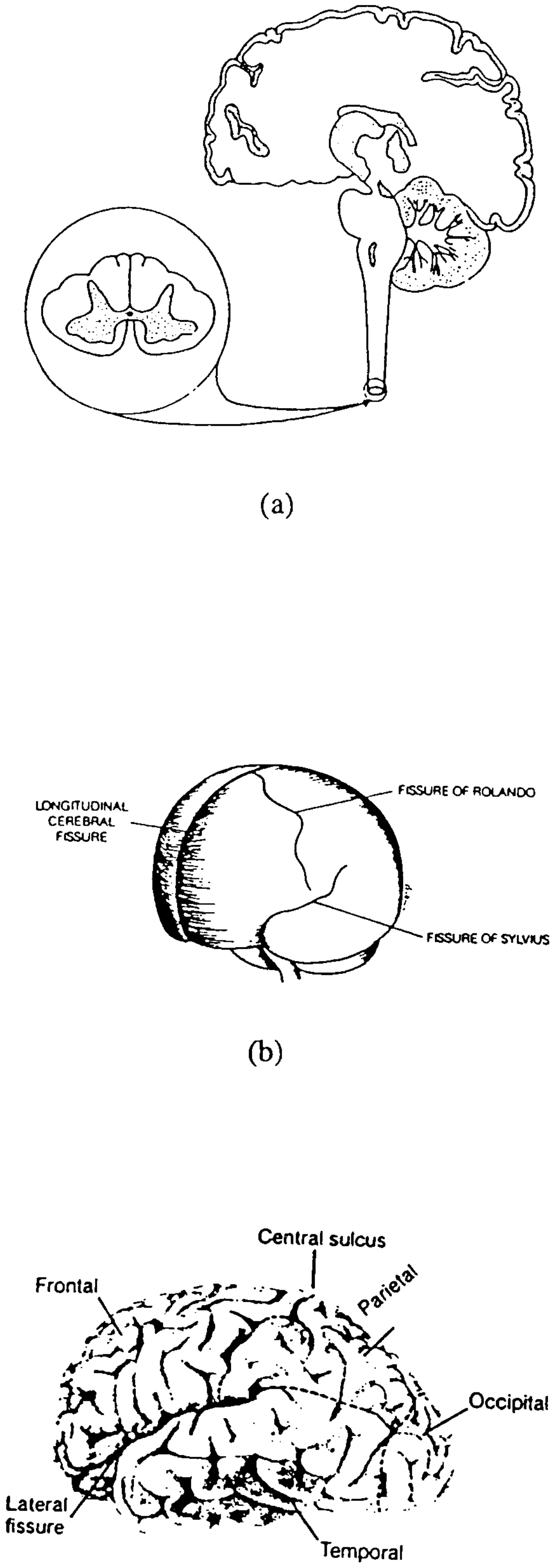
be controlled by the left hemisphere.

The above discussion indicates that the left hemisphere contributes more to highly structured linguistic functions than the right. Therefore, the question would be how these separate components are integrated into a moment of speech. The corpus callosum, which connects the hemispheres along with four small commissural pathways, as shown in Figure 2.3-c, supplies higher level coordination of propositional and prosodic components between the right and left hemispheres. Except for the corpus callosum, the subcortical motor components, mentioned later, are thought to be important components which contribute to the integration of the speech production between two hemispheres.

Cerebral Cortex:

The cerebrum is basically composed of gray and white matter. The organization of gray and white matter in the cerebrum is reversed in the spinal cord and brainstem, as shown in Figure 2.4-a. Gray and wrinkled matter covers the outer part of the cerebral hemispheres. The organization of gray layers is called the cerebral cortex. It is made up almost exclusively of the cell bodies of neurons. The white matter underlies the cortical mantle. The cerebral cortex is responsible for goal-directed behavior. From the structure of the cerebral cortex, a basic neurophysiologic plan is reflected. The cerebral cortex and its various areas are considered neuronal pools since they primarily contain the cell bodies of neurons.

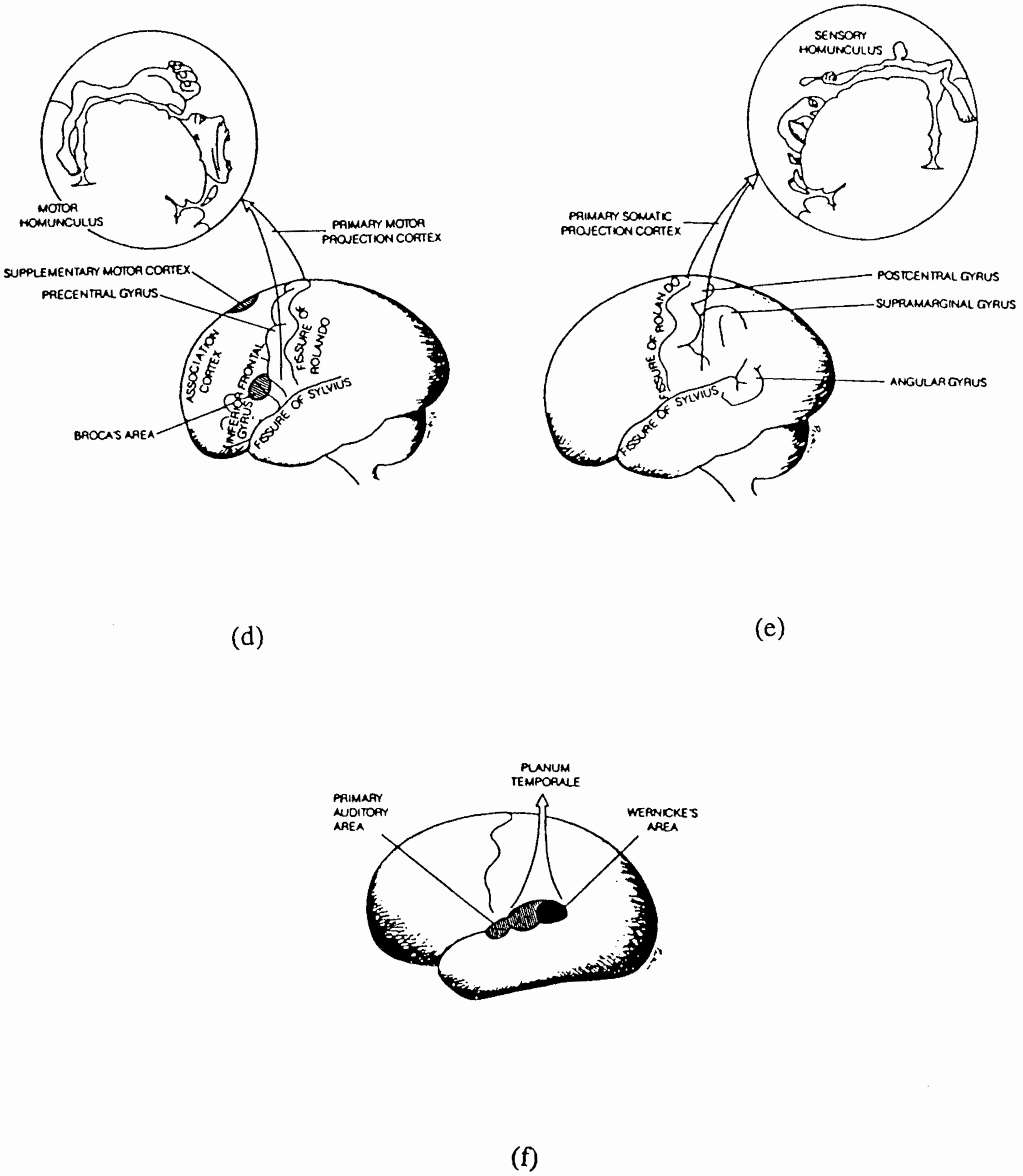
The cerebral cortex is about an eighth of an inch thick and has six layers of different types of cells. The layers vary in thickness and arrangement throughout the cortex. The fundamental structure of the six-layer cortex is similar, but has regional differences. The cortex is composed of the following four nerve cells: ascending axons, descending axons, horizontal axons, and short axons. The cortex has folded in on itself to form ridges (called gyruses), shallow depressions (called sulcuses), and deep depressions (called fissures). More than two thirds of the cortex is buried inside the depressions and is not visible from the exterior surface of the brain. There are three important cerebral fissures, as shown in Figure 2.4-b. The longitudinal cerebral fissure separates the two hemispheres. Each hemisphere is subdivided into four lobes: frontal, parietal, temporal, and occipital, which are named after the bones of the cranium that lie on top of them, as shown in Figure 2.4-c. The major landmarks used to define these lobes are provided by the Sylvian fissure (also called the lateral fissure) and the fissure of Rolando (also called the central sulcus). The fissure of Rolando (central sulcus) divides the frontal lobe, which


LONGITUDINAL CEREBRAL FISSURE
FISSURE OF ROLANDO
FISSURE OF SYLVIUS
Central sulcus
Frontal
Parietal
Occipital
Lateral fissure
Temporal

(a)

(b)

(c)



(d)

(e)

(f)

Figure 2.4 The main functions and anatomy structure of the cerebral cortex, from Perkins and Kent (1986) and Kolb and Whishaw (1990). (a) The arrangement of gray and white matters in the brain and spinal cord. (b) Three major cerebral fissures separating the left and right hemispheres. (c) Four lobes in each hemisphere. (d) The anatomical structure of the frontal lobe and the motor homunculus of the primary motor cortex in the left hemisphere. (e) The anatomical structure of the parietal lobe and the motor homunculus of the primary somatic sensory cortex in the left hemisphere. (f) The anatomical structure of the temporal lobe in the left hemisphere.

mostly serves motor functions, and parietal lobe, which contains the primary somatic sensory cortex. The Sylvian fissure (lateral fissure), the most prominent landmark of the cortex, separates the temporal lobe from the frontal and parietal lobes.

Overall, the motor planning function becomes more abstract, the further away from (anterior) the fissure of Rolando. Reciprocal innervation of flexors and extensors may occur when the motor cortical areas are stimulated. Denny-Brown (1966) studied the functions of different areas of the motor cortex. He found that the frontal (or parietal) cortex is responsible for the production of flexion, e.g., withdrawing and grasping. He also discovered that the parietal cortex, which receives broad somatic sensory input, is concerned with extension movement patterns. The four lobes of the brain are devoted to the following functions in brief: the frontal lobe is devoted to motor control, short term memory, planning, strategy formulation, problem solving, and personality; the parietal lobe to somatic processing, spatial processing, attention, and some language functions; the temporal lobes to language, visual feature recognition, audition, and memory; occipital lobe to processing visual input. Relative to the discussion of speech functions, there are three important lobes: frontal lobe, parietal lobe, and temporal lobe. These are now addressed in more detailed.

(i) Frontal Lobe:

The frontal lobe, the portion anterior to the fissure of Rolando of the cerebrum and above the fissure of Sylvius (as shown in Figure 2.4-c), is devoted to motor control, short term memory, planning, strategy formulation, problem solving, and personality. The primary motor cortex, immediately anterior to the fissure of Rolando in the precentral gyrus, contains a preponderance of pyramidal motor cells that connect to the body. Figure 2.4-d shows the homunculus for the point-by-point representation of muscles of the body in the primary motor cortex. In the cortex, there is a vast region, especially in the frontal lobe, which does not work for any specific function. These areas are associated with the place where the neural signals are integrated for purposeful plans. One of the most famous planning centers in the frontal lobe is Broca's area, located in the inferior frontal gyrus of the left hemisphere (sometimes in the right hemisphere). This area is the place where the plans are organized for motor speech articulation. The supplementary motor cortex, located in the upper left frontal lobe of the dominant hemisphere, assists in motor planning of propositional speech.

(ii) Parietal Lobe:

Anatomically, the parietal lobe, which is the portion posterior to the fissure of Rolando and above the fissure of Sylvius (as shown in Figure 2.4-c), is devoted to somatic processing, spatial processing, attention, and some language functions. The postcentral gyrus, which is immediately posterior to the fissure of Rolando, is the primary somatic sensory cortex and translates the signal from the sense organs. The sensory homunculus, shown in Figure 2.4-e, matches the motor homunculus closely, as shown in Figure 2.4-d. Two other areas of some importance to speech in the parietal lobe are the supramarginal gyrus and angular gyrus.

(iii) Temporal Lobe:

The temporal lobe is separated from each cerebral hemisphere by the fissure of Sylvius (lateral fissure), as shown in Figure 2.4-c. This lobe is devoted to language, visual feature recognition, audition, and memory. The temporal lobe contains the primary auditory cortex and the major language formation center called Wernicke's area, as shown in Figure 2.4-f. Wernicke's area is located on the planum temporale, on the upper surface of the lobe inside the Sylvian fissure, which has been found to be mainly concerned with articulatory processing. Further, the Wernicke's area is near the intersection of the temporal, parietal, and occipital lobes which may also be significant for language. This specific location would seem to work for coordination of auditory, visual, and bodily senses.

Subcortical Mechanisms:

The subcortical centers involved in motor-speech processing are regional networks within larger neural networks that are connected together in loops. The basic components include basal ganglia, cerebellum, thalamus, limbic system, and reticular formation, as shown in Figure 2.5.

(i) Basal Ganglia:

The basal ganglia (located at the top of the brainstem) is the largest of the subcortical motor centers in the extrapyramidal system (which will be discussed later). The main function of the basal ganglia is to generate motor programs. The coordination of basal ganglia and cerebellum may contribute much to speech articulation. Yahr (1976) had discussed many abnormal manifestations caused by the problems of basal ganglia.

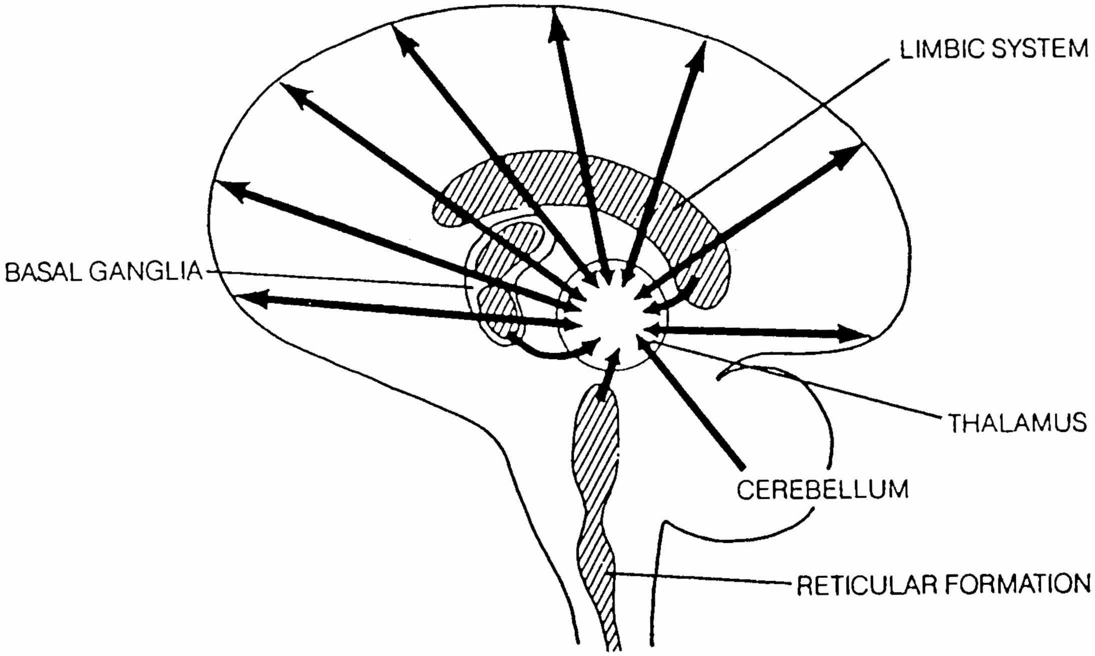


Figure 2.5 Subcortical mechanisms, from Perkins and Kent (1986). The basic components include basal ganglia, cerebellum, thalamus, limbic system, and reticular formation. They are regional networks within larger neural networks that are connected together in loops. The thalamus plays a pivotal role in motor-speech processing and has extensive connections to the right hemisphere, as well as being the relay station between the cortex and limbic system, basal ganglia, and cerebellum. The arrows show the relations between thalamus and other basic components.

(ii) Cerebellum:

The cerebellum is situated in the posterior fossa of the skull, astride the brain stem. It receives input from the spinal cord, the brain stem, and the cerebrum. It distributes its output to these structures to modulate motor activities arising there. The portions of the cerebellum which are most necessary for skilled and voluntary movements (e.g., speech) receive the major afferent signals from motor areas of the cerebral cortex. The same portions of the cerebellum direct fibers back to the motor areas of the cerebral cortex. It is also true that various areas of the cerebellum receive fibers from the spinobulbar, vestibuloreticular, and extrapyramidal levels and direct back to them. Thus, the cerebellum has been described as the "comparators". It is used for detecting a discrepancy between the intent and the feedback information. With extensive connections from the entire cortex, the cerebellum is apprised of plans for voluntary action; with brainstem circuits, it is apprised of plans for reflexive action. With these inputs, and with outputs mainly back to the motor cortex for speech, the cerebellum checks the difference between the response and the command, sends modified impulses back to the motor centers which sends these impulses out to the muscles. Therefore, the tasks of the cerebellum mainly inhibit rather than facilitate excitability of neurons.

Lesions to the posterior midline portion (posterior of the cerebellum vermis, flocculonodular lobe) impair the equilibrium required to sit or stand. Lesions of the anterior portion of the cerebellum result in gait ataxia. The right portion (hemisphere) of the cerebellum receives input from the left sensory and motor cortex. Lesions of the right cerebellum impair the coordination of skilled movements of the right extremities (limb ataxia). The left cerebellum has the same influence to the right sensory and motor cortex and left extremities. The area of the cerebellum midway between the anterior and posterior vermis is the most likely primary locus for the coordination of the speech mechanism. This general region has been named as the "midportion" of the cerebellum. Lesions of the midportion of the cerebellum impair the coordination of the speech mechanism.

In summary, the cerebellum is considered as an organ that does not initiate posture (or movement) but rather modulates movements originating elsewhere. If the cerebellar functions are intact, the cerebellum should not originate movements or movement patterns, but is supplied with information concerning all levels of motor activity.

(iii) Thalamus:

The thalamus is a key structure in the procedure of speech and language processing. It has extensive connections to the right hemisphere and is the relay station between the cortex and limbic system, basal ganglia, and cerebellum. It plays a pivotal role in the integration of the propositional and emotional component of speech and also in arousal, attention and short-term memory. All parts of the cortex are extensively connected to the thalamus. Therefore, cortical and thalamic functions are inseparable. The arrows in Figure 2.5 show the relationships between thalamus and other basic components.

(iv) Limbic System:

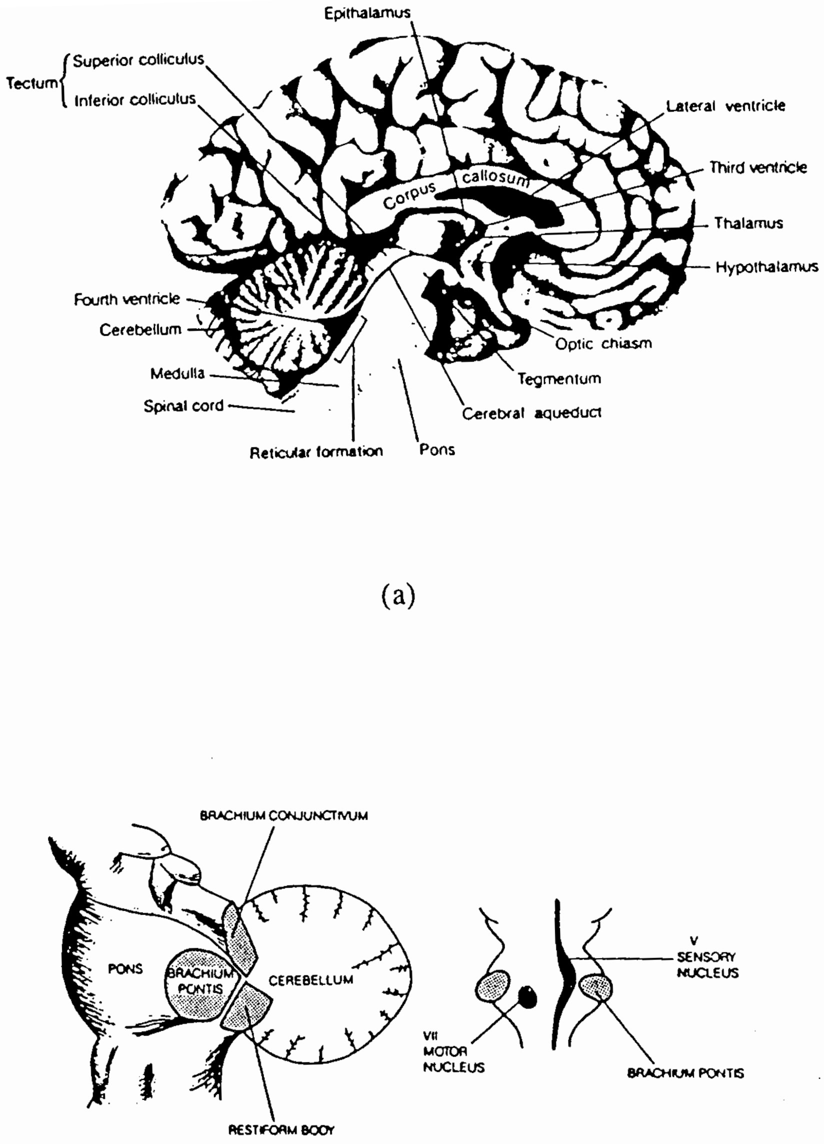
The limbic system encircles more or less the basal ganglia and connects to the hypothalamus. The limbic system intercedes information about internal states, e.g., hunger, fear, pleasure, and sex. It also receives information from the prefrontal cortex, primary sensory areas, and their related association areas; therefore, it can color perceptions emotionally. The limbic system is also connected back to the prefrontal cortex through the thalamus. Thus, the prefrontal cortex executes control over emotional expression originating from the limbic system. Non-human vocalization and human emotional expressions, e.g., laughing and crying, are controlled by the limbic system. In addition, the limbic system and thalamus are formed together to formulate a network for originating what is to be said in the early stage of an utterance.

(v) Reticular Formation:

The reticular formation connects with the pyramidal (mentioned later), extrapyramidal (mentioned later), and sensory systems. The reticular system plays a role in integrating the separate activities of the nervous system into a whole. It monitors the stimuli of our senses, accepting those of importance to us and rejecting the rest. The reticular system, then, seems to be at the neurological pivot of speech, especially integrating of the emotional and propositional components.

Pons:

Pons is the base of the brainstem just above the medulla, as shown in Figure 2.6-a. The most distinguishing segment of the pons is the brachium pontis that joins the two cerebellar hemispheres through the brain stem, as shown in Figure 2.6-b. It connects the motor planning centers of the brain to the cerebellum. The pons also contains nuclei for



(a)

(b)

Figure 2.6 The anatomy of brain and pons. (a) The medial view through the center of the brain showing structure of the brainstem, from Kolb and Whishaw (1990). (b) The anatomical structure of the pons, from Perkins and Kent (1986).

the cranial nerves. At the lower half of the pons and the medulla, the motor cranial nerve nuclei, e.g., V and VII which are important in motor speech, are located.

Upper Motor Neurons (Pyramidal and Extrapyramidal Motor Systems):

These two systems deliver signals to cranial and spinal nerves for innervation of skeletal muscles, e.g. limb muscles and speech muscles. Figure 2.7-a shows that the pyramidal system is constructed in "parallel". It delivers the signal directly from the motor cortex to lower motor neurons. The pyramidal (direct) system is mainly excitatory and mediates volitional movement. This system is chiefly responsible for discrete, accurate, quick, and phasic movements. It is also most concerned with spatially oriented movements. The main contribution to the direct system is made by the corticobulbar and corticospinal neurons located chiefly in the motor cortex in front of the central fissure of Rolando. Experimental lesions of the pyramidal system produce weakness with loss of skilled movements, absence of abdominal reflexes, decreased tone, and positive Babinski signs. The Babinski sign is defined as: an upward flexion, especially of the big toe, and outward fanning of the toes when the lateral planar surface of feet is stimulated.

Figure 2.7-b shows that the extrapyramidal system is constructed in "series". It provides indirect connections of the motor cortex to lower motor neurons. The extrapyramidal system is both inhibitory and excitatory and mediates various reflexive responses, e.g., postural adjustment and emotional expression. The main contribution to the indirect system is made by the premotor areas located in front of the primary motor area of cortex. Lesions and removal of the premotor cortex produce spasticity, increased muscle stretch reflexes, and weakness.

The pyramidal system, with its direct motor translation from cortex to lower motor neurons, is a fast route for fine coordination of speech and the discrete skilled aspects of an act. However, the pyramidal system produces only isolated segments of movement. For these smooth movements, the pyramidal system must cooperate with the extrapyramidal system, which picks up neural information for reflexive adjustment at every motor station from the cortex to the spinal cord.

Lower Motor Neurons:

The lower motor neuron is located in the somatic motor nuclei of the cranial nerves and in the anterior horns of the spinal cord. It uses synapses to connect the spinal cord and brainstem from the pyramidal and extrapyramidal motor systems, as shown in

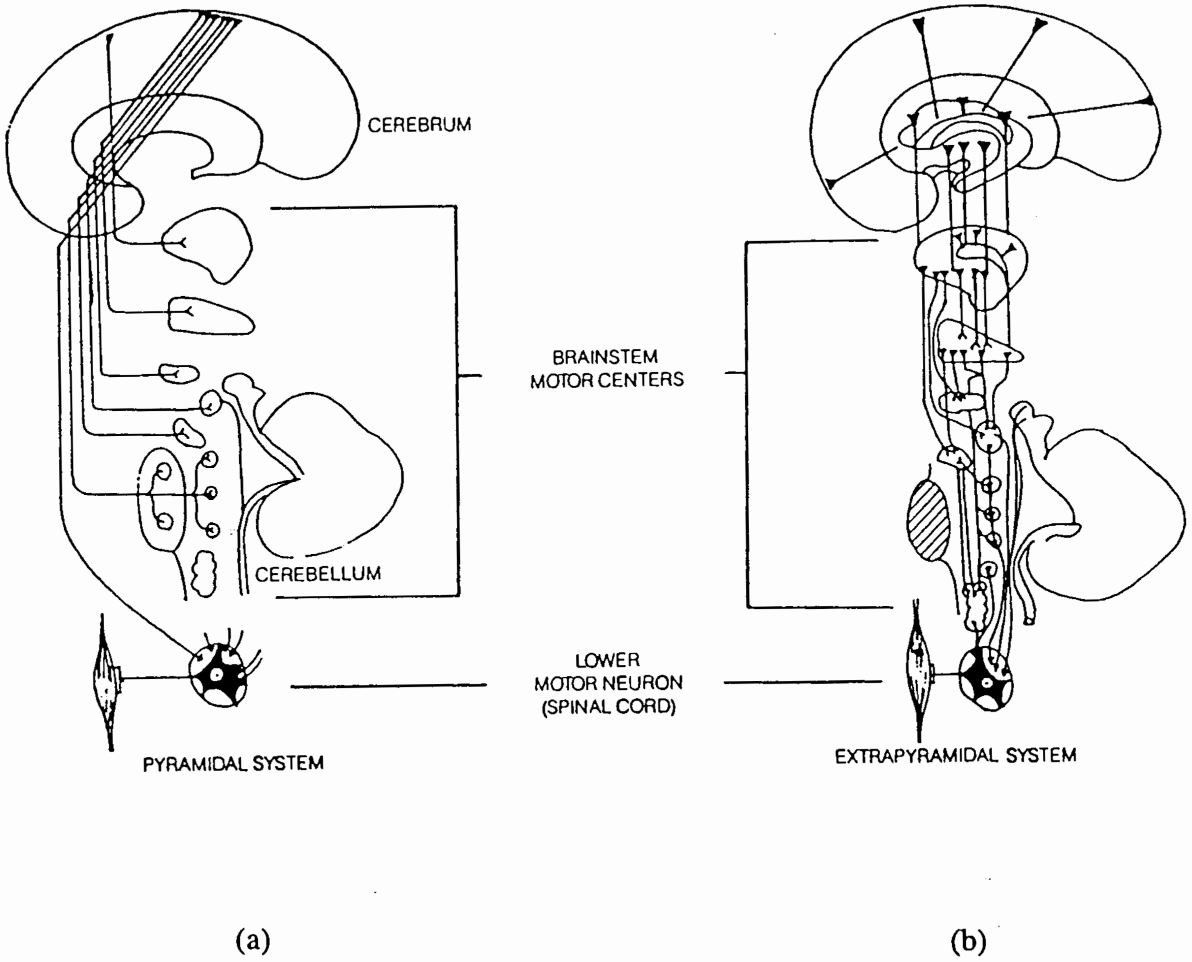


Figure 2.7 Upper motor neuron, from Perkins and Kent (1986). (a) The pyramidal (direct) system transmits directly the signal from the cerebellar cortex motor areas to the lower motor neuron, including the cranial nerves and the anterior horns of the spinal cord. (b) The extrapyramidal (indirect) system transmits indirectly the signal from the cerebellar cortex motor areas to the lower motor neuron, including the cranial nerves and the anterior horns of the spinal cord.

Figure 2.8. Its axon leaves the central nervous system directed by a peripheral nerve to skeletal muscle fibers. It may be referred to as "the final common pathway" because stimulus to movement (or muscle contraction) must finally use the lower motor neurons to produce the movement.

If the lower motor neurons are fired, then the muscle will contract. The faster the firing to a fiber is, the more fibers are innervated simultaneously and the stronger the muscle contraction is. Muscle tone is maintained by the balance of tension between opposing muscles. Tone is low when relaxed and is high when heedful. If a quick response is needed, slack must be minimal; therefore the muscle tone must be high. The alpha motor neuron is used to innervate the muscle tone and the gamma motor neuron, which provides high-speed feedback, adjusts the background of muscle tone against all movements which are innervated by alpha motor neuron, as schematized in Figure 2.8. Both of these two motor neurons are located in the anterior horns and cranial nerve nuclei.

Vocal Tract:

Figure 2.9-a shows the anatomical structure of the vocal tract consisting of the throat, mouth, and nose. There are three cavities in the vocal tract: pharyngeal cavity (throat), nasal cavity (nose), and oral cavity (mouth), as schematized in Figure 2.9-b. The boundary between oral and pharyngeal cavities is dependent on the position of the vocal tract constriction. The nasal cavity can be closed from the remainder of the vocal tract by the soft palate (velum). The movements of the lips, cheeks, tongue body, and laryngopharynx change the shape of the vocal tract. The sound sources for generating the speech are from the vibration of vocal folds (voiced sound), turbulence noise produced by the pressurized air flowing through a narrow constriction (voiceless sound), or combining both (voiced sound with turbulence noise). The three cavities are used as the resonators to filter the sources and to produce the utterances. In general, the movement control of the tongue body, lips, cheeks, and laryngopharynx influence the vocal tract shape. The larynx function controls whether or not there is glottal vibration and controls the type of sound that is generated near the glottis. The velum controls the nasalization of sound. Any dysfunction of these parts will cause the utterance output to be abnormal.

There are twelve pairs of cranial nerves which emerge from the base of the skull. Each cranial nerve is marked as a Roman numeral and a name. Not all of them are relevant to the speech function. Cranial nerve V (serving the mandibular and tongue

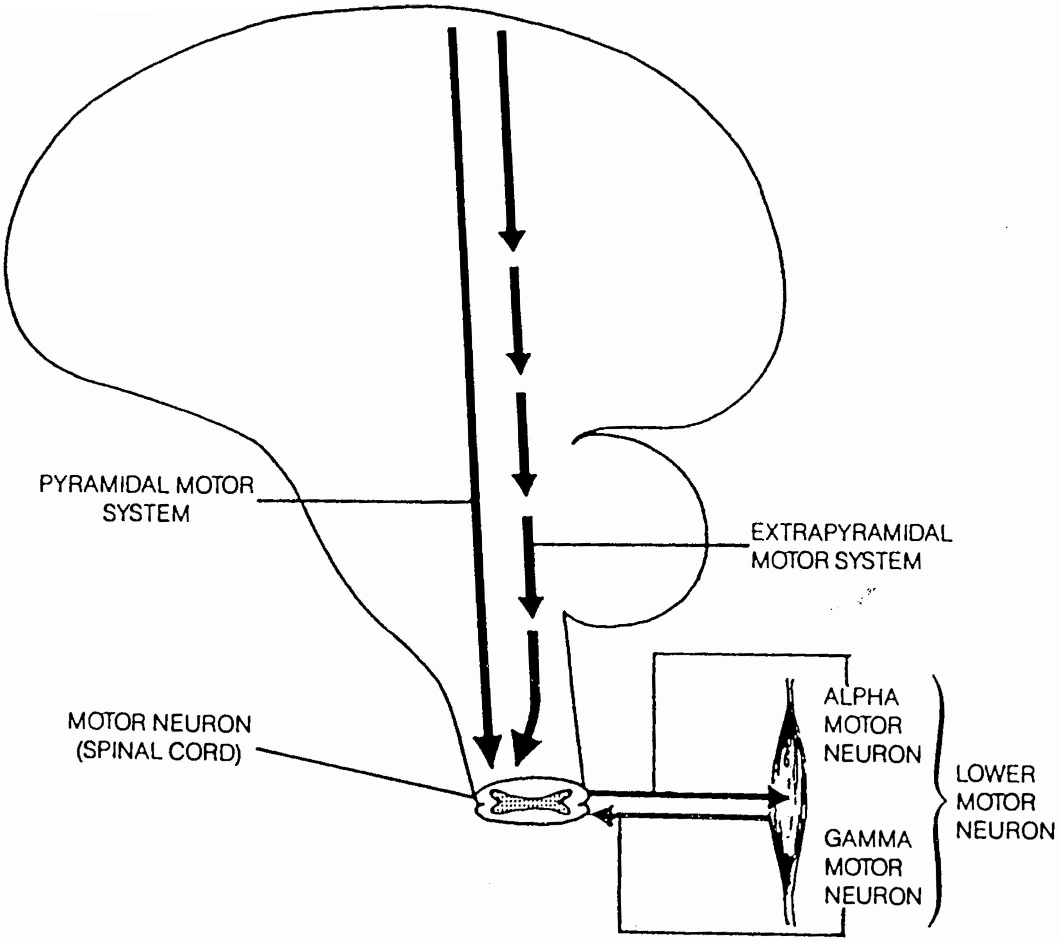


Figure 2.8 Lower motor neuron, from Perkins and Kent (1986). Its axon leaves the central nervous system to project by a peripheral nerve, including cranial and spinal nerves, to skeletal muscle fibers. The motor signal is transmitted from the cerebellar cortex through pyramidal and extrapyramidal motor systems and the lower motor neuron to the muscle fibers. The alpha motor neuron is used to innervate the muscle tone and the gamma motor neuron is used to regulate the background of muscle tone.

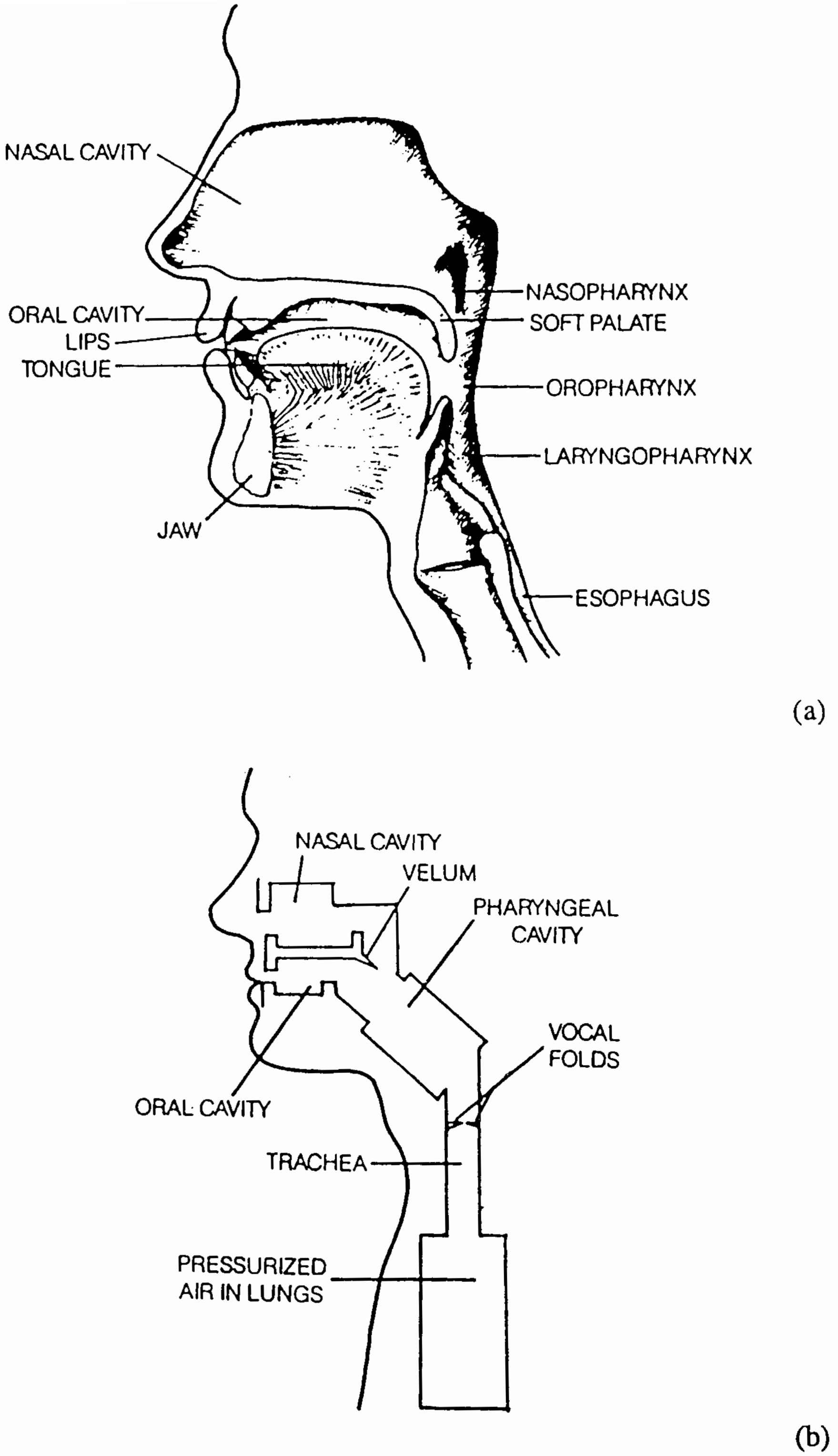


Figure 2.9 The vocal tract and the corresponding acoustic model, from Perkins and Kent (1986). (a) The basic anatomy structure of the vocal tract. (b) The acoustic model for the vocal tract.

muscles and sensory to face and head), VII (sensory for tongue and motor for facial expression), VIII (sensory for audition and balance), IX (sensory to tongue and pharynx and motor to pharynx), X (complex nerve with sensory or motor fibers, or both, to larynx, respiratory system, gastrointestinal system, and cardiac system), XI (motor to shoulder, arm, and throat which includes the pharynx, soft palate, and neck), and XII (motor to tongue, and possible position (proprioception) sense) are relevant for controlling the vocal tract and lungs. In addition, the thoracic nerves (T1-12) of the spinal nerves also influence speech because they serve respiration by innervation of the muscles of the chest wall. Therefore, if there is any damage to these nerves, the vocal tract function may be impaired and the speech output may become abnormal.

Inspiration and Expiration Systems:

The power for the sound is generated by the coordination of the chest, abdomen and back musculature, which produce an air-flow through the trachea. The principal muscles of inspiration are the diaphragm and the external intercostal muscles. The primary muscles of expiration are the abdominal muscles, but internal intercostals and other chest and back muscles also contribute. Trauma or surgery that alters the structure or function of these muscles impairs the power source of the voice, e.g., asthma.

Stages of Speech Processing:

Figure 2.10 shows a more realistic view of the speech processing procedure. As shown in Figure 2.10-a, it has been proposed that an early stage of an utterance is the formulation in the limbic network (including limbic system and thalamus) for the configuration of what is to be said. Furthermore, the cognitive and emotional components of an utterances, which may receive contributions from the right hemisphere, are also integrated into the formulation in the beginning of the first step. In the second step, as shown in Figure 2.10-b, the formulation then extends out through linguistic planning networks until it is ready for motor implementation. This action is accomplished by the networks involving cortical association areas, basal ganglia, Wernicke's area, Broca's area, and cerebellum. Then, the third step, as shown in Figure 2.10-c, shows that cerebellar processing is incorporated through thalamic networks with precentral motor networks for the generation of motor speech. In the final stage, as shown in Figure 2.10-d, the motor signal propagates from motor cortex through the motor system networks (including pyramidal motor system and extrapyramidal motor system) to lower motor neurons (including alpha motor neuron, gamma motor neuron) and muscle fibers for execution of articulatory movement.

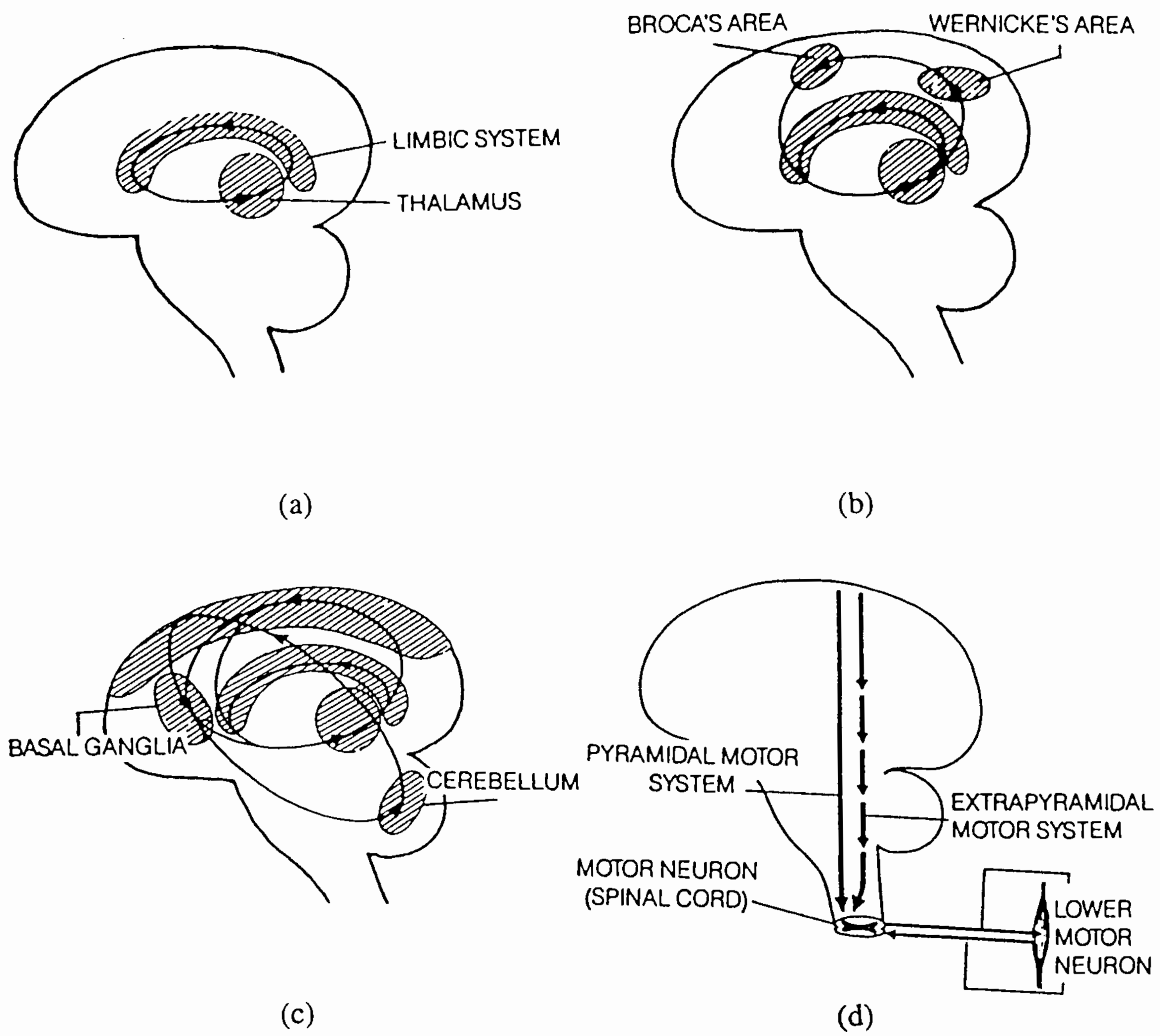


Figure 2.10 The stages of the speech processing, from Perkins and Kent (1986). In the first step, Part (a), it has been proposed that an early stage of an utterance is the formulation in the limbic network, including limbic system and thalamus. In the second step, Part (b), the formulation then extends out through linguistic planning networks until it is ready for motor implementation. This action is accomplished by the networks involving cortical association areas, basal ganglia, Wernicke's area, Broca's area, and cerebellum. Then, the third step, Part (c), shows that cerebellar processing is incorporated through thalamic networks with precentral motor networks for the construction of motor speech programs. In the final stage, Part (d), the motor signal reaches from motor cortex through the pyramidal and extrapyramidal motor systems to lower motor neurons, including spinal cord, alpha and gamma motor neuron, and muscle fibers, for execution of articulatory movement.

2.3 Dysarthric Speech:

Dysarthria is a group of speech disorders resulting from disturbances in the muscular control of the speech mechanism due to neuromuscular disease. Dysarthria results in a weakness, slowness or incoordination of speech and may involve several or all of the basic processes of speech: respiration, phonation, resonance, articulation, and prosody (Yorkston, 1988). Dysarthria may accompany aphasia and apraxia. Aphasia is defined as a kind of "impairment, due to brain damage, of the capacity to interpret and formulate language symbols; a multimodal loss or reduction in decoding conventional meaningful linguistic elements" (Yorkston, 1988) [p. 60]. Apraxia is defined as "an articulatory disorder resulting from impairment, as a result of brain damage, of the capacity to program the positioning of speech muscles and the sequencing of muscle movements for the volitional production of phonemes" (Yorkston, 1988) [p. 60]. In the present research project, in order to minimize the complexity of the speech symptoms, the dysarthric subjects participating in the project do not suffer from the symptoms of apraxia or aphasia. In short, the special subjects required in this project are individuals who have speech problems due to the disturbances in the muscular control of the speech mechanism. These subjects' auditory comprehension and reading skills are intact.

2.3.1 Classification of Dysarthrias

Darley et al. (1975) described dysarthria as follows:

> Dysarthrias can be classified according to age of onset (congenital, acquired); etiology (vascular, neoplastic, traumatic, inflammatory, toxic, metabolic, degenerative); neuroanatomic area of impairment (cerebral, cerebellar, brain stem, spinal; or central, peripheral); cranial nerve involvement (V, VII, IX - X, XII); speech process involved (respiration, phonation, resonance, articulation, prosody); or disease entity (parkinsonism, myasthenia gravis, amyotrophic lateral sclerosis, etc.). Probably most useful clinically are systems of classification reflecting neuroanatomic and neurophysiologic considerations: what part of the central or peripheral nervous system is implicated and what distinctive behavior results? [p. 12]

Luchsinger and Arnold (1965) distinguished six dysarthrias from an anatomic viewpoint, due to lesions of (i) the precentral motor cortex, (ii) pyramidal tract, (iii) extrapyramidal system, (iv) frontocerebellar tract, (v) cerebellar coordination centers, and (vi) bulbar nuclei of cranial nerves. Darley et al. (1968) provided seven groups with perceptually based descriptions of the dysarthrias, given in Table 2.1. The designation for each type of dysarthria is listed at the beginning of each group, and speech and nonspeech impairments are then described. Several years later, Darley et al. (1975) provided another classification of dysarthrias based on the neurological symptoms, given in Table 2.2. The neuromuscular condition is specified and the supplementary neuroanatomic or neurologic designations indicate the probable origin. LaPointe (1994) summarized the most common dysarthria classifications using five categories. The corresponding designation and explanation for each type of dysarthria is listed in Table 2.3.

The different subjects in the present study exhibited three types of dysarthria: spastic dysarthria, athetoid (hyperkinetic) dysarthria, and ataxia dysarthria. Before giving a detailed discussion of these three specific types of dysarthria, a brief review of the cerebral palsy will be introduced.

2.3.2 Cerebral Palsy

The term "cerebral palsy" refers to disorders of movement resulting from damage to the brain at some time during the period of growth and development. Other deficits in function may also occur due to damage to parts of the brain other than the motor cortex. The manifestations of cerebral palsy depend upon the location and severity of the brain damage. Although the locations are not precisely known, they can be inferred from the observed complex of symptoms. For this reason, some authorities feel uncomfortable with the term, since it does not truly describe a discrete entity, its etiology, or its anatomical location. Easton and Halpern (1981) have stated that cerebral palsy is itself:

> ... not progressive, but manifestations vary in expression with the maturity of the brain and body... Basically, the motor or movement problem in cerebral palsy is lack of control of the muscles rather than muscle weakness. This lack of control is often expressed by a failure of inhibition of some specific central nervous system reflexes. These reflexes involving muscle movement, in increasing order of complexity, are: stretch, crossed extension, long spinal, symmetrical tonic neck, asymmetrical tonic neck, vestibular, and startle [p. 137].

Yorkston (1988) also defined cerebral palsy as "a nonprogressive motor disorder that stems from an insult to the cerebral level of the central nervous system during the prenatal

Table 2.1 A perceptually based descriptions of the dysarthrias, adapted from Darley et al. (1968). The speech and nonspeech symptoms are mentioned in each type of dysarthria.

(i) Pseudobulbar Palsy (Spastic Dysarthria):
Speech symptoms: speech slow and labored, the articulation being rather consistently imprecise (especially on more complicated groups of consonant sounds), low and monotonous pitch, harsh and often strained or strangled-sounding voice quality, and hypernasality (but usually no audible nasal emission).
Nonspeech symptoms: increase of deep tendon reflexes, appearance of the sucking reflex, increased jaw jerk, sluggish tongue movements, and activity of accessory respiratory musculature.

(ii) Bulbar Palsy (Flaccid Dysarthria):
Speech symptoms: hypernasality with associated nasal emission of air during speech as its most prominent speech symptom, audible inhalation, breathy exhalation, air wastage being manifested also in shortness of phrases, often imprecise articulation on either or both of the following two bases: (1) consonants may be weak through failure to impound sufficient intraoral breath pressure because of velopharyngeal incompetence (2) immobility of tongue and lips because impairment of the hypoglossal and facial nerves prevents normal production of vowels and consonants.
Nonspeech symptoms: fasciculation and atrophy of the tongue, reduced rate if alternating motion of tongue and lips, poor elevation of the soft palate and nasal alar contraction and grimacing as the patient tries to compensate for velopharyngeal incompetence.

(iii) Amyotrophic Lateral Sclerosis (Combined Spastic and Flaccid Dysarthria):
Speech symptoms: a developing effect on speech. In an earlier stage either spastic or flaccid speech and nonspeech signs predominate; in an advanced stage both sets of features (spastic and flaccid dysarthria) described above are present. It is a progressive degenerative disease.

(iv) Cerebellar disorders (Ataxic Dysarthrias):
Speech symptoms: producing one of two patterns of speech deviation, the two seldom appearing concurrently: (1) intermittent disintegration of articulation, together with dysrhythmia and irregularities of pitch and loudness in performing tests of oral diadochokinetic rate (2) altered prosody involving prolongation of sounds, equalization of syllabic stress (by undue stress on usually unstressed words and syllables), and prolongation of intervals between syllables and words.

(v) Parkinsonism (Hypokinetic Dysarthria):
Speech symptoms: reducing vocal emphasis, peaks and valleys of pitch and variations of loudness being flatted out monotonously, short rushes of speech (separated by illogically placed pauses), the speech rate being variable (often accelerated), blur consonant articulation in contextual speech and syllable repetition as muscles fail to go through their complete excursion, difficulty in initiating articulation (shown by repetition of initial sounds and inappropriate silences), breathy voice, and weak loudness even to inaudibility.

(vi) Dystonia (Hyperkinetic Dysarthria):
Speech symptoms: voice stoppages, disintegration of articulation, excessive variations of loudness, distortion of vowels, and normal prosody altered by slowing of rate, reduction in variations of pitch and loudness, prolongation of inter-word intervals, and interposition of inappropriate silences.
Nonspeech symptoms: involuntary bodily and facial movements.

(vii) Choreoathetosis (Hyperkinetic Dysarthria):
Speech symptoms: abnormal breathing cycle, sudden exhalatory gusts of breath, bursts of loudness, elevations of pitch, disintegration of articulation, increasing loudness level, prolonging pauses, and equalizing stress on all syllables and words.

Table 2.2 The classification of dysarthrias based on the neuromuscular condition as the following six groups, adapted from Darley et al. (1975). The designation is listed in the beginning of each group and the explanation for the corresponding type of dysarthria is given at the right.

	Designation:	Explanations:
(i)	Flaccid Dysarthria:	lower motor neuron lesion.
(ii)	Spastic Dysarthria:	bilateral upper motor neuron lesion.
(iii)	Ataxia Dysarthria:	cerebellar or cerebellar pathway lesion.
(iv)	Hypokinetic Dysarthria: (a) in parkinsonism	extrapyramidal lesion.
(v)	Hyperkinetic Dysarthria: (a) in chorea: quick hyperkinesia (b) in dystonia: slow hyperkinesia (c) others	extrapyramidal lesion
(vi)	Mixed Dysarthria: (a) spastic-flaccid in amyotrophic lateral sclerosis (b) spastic-ataxic-hypokinetic in Wilson's disease (c) variable in multiple sclerosis (d) others	lesions of multiple systems

Table 2.3 The most common dysarthria classifications, adapted from LaPointe (1994). The designation is listed in the beginning of each group and the explanation and the speech impairments for the corresponding type of dysarthria are followed.

	Designation:	Explanations:
(i)	Spastic Dysarthria:	caused by damage to the bilateral motor strip in the brain and resulting in poor articulation and strain / strangled voice.
(ii)	Ataxia Dysarthria:	caused by damage to the cerebellum and resulting in irregular speech and syllable repetition.
(iii)	Flaccid Dysarthria:	originating from neurological involvement in the brainstem and resulting in poor articulation and hypernasality.
(iv)	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.
(v)	Hypokinetic Dysarthria:	usually evolving from Parkinson's disease and resulting in reduced loudness and speech rate.

or perinatal period" [p. 80]. Although the definition of cerebral palsy has the term nonprogressive, the manifestations of this primarily motor disorder will change as the body changes through growth.

The development of language and communication skills from infancy depends upon the four factors mentioned in the beginning of this chapter: an intact hearing, an intact brain, an environment of interactions between peers and adults, and sufficient motor skills to communicate. The person with cerebral palsy may be deficient in all or part of these four functions from birth (Easton and Halpern, 1981). Furthermore, the cerebral palsy that results from damage to the basal ganglia and cerebellum causes more impairment of speech control than does injury to the motor cortex (Perkins and Kent, 1986).

Erenberg (1984) has studied the average percentage of four classes of cerebral palsy: (i) spastic (75%), (ii) athetoid (5%), (iii) dystonic athetoid or ataxia (10%), and (iv) a mixed variety (10%). The spastic variety accounts for the vast majority of cases. Within the spastic category are the subclasses of hemiparesis (30%), diparesis (25%), and quadriparesis (20%). Darley et al. (1975) also used "spastic", "athetoid", "ataxic", and other terms to indicate the predominant neurologic features of various clinical types of cerebral palsy. Easton and Halpern (1981) listed nine symptoms of cerebral palsy which are included in the concurrent symptomatolgic classification accepted by the American Academy for Cerebral Palsy and Developmental Medicine. These are: (i) spasticity, (ii) athetosis (nontension and tension), (iii) ataxia, (iv) dystonia, (v) atonia, (vi) rigidity, (vii) tremor, (viii) mixed, and (ix) undetermined. Of the above, spasticity, ataxia, and athetosis are the main movement disorders. These symptoms may occur alone or in conjunction with others and may affect one or more body parts.

It is also commonly believed that speakers with cerebral palsy have difficulty in "coordinating" the processes of articulation, phonation, and respiration. Easton and Halpern (1981) mentioned, "The average incidence of oral communication disorders among cerebral palsied individuals has been estimated at 70-80 percent" [p. 146]. Wolfe (1950) indicated that 70 percent of individuals with cerebral palsy show some degree of dysarthria. The review shows that speech disorders are a common symptom of this neurological disorders.

Since spasticity and athetosis are the most important neurologic features of various clinical types for cerebral palsy, and since all of the cerebral palsy subjects who participate in this project belong to these two kinds of cerebral palsy, these two types of dysarthria are discussed in the following sections.

(i) Athetoid Cerebral Palsy

Etiology and General Considerations:

Athetosis belongs to a kind of hyperkinetic dysarthria (Darley et al., 1968 and 1975, and LaPointe, 1994). The condition occurs as one form of cerebral palsy. Athetosis is associated with lesions in the mid-brain tegmentum, subthalamic nucleus, ventral lateral thalamus, pallidum, striatum, and cortex. However, the special limited location has not been exactly determined (Darley et al. 1975). LaPointe (1994) also defined it as a disease caused by damage throughout much of the extrapyramidal motor system. Bulbar musculature will participate if the athetosis is bilateral (Darley et al. 1975).

Salient Nonspeech Features:

Darley et al. (1975) have described Athetosis as follows: "repetitive twisting, writhing movements that slowly blend one into another. Slow flexion of individual fingers, for example, blends into overextension of the wrist, followed by spreading and extension of the fingers and rotation of the wrist. The movements affect distal muscles most prominently" [p. 211].

Salient Speech Features:

Articulation problems are found to be significant in a group of 58 athetoid children studied by Palmer (1952). Darley et al. (1975) mentioned, "The slow, wriggling, writhing, and involuntary movements of athetosis can be expected to interfere with speech production" [p. 216]. In addition, Purves-Stewart and Worster-Drought (1952) stated, "The grimaces of the face and the involuntary movements of the tongue interfere with articulation. Moreover, irregular spasmodic contractions of the diaphragm and other respiratory muscles give the voice a curiously jerky or groaning character, due to sudden interruption of breathing" [p. 238]. Nielsen (1951) reported, "the tongue and muscles of mastication and articulation are affected so that dysphagia is prominent and speech is explosive. Words may be cut in half by unexpected movement of the larynx" [p. 202]. Berry and Eisenson (1956) noted that:

> The individual who suffers from a constant, involuntary shift of tonus from one set of muscles to another, from an overflow of stimuli to muscles unrelated to the activity, and from a lack of direction and kinesthetic perception of movement, particularly noticeable in the tongue or jaw, may have completely unintelligible speech or no speech (mutism). On the other hand, the writer has met "medically-diagnosed athetoids" v/hose speech exhibits only a slight deviation in rhythm or articulation [p. 358].

Rutherford (1944) and Leith (1954) concluded that athetoid individuals demonstrate a monotonous pitch level. Clement and Twitchell (1959) described the voice of athetoid speakers as low in pitch with sudden uncontrolled rising inflections and exhibiting a forced and throaty speech.

In summary, the 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.

(ii) Spastic Cerebral Palsy

Etiology and General Considerations:

Spasticity is a kind of disorder of the upper motor neurons. It is caused by damage to the bilateral upper motor neurons in the brain (Darley et al., 1975 and LaPointe, 1994). Two classes of symptoms emerge from upper motor neuron lesions: the negative and the positive. Negative symptoms are the losses of functions that are a direct effect of the damage, e.g., paralysis of a voluntary movement. Positive symptoms are the evidences of overactivity, e.g., spasticity and hyperactive reflexes (Darley et al., 1975).

Salient Nonspeech Features:

Spastic muscles are stiff, move sluggishly through a limited range, and tend to be weak. The muscles have increased tone with heightened deep tendon reflexes (Dorland's Illustrated Medical Dictionary, 1981). Darley et al. (1975) also mentioned, "diseases of the upper motor neurons affect voluntary movements - and thus speech - by four major abnormalities of muscular functions: spasticity, weakness, limitation of range, and slowness of movement" [p. 131]. In brief, the nonspeech symptoms for the spastic are (i) increase of deep tendon reflexes, (ii) appearance of the sucking reflex, (iii) increased jaw jerk, (iv) sluggish tongue movements, and (v) activity of accessory respiratory musculature (Darley et al., 1968).

Salient Speech Features:

Disease of the upper motor neurons damages the part of movements which are necessary for efficient speech production. Therefore, speech becomes slow and seems to emerge effortfully (Darley et al. 1975). Parker (1956) summarized the overall effect of spastic dysarthria from one of his subject: "his speech is slow, rasping, labored, and each word is prolonged. It is dominant in lower tones and hardly intelligible. It is like the stiff gait of a spastic patient moving with might and main but progressing ineffectually under heavy internal difficulties" [p. 163]. Zentay (1937) differentiated spastic dysarthria from other dysarthrias as follows: "it is a slow dragging speech with indistinct articulation, apparently requiring a great deal of effort and often accompanied by facial distortions. At times the slowness alternates with explosiveness" [p. 150]. LaPointe (1994) also mentioned that the spastic dysarthrias result in poor articulation and strain / strangled voice. In brief, the speech symptoms of the spastic dysarthrias are: (i) slow and labored speech, (ii) consistently imprecise articulation (especially on more complicated groups of consonant sounds), (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).

(iii) Comparison of Athetosis with Spastic

Lencione (1953) studied 129 educable cerebral palsied children between the ages of 8 and 14. There were 45 subjects who were predominantly of athetoid type. She found the athetoid children to be significantly less proficient than the spastic children in producing the phonemes tested. In addition, the athetoids were also significantly less intelligible than the spastic children. Further, Byrne (1959) found that a group of 29 athetoid quadriplegic children produced vowels and consonants less accurately than a group of 32 spastic quadriplegic children, but none of the differences between the groups reached statistical significance. Farmer (1975) studied the speech production of five athetotic and five spastic cerebral-palsied speakers and found that the athetoid group produced longer and more variable voice onset times than the spastics. Rutherford (1944) also reported that the athetoid group (48 children) has more loud voices, more low-pitched voices, more monotonous voices, and more breathy voices, than the spastic group (74 children). Platt et al. (1980a) also mentioned the following findings:

> In this adult sample, athetoids produced twice as many incomprehensible words as spastics, and their maximum rate of syllable generation was about one third as fast as normals, while spastics were about half as fast as normals. With regard to specific phonemic features, spastics and athetoids appeared to different only in terms of amount of accuracy, and not in terms of any distinctive phonetic difference. In general, it appears that the neuro-muscular limitations imposed on the dynamic process of speech production in cerebral palsy are more severe in athetosis than in spasticity [p. 38].

2.3.3 Cerebellar Ataxia:

Etiology and General Considerations:

Ataxic dysarthria results from damage to the cerebellum or cerebellar pathway (Darley et al., 1975 and LaPointe, 1994). The main functions for the cerebellum have been summarized in the previous section. Most frequently ataxic speech occurs in the presence of localized or generalized cerebellar damage. Darley et al. (1975) described the localized and generalized cerebellar damage as follows:

> Localized damage to the cerebellum results from tumors, multiple sclerosis, the toxic effect of alcohol, strokes, or trauma. Such localized damage may selectively impair equilibrium, gait, unilateral limb coordination, or speech. With more extensive lesions, a combination of impairments will occur. Generalized damage to the cerebellum may follow inherited or sporadic neuronal degeneration, encephalitis, exposure to toxins, cancer of the lung, or disseminated demyelinating or vascular lesions. Although generalized damage may initially affect only one cerebellar function, eventually gait, limb coordination, speech, and equilibrium all become impaired [p. 151].

Salient Nonspeech Features:

The salient features of ataxia are: (i) inaccuracy of movement, (ii) slowness of movement, and (iii) hypotonia. Voluntary movements are slow. The force, range, timing, and direction of movements are also inaccurate. The patient's equilibrium for sitting and gait may be impaired, e.g., producing staggering. Eye movements may be irregular and jerky. Nystagmus is commonly observed. Extremity tremor may occur and increase to a maximum toward the end of the movements (Darley et al., 1975).

Salient Speech Features:

The patients with functional impairment of the cerebellum and its fiber tracts display breakdowns of the highly coordinated muscular adjustments necessary for speech. Descriptive terms are used widely in the literature: slow, staccato, slurred, jerky, forced, explosive, irregular, interrupted, drawling, monotonous, asynergic, labored, and scanning. The scanning speech is the most frequent disorder for the ataxic dysarthrias (Darley et al., 1975). Walshe described scanning speech as follows: "Each component syllable in a word is pronounced deliberately and slowly as though it were a separate word, and is not run together with its fellow syllables as in normal speech. In severe cases the disorder may render speech almost unintelligible." (Walshe, 1970) [p. 29]. Darley et al. (1969b) found that the most prominent problem in the speech of the Mayo clinic subjects with cerebellar lesions was imprecision of consonant articulation. Two other characteristics (excess and equal stress and irregular articulatory breakdown) were also apparent. Darley et al. (1975) also observed, "one might reasonably expect discrete and repetitive movements of the oral speech mechanism to be performed somewhat slowly and probably irregularly, since structures affected by cerebellar diseases tend to have reduction of muscle tone - hypotonia - and cerebellar ataxia is characterized by inaccuracy and slowness" [p. 164]. Kammermeier (1969) measured the fundamental vocal frequency, vocal frequency variability, intensity variability, oral reading rate, and durational aspects (such as mean length of syllables and phonation segments) for five clinical groups of dysarthric patients: spastic, flaccid, ataxic, hypokinetic, and hyperkinetic dysarthrias. From a group of eight patients with ataxic dysarthria, their mean vocal frequency, measured from analysis of an oscillographic tracing of a recorded 27-syllable sentence, was 135.0 Hz, which was the highest one in the five groups of dysarthrias. Kent and Netsell (1975) showed that the physiological abnormalities of ataxic dysarthria are slow articulatory movements, errors of direction and range of articulatory movements, and reduced variety of fundamental frequency contour.

Yorkston and Beukelman (1981b) mentioned three perspectives for studying the characterisitics of ataxic dysarthria: perceptual speech characteristics, physiological characteristics, and acoustic characteristics. They also made a review about the ataxic dysarthria from a variety of viewpoints. Simmons (1982) described the speech characteristics of ataxia as follows: excess and equal stress, slow rate, monoloudness, and monopitch. Darley et al. (1975) gave the following summary of the characteristics of ataxia speech:

In summary, ataxic dysarthria is characterized by the following irregularities: marked breakdown of articulation involving both consonants and vowels, alteration in the prosodic aspects of speech so that word and syllabic stresses tend to be equalized, with excessive stress placed on usually unstressed words and syllables, and with prolongation both of phonemes and of the intervals between them, dysrhythmia of speech and syllable repetition, generally slow rate, and some harshness of voice along with monotony of pitch and loudness, occasionally broken by patterns of excessive loudness variation [p. 169].

2.4 Speech Recognition and Speech Synthesis

Speech is one of the most essential tools for human communication. However, machines prefer the symbols exchanged in printed form through a computer terminal. If computers can communicate like humans, their value and ease of use would increase (Flanagan, 1982). Although there are many input channels for the operation of computers, as illustrated in Figure 1.1, speech input and output are still great potential communication channels for the computer. The speech chain shown in Figure 2.11 represents the sequences for speech communication (Denes and Pinson, 1973). Any utterance produced from a speaker should pass through the same series of steps, from the linguistic level to the physiological level, to produce the sound waves on the acoustic level. The acoustic signal is produced from resonances of vocal tract, which is controlled by a set of muscles. The basic theory has been reviewed in Section 2.2. When the listener's and speaker's ears receive the acoustic signal from the air, the signal is translated into the brain by the nervous system and its transformation is the inverse of speech production. Finally, a listener can understand what the speaker would like to express and the speaker can hear the sound signal pronounced by himself (or herself) to calibrate the output speech. Therefore, the basic procedure of speech communication is complete. If human listeners are replaced by a computer, this replacement part is called a speech recognition system. On the other hand, if human speakers are replaced by a computer or the equivalent speech output system, the replacement part is called a speech synthesis system.

In particular, speech recognition can accelerate the typing speed or automation of computer operations in the future. Further, a speech synthesizer can be used as the output channel for a computer to translate text to speech, e.g., in a multi-media system. However, making computers talk and listen to human's commands would still depend on the economical implementation of the speech recognition and synthesis. Fortunately, advances in integration of computers, electronics, and the relevant hardware / software equipment has made the cost of the speech recognition and synthesis commercially feasible. Since the speech recognition function is the main issue discussed in this project, the following sections focus mainly on the study of speech recognition.

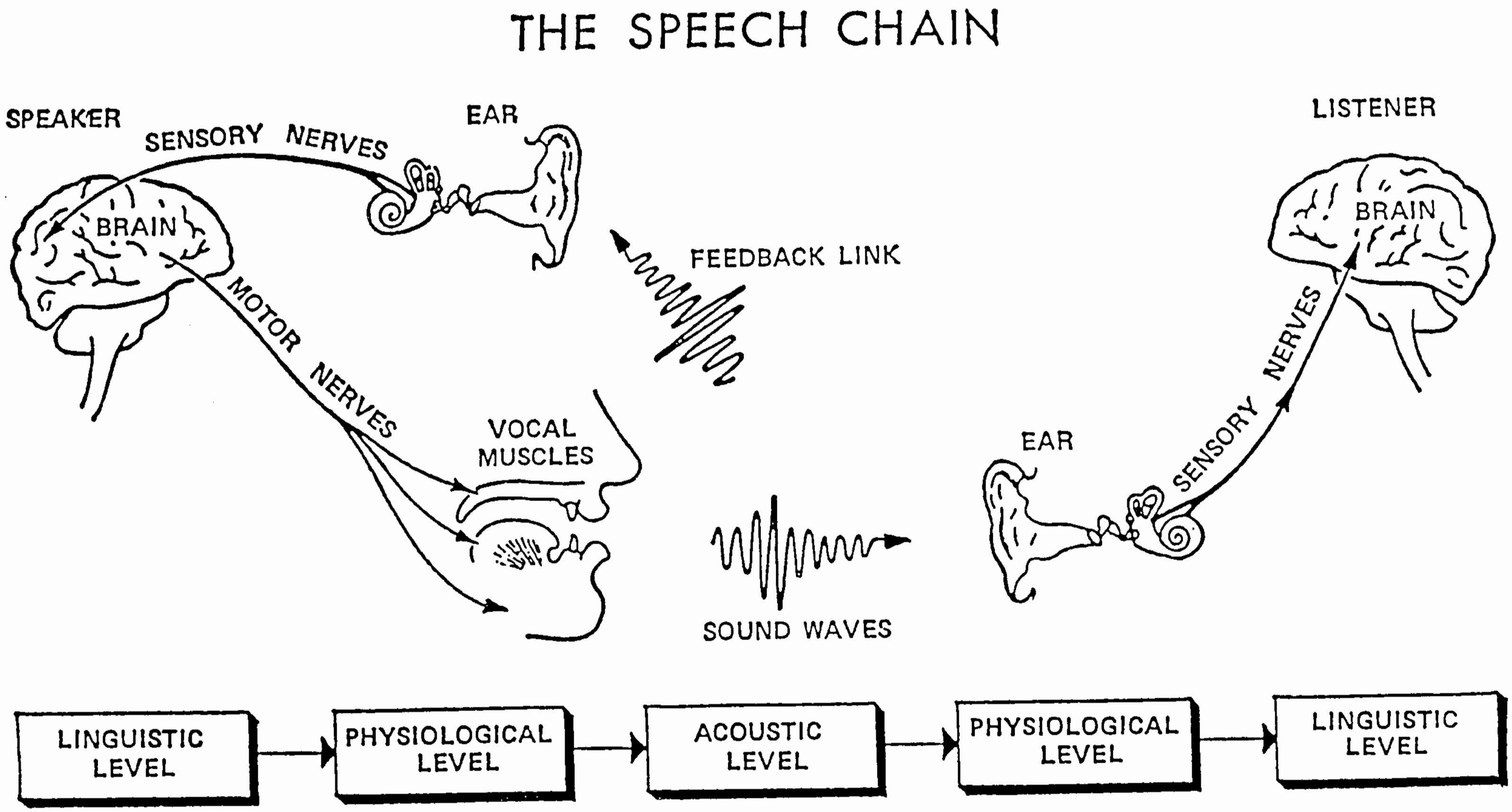


Figure 2.11 Speech chain, from Denes and Pinson (1973). The different steps in which a spoken language translates from the mind of the speaker to the mind of the listener are illustrated.

2.4.1 Classification of Speech Recognition Systems

Automatic speech recognition (ASR) has been researched for at least 40 years (Rabiner and Juang, 1993). In recent years, recognition of a limited vocabulary of single words has found a number of applications (Dixon and Martin, 1979; Flanagan, 1982). Speech recognition vocabularies of up to 30,000 words are possible (Baker, 1989). There are many commercial products in the market, e.g., Dragon System, BBN, Articulate System, and so on. Each product has its own market share and computer applications; for example, the Dragon dictate of Dragon System is on a PC base and the Power-Secretary of Articulate system is on a Macintosh base.

Speech recognition systems can be classified based on the vocabulary size, e.g., 20 words or up to 100,000 words, the user enrollment requirement, e.g., speaker-dependent or speaker-independent, and speech mode, e.g., continuous or isolated speech recognizers (Flanagan, 1982). Speaker-dependent recognizers need to be trained by the specific speakers before they can be used. This training usually takes much time for the customers before the system can be used. By contrast, speaker-independent recognizers can be applied to any speaker (whose speech pattern is close enough to the base-model built up in the machine) without training, but the recognition accuracy is still a problem since not every person has the same (or similar) speech patterns. On the other hand, a continuous speech recognition system lets the users feel more natural and comfortable when they are using the system. However, because there is phonological variation due to the context, prosody, and syllable structure, the difficulty and complexity of continuous speech recognizers are greater than for isolated word recognizers. The recognition accuracy for continuous recognizers is naturally lower than for isolated word systems. In particular, speaker-independent continuous systems are strongly influenced by the language models that are incorporated in the systems. Thus, automatic recognition of unconstrained fluent speech by any speaker on any subject (the speaker-independent recognizers with large vocabulary size) is not close to being commercialized. Present understanding of the "trades" between ease of use and recognition accuracy may be made by using either a small vocabulary size and a speaker-independent isolated-word system (e.g., only the digits 0 - 9) or large vocabulary size and speaker-dependent continuous-speech system. The prices of speech recognizers are very different based on the function and the complexity of the speech recognition system that the customers want, e.g., $300 - $10,000 or more.

2.4.2 Speech Recognition Algorithm

In brief, an ASR algorithm is used to convert the speech signal into recognized words. Rabiner and Juang (1993) and Lee and Reddy (1989) have shown the detailed theory for the speech recognizers. Figure 2.12 shows the rough procedure of ASR. The speech signal which is sampled by the ASR as the input is produced by a human speaker. Then, the ASR system produces a series of recognized words which can be used as the final output or the input to natural language processing. The main components in ASR are shown schematically in Figure 2.13. The speech signal produced by the speaker forms the input. Then, the input signal is processed in the "Representation" section. In this section, several issues are considered, e.g., signal representation (conventional Fourier representation, Mel-Scale cepstral representation, and so on), feature extraction (spectral and energy vectors and their time derivatives), and size of measurement units (fixed-size frames or variable-length segments). In the next step, the "Search" section is used to compare the input signal patterns with the acoustic models, lexical models, and language models by using a different search technique, e.g., using the Viterbi algorithm and the Baum-Welch algorithm. All of the acoustic models (e.g., Hidden Markov model (HMM) and Dynamic Warping model (DWM)), lexical models (e.g., single phonemic base form), and language models (e.g., word-pair model and N-gram model) are built up from training data. After the search section, the recognized words are output from the ASR system.

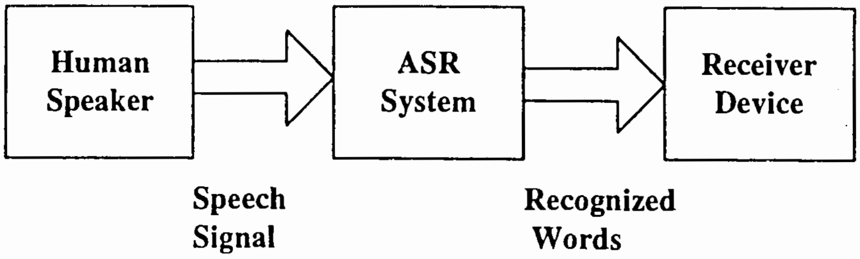


Figure 2.12 Automatic speech recognition (ASR) configuration.

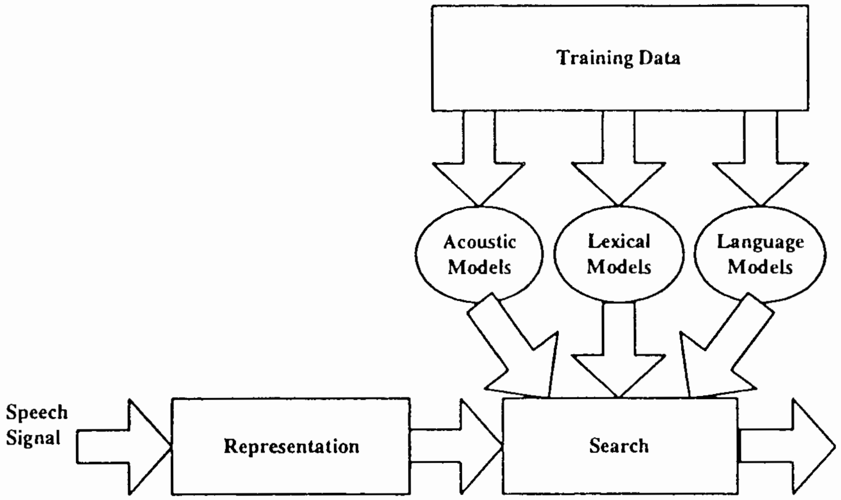


Figure 2.13 Major components in the ASR system.

The acoustic model is the most important and the most widel;/ used of the three models mentioned above. In the acoustic model, different statistical models are employed in the speech recognition systems, e.g., Hidden Markov model (HMM), Dynamic Time-Warping (DTW) model, or Neural-Network model. The HMM is the most popular one in current speech recognition systems. HMMs are powerful statistical methods of characterizing the spectral properties of the frames of a pattern. This model has been used for a variety of speech related applications, e.g., phoneme or word recognition, speaker identification, and language identification. Rabiner and Juang (1986) have written a tutorial paper with a brief introduction about the theory and application of HMMs. Figure 2.14 shows a example for a word "put", represented by an HMM. Each state corresponds to some phonetic event and each event could be skipped. The arrow direction represents the possible translation path of the states. In practice, it is very hard to choose the right number of events to represent the whole word. Therefore, researchers usually choose more states than are needed for the longest word. Then, they apply the same model for all words (Lee and Reddy, 1989). However, for a large vocabulary size, this model is not practical because of the amount of training data and memory storage tht is needed. Thus, a phone or phoneme model is used. Figure 2.15 shows an example of using HMM to represent three states of a phoneme: the transition into the phoneme, the steady state portion, and the transition out of the phoneme. Using this model as in Figure 2.15 can reduce the training size of the vocabulary from a large set of words to a smaller number of set of words which consist all of the phonemes. This model has been used widely in large

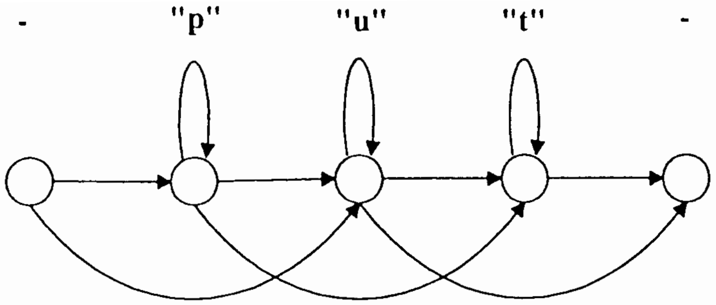


Figure 2.14 An HMM representing "put".

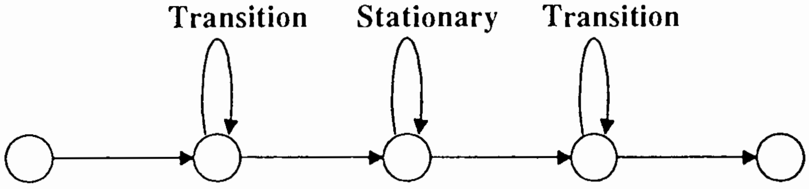


Figure 2.15 An HMM for a phoneme, from Lee and Reddy (1989).

vocabulary size speech recognition systems such as Dragon Dictate System and Power Secretary System.

For an isolated-word speech recognition system, the HMM training procedure can be implemented by using Baum-Welch reestimation formulas (also called forward-backward algorithm). The first step is to collect many examples for each word in the vocabulary. The actual utterances are represented as speech patterns, based on their spectrum analysis, and are stored as templates. Then, the second step is to train the HMM for each word from all of the examples. In the training procedure, Baum-Welch reestimation formulas are used to predict the physical model of the corresponding word. After the training models are obtained, the input test word can be scored against each of the models. A forward or backward algorithm is applied in the scoring procedure. In the recognition step, assuming only the acoustic model is used, the training model with the highest probability is chosen as the recognized word. In some of the recognition systems, the users can define the "threshold" value to specify the minimum recognition confidence score below which the system should reject an utterance, even though this utterance has the highest score. For a continuous speech recognition systems, the procedure is similar to the isolated-word system, but it needs to include more consideration, such as lexical and language models.

2.4.3 Application of the Speech Recognition

From the brief explanation given above, the performance of a recognition system will highly depend on the consistency of the training and test tokens (utterances). If the training and test data are highly consistent, the recognition accuracy for the test utterances would be higher than it would be if the data were not consistent. Because of this property, speech recognizers can be used not only as a dictation tool but also as an assessment tool to diagnose the consistency of a speaker's utterances. In this project, Dragon Writer-1000 is chosen as the commercial speech recognizer. This unit is an isolated-word recognizer, middle vocabulary size (up to 1000 words), speaker-dependent (training before use), and uses HMM as the recognition model.

Raghavendra et al. (1994), Sy and Horowitz (1993), Goodenough-Trepagnier et al. (1992), Coleman and Meyers (1991), Schmitt and Tobias (1986), and Rodman (1985) have studied the possibility of using speech recognizer systems as an input channel for individuals with disabilities, especially focusing on individuals with dysarthria. However, their research has found that persons with dysarthria have some difficulty in using the speech recognition system since their speech has malfunctions. Therefore, for people with dysarthria using the speech input channel, finding a way to improve their speech recognition accuracy is an important issue.

If a speech recognition system can recognize dysarthric speech with sufficient accuracy for the aspects of speech production that are consistent, then it could be used as an interface to type messages or to send signals to a speech synthesizer, which would then translate the difficult-to-understand speech into a more intelligible speech output. By providing voice entry to executive work stations for individuals who are either unwilling or incapable of using keyboards effectively, a variety of services become available instantly.

Chapter 3 Analysis Methods and Procedure

3.1 Experimental Procedure and Rationale

The procedure which forms the basic concept for this project can be represented using the flow chart of Figure 3.1. The flow chart is comprised of twc components. Part 1 includes the study of the characteristics of the speech for each speaker and selection of an initial word list for use with the recognizer. Part 2 is a test of the speech recognizer performance with the initial word list and revision of that list.

In Part 1, the characteristics of the subject's speech, including the accuracy of pronunciation, the pattern of errors, and speech consistency, are obtained from the intelligibility test (I.T.), transcription test (T.T.), recognition test (R.T.), and acoustic analysis (A.A.), as described later. The 70-word list adapted from Kent et al. (1989) is used as the diagnostic word list. Every subject recites the 70-word diagnostic list once, and these words are used as input for the I.T. and T.T. (to assess the accuracy of speech production). In addition, ten tokens of each word in this list are spoken as input for the speech recognizer for R.T. (to analyze the speech consistency). Analysis of the speech characteristics (including I.T., T.T., R.T., and A.A.) and assessment of user preference for particular words are used to select a potential computer command vocabulary, e.g., 50 words, from a standard word list of up to 1000 words. Particular attention is paid in acoustic analysis to interpreting the articulatory errors, unstable events and the pathological features of a subject's speech which make it difficult for a recognizer to tell words apart or recognize a word for which it has a previously trained model. Presumably, lists of words which avoid these problems will achieve high recognition accuracy. From all of this information, some of the speech characteristics and impairments for each subject can be determined and used as a basis for choosing the final word list. These speech characteristics can help to guide efficiently the initial selection of the word list for the speech recognizer. Moreover, other acoustic information and the user's preferred words or other nonvocal sounds can be included in the test vocabulary.

Part 2 is the test of the performance of the recognizer. The purpose of this test is to determine the performance or accuracy of the particular commercially available speech recognizer under consideration, e.g., Dragon Writer-1000, when the subject uses words from the draft short list. This list, derived in Part 1, may be modified with the goal of improving recognition performance. Since all of the analyses and tests mentioned in Part

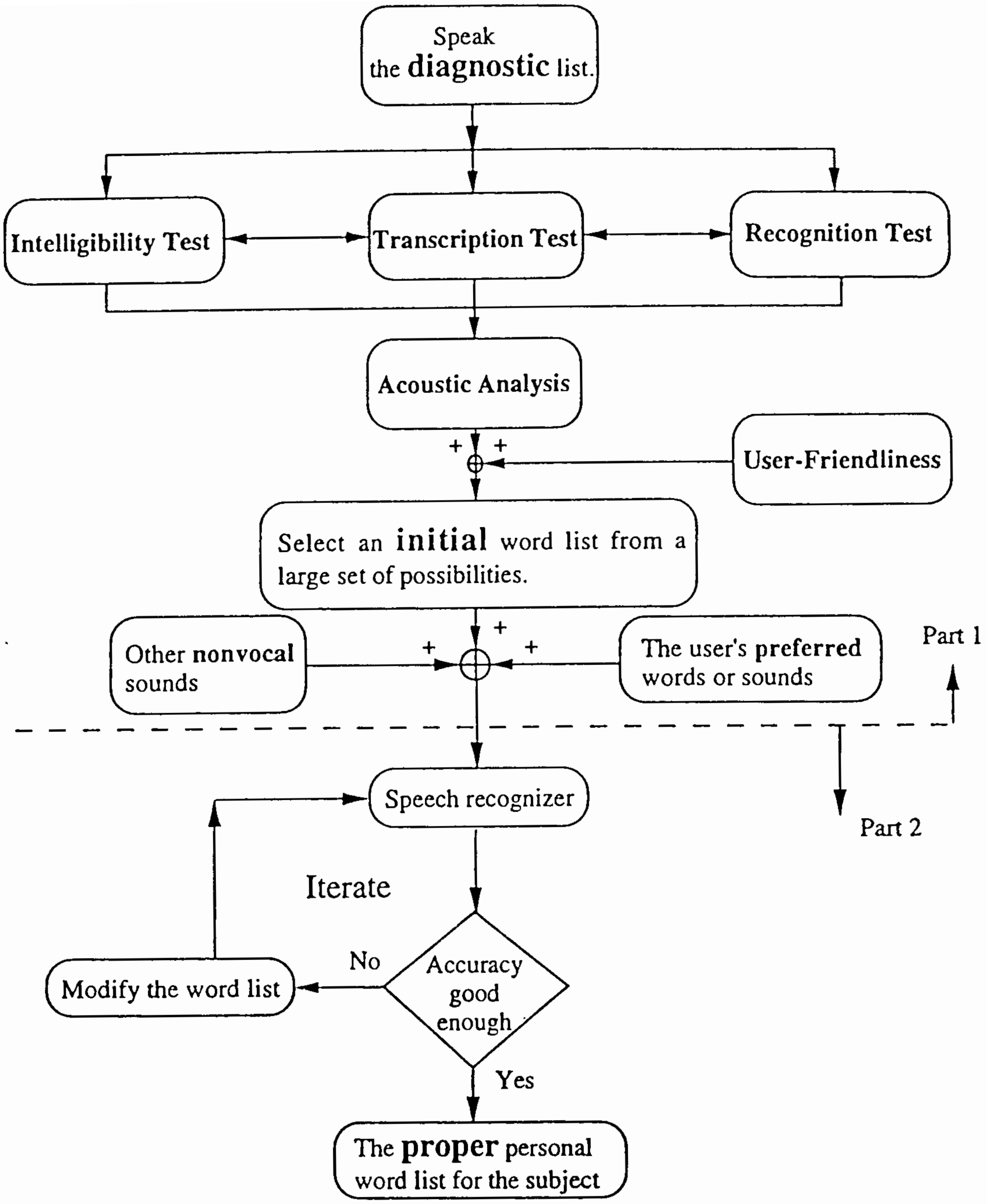


Figure 3.1 Flow chart showing the process by which evaluation of a speaker's utterances leads to selection of words to use for computer control. In Part 1, the characteristics of the subject's speech are obtained from the intelligibility test, transcription test, recognition test, and acoustic analysis. In Part 2, the recognition performance for the initial word list, which is drafted from Part 1, is tested. Optimal performance will be achieved after several iterations in which the initial word list is modified.

1 still lack enough data to cover all of the phonetic problems for each dysarthric speaker, the initial word list needs to be modified as a result of actual recognition tests. Optimal performance will be arrived at after several iterations in which the word list is modified. During each iteration, words which are highly confusable or still give articulation difficulty to these subjects will be replaced with other easily distinguishable words by using the speech characteristics found in Part 1. However, when a word is modified, the consistency of pronunciation, the distinctiveness of the words, and ease of computer use all need to be considered simultaneously and carefully.

3.2 Speech Recording and Equipment

Speech recordings for use in all tests are made in a quiet room with two microphones. The first microphone, a headset microphone, is placed at a constant mouth-to-microphone distance of 2 cm to sample the speech signal for the speech recognizer; the second microphone, an omnidirectional microphone, records the speech signal on the tape recorder from a distance of 10 cm. Most of the dysarthric subjects require straps tied on their heads to hold the headset microphone on and maintain an appropriate distance. The strap is adjusted to firmly hold the headset microphone but not to interfere with head movement for speech

For the intelligibility test (I.T.), only the speech on the tape recorder is used. Every subject practices several words which are written on index cards and are adapted from a 70-word list published by Kent et al. (1989). The word list is given in the first column of Appendix 3.1. Subjects can rehearse these words until they pronounce them as naturally as possible. Every word is pronounced only one time for recording on the tape recorder. These words are presented to the subject in random order. Each word is used to analyze the pattern of phonetic errors (Kent et al., 1989) for each individual speaker.

For the recognition test (R.T.), the Dragon Writer-1000, whose software development library provides tools that allow for the creation of models from training tokens and for the recognition test for each word, is used in this project. This device has been chosen as the commercial speech recognizer for the IBM PC. Before the actual R.T., every subject can first practice several words from the 70-word list until he or she pronounces them as they were in the I.T. Each word is displayed on the computer monitor in a random sequence. The utterance is digitized into the speech recognizer and is also recorded on the tape recorder. Both microphones are used to record the subjects' utterances. The gain of the headset microphone is automatically adjusted and the speaker is informed if the voice is not loud enough or too loud to be digitized. Each subject's utterance is counted as one token. In the actual recordings, the subjects can repeat or skip the words which give them much difficulty in pronouncing. After the subjects follow the instruction of the computer and finish pronouncing all of the 70 words, another repetition is started, but with the 70-word list presented in a different random sequence. In total, ten repetitions are used in this data collection, i.e., ten tokens are obtained for each word. These tokens are subsequently used for computer training and testing. Five of the ten tokens are used to build up the language models for the recognizer and five of them are used to test the recognition accuracy.

Two of the eight subjects, CH and GG, have a text reading problem, because of dyslexia. Therefore, in the data collection, these two subjects need one more person who can read the words shown on the cards (for the intelligibility test) or a computer monitor (for the recognition test). In addition, they need a earphone headset which is used to bring the reader's speech to their ears. Although these two subjects with dyslexia pronounce the words by following the reader's speech, they try to pronounce the utterances as they do usually.

3.3 Analysis Methods and Rationale

The VAX computer system in the Speech Laboratory at MIT is used to digitize the recorded speech and conduct the acoustic analysis. The sampling rate is chosen as 16 kHz with 16-bit A/D board and 7.5 kHz low-pass filter. The intelligibility test (I.T.) is used to study the general speech impairments, including the consonants in word-initial and -final positions and the vowels. However, the transcription test (T.T.) and recognition test (R.T.) specially focus on the consonants in word-initial position. All of the subjects perform with a greater or lesser degree of pronunciation difficulty or abnormality and most of the subjects have particular difficulty with the consonants in word-initial position. Platt et al. (1980b) showed that many consonant errors happen in word-initial position for the 50 cerebral palsy subjects that were studied. Byrne (1959) and Hixon and Hardy (1964) also mentioned the articulatory accuracy problems for the voiceless consonants in word-initial position. Further, Coleman and Meyers (1991) showed that vowels in an h-d environment and in words defined as easy and difficult for speakers with dysarthria have significantly higher accuracy by the speech recognizer than do consonants followed by a neutral vowel. Furthermore, the vowels and final consonants appear to be pronounced more easily than the word-initial consonants after the consonant in word-initial position is released. There are often many pre-voicing noises, heavy breathing noises, or saliva noises especially before the word-initial consonants. All of these factors will cause the speech recognizer to become confused and consider these extra noises as input commands or utterances. Therefore, the data from word-initial consonants provide important information that is a basis for improving the speech recognition accuracy.

3.3.1 Intelligibility Test (I.T.)

The I.T. is a test with closed response sets, designed by Kent et al. (1989) at the University of Wisconsin at Madison. This test has the following advantages: (i) the phonetic contrasts are selected to be sensitive to the articulatory problems experienced by speakers with dysarthria, (ii) the results can be correlated with attributes derived from acoustic analysis, and (iii) the procedures of data analysis and collection can be standardized and written as a package of computer programs. Therefore, this special closed test method was chosen as the main tool in assessing intelligibility of the subjects. This test utilizes a 70-word list adapted from Kent et al. (1989) (the first column of Appendix 3.1). In Appendix 3.1, each row includes a target item and three alternatives that differ from the target in one, or occasionally two, phonetic features for the listeners. The purpose of this test is to assess the adequacy of production (or articulatory position accuracy) of different classes of speech sounds by dysarthric individuals. In this test, five judges who are native English speakers and not familiar with dysarthric speech listened to the recordings from the dysarthric speakers. One test from each of the eight speakers with dysarthria is presented to the judges. During the test (I.T.), judges are given a reference list of four possible response words for each item. These four words are presented in the same form as Appendix 3.1 except the sequence for each row is randomized. The judges mark the word in the four-word group that is most similar to what they hear.

The overall accuracy of I.T. is counted as the total number of correct responses across the five judges divided by the total number of the 70-word list tries for the same five judges, 350 (= 5 * 70). The percentage error for a particular type of contrast is calculated as the total number of misarticulations for that contrast across five judges divided by the total number possibilities for that contrast in the entire list of items. For example, Table 3.1 shows that the number of opportunities for a Front - Back Vowel type of error is 11. Then, the total number possibilities for the Front - Back Vowel type of error across five judges is 55, (= 11 * 5).

From Kent's paper, nineteen phonetic contrast groups, listed in Table 3.1, are derived from this list to study the phonetic errors that speakers might make. For example, the four-choose-one list for the target word "bad" is "bad", "bed", "bat", and "pad". The corresponding phonetic errors in Table 3.1 are defined as follows: (i) if the judges choose "bed" as the actual response, it is scored as a vowel duration error; (ii) if the judges choose "bat" as the actual response, it is considered as a final voicing contrast error; (iii) if the judges choose "pad" as the actual response, it is considered as an initial voicing contrast error. The detailed word pairs for the 19 phonetic contrasts are shown in Appendix 3.2. The number of opportunities for each type of error is listed after the corresponding item in Table 3.1. Because judges served in this test, each number of opportunities in Table 3.1 should be multiplied by five to give the total number of tries for the corresponding error contrast. The data will be presented as the mean error percentage of all tries of a particular contrast. The phonetic contrasts were selected to be sensitive to the articulatory problems experienced by speakers with dysarthria and also had the advantage that there were reasonably well-defined correlations with acoustic properties. Kent et al. argued that the acoustic level of analysis is crucial because the test not only indexes the severity of the dysarthria but may also serve to identify the locus of the intelligibility deficit.

For purposes of this study, the 19 contrasts are further divided into 40 contrasts, as given in Table 3.2. The 40 phonetic contrast groups focus not only on the articulatory confusion problems but also on the confusion error causality, e.g., voiced consonants confused with voiceless consonants and voiceless consonants confused with voiced consonants are counted as two different types of errors. Moreover, some error types are split into subcategories. For example, the 7th type of error in Table 3.1, Consonant Place, is split into three subgroups in Table 3.2: Alveolar - Other Consonant Place, Velar - Other Consonant Place, and Labial - Other Consonant Place. These new error groups can support more detailed information about the consonant confusion errors at alveolar, velar, and labial positions than the original group in Table 3.1 which only shows the rate of the consonant confusion errors. Thus, the 40 phonetic contrast groups can show more detailed impairment information than the original 19. By using the same example as before, the four-choose-one list for the target word "bad" is "bad", "bed", "bat", and

Table 3.1: Nineteen phonetic contrast groups adapted from Kent et al (1989). Items 1 to 19 describe the acoustic - phonetic error contrasts. These pairs show the target and actual response that were scored as an error. The number of opportunities for each type of error in the 70-word list is given after the corresponding item.

1. Front - Back Vowel	(11)	11. Stop - Affricate	(6)
2. High - Low Vowel	(12)	12. Stop - Nasal Consonant	(10)
3. Vowel Duration	(11)	13. Initial Glottal Consonant - Null	(11)
4. Initial Voicing Contrast	(9)	14. Initial Consonant - Null	(14)
5. Final Voicing Contrast	(11)	15. Final Consonant - Null	(9)
6. Alveolar - Palatal Place	(8)	16. Initial Cluster - Singleton	(12)
7. Consonant Place	(10)	17. Final Cluster - Singleton	(12)
8. Other Fricative Place	(17)	18. / r / - / l /	(10)
9. Fricative - Affricate	(10)	19. / r / - / w /	(8)
10. Stop - Fricative	(19)		

Table 3.2: Forty contrast groups. Items 1 to 40 describe the acoustic - phonetic error contrasts. These pairs show the target (before "-") and actual response (after "-") that were scored as an error. The number of items on which each type of error is based is listed after the corresponding item.

1. Front - Back Vowel	(8)	21. Affricate - Fricative Consonant	(6)
2. Back - Front Vowel	(3)	22. Glottal Fricative - Affricate Consonant	(1)
3. High - Low Vowel	(11)	23. Fricative - Stop Consonant	(19)
4. Low - High Vowel	(1)	24. Affricate - Stop Consonant	(6)
5. Long - Short Vowel Duration	(8)	25. Stop - Nasal Consonant	(8)
6. Short - Long Vowel Duration	(3)	26. Nasal - Stop Consonant	(2)
7. Init. Voiced - Voiceless Contrast	(3)	27. Initial Glottal Consonant - Null	(8)
8. Init. Voiceless - Voiced Contrast	(6)	28. Initial Null - Glottal Consonant	(3)
9. Final Voiced - Voiceless Contrast	(4)	29. Initial Consonant - Null	(9)
10. Final Voiceless - Voiced Contrast	(7)	30. Initial Null - Consonant	(5)
11. Alveolar - Palatal Place	(6)	31. Final Consonant - Null	(8)
12. Palatal - Alveolar Place	(2)	32. Final Null - Consonant	(1)
13. Alveolar - Other Consonant Place	(2)	33. Initial Cluster - Singleton	(12)
14. Velar - Other Consonant Place	(4)	34. Final Cluster - Singleton	(3)
15. Labial - Other Consonant Place	(4)	35. Final Singleton - Cluster	(9)
16. Alveolar - Other Fricative Place	(6)	36. / r / - / l /	(7)
17. Palatal - Other Fricative Place	(1)	37. / l / - / r /	(2)
18. Labial - Other Fricative Place	(2)	38. / w / - / l /	(1)
19. Glottal - Other Fricative Place	(8)	39. / r / - / w /	(6)
20. Fricative - Affricate Consonant	(3)	40. / w / - / r /	(2)

"pad". The corresponding phonetic errors in Table 3.2 are defined as follows: (i) if the judges choose "bed" as the actual response, it is scored as a long vowel confusion with short vowel error. (ii) if the judges choose "bat" as the actual response, it is considered as a final voiced consonant confusion with voicing consonant error. (iii) if the judges choose "pad" as the actual response, it is considered as an initial voiced consonant confusion with voiceless consonant error. The detailed word pairs for the 40 phonetic contrasts are shown in Appendix 3.3. The data will be presented in the same format as the 19 contrast groups.

These classifications in Tables 3.1 and 3.2 help to determine the possible functional problems experienced by each dysarthric speaker in an efficient way, although these error groups may be too fine grained and based on too little data (e.g., the Low - High Vowel contrast in Table 3.2 has only one error pair in the test). Since this intelligibility test method is a standard routine for all speakers, the whole assessment steps have been written in a C-language program on the IBM PC to operate automatically. The computer output files can show the phonetic error distribution and the confusion pairs for the corresponding subject's speech. This labor-saving program can save much time in the data analysis. In addition to understanding the speech characteristics for each subject's speech, Chi-Square tests for each subject's I.T. have also been made for studying the inter-judge consistency.

Figure 3.2 is an example of a display of the phonetic errors in terms of 19 and 40 phonetic contrast groups for one of the subjects, JF. The 19 (or 40) contrasts listed in Table 3.1 (or 3.2) are ordered along the x-axis in Figure 3.2-a (or 3.2-b), and the height of each bar indicates the mean error percentage of that specific error contrast. Figure 3.2-a shows that this specific subject's two most frequent errors are vowel duration (11%) and Final Cluster - Singleton contrast (8%). Figure 3.2-b shows further information about this subject's speech problems: the two most frequent errors are final clusters confusion with singleton consonants in word-final position (33%) and long vowel confusions with the short vowels (15%). The second figure demonstrates not only the main types of errors but also indicates the cause of the errors. Therefore, the 40 phonetic contrast groups can show more information for the specific subject's speech than the 19 groups. Detailed discussion of these data for each subject is given in Chapter 4.

The format for presenting I.T. results is illustrated for one speaker in Tables 3.3, 3.4, and 3.5. I.T. in Table 3.3 shows the percentage accuracy for the intelligibility test

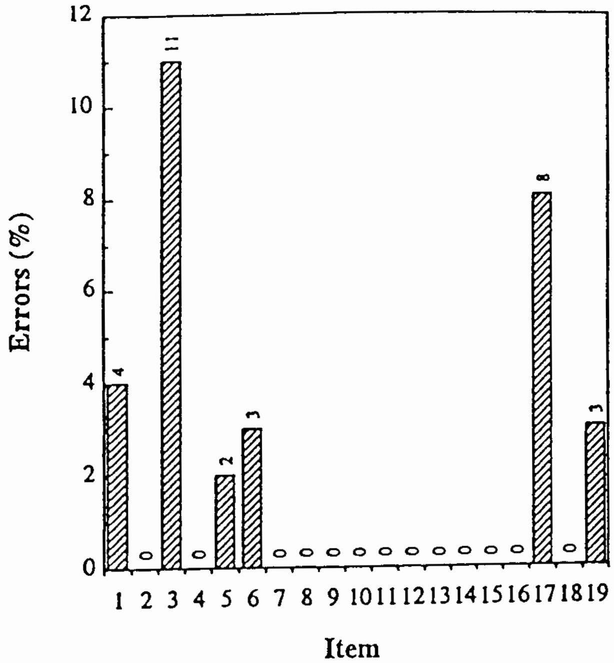


Figure 3.2-a Format of the intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution in this example shows the two most frequent errors are vowel duration (11%) and Final Cluster - Singleton (8%) for one of the eight subjects, JF

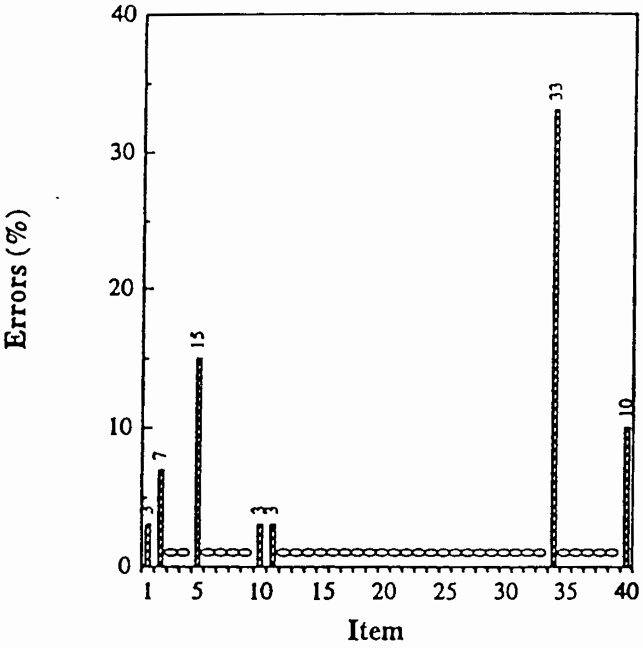


Figure 3.2-b Format of the intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution in this example shows the top four most frequent errors are Final Cluster - Singleton (33%), Long - Short Vowel Duration (15%), /w/ - /r/ (10%), and Back - Front Vowel (7%) for one of the eight subjects, JF.

Table 3.3 Example of the results of the intelligibility test (I.T.), recognition test (R.T.), and acoustic analysis (A.A.) for one of the eight subjects. Results are given for the intelligibility test (I.T.(%)), Chi-Square test score of I.T. (I.T.-X2), Chi-Square test score of T.T. (T.T.-X2), recognition test (R.T.(%)) for the 70-word diagnostic list, recognition test (R.T.-35(%)) for 35 words which are randomly picked from the 70-word list, average duration of words in the 70-word diagnostic list relative to the durations for a normal subject (A.D.), recognition rejection rate (Rej.(%)) for the 70-word diagnostic list, the recognition accuracy for the word list based on the 52-new word list (New_L(%)), the recognition accuracy for 35 words which are chosen from the 52-new word list (New_L-35(%)), the final recognition improvement based on the 70-word diagnostic list (Improve(%)), and the recognition improvement based on the 35-word base (Improve-35(%)).

Subj. \ Status	I.T.(%)	I.T.-X2	T.T.-X2	R.T.(%)	R.T.-35(%)	A.D.	Rej.(%)	New_L(%)	New_L-35 (%)	Improve (%)	Improve-35 (%)
JF	95	3.5	0.37	73	79	1.8	13	82	85	9	6

Table 3.4: Format used for comparison of the I.T. and the R.T. for one of the eight subjects. Row I.T. gives the results of I.T. Row R.T. gives the results of R.T. The different columns give the results of the I.T. and the R.T. with all of the 70 words (All), words with an initial sonorant (SONOR_IN), without any obstruent at all (WTOT_OB), obstruent (OBS_IN), stop consonant (STOP_IN), clusters (CLUS_IN), fricative / affricate (FR_AF_IN), /f/ (/f/_IN), /s/ (/s/_IN), /sh/ (/sh/_IN), and /ch/ (/ch/_IN) in word-initial position.

Sub/Ac(%)		All	SONOR_IN	WTOT_OB	OBS_IN	STOP_IN	CLUS_IN	FR_AF_IN	/f/_IN	/s/_IN	/sh/_IN	/ch/_IN
JF	I.T.	95	97	100	94	95	83	93	95	90	93	100
	R.T	73	76	91	74	76	70	73	80	69	73	80

Table 3.5: Comparison format of the I.T. and the R.T. for the STOP_IN group for one of the eight subjects. Row I.T. gives the results of I.T. Row R.T. gives the results of R.T. The different columns give the results for the I.T. and the R.T. with all of the 70 words (All), all of the stop consonants (STOP_IN), and labial (LABIAL), alveolar (ALVEOLAR), and velar (VELAR) stop consonants in word-initial position.

Sub \ Ac (%)		ALL	STOP_IN	LABIAL	ALVEOLAR	VELAR
JF	I.T.	95	95	90	100	100
	R.T	73	76	70	93	75

(with a reference list of four words). Table 3.3 also shows the Chi-Square test score of I.T. (based on df (degree of freedom) = 4 and α = 0.05, where the df is equal to (column number, i.e., five judges, - 1) * (row number, i.e., accuracy and error rate, - 1) and α is equal to (1 - confidence interval, i.e., 95%), refer to Sincich (1987) [p. 714]. A detailed breakdown of the I.T. is shown in Tables 3.4 and 3.5. The categories selected for more detailed examination in Tables 3.4 and 3.5 are based in part on analysis of the data across subjects, in order to highlight the features that show both high and low error rates. In the format of Table 3.4, all of the 70 diagnostic words with consonants in word-initial position are divided into two groups: (i) the words with sonorants in word-initial position (SONOR_IN) and (ii) the words with obstruents in word-initial position (OBS_IN). Additionally, the words without any obstruent consonant at all (WTOT_OB), which represent a special case of the words with sonorants or vowels in word-initial position, e.g., "air" and "row", are included in the comparison of SONOR_IN and WTOT_OB. The group, OBS_IN, is split again into two sub-groups: (i) the words with stop consonants in word-initial position (STOP_IN) and (ii) the words with fricative / affricate consonants in word-initial position (FR_AF_IN). In addition, the words with initial clusters (CLUS_IN) represent a special case of a two-consonant unit with stop or fricative consonants in the initial position, e.g., "bl", "sl", "sp", and "st". A further study of the FR_AF_IN is performed in the following groups: (i) the words with /f/ consonant in word-initial position (/f/_IN), (ii) the words with /s/ consonant in word-initial position (/s/_IN), (iii) the words with /sh/ consonant in word-initial position (/sh/_IN), and (iv) the words with /ch/ consonant in word-initial position (/ch/_IN). All of the I.T. data are compared with the recognition test data (R.T.) in Table 3.4. The format in Table 3.5 shows the detailed study of the STOP_IN group. The STOP_IN group is split into LABIAL, ALVEOLAR, and VELAR sub-groups, corresponding to the performance of the words with labial, alveolar, and velar stop consonants in word-initial position. These sub-groups show the detailed I.T. and R.T. accuracy for words with different types of stop consonants in word-initial position. The word list for each group mentioned above is given in Appendix 3.4.

Because the I.T. can not cover all of the phonetic problems for each dysarthric speaker and judges are limited to choose one in the four-choosing-one reference list, a further detailed test (a transcription test) is introduced as a follow-up study of the detailed phonetic misarticulation problems. The transcription test can be used to study in depth the important impaired features (e.g., consonants in word-initial position) that are found in the

I.T. The detailed I.T. data interpretations are described in each subject's I.T. section of Chapter 4. The final summary across all eight subjects will be discussed in Chapter 5.

3.3.2 Transcription Test (T.T.)

The same speakers' utterances are used in this test. Two judges, experienced in transcription, transcribe the dysarthric speech. The judges are allowed to replay the token as often as they like. The purpose of this test is to obtain a detailed transcription of the sounds in words produced by each dysarthric individual. The consonants are considered only in word-initial position for the present analysis. The data are presented in the form of confusion matrices. From this test, detailed estimates are obtained of front - back vowel accuracy, obstruent and sonorant accuracy in word-initial position, voiced-voiceless accuracy for consonants in word-initial position, and place accuracy for obstruents in word-initial position. Table 3.6 is an example of the format that is used for displaying these data. The response percentage for each feature (or feature group) is expressed as a percentage of the total number of stimuli with the corresponding segment type by the specific dysarthric speaker. The total number of the tries for type is listed beside (or under) the corresponding item in Table 3.6. The corresponding word list for each type is presented in Appendix 3.5.

The obstruent and sonorant scores at the top of the table are the percentage of obstruent and sonorant consonants that were identified correctly in every feature. In the confusion matrices of Table 3.6, the score represents the percentage of times that a particular feature was identified. For the vowel matrix, only front, back, and diphthong vowels are presented. The vowel targets (T.) are listed in the left column of the vowel matrix. The vowel responses (R.) could be front, back, middle, diphthong, or disappearance ([]) (as indicated at the top of the matrix). For the voiced-voiceless matrix, both obstruent and sonorant consonants in word-initial position are included. The "+" and "-" signs represent the corresponding voiced and voiceless consonants. The [] item includes the disappearance of the corresponding phonetic response. For the matrix representing place for obstruents in word-initial position, there are four targets (labial, alveolar, palatal, and velar) and six responses (labial, dental, alveolar, palatal, velar, and []. The [] item includes the disappearance of the corresponding phonetic response or other responses not included in any of the six response items mentioned above, e.g., glottal and uvular fricative consonants.

Obstruent Accuracy:(76) 95%

Sonorant Accuracy: (54) 81%

Vowels

T. \ R.	Front	Back	Middle	Diph.	[]
Front (94)	100%				
Back (34)		100%			
Diph. (12)				100%	

Voicing

T. \ R.	+	-	[ ]
+ (52)	100%		
- (78)	1%	99%	

Place for Obstruents

T. \ R.	Lab.	Dent.	Alv.	Pal.	Vel.	[ ]
Lab. (24)	100%					
Alv. (32)			97%	3%		
Pal. (12)			8%	92%		
Vel. (8)					100%	

Table 3.6 Format for summarizing the transcription data for vowels, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position for one of the eight subjects, JF. The accuracies of obstruent and sonorant at the top are percent correct identification of the obstruent and sonorant consonants in word-initial position. The left column of each matrix represents the target items and the response items are listed across the top. The number listed under or beside each item is the total number of trials. Since two judges served in T.T., all of the trial numbers have been multiplied by two.

In addition to understanding the speech characteristics for each subject's speech, Chi-Square tests for each subject's T.T. have also been made for studying the inter-judge consistency. The Chi-Square test values indicate the degree of consistency of the inter-judge responses for T.T. Table 3.3 shows Chi-Square test score of T.T. for one subject (based on df = 4 and $\alpha = 0.05$).

The advantage of the T.T. is that the judge need not be restricted by choosing only one item from the reference list. During this test, the judges are allowed to write down exactly what they hear. Thus, the transcription data might show more phonetically detailed descriptions for each of the speakers with dysarthria than are captured by the I.T. The I.T. is limited by the fact that one item must be selected from the four words in the reference list and the exact phonetics heard by the listeners cannot always be represented. From the study of dysarthric speech, these speakers sometimes misarticulate completely and pronounce utterances that are very different from the normal ones. Therefore, a detailed phonetic analysis becomes necessary and important. Additionally, the response of I.T. sometimes can not reflect the real articulatory mistakes, e.g., if the speaker articulates an abnormal /sh/ as /x/ for the word "shoot" but pronounces correct /u/ and /t/, the judges would probably choose "suit" as the responses when the reference list is "shoot", "suit", "sheet", and "shot". Since there is no word that fits the /x/ in word-initial position, the judges might choose the closest answer. However, this kind of error in the I.T. would count as palatal - alveolar confusion error, but it would count as Palatal - Velar type of error in T.T. because /x/ is a velar fricative consonant. As a result, these two kinds of analysis can lead to different conclusions. Although T.T. sometimes appears to give more accurate information than the I.T., the disadvantage of T.T. is that the judges need to be familiar with transcription and it is more time-consuming to analyze the data than the I.T. It also is hard to score the transcriptions automatically by the computer since there is not a closed answer and format for T.T. From the clinical view, the T.T. is an accurate method but may not be an efficient one. Therefore, both I.T. and T.T. have advantages and disadvantages. In this project, I.T. is used as a quick tool to obtain the main malfunctions for the specific subject with dysarthria and, then, T.T. is applied to study in depth the main impaired features that are found in I.T. The detailed T.T. data and interpretation are reported in each subject's T.T. section of Chapter 4. The final summary across all eight subjects will be discussed in Chapter 5.

3.3.3 Recognition Test (R.T.)

In the I.T. and T.T., the accuracy of pronunciation and the pattern of errors are explored, but not the consistency with which a particular word is pronounced because (i) only one utterance for each word was collected in I.T. and T.T.; (ii) human perception is hard to distinguish the utterances with only a small amount of duration difference. There is both anecdotal and research evidence to indicate that some consistency exists in the speech production of speakers with dysarthria (Neilson and O'Dwyer, 1984). If speakers with dysarthria are sufficiently consistent and have enough variation in their speech so that they can articulate a number of sounds which can be differentiated, then the speech recognizer system may be useful to them (Coleman and Meyers, 1991). The recognition test (R.T.) using a speech recognizer supplies information about the subject's speech consistency. This test measures the variability in a particularly relevant way, i.e., by determining how often an utterance falls within the "boundary" defined by a language model derived from five earlier utterances of the same word by the same speaker.

Before the actual R.T. data collection, each subject first practices several words of the 70-word list, which are displayed on the computer monitor, until they have become familiar with the instruction of the computer. The subjects can choose an adequate font size for the target word shown on the computer monitor in order to reduce visual errors. In the actual recordings, the subjects can repeat or skip words which give them great difficulty in pronunciation. The tester can have the subject redo an item if some extraneous noise interfered such as cough or environmental noise. Ten tokens of each word are used as the input of the Dragon Writer-1000 and the tape recorder. These ten tokens are collected in two or three different sessions which are separated by at least one or two weeks in order to take into account possible variations in the subject's speech patterns over time. Then, these ten tokens are mixed together by choosing three tokens from the first collection session and two tokens from the second collection session in order to form the first five-token set. The remaining five tokens form the second five-token set. The first five-token set of each word is used to form a language model and the second five-token set is used for model testing. The tolerance for model testing defined in the Dragon Writer-1000 is set at 25. The tolerance is specified as a number between 1 and 100. A value of 100 makes the Dragon Writer-1000 very tolerant of variability in the subject's speech, but also makes it more prone to mistakenly recognize extraneous noise as real commands. At the other extreme, a value of 1 makes the Dragon Writer-1000 reject almost all noise, but also reject most valid utterances. The normal tolerance level for the

Dragon Writer-1000 is 50. If the speech recognizer frequently hears noises, the tolerance can be reduced to the range of 20 or 30. Since the subjects with dysarthria usually produce a lot of noises, e.g., saliva noise, deep breathing noise, and the wheelchair noise coming from the involuntary body movements, the recognizers often misrecognized these noises as input commands. Therefore, in order to reduce misrecognition of these noises, the tolerance is reduced to 25, which is set from the practical test experience for the users with dysarthria in our laboratory. Furthermore, when the computer makes an error in recognition, it either does not recognize the item at all or incorrectly recognizes it as another item. The correct and incorrect recognitions are discussed in later chapters.

The recognition score for each subject is listed in the format of Table 3.3. The average recognition accuracy for the 70-word diagnostic list, R.T.(%), is the percentage recognized correctly. It reveals the consistency of pronunciation of the words for the specific subject. The average recognition accuracy for the 35-word list, R.T.-35(%), which is randomly picked from the 70-word diagnostic list, is also the percentage recognized correctly on the base of 35 words from the 70-word list. The purpose of getting R.T.-35(%) is mentioned in Section 3.4. The Rej.(%) item in Table 3.3 shows the speech recognition rejection rate, i.e. the percentage of the words which are unrecognized by the language models built up previously. Tables 3.4 and 3.5 illustrate how individual consonant groups perform in the R.T. The groups are the same as those used in analyzing the I.T. When the results of the R.T. are combined with the results of the previous two tests, I.T. and T.T., a more complete picture of the speech characteristics for each individual can be obtained.

The detailed study of the R.T. performance for each subject will be reported in Chapter 4. The overall findings across these eight subjects are discussed in Chapter 5.

3.3.4 Acoustic Analysis (A.A.)

Perceptually-based assessments, including I.T. and T.T., have limited analytic power for determining which aspects of the speech motor patterns are affected. Perceptual judgments are difficult to standardize over time and across different settings. Moreover, the perceptual rating systems are often difficult to replicate. In addition, although R.T. can supply information about the speech consistency for a specific subject, it is difficult to interpret the results in term of specific production errors. Acoustic analysis is particularly a reliable means of documenting speech changes (Simmons 1982; Kent et al. 1979). Analysis of speech spectrograms or other types of acoustic analysis can help to define the speech characteristics for the specific subject. Examination of the spectrograms and formant contours, spectra, utterance durations, and speech waveforms allow for a more detailed and objective description of the acoustics and speech consistency, and permit an interpretation in terms of the underlying articulatory events. Furthermore, acoustic analysis captures certain acoustic events corresponding to articulatory anomalies which are not found in normal speech and therefore would not be generally detected or described in a typical listening or transcription test. For example, unpredictable extra formants appearing on the spectrograms of the dysarthric speech are very common, as will be shown in Chapter 4 for Subjects JS, JR, and GG. The positions of the extra formants and their variation with time could potentially provide some insight into anomalies in vocal tract shapes that are producing the sounds.

Acoustic analysis can also be used to explore the reasons for conflicting results from the first two tests (the I.T. and T.T.) and could examine the lack consistency of the subjects in producing the words that cause confusion errors in the recognition test. The acoustic analysis provides quantitative evidence for the data derived from the recognition test. Each vowel or consonant has its specific articulatory configuration and spectral characteristics. All phonetic features occurring in language have their roots in acoustic attributes with these characteristics (Stevens, 1972 and 1989). Examination of the time-varying spectra can show the articulatory control variety of a specific utterance. The formant positions reveal the vocal-tract shape and the interaction between the vocal-fold vibrations and the supraglottal system (Stevens, 1972 and 1989). The formant contours show the movements of the articulatory system. Further, analysis of the fundamental frequency shows the ability to control the glottis. Comparison can also be made between the average duration of the 70 words for each subject and the duration for a subject who does not have dysarthria, MJ in this case. Durations of individual speech events within each word can be compared with those of the normal speaker, MJ. The A.D. item in Table 3.3 shows the average duration of the 70 words for each subject's speech compared with the normal subject's (MJ's) speech (478 ms). The A.D. values reflects the subjects' utterance prolongation compared with that of MJ's value. The relationship between the average duration of each subject's 70-words and their recognition rejection rates will be discussed in Chapter 5.

This project will focus especially on acoustic analysis to explore its potential contribution to the study of the dysarthric speech accuracy and consistency by using observations of the spectrogram, spectra, fundamental frequency, and average duration. Additionally, acoustic theories of speech production will be applied to predict the articulatory constriction position from the acoustic data (Stevens, 1989). In this analysis, the normal subject's (MJ's) speech is used as a control to compare with the speech of the dysarthric individuals. Examples of acoustic analysis and its interpretation are given in each subject's A.A. section of Chapter 4. The final discussion of A.A. across all eight subjects will be summarized in Chapter 5.

3.4 Evaluation and Application Tests

A successful new word list needs to be carefully considered for not only the speech characteristics (obtained from the three tests and acoustic analysis mentioned above) for each individual but also the practical computer application and user-preferable interface. This integrated design can help these speakers with dysarthria use the computer with ease and communicate with others by their own speech. It must be easily operated and the words must be readily remembered by these dysarthric users, in addition to achieving consistently acceptable recognition rates. Moreover, the new word list should also be usable with speech recognizers on different commercial computer systems, e.g., IBM PC and Macintosh, since the goal of this project is to derive a method which is generalizable to any speaker and recognizer.

By using the Dragon Writer-1000, the application of speech dictation has been made on the IBM PC edit-package, e.g., PE2 and Lotus 123, by all of the subjects except of GG. In real applications, the Dragon Writer-1000 allows the computer to "talk". A text-to-speech synthesizer, Speech-Plus Prose 4000, can support the Dragon Writer-1000 with a text-to-speech output. If this synthesis capability is available, the subjects can hear and check the speech recognizer responses. On the other hand, these subjects can also use the speech synthesis to translate their difficult-to-understand speech into a more recognizable form. Thus, these speakers with dysarthria can communicate with others by using their own speech. In addition to the Dragon Writer-1000, another commercial speech recognizer, Voice Navigator, used with the Macintosh, is also applied in this project since some of the subjects often use the Macintosh. However, this recognizer, Voice Navigator, is only used to evaluate whether the final word list can be still helpful for the subjects to dictate to the Macintosh and to replace the function of the mouse. This exercise demonstrates that the subjects could use different computer systems and interfaces by means of the new speech input channel.

Because it is important for the computer interface to be matched to the preference of the users, an easy-to-remember alphabet spelling system is one of the basic criteria for designing the new word list. Therefore, a user-preferable alphabet spelling system, such as "alpha" represents "a", and "bravo" is for "b" in the military alphabet spelling system, is a factor in the word design consideration. However, because of the limitations of the impaired speech, the words used in the military spelling system must be re-designed to fit the speech characteristics of each subject. In this special requirement, some of the words in this alphabet spelling system need to use characters which may conflict with the criteria obtained from the Part 1 of Figure 3.1. For example, words beginning with "s" generally have poor recognition performance, but the words with "s" in word-initial position may be preferred to represent the character "s" in the alphabet spelling list (in order to keep the consistency of the alphabet spelling system), e.g., "sit". In designing the word list, although the initial consonants for some words do not fit the speech criteria of the specific subject, attention can be given to the vowels and the consonants that follow so that confusion with other words is minimized in spite of the variation in word-initial consonants. This is one of the most challenging aspects of the design of the new list. If the recognition performance of these specific words still can not be improved after several tries, a word which does not fit the criteria of designing the alphabet spelling system is chosen. For example, "kiss" might be used to represent the character "s" since the word ends with "s".

Furthermore, within the constraints imposed by the ease of computer use and speech consistency issues, the newly designed words or utterances should be different enough so that they are not confused with each other. Thus, when a new word is designed and put into the word list, the consistent manner, phoneme distinction over the whole list, and ease of computer use all need to be considered simultaneously and carefully. Coleman and Meyers (1991) also mentioned, "In order for the signal to be easily recognized, the speaker must do two things: produce signals that are different enough not to be confused among each other; and produce individual signals in a consistent manner each time so they can be matched to the stored templates" [p. 41].

After generating a successful personal word list for the specific subject, including consideration of the speech characteristics for each individual and the user-preference issues, a practical computer input channel based on speech recognition for the dysarthric speaker can be designed and compared with the traditional method which this subject usually uses to control his / her computer. Two sentences, "I am brilliant." and "No one here will run away." (including "."), are used to make a comparison between each subject's dictation and typing speed on the IBM PC edit-package, PE2. The recognition accuracy for each subject's initial word list is listed in the New_L(%) item of Table 3.3. Further, because the vocabulary size of the 52-word list and the 70-word diagnostic list is different, the recognition accuracy for 35 words, New_L-35(%), which are picked from the 52-word list (most are the words with sonorants in the word-initial position and the words without any obstruent), is listed to make a comparison with R.T.-35(%) on the same vocabulary size base (35). The "Improve(%)" item of Table 3.3 shows the percentage of improvement (difference between New_L(%) with R.T.(%)) these subjects have achieved by using the new initial word list. The "Improve-35(%)" item of Table 3.3 shows the percentage of improvement (comparing New_L-35(%) with R.T.-35(%)) these subjects have achieved for basing on the same vocabulary size. The detailed data for each subject will be presented in Chapter 4. The relation between the I.T. and the percentage recognition improvement across all subjects will be discussed in Chapter 5.

Chapter 4 Individual Subjects: Data and Experiments

Nine speakers served as subjects for this study: eight speakers with dysarthria (four males and four females) and one normal speaker (male) as a control. The basic background, history, and types of dysarthria of the impaired speakers are listed in Table 1.1. The speakers with dysarthria range in age from 22 to 61 years. All are native speakers of English and have no apparent cognitive deficits. Six of the eight dysarthric subjects (JF, JS, DW, ES, JR, and GG) have the speech impairment of cerebral palsy (CP), one subject (CH) has CP with surgery damage on her nervous system, and one subject (MW) has the speech impairment of ataxia. The range of education for these subjects are from 5th grade to master's degree. Four subjects (JF, DW, ES, and CH) use their right or left index fingers, one (JR) uses her right thumb, one (GG) uses a pencil grasped by her right or left fingers, and one (JS) uses his nose, to key in the computer. Two of the eight subjects, CH and GG, have dyslexia, a text reading problem. They need an assistant to help their reading.

For purposes of assessment, four analysis tools described in Chapter 3 are applied to analyze the data from the specific dysarthric individuals: (1) intelligibility test (I.T.) (2) transcription test (T.T.) (3) recognition test (R.T.), and (4) acoustic analysis (A.A.). The scores of the I.T. and R.T. for the 70-word list, average duration relative to a subject who does not have dysarthria (A.D.), recognition rejection rate (Rej.(%)), the recognition accuracy for the new word list (New_L(%)), and the final recognition improvement based on the 70-word diagnostic list (Improve(%)) are listed for each subject in Table 4.1. On the other hand, Table 4.1 also shows the recognition accuracy (R.T.-35(%)) based on the 35 words randomly picked from the 70-word list. The table also indicates the recognition accuracy (New_L-35(%)) based on the 35 words chosen from the 52-word list (first choosing the words with sonorants in the word-initial position and the words without any obstruent and then the words with obstruents), and the recognition accuracy improvement (Improve-35(%)) based on the same vocabulary size (35) comparison for the 52-word and 70-word lists. Chi-Square tests for each subject's I.T. have also been made for studying the inter-judge consistency (I.T.-X2). Table 4.2 shows the comparison data of the I.T. and the R.T. in ten subgroups, which have been described in Chapter 3, for all of the eight subjects. Table 4.3 shows the comparison data of the I.T. and the R.T. for four subgroups in the STOP_IN group, which have been mentioned in Chapter 3, for all of the eight subjects. A summary of the salient aspects of the findings and recognition accuracy improvements for eight subjects with dysarthria (JF, MW, JS, DW, ES, JR, CH, and GG)

Table 4.1 Test conclusions for all of the subjects. The scores and results of each test are listed below: the I.T. (I.T.(%)) and Chi-Square test value of I.T. (I.T.-X2) for the 70-word list, Chi-Square test value of T.T. (T.T.-X2) for the 70-word list, the R.T. (R.T.(%)) for the 70-word list, the recognition accuracy R.T.-35(%) for the 35 word list adapted from the 70-word list, average duration relative to a subject who does not have dysarthria (A.D.), recognition rejection rate (Rej.(%)) based on the 70-word diagnostic list, the recognition accuracy for the new word list (New_L(%)), the recognition accuracy for the 35 word list adapted from the 52-word list (New_L-35(%), the final recognition improvement based on the 70-word diagnostic list (Improve(%)), and the final recognition improvement based on the 35 word list comparison between New_L-35(%) and R.T.-35(%). A subject who does not have dysarthria (MJ) is used as a control. "x" means no data available.

Subj. \ Status	I.T.(%)	I.T.-X2	T.T.-X2	R.T.(%)	R.T.-35(%)	A.D.	Rej.(%)	New_L(%)	New_L-35 (%)	Improve (%)	Improve-35 (%)
JF	95	3.5	0.37	73	79	1.8	13	82	85	9	6
MW	82	5.8	0.11	71	79	1.3	14	86	82	15	3
JS	60	1.9	5.55	37	44	1.8	29	75	74	38	30
DW	57	3.6	0.62	30	36	2.5	43	53	57	23	21
ES	97	4.1	0.07	80	88	1.6	10	93	91	13	3
JR	64	13.4	0.34	59	64	1.7	17	73	78	14	14
CH	61	3.2	8.84	26	34	3.0	33	59	61	33	27
GG	89	1.4	0.01	63	x	1.6	17	x	x	x	x
MJ	x	x	x	95	x	1	0.3	99	x	4	x

Table 4.2: List of data from the I.T. and the R.T. for all of the eight subjects.

Row I.T. gives the the results of I.T. Row R.T. gives the results of R.T. Columns All - /ch/_IN give the results for the I.T. and R.T. with all of the 70 words, sonorant, without any obstruent at all, obstruent, stop consonant, fricative / affricate, /f/, /s/, /sh/, /ch/, and cluster in word-initial position.

Subj. \ Acc. (%)		All	SONOR_IN	WTOT_OB	OBS_IN	STOP_IN	CLUS_IN	FR_AF_IN	/f/_IN	/s/_IN	/sh/_IN	/ch/_IN
JF	I.T.	95	97	100	94	95	83	93	95	90	93	100
	R.T.	73	76	91	74	76	70	73	80	69	73	80
MW	I.T.	82	81	90	81	89	73	75	80	65	87	100
	R.T.	71	78	80	67	79	57	59	25	65	80	60
JS	I.T.	60	65	74	54	69	53	43	60	31	47	73
	R.T.	37	39	49	32	37	27	29	30	23	40	40
DW	I.T.	57	74	83	45	65	33	32	80	23	33	7
	R.T.	30	34	37	26	29	23	23	40	20	33	7
ES	I.T.	97	99	97	96	100	97	94	95	98	93	73
	R.T.	80	88	86	72	81	70	65	75	62	67	67
JR	I.T.	64	67	97	62	76	47	53	95	38	27	87
	R.T.	59	74	71	49	65	47	39	45	31	47	60
CH	I.T.	61	66	80	58	64	50	56	55	52	73	47
	R.T.	26	49	60	17	24	23	13	0	15	20	13
GG	I.T.	89	82	89	92	96	100	89	90	94	100	53
	R.T.	63	71	86	55	63	50	50	35	52	40	73

follows. A subject who does not have dysarthria (MJ) will be used as a control. For the purpose of testing the possibility of the new word list used in the commercial edit-package, e.g., PE2 on IBM PC, and making the computer input be user-preferable, an easy-to-remember alphabet spelling system, e.g., "alpha" represents "a" and "bravo" is for "b", is one of the basic criteria for designing the new word list. The dictation speed comparison with the typing speed for all of the eight subjects is listed in Table 4.4. In the following sections, the specific findings for each subject's speech will be discussed. At the end of this chapter, a brief summary will be given for some of the attributes of the data that are common to all of these subjects and are helpful for designing the new word list for each subject.

4.1 Subject 1 (JF)

4.1.1 Background and Symptoms

JF, who is 61 years old, has a high school diploma. His mother bore him at home and had difficulty in childbirth. The doctor paid more attention to saving his mother's life first and devoted less attention to taking care of him. Three days later, when JF's mother gave him a bath, she found that JF moved abnormally. His neuromotor condition is characteristic of spastic cerebral palsy: the muscles are stiff and the movements awkward. The muscles have increased tone with heightened deep tendon reflexes (Dorland's Illustrated Medical Dictionary, 1981). Both his hands and his legs move more inwards than outwards. His neck has involuntary movements. In the typing test, he can only type by using his left index finger with the right hand holding the left. However, JF's speech sounds more normal and less throaty than most of the other subjects, at least to the unfamiliar listener. On the other hand, the intelligibility of his speech is variable. Most of the time, his speech is intelligible, but sometimes it becomes confusing.

4.1.2 Analysis and Discussion

(i) Intelligibility Test (I.T.)

Table 4.1 shows that JF's intelligibility test, I.T., is 95%. The Chi-Square test value of I.T. is 3.5 (< 9.49 based on $df = 4$ and $\alpha = 0.05$), refer to Sincich (1987) [p. 715].

Table 4.3: List of data from the I.T. and the R.T. for the STOP_IN group in all of the eight subjects.

Row I.T. gives the the results of I.T. Row R.T. gives the results of R.T. Columns All - VELAR give the results for the I.T. and R.T. with all of the 70 words, all of the stop consonants, and labial, alveolar, and velar stop consonants in word-initial position.

Subj. \ Acc. (%)		All	STOP_IN	LABIAL	ALVEOLAR	VELAR
JF	**I. T.**	95	95	90	100	100
	R. T.	73	76	70	93	75
MW	**I. T.**	82	89	93	100	75
	R. T.	71	79	68	93	90
JS	**I. T.**	60	69	75	47	75
	R. T.	37	37	38	33	40
DW	**I. T.**	57	65	73	80	40
	R. T.	30	29	35	33	15
ES	**I. T.**	97	100	100	100	100
	R. T.	80	81	70	100	90
JR	**I. T.**	64	76	70	87	80
	R. T.	59	65	68	80	50
CH	**I. T.**	61	64	60	53	80
	R. T.	26	24	28	27	15
GG	**I. T.**	89	96	93	100	100
	R. T.	63	63	53	73	75

Table 4.4: Dictation speed comparison with the typing speed for all of the eight subjects.

Row Typing gives the typing speed in seconds or minutes. Row Speech gives the dictation speed in seconds or minutes. The "x" sign represents this subject has not done the specific test.

(i) Sentence I: "I am brilliant."

Time \ Subject	JF	MW	JS	DW	ES	JR	CH	GG
Typing	11"	27"	1'20"	24"	31"	33"	1'05"	x
Speech	36"	33"	48"	56"	31"	34"	1'21"	x

(ii) Sentence II: "No one here will run away."

Time \ Subject	JF	MW	JS	DW	ES	JR	CH	GG
Typing	26"	1'10"	1'52"	40"	1'09"	59"	1'08"	x
Speech	1'07"	1'15"	1'40"	1'37"	1'50"	1'33"	4'16"	x

The Chi-Square test value indicates that the listening responses between each of the five judges for JF's speech are not significantly different. Figure 4.1-a shows the summary of results from the intelligibility test based on contrast analysis of 19 phonetic groups. The percentage error for a specific contrast was computed as the total number of errors made across the judges for that contrast divided by the total number of trials in which that specific contrast was available as a response for the judges. The two most frequent errors are vowel duration (11%) and Final Cluster - Singleton (8%). Figure 4.1-b shows the summary of results from the intelligibility test based on contrast analysis of 40 phonetic groups. The top four most frequent errors are Final Cluster - Singleton (33%), Long - Short Vowel Duration (15%), /w/ - /r/ (10%), and Back - Front Vowel (7%). Because JF's speech is not seriously impaired (I.T.: 95%), the error distributions have not shown very serious confusion errors or special articulatory problems except for the control of vowel duration and the final cluster - singleton consonants.

From Table 4.2, the I.T. results for JF indicate that the average accuracy for the 70 words is 95%. Further, groups SONOR_IN (97%), WTOT_OB (100%), and OBS_IN (94%) show that JF has the most severe problem for words with an obstruent consonant in word-initial position. The words without any obstruent consonant (100%) have the best I.T. performance. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The CLUS_IN (83%) group has the worst performance in these three sub-groups. The FR_AF_IN (93%) group shows a position control problem similar to the STOP_IN (95%) group. More detailed study of the FR_AF_IN group shows that the /s/_IN (90%) has the lowest accuracy in this FR_AF_IN group. /sh/_IN (93%) also is lower than the I.T. average of the 70-word list (95%). Table 4.3 shows the detailed I.T. analysis for the STOP_IN group which includes LABIAL (90%), ALVEOLAR (100%), and VELAR (100%) stop consonant groups. It appears that the words with velar and alveolar stop consonants in word-initial position have better I.T. performance than the words with labial consonants.

In summary, this subject has particular difficulties with obstruent consonants in word-initial position. The words with initial clusters and /s/ in word-initial position have especially low accuracy in comparison with the average intelligibility of the entire word list. Final cluster - singleton contrast errors show JF has final cluster control difficulty. Further, the mild vowel duration errors show that JF cannot always make good distinctions between long and short vowels. The words without any obstruent consonant have the best I.T. performance.

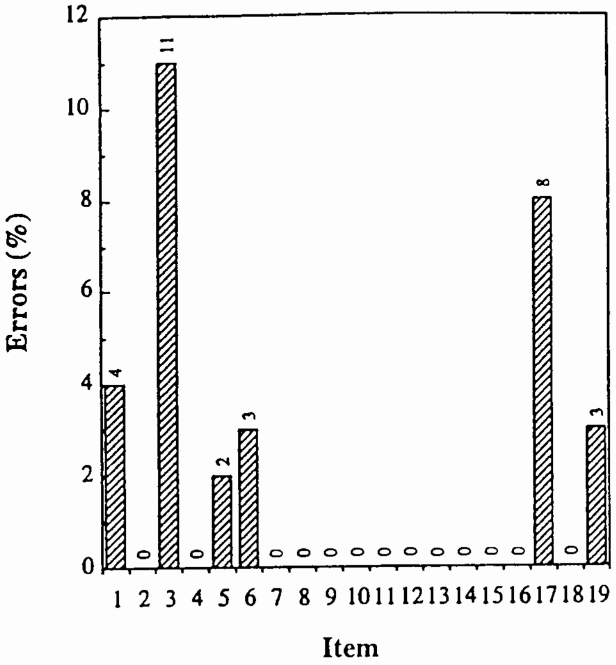


Figure 4.1-a JF's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution shows the two most frequent errors are vowel duration (11%) and Final Cluster - Singleton (8%).

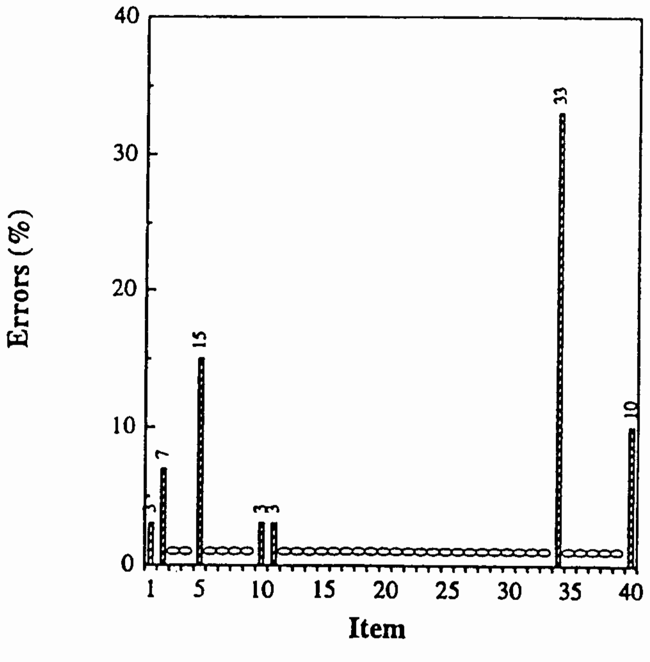


Figure 4.1-b JF's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution shows the top four most frequent errors are Final Cluster - Singleton (33%), Long - Short Vowel Duration (15%), /w/ - /r/ (10%), and Back - Front Vowel (7%).

(ii) Transcription (T.T.)

From Table 4.1, the Chi-Square test value of T.T. is 0.37 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the transcription responses between two judges for JF's speech are not significantly different. From Table 4.5, the T.T. results for JF indicate the average accuracy judged by two experienced judges for the obstruent (95%) and sonorant (81%) consonants, identified correctly in every feature, in word-initial position. This result is slightly different from the findings in the I.T. Vowel, voicing, and obstruent place are presented in the form of confusion matrices. The vowel confusions, which only focus on one feature identification, show that this speaker makes good distinctions between front vowels

Obstruent Accuracy:	95%
Sonorant Accuracy:	81%

Vowels

T. \ R.	**Front**	**Back**	**Middle**	**Diph.**	**[]**
Front	100%				
Back		100%			
Diph.				100%	

Voicing

T. \ R.	+	-	[ ]
+	100%		
-	1%	99%	

Place for Obstruents

T. \ R.	**Lab.**	**Dent.**	**Alv.**	**Pal.**	**Vel.**	**[ ]**
Lab.	100%					
Alv.			97%	3%		
Pal.			8%	92%		
Vel.					100%	

Table 4.5 JF's transcription data for vowel, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position. The accuracies of obstruent and sonorant consonants are percent correct identification of the obstruent and sonorant consonants in word-initial position. The target items are listed in the left column and the response items are displayed across the top.

(100%), back vowels (100%), and diphthongs (100%). The confusions associated with the voiced - voiceless contrast, which addresses only one feature identification, show that this speaker does not have a glottal control problem for voiced-voiceless consonants in word-initial position. The confusions for place of articulation for obstruents, which are based only one feature identification, indicate that palatal position control (92%) is the most serious one compared to labial (100%), alveolar (97%), and velar (100%) control. However, the palatal position control error is mild. A detailed examination of the errors for obstruents shows that all of the alveolar and palatal confusion errors come from fricative / affricate consonants but not from stop consonants.

Overall, obstruent accuracy (95%) is better than sonorant accuracy (81%) for consonants in word-initial position. The data for obstruents in word-initial position show that this subject does not have much difficulty with consonants that require pressure buildup behind a constriction in the vocal tract. However, the alveolar and palatal confusions indicate a slight alveolar and palatal position control error (Table 4.5). This observation is consistent with the results from Figure 4.1-a, which also show very mild average percentage errors for alveolar - palatal place (3%), and with the results from Table 4.2, which show the /s/_IN (90%) , /sh/_IN (93%), and CLUS_IN (83%) all are lower than the average accuracy (95%).

(iii) Recognition Test (R.T.)

From Table 4.2, the R.T. results for JF indicate that the average accuracy for the 70 words is 73% compared with the I.T. of 95%. Further, groups SONOR_IN (76%), WTOT_OB (91%), and OBS_IN (74%) show that JF has the worst articulatory consistency control problem for words with an obstruent consonant in word-initial position. This is consistent with the finding of the I.T. described above. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The CLUS_IN (70%) group has the worst performance in these three sub-groups. The FR_AF_IN (73%) group shows a slightly more serious consistency control problem than the STOP_IN (76%). More detailed study of the FR_AF_IN group shows that the /s/_IN (69%) has the lowest accuracy in this FR_AF_IN group. The STOP_IN is a little higher than the average R.T. accuracy (73%). Table 4.3 shows the detailed R.T. analysis for the STOP_IN group which includes LABIAL (70%), ALVEOLAR (93%), and VELAR (75%) stop consonant groups. It appears that the words with velar and alveolar consonants in word-initial position have better recognition performance than labial consonants, and also better than the R.T. average of the 70-word list (73%). On the other hand, the R.T. rejection rate of the 70-word list is 13% (Table 4.1) compared with the normal subject's 0.3%.

In summary, this subject has particular consistency control difficulties with initial clusters (70%) and /s/ in word-initial position (69%) in comparison with the average R.T. of the entire word list (73%). The words without any obstruent consonant (91%) have the best R.T. performance.

(iv) Acoustic Analysis (A.A.)

Figures 4.2-a and 4.2-c show JF's spectrograms corresponding to "bad" and "sticks". Figures 4.2-b and 4.2-d show MJ's, the normal subject's, spectrograms corresponding to "bad" and "sticks". Figure 4.2-a (300 to 770 ms) shows that the first formant, F1 (around 500 Hz), is too low in comparison with MJ's F1 (around 700 Hz) in Figure 4.2-b (120 to 430 ms). In addition, Figure 4.2-a (300 to 770 ms) also shows that JF's second formant, F2 (around 1900 Hz), is too high in comparison with MJ's F2 (around 1700 Hz) in Figure 4.2-b (120 to 430 ms). Peterson and Barney (1952) noted that the average F1 and F2 of /æ/ for male speakers are about 660 and 1720 Hz and the average F1 and F2 of /ɛ/ are about 530 and 1840 Hz over 76 speakers. Therefore, JF's /æ/ of "bad" is closer to /ɛ/ than /æ/. This is the reason why four of the five judges confused JF's "bad" with "bed" in the closed intelligibility test, I.T. This observation is consistent with Figures 4.1-a, which gives some evidence of vowel duration confusion (11%) (Kent et al. (1989) defined the /æ/ and /ɛ/ as vowel duration contrast.).

On the other hand, as was discussed in the I.T., the final cluster percentage error (Figure 4.1-a) is the second highest (8%) compared with the other contrasts. Further, Figure 4.1-b shows that the Final Cluster (goal) - Final Singleton (response) (33%) is the main confusion error. More detailed study indicates that "sticks" is the only word which is found to have final cluster errors. Figure 4.2-c shows that the final /s/ is missing and there is only the release and aspiration parts of the /k/ (970 to 1110 ms) for JF's "sticks". Figure 4.2-d shows MJ's spectrogram of "sticks". The release of /k/ starts around 410 ms and is followed by the short frication of /k/ and the /s/ (430 - 630 ms) with strong high-frequency noise. For this reason, all of the five judges heard only /k/ rather than /ks/ in word-final position, and chose the response "stick" for target "sticks". This interpretation can be also used to explained the conflicting finding between CLUS_IN (86%) in Table 4.2 and the

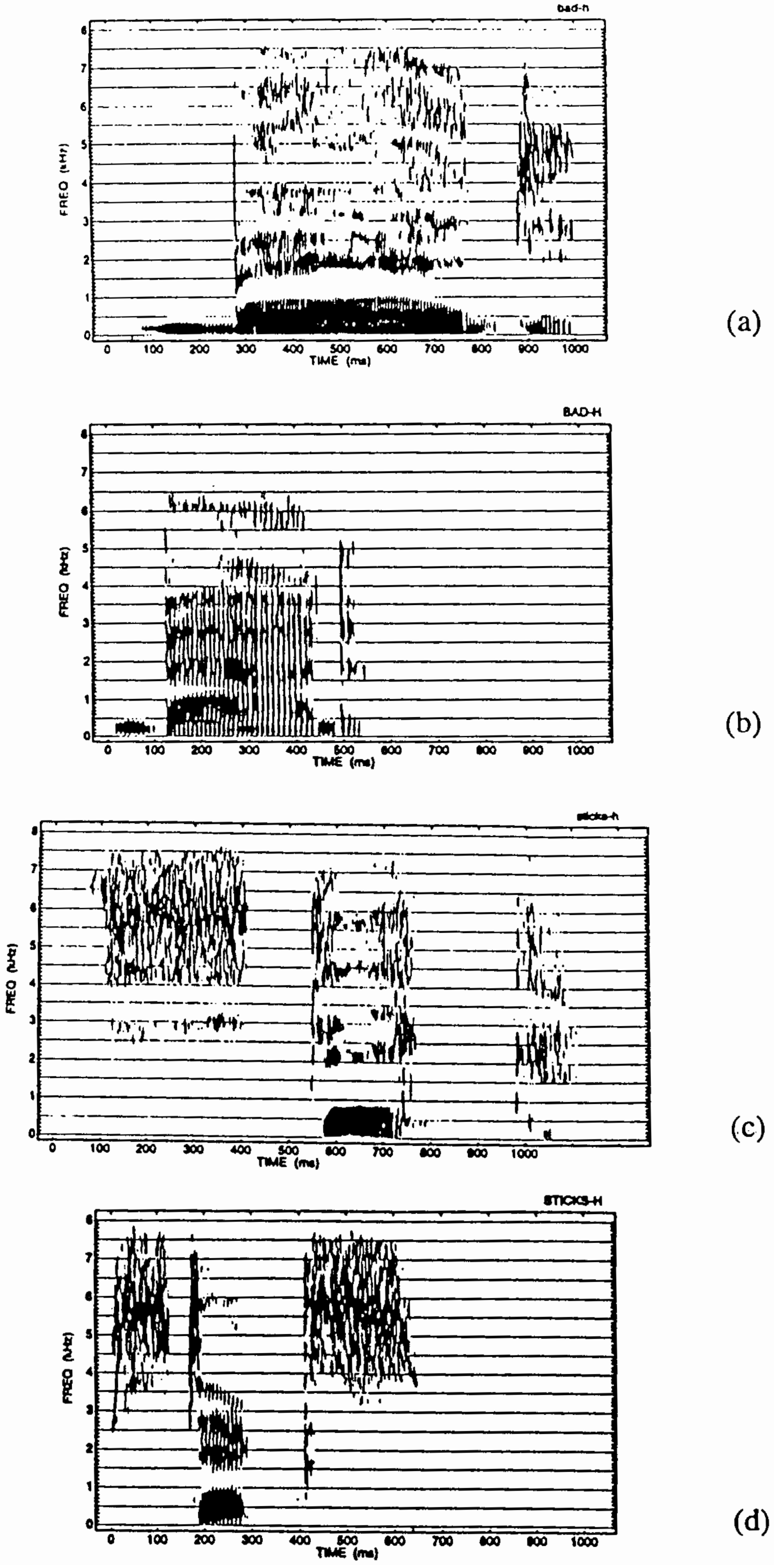


Figure 4.2: JF's spectrograms of "Bad" (4.2-a) and "Sticks" (4.2-c) compared with MJ's spectrograms of "Bad" (4.2-b) and "Sticks" (4.2-d).

error of Initial Cluster - Singleton (0%) in Figure 4.1-a. Although JF's problem is the final cluster for "sticks", "sticks" also is counted as a member in CLUS_IN and /s/_IN groups in Table 4.2, since "sticks" contains the initial cluster /st/ which has /s/ in the beginning. Therefore, the final cluster error of "sticks" also is counted as the errors of CLUS_IN and /s/_IN in Table 4.2, but this confusion error in fact comes from the Final Cluster - Singleton error in Figure 4.1-a.

Figures 4.2-a and 4.2-c show that there is mild aspiration noise in the vowel especially above 2.5 kHz, indicating that the subject's speech is breathy. However, JF's speech is not so breathy as JS', who is introduced later.

The average duration measurement based on the spectrograms for these 70 diagnosis words is 840 ms compared with MJ's 478 ms. The prolongation ratio is 1.8; that is, the duration of JF's words is 1.8 times that of the subject who does not have dysarthria, MJ, on the average.

4.1.3 Findings and Design Criteria

In summary, the four tests described above (I.T., T.T., R.T., and A.A.) show that JF has mild problems in producing fricatives / affricates in word-initial position (especially /s/ and clusters), clusters in final position, and vowel duration. Both the I.T. and the R.T. show that the words without any obstruent consonants and the words with sonorant consonants in word-initial position have a better performance than the words with obstruents in word-initial position. This observation is important for the design of a new word list. Furthermore, the words with stop consonants in word-initial position also have a fair performance. The words with alveolar and velar stop consonants in word-initial position have a better performance than the ones with labial stop consonants. Moreover, the problems with vowel distinctions, particularly vowel duration, should also be considered in the word-list design. Vowels that are at extreme positions in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, should be chosen to let the vowels be as distinctive as possible and to avoid lax - tense confusions.

4.1.4 Improvement of Speech Recognition Accuracy

From the findings described above for JF, a new word list with 52 words was designed for him to dictate to the computer. Table 4.1 shows that the recognition accuracy for the new word list improved from 73% (R.T.(%)) based on the 70 diagnosis words to 82% (New_L(%)) based on the new 52-word list (or 79% for R.T.-35(%) and 85% for New_L-35(%) based on the same vocabulary size, 35). An improvement of 9% (Improve(%)) or 6% (Improve-35(%)) has been achieved. After designing this new list and following Part II of the flow chart in Figure 3.1, some of the words which were still unsatisfactory were modified. The final list is in Appendix 4.1.

Two sentences, "I am brilliant." and "No one here will run away." were dictated by following the alphabet spelling, e.g., "air" is for "a" and "bad" is for "b". JF's dictation speed for these two sentences was compared with his usual typing speed, using his left index finger with the right hand holding the left. Table 4.4 shows the results of the test based on the average of three tries. His average typing time for sentence I is 11 seconds compared with a dictating time of 36 seconds; for sentence II, the typing time is 26 seconds compared with a dictating time of 1 minute and 7 seconds. These data show that JF's typing speed is three times faster than his speech dictation. However, the speech input can still be a useful tool for him to control the computer. He can use simultaneously typing and speech dictation methods to reduce the over-load of using his left index finger and hands. For example, he can use his left index finger for typing characters and his speech dictation for changing different word fonts and sizes or choosing software function commands.

4.2 Subject 2 (MW)

4.2.1 Background and Symptoms

MW, who is 38, has a bachelor's degree. MW's motor control was not unusual until he was 1-1/2 years old. When he tried to learn to walk, his parents found that he could not keep balance. This is characteristic of telangiectasia: failure of muscular coordination and irregularity of muscular action usually appearing when the child attempts to walk. He needs a T-board and inclines his body forward to support stably his right and left palms when he controls the computer. Otherwise, because of the tremor and the involuntary movements of his hands, he can not type the correct key accurately. Furthermore, because of the inclination of his body and head, he can not watch the monitor and keyboard simultaneously. He feels pain and is easily fatigued in typing or programming jobs. 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). The overall intelligibility of his speech is somewhat less than that of JF. MW's cognitive and linguistic abilities are intact.

4.2.2 Analysis and Discussion

(i) Intelligibility Test (I.T.)

Table 4.1 shows that MW's intelligibility for the I.T., is 82%. The Chi-Square test value of I.T. is 5.8 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the listening responses between each of the five judges for MW's speech are not significantly different. Figure 4.3-a shows the summary of results from the intelligibility test based on contrast analysis of 19 phonetic groups. The percentage error for a specific contrast was computed as the total number of errors made across the judges for that contrast divided by the total number of trials in which that specific contrast was available as a response for the judges. The five most frequent errors are Final Voicing Contrast (18%), Alveolar - Palatal Place (13%), Other Fricative Place (12%), Fricative - Affricate (12%), High - Low Vowel (10%). There are some errors for vowel duration (8%). Figure 4.3-b shows the summary of results from the intelligibility test based on contrast analysis of 40 phonetic groups. The five most frequent errors are Alveolar - Other Fricative Place (33%), Final Voiceless - Voiced Contrast (26%), Short - Long Vowel Duration (20%), Alveolar - Palatal Place (17%), and Velar - Other Consonant Place (15%). The Fricative - Affricate Consonant (13%) and Affricate - Fricative Consonant (13%), along with Alveolar - Other Fricative Place and Alveolar - Palatal Place, indicate that this speaker has some problems with alveolar and palatal control, especially for fricatives and affricates. A more detailed analysis indicates that most of the alveolar confusion errors come from fricative consonants but not from stop consonants. The percentage error for Final Voiceless - Voiced Contrast shows that this

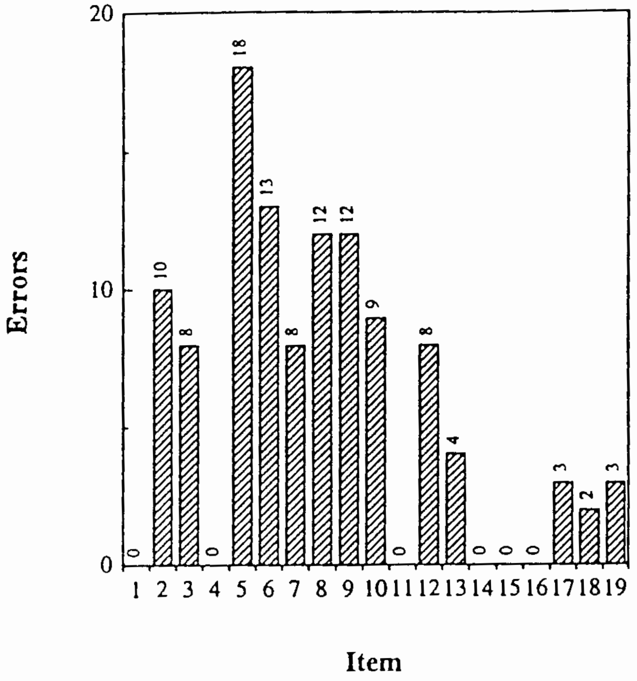


Figure 4.3-a MW's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution shows the five most frequent errors are Final Voicing Contrast (18%), Alveolar - Palatal Place (13%), Other Fricative Place (12%), Fricative - Affricate (12%), High - Low Vowel (10%).

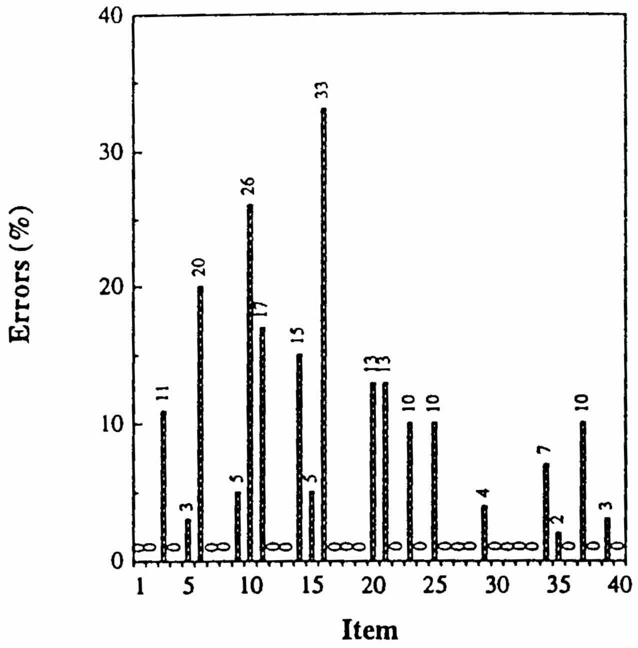


Figure 4.3-b MW's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution shows the five most frequent errors are Alveolar - Other Fricative Place (33%), Final Voiceless - Voicing Contrast (26%), Short - Long Vowel Duration (20%), Alveolar - Palatal Place (17%), and Velar - Other Consonant Place (15%).

subject has problems with glottal control or with adjustment of vowel duration before voiced and voiceless consonants. Additionally, MW has also problems of vowel duration control.

From Table 4.2, the I.T. results for MW indicate that the average accuracy for the 70 words is 82%. Further, groups SONOR_IN (81%), WTOT_OB (90%), and OBS_IN (81%) show that the words without obstruents are the best and the words with sonorants in word-initial position are roughly equal to the words with obstruents in word-initial position. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The CLUS_IN (73%) group has the worst performance in these three sub-groups. The FR_AF_IN (75%) group shows a more serious position control problem than the STOP_IN (89%). More detailed study of the FR_AF_IN group shows that the /s/_IN (65%) has the lowest accuracy in this FR_AF_IN group. Table 4.3 shows the detailed I.T. analysis for the STOP_IN group which includes LABIAL (93%), ALVEOLAR (100%), and VELAR (75%) stop consonant groups. It appears that the words with labial and alveolar stop consonants in word-initial position have better I.T. performance than velar consonants, and also better than the I.T. average of the 70-word list (82%).

In summary, this subject has particular difficulties with obstruent consonants in word-initial position. The fricatives and affricates, particularly for the alveolar and palatal positions, in word-initial position appear to be serious control problems. The words with initial clusters and /s/ in word-initial position have low I.T. accuracy in comparison with the average intelligibility of the entire word list. Further, the final voiced - voiceless contrast shows that MW has problems with glottal control or with adjustment of vowel duration before voiced and voiceless consonants. The mild vowel duration errors show that MW cannot make good distinctions between long and short vowels. The words without any obstruent consonant have the best I.T. performance.

(ii) Transcription Test (T.T.)

From Table 4.1, the Chi-Square test value of T.T. is 0.11 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the transcription responses between two judges for MW's speech are not significantly different. From Table 4.6, the T.T. results for MW indicate the average accuracy judged by two experienced judges for the obstruent (78%) and sonorant (85%) consonants,

Obstruent Accuracy: 78%

Sonorant Accuracy: 85%

Vowels

T. \ R.	Front	Back	Middle	Diph.	[]
Front	88%	3%		9%	
Back		100%			
Diph.				100%	

Voicing

T. \ R.	+	-	[ ]
+	100%		
-	1%	99%	

Place for Obstruents

T. \ R.	Lab.	Dent.	Alv.	Pal.	Vel.	[ ]
Lab.	96%		4%			
Alv.	3%	41%	56%			
Pal.				100%		
Vel.					100%	

Table 4.6 MW's transcription data for vowel, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position. The accuracies of obstruent and sonorant consonants are percent correct identification of the obstruent and sonorant consonants in word-initial position. The target items are listed in the left column and the response items are displayed across the top.

identified correctly in every feature, in word-initial position. This is consistent with the findings in the I.T. Vowel, voicing, and obstruent place are presented in the form of confusion matrices. The vowel confusions, which only focus on one feature identification, show that this speaker makes good distinctions between front vowels (88%), back vowels (100%), and diphthongs (100%). The confusions associated with the voiced - voiceless contrast, which addresses only one feature identification, show that this speaker does not have a glottal control problem for voiced-voiceless consonants in word-initial position (only 1% error for voiceless consonants confused with voiced ones). The confusions for place of articulation for obstruents, which are based only one feature identification, indicate that alveolar position control (56%) is the most serious one compared to labial (96%), palatal (100%), and velar (100%) control. Most of the alveolar errors are made with fricative consonants rather than stops.

Overall, obstruent accuracy (78%) is worse than sonorant accuracy (85%) for consonants in word-initial position. The data for obstruents in word-initial position show that this subject does not have much difficulty in pronouncing these obstruent consonants except in the alveolar position. The alveolar confusions indicate a serious alveolar position control error (Table 4.6). This observation is consistent with the results from Table 4.2, which show a low average percentage accuracy for /s/_IN, and Figures 4.3-a and 4.3-b which both show a high percentage error for the alveolar position control.

(iii) Recognition Test (R.T.)

From Table 4.2, the R.T. results for MW indicate that the average accuracy for the 70 words is 71% compared with the I.T. of 82%. Further, groups SONOR_IN (78%), WTOT_OB (80%), and OBS_IN (67%) show that MW has the worst articulatory consistency control problem for words with an obstruent consonant in word-initial position. This is consistent with the finding of the I.T. described above. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The CLUS_IN (57%) group has the worst performance in these three sub-groups. The FR_AF_IN (59%) group shows a more serious consistency control problem than the STOP_IN (79%). More detailed study of the FR_AF_IN shows that /f/ in word-initial position (25%) has the lowest accuracy in this group; /s/ in word-initial position (65%) and /ch/ in word-initial position (60%) are both worse than the average R.T. accuracy (71%). The STOP_IN accuracy (79%) is higher than the average R.T. accuracy (71%). Table 4.3 shows the detailed R.T. analysis for the STOP_IN group which includes LABIAL (68%), ALVEOLAR (93%), and VELAR (90%) stop consonant groups. It appears that the words with alveolar and velar consonants in word-initial position have better recognition performance than labial consonants, and also better than the R.T. average of the 70-word list (71%). This is not very consistent with the findings in the I.T. On the other hand, the R.T. rejection rate of the 70-word list is 14% (Table 4.1) compared with the normal subject's 0.3%. It shows that MW's speech has a mildly inconsistent control problem.

In summary, this subject has particular consistency control difficulties with initial clusters, and fricatives / affricates in the initial-position (for /f/, /s/, and /ch/ in word-initial position) in comparison with the average R.T. of the entire word list. Moreover, the words without any obstruent consonant (80%) have the best R.T. performance.

(iv) Acoustic Analysis (A.A.)

Figures 4.4-a, 4.4-b and 4.4-c show spectrograms corresponding to "sin", "coat" and "write". Figure 4.4-a (50 to 200 ms) shows that the /s/ in "sin" looks like /theta/ because the noise distribution is weak compared with /s/. This weak noise amplitude is probably the source of the confusion for /s/ with /theta/ and the high error percentage for Alveolar - Other Fricative Place (33%) (Figure 4.3-b). Furthermore, the alveolar place for obstruents in Table 4.6 indicates that there is a high degree of alveolar - dental confusion (44%). The low-frequency energy at 300 Hz (Figure 4.4-a) in the fricative consonant probably comes from airflow hitting the microphone.

As was discussed in the intelligibility test, the Final Voicing Contrast (Figure 4.3-a) is the worst one (18%) compared with the other contrasts. Close inspection of the Test 1 data indicates, for example, that judges responded to the target "coat" with the response "code". Detailed examination of the transcription data indicates why the judges chose the response "code" for the target "coat" (i.e., a voiced consonant in word-final position). During transcription, the judge heard a glottal stop /?/ in place of the final /t/. This is supported by the acoustic data (Figures 4.4-b and 4.4-c). Figures 4.4-b and 4.4-c indicate that the final /t/ is missing and there are a corresponding prolongation of the vowels, /o/ and /ai/, and the final glottal stop replaces the final /t/. For these reasons, the judges heard /t/ as /d/ in word-final position and chose the response "code" for target "coat". The vowel lengthening might lead to interpretation of the final consonants as voiced.

The average duration measurement based on the spectrograms for these 70 diagnosis words is 607 ms compared with MJ's 478 ms. The prolongation ratio is 1.3; that is, the duration of MW's words is 1.3 times that of the subject who does not have dysarthria, MJ, on the average.

4.2.3 Findings and Design Criteria

In conclusion, the four tests described above (I.T., T.T., R.T., and A.A.) show that MW has control problems in producing fricatives and affricates in word-initial position (especially in the alveolar position, e.g., /s/ as "sin"), distinctions in vowel duration, consonant clusters (both in word-initial and -final positions), and voicing contrast in word-final position. Both the I.T. and the R.T. show that the words without any obstruent

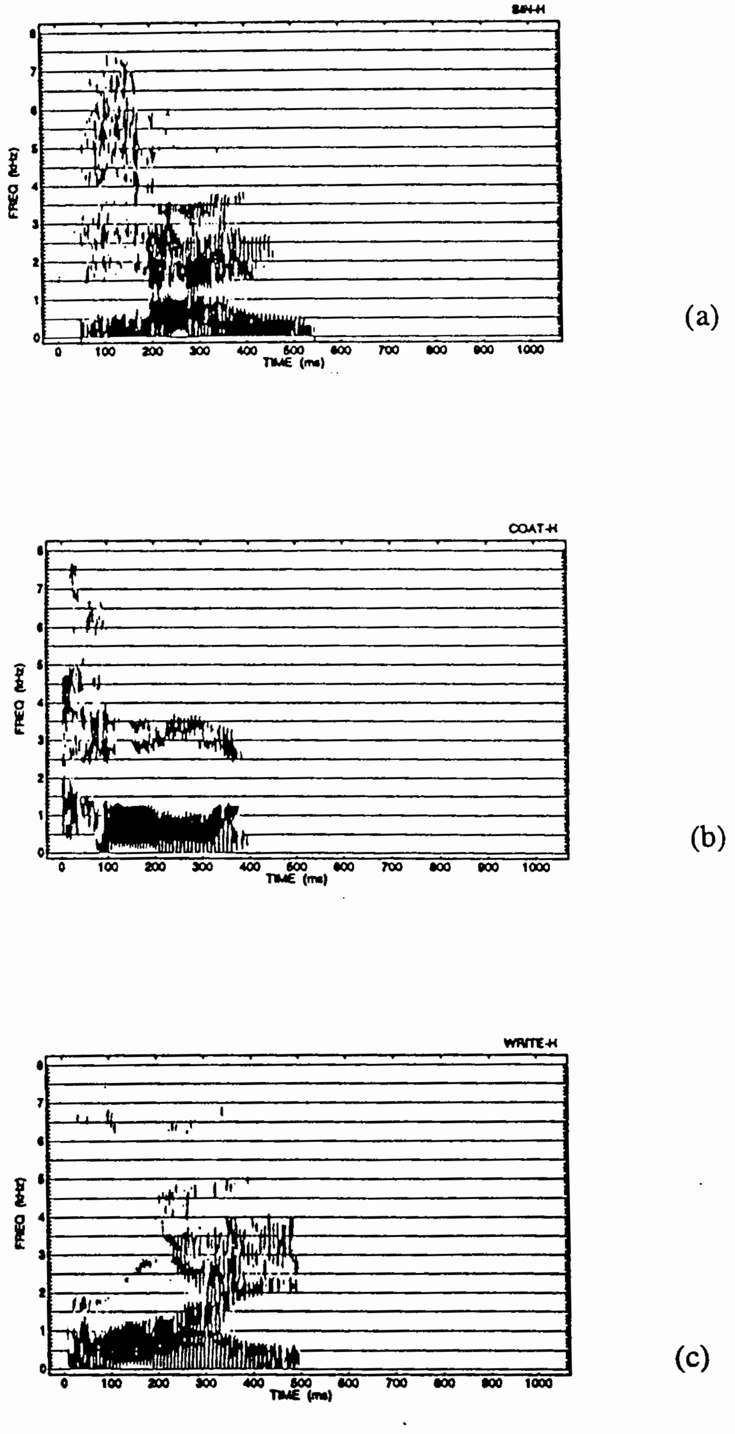


Figure 4.4: MW's spectrograms of "Sin" (4.4-a), "Coat" (4.4-b), and "Write" (4.4c).

consonants and the words with sonorant consonants in word-initial position have a better performance than the words with obstruents in word-initial position. This observation is important for the design of a new word list. Furthermore, the words with stop consonants in word-initial position also have a fair performance. The words with alveolar and velar stop consonants in word-initial position have a better recognition performance than the ones with labial stop consonants. On the other hand, the problems with vowel distinctions, particularly vowel duration, should also be considered in the word-list design. Vowels that are at extreme positions in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, should be chosen to let the vowels be as distinctive as possible and have no lax - tense confusion since there are some errors for vowel duration and position.

4.2.4 Improvement of Speech Recognition Accuracy

From the findings described above for MW, a new word list with 52 words was designed for him to dictate to the computer. Table 4.1 shows that the recognition accuracy for the new word list has been improved from 71% (R.T.(%)) based on the 70 diagnosis words to 86% (New_L(%)) based on the new 52-word list (or 79% for R.T.-35(%) and 82% for New_L-35(%) based on the same vocabulary size, 35). An improvement of 15% (Improve(%)) or 3% (Improve-35(%)) has been achieved. After designing this new list and following Part II of the flow chart in Figure 3.1, some of the words which were still unsatisfactory were modified. The final list is in Appendix 4.2.

Two sentences, "I am brilliant." and "No one here will run away." were dictated by following the alphabet spelling, e.g., "air" is for "a" and "bad" is for "b". The dictation speed for these two sentences was compared with his usual typing method, using fingers with a hand stand T-board for support. Table 4.4 shows the results of the test. His average typing time for sentence I is 27 seconds compared with a dictating time of 33 seconds; for sentence II, the typing time is 1 minute and 10 seconds compared with a dictating time of 1 minute and 15 seconds. These data show that MW's typing speed is almost equal to speech dictation speed; therefore the speech input can be a useful tool for him to control the computer. When he is using the speech recognition system, MW can watch simultaneously the computer monitor and keyboard when he is doing typing or programming jobs. Further, he has successfully utilized the speech recognizer systems in both IBM PC and Macintosh to replace the computer mouse function and to help him to program or type easily.

4.3. Subject 3 (JS)

4.3.1 Background and Symptoms

JS, who is 48, is studying for a bachelor's degree. At birth, JS' umbilical cord was wrapped around his neck. Respiration ceased for approximately 5 minutes, causing damage to that part of the cerebellum controlling motor and speech coordination. His motor control is characteristic of athetosis: 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 the serious tremor and involuntary movement of his hands, he can not even type or use a mouse (or joystick) easily. He uses his nose touching the keyboard to type his reports and to do analysis jobs with the computer. His speech impairment is indicative of poor respiratory control, exhibiting a forced, throaty voice quality and a large range of jaw movements.

This subject's speech is non-functional 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. His cognitive and linguistic abilities are intact.

4.3.2 Analysis and Discussion

(i) Intelligibility Test (I.T.)

Table 4.1 shows that JS' intelligibility for the I.T. is 60%. The Chi-Square test value of I.T. is 1.9 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the listening responses between each of the five judges for JS' speech are not significantly different. Figure 4.5-a shows the summary of results from the I.T. based on contrast analysis of 19 phonetic groups. The percentage error for a specific contrast was computed as the total number of errors made across the judges for that contrast divided by the total number of trials in which that specific contrast was available as a response for the judges. The seven most frequent errors are Initial Voicing Contrast (29%), Alveolar - Palatal Place (25%), /r/ - /l/ (24%), Fricative - Affricate (22%), Vowel Duration (18%), Consonant Place (18%), and Initial Glottal - Null (18%). Figure 4.5-b shows the summary of results from the intelligibility test

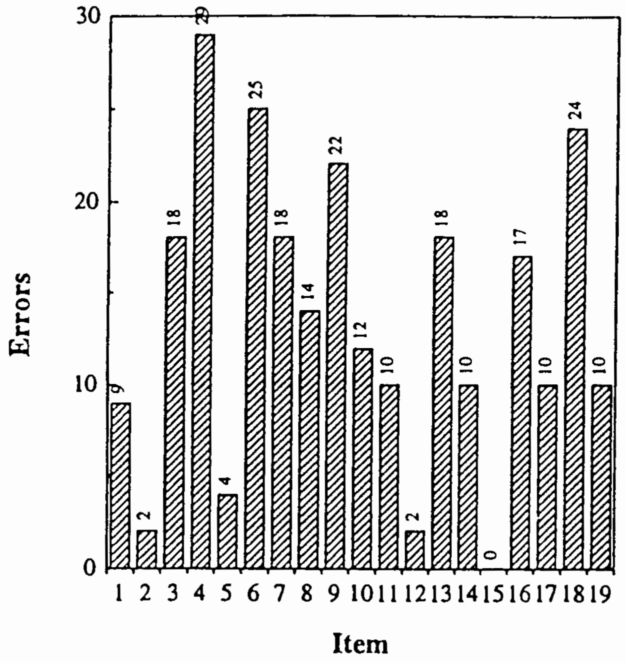


Figure 4.5-a JS' intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution shows the seven most frequent errors are Initial Voicing Contrast (29%), Alveolar - Palatal Place (25%), /r/ - /l/ (24%), Fricative - Affricate (22%), Vowel Duration (18%), Consonant Place (18%), and Initial Glottal - Null (18%).

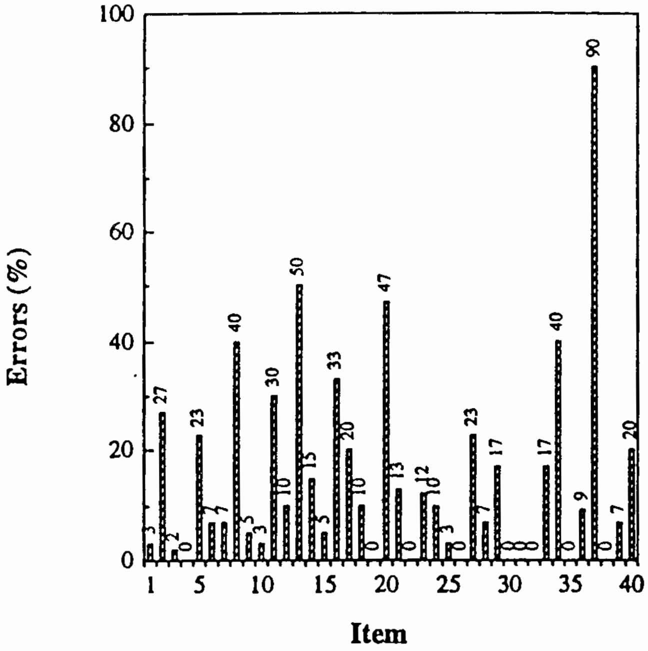


Figure 4.5-b JS' intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution shows the five most frequent errors are /l/ - /r/ (90%), Alveolar - Other Consonant Place (50%), Fricative - Affricate Consonant (47%), Init. Voiceless - Voicing Contrast (40%), and Final Cluster - Singleton (40%).

based on contrast analysis of 40 phonetic groups. The five most frequent errors are /l/ - /r/ (90%), Alveolar - Other Consonant Place (50%), Fricative - Affricate Consonant (47%), Init. Voiceless - Voiced Contrast (40%), and Final Cluster - Singleton (40%). The Alveolar - Palatal Place (30%) and Alveolar - Other Fricative Place (33%), along with Alveolar - Other Consonant Place and Fricative - Affricate Consonant, indicate that this speaker has some problems with alveolar control, especially for fricatives. In addition, JS might also have some problems with tongue-blade control because of the high percentage errors of /l/ - /r/ as well as the errors with alveolar consonants just noted. The most frequently occurring percentage error for /l/ - /r/ (90% in Figure 4.5-b) indicates that this subject has particular difficulty shaping the tongue blade to produce a lateral consonant /l/. A more detailed review of the data indicates that most of the alveolar confusion errors come from fricative consonants but not from stop consonants. It shows again that the fricative consonants are the main problems for JS' speech. Moreover, the percentage error for Init. Voiceless - Voiced Contrast (40% in Figure 4.5-b) indicates that this subject has problems with glottal control for word-initial position.

From Table 4.2, the I.T. results for JS indicate that the average accuracy for the 70 words is 60% when the listeners have a reference list. Groups SONOR_IN (65%), WTOT_OB (74%), and OBS_IN (54%) show that JS has the most severe problem for words with an obstruent consonant in word-initial position. The words without any obstruent consonant (74%) have the best I.T. performance. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (43%) group shows the most serious position control problem in these three sub-groups. More detailed study of the FR_AF_IN group shows that the /s/_IN (31%) and /sh/_IN (47%) are the most serious two in the FR_AF_IN group. Table 4.3 shows the detailed I.T. analysis for the STOP_IN group which includes LABIAL (75%), ALVEOLAR (47%), and VELAR (75%) stop consonant groups. It appears that the labial and velar stop consonants in word-initial position have better I.T. performance than alveolars and that they are above the I.T. average of the 70-word list (60%).

In summary, this subject has particular difficulties with obstruent consonants in word-initial position. JS' tongue-blade control is impaired. The alveolar position control in word-initial position has particular impairments, especially for fricatives. The words with /s/ in word-initial position, /sh/ in word-initial position, and initial clusters have low accuracy in comparison with the average intelligibility of the entire word list (60%). Further, initial voiced - voiceless contrast and initial glottal - null errors show JS has some problems with glottal control. The words with stop consonants in word-initial position have fair performance compared to the I.T. average of the 70-word list. Finally, the vowel duration errors show that JS cannot make good distinctions between long and short vowels. The words without any obstruent consonant have the best I.T. performance.

(ii) Transcription Test (T.T.)

From Table 4.1, the Chi-Square test value of T.T. is 5.55 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the transcription responses between two judges for JS' speech are not significantly different. However, the inter-judge consistency for T.T. is worse than other subjects (except CH). From Table 4.7, the T.T. results for JS indicate the average accuracy judged by two experienced judges for the obstruent (21%) and sonorant (54%) consonants, identified correctly in every feature, in word-initial position. This is consistent with the findings in the I.T. Vowel, voicing, and obstruent place are presented in the form of confusion matrices. The vowel confusions, which only focus on one feature identification, show that this speaker makes better distinctions for front and back vowels than diphthongs, (front vowel: 84%, back vowel: 79%, and diphthong: 42%). There are 21% errors for back vowels confused with front vowels. This is supported by Figure 4.5-b, Back - Front Vowel (27%). The confusions associated with the voiced - voiceless contrast, which addresses only one feature identification, show that this speaker has a glottal control problem for voiceless (goal) - voiced (response) consonants in word-initial position (42% errors for voiceless consonants confused with voiced ones and 12% of the voiceless consonants disappeared), but the voiced consonants in word-initial position are usually judged to have the correct voicing (98% accuracy). The confusions for place of articulation for obstruents, which are based only on one feature identification, indicate that alveolar position control is the most serious one (22%) compared to labial (46%), palatal (75%), and velar (87%) control. With more detailed study of T.T., it shows that most of the alveolar errors are made in fricative consonants, but not stops.

Overall, obstruent accuracy (21%) is worse than sonorant accuracy (54%) for consonants in word-initial position. The data for obstruents in word-initial position show that this subject has more difficulty in pronouncing the obstruent consonants at the anterior part of the vocal tract than at the posterior position. The alveolar confusions indicate a serious alveolar position control error, especially for the fricative consonants (Table 4.7). This is consistent with the results from Table 4.2, which show a low average

Obstruent Accuracy: 21%

Sonorant Accuracy: 54%

Vowels

T. \ R.	Front	Back	Middle	Diph.	[]
Front	84%	13%	2%	1%	
Back	21%	79%			
Diph.	25%	33%		42%	

Voicing

T. \ R.	+	-	[ ]
+	98%	2%	
-	42%	46%	12%

Place for Obstruents

T. \ R.	Lab.	Dent.	Alv.	Pal.	Vel.	[ ]
Lab.	46%		17%		25%	13%
Alv.	9%	25%	22%	13%	9%	22%
Pal.	8%			75%	8%	8%
Vel.	13%				87%	

Table 4.7 JS' transcription data for vowel, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position. The accuracies of obstruent and sonorant consonants are percent correct identification of the obstruent and sonorant consonants in word-initial position. The target items are listed in the left column and the response items are displayed across the top.

percentage accuracy for /s/_IN, and Figures 4.5-a and 4.5-b which both show a high percentage error for the alveolar position control.

(iii) Recognition Test (R.T.)

From Table 4.2, the R.T. results for JS indicate that the average accuracy for the 70 words is 37% compared with the I.T. of 60%. Further, groups SONOR_IN (39%), WTOT_OB (49%), and OBS_IN (32%) show that JS has the worst articulatory consistency control problem for words with an obstruent consonant in word-initial position. This is consistent with the finding of the I.T. described above. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The CLUS_IN (27%) group has the worst performance in these three sub-groups. The FR_AF_IN (29%) group shows a more serious consistency control problem than the STOP_IN (37%). More detailed study of the FR_AF_IN shows that /s/ in word-initial position (23%) has the lowest accuracy in this group. /f/ in word-initial position (30%)

also is worse than the average R.T. accuracy (37%). The accuracy of STOP_IN group is a little lower than the average R.T. accuracy (37%). Table 4.3 shows the detailed R.T. analysis for the STOP_IN group which includes LABIAL (38%), ALVEOLAR (33%), and VELAR (40%) stop consonant groups. It appears that the labial and velar stop consonants in word-initial position have slightly better recognition performance than alveolar ones, and also slightly better than the R.T. average of the 70-word list (37%). This is consistent with the findings in the I.T. On the other hand, the R.T. rejection rate of the 70-word list is 29% (Table 4.1) compared with the normal subject's 0.3%. It shows that JS' speech is very inconsistent.

In summary, this subject has particular consistency control difficulties with /s/ (23%), /f/ (30%), and initial clusters (27%) in word-initial position in comparison with the average R.T. of the entire word list (37%). Moreover, the words without any obstruent consonant (49%) have the best R.T. performance. The words with stop consonants in word-initial position have fair performance compared with the R.T. average of the 70-word list.

(iv) Acoustic Analysis (A.A.)

Figures 4.6-a, 4.6-b, 4.6-c and 4.6-d show the spectra corresponding to /f/ in "feed", /sh/ in "shoot", /s/ in "sell", and /s/ in "sin". The dashed line indicates the theoretical shape of the corresponding spectra for normal speech. The spectrum of the /f/ in Figure 4.6-a shows that the spectrum energy is flat in the high-frequency portion, similar to the theoretical prediction for /f/. From the acoustic view, it shows that this labial fricative is only slightly abnormal. Table 4.7 indicates also that labial position control for the obstruent consonants (including stop and fricative / affricate consonants) is moderately good (46% accuracy compared to 22% accuracy for alveolars).

The spectrum of /sh/ in Figure 4.6-b indicates that the match to the theoretical prediction is not very good; however, the basic spectral shape of /sh/ can be seen. The spectrum for /s/ in Figure 4.6-c is closer in shape to the speaker's realization of /sh/ in Figure 4.6-b than to the shape of the theoretical prediction of /s/. However, the speaker's realization of /s/ in Figure 4.6-d results in an overall shape which is a fairly good match to the theoretical prediction, but the resonance peak position is too low (3 kHz compared to the predicted 4.5 kHz). It appears that JS' tongue tip is placed too far back for this /s/. This observation is consistent with the confusions presented in Table 4.7 for place of

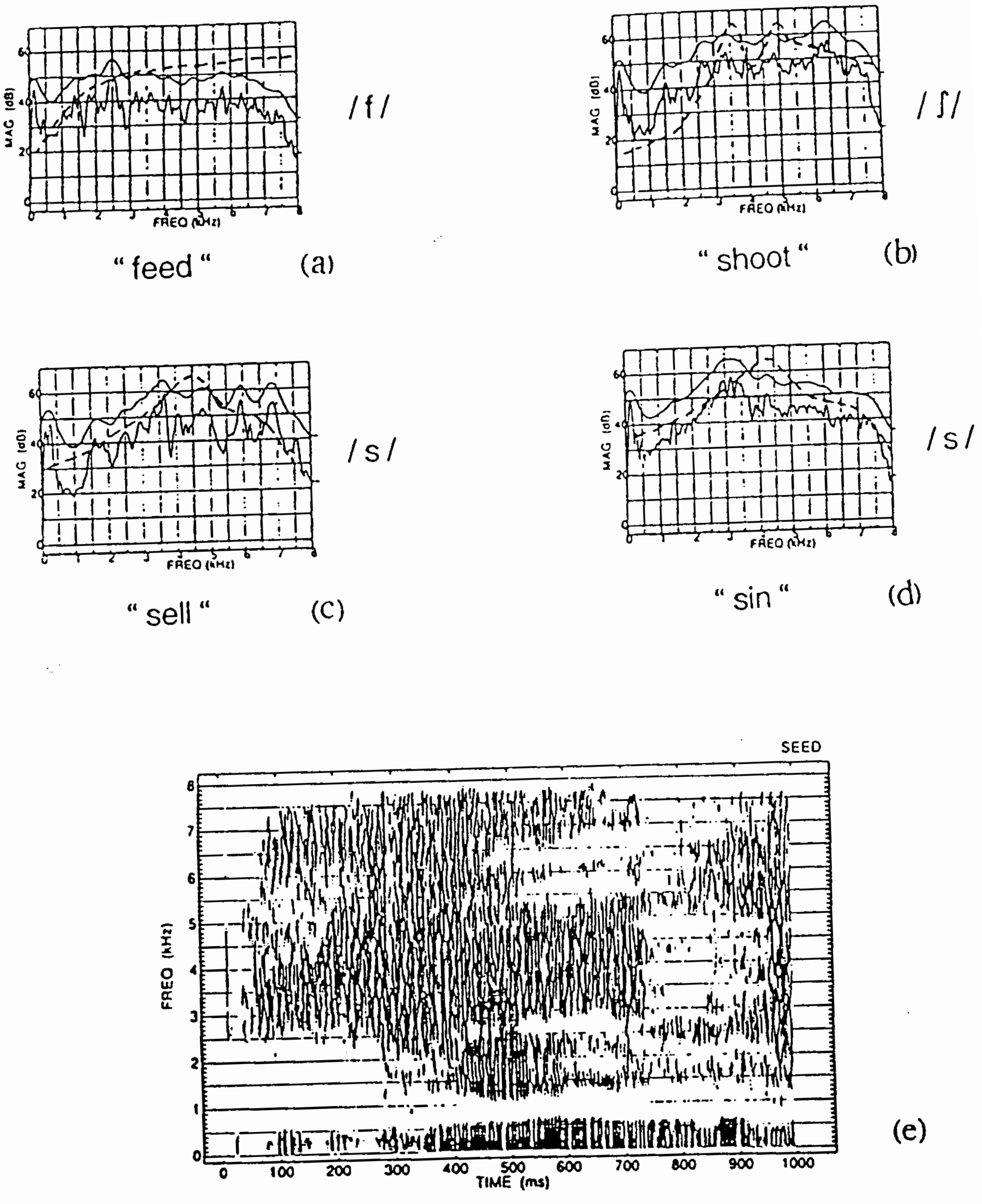


Figure 4.6: JS' spectra for /f/ (4.6-a), /sh/ (4.6-b), /s/ (4.6-c and 4.6-d), and spectrogram for "seed" (4.6-e). The dashed lines superimposed on the spectra are the expected theoretical spectrum shapes.

obstruent constriction. Table 4.7 shows alveolar position control to be highly unstable and confused with palatal (13%) and dental consonants (25%). This accounts for the low accuracy for /s/ found in Table 4.3 for this subject (JS). These four figures depict a high-amplitude first harmonic indicating a voiced consonant consistent with the voiceless-voiced confusions presented in Table 4.7. The voiceless-voiced confusion matrix indicates a high error rate of 42% (Table 4.7). Figure 4.6-e shows that there is severe aspiration noise in the vowel, especially above 2.5 kHz, indicating that the subject's speech is breathy.

Acoustic analysis has also been applied to study the conflict between the I.T. and R.T. for JS' speech. For example, Table 4.8-a shows the I.T. and R.T. accuracy for two words, "steak" and "side", which have completely different performance for these two tests. "Side" has 100% accuracy for human listeners (I.T.) but 0% accuracy for the speech recognizer (R.T.). "Steak" has inverse results: 20% accuracy for human listeners (I.T.) but 100% accuracy for the speech recognizer (R.T.). Detailed acoustic study of these two words shows that JS' /s/ for these two words has a particularly poor performance, including the constriction position variance mentioned above and the articulatory duration variance (Table 4.8-b). Table 4.8-b shows the /s/ of "side" has 159 ms average duration with 139 ms standard deviation over the ten tokens collected for the R.T. (compared to MJ's 160 ms average duration with 20 ms standard deviation, Table 4.8-c). Moreover, vowel comparison for JS' and MJ's "side" gives a better consistency than the /s/ (407 ms average duration with 23 ms standard deviation for JS compared with 297 ms average duration with 10 ms standard deviation for MJ). From these comparisons, it appears JS' /s/ for "side" has strong duration and position control problems. The long duration of /s/ leads the human listeners still to recognize it as /s/ but the large duration variance of /s/ causes the speech recognizer to become confused or to reject the tokens. This reason might be used to explain why there is so much difference in performance between the I.T. and R.T. This type of analysis can be applied to the other word, "steak". The I.T. accuracy for "steak" is low but the R.T. accuracy is high. Table 4.8-b shows that the average duration of /s/ for JS is only 23 ms with 4 ms standard deviation (compared with 124 ms average and 12 ms standard deviation for MJ). The /s/ of "steak" for JS seems too short to be distinguished by human listeners but is good for the recognizer. The speech recognizer would "neglect" the existence of this short /s/. Especially, JS' /s/ has poor position performance. It would be of great help for the recognizer to "neglect" this abnormal /s/. Therefore, the performance of R.T. is much higher than that of I.T. This

Table 4.8 Comparison of the I.T. and R.T. accuracy (4.8-a) and duration measurement for JS' (4.8-b) and MJ's (4.8-c) "Steak"s and "Side"s. The item "Average" and "Std. Dev." in (4.8-b) and (4.8-c) represent correspondingly the average duration and the standard deviation of the specific consonant or vowel in unit of millisecond.

(a)

	I.T. Accuracy (%)	R.T. Accuracy (%)
"Side"	100	0
"Steak"	20	100

(b)

	"Side"		"Steak"	
JS	/s/ (ms)	Vowel (ms)	/s/ (ms)	Vowel (ms)
Average:	159	407	23	321
Std. Dev.:	139	23	4	31

(c)

	"Side"		"Steak"	
MJ	/s/ (ms)	Vowel (ms)	/s/ (ms)	Vowel (ms)
Average:	160	297	124	169
Std. Dev.:	20	10	12	11

study illustrates the value of acoustic analysis in interpreting the properties of abnormal speech.

The average duration measurement based on the spectrograms for these 70 diagnosis words is 868 ms compared with MJ's 478 ms. The prolongation ratio is 1.82; that is, the duration of JS' speech words is 1.82 times that of the subject who does not have dysarthria, MJ, on the average.

4.3.3 Findings and Design Criteria

In summary, the four tests described above (I.T., T.T., R.T., and A.A.) show that JS has particular control problems in alveolar obstruents in word-initial position (especially for fricatives), tongue blade, glottis, initial clusters, and vowel duration. Both the I.T. and the R.T. show that the words without any obstruent consonants and the words with sonorant consonants in word-initial position have a better performance than the words with obstruents in word-initial position. This observation is important for the design of a new word list. Furthermore, the words with stop consonants in word-initial position have a fair performance compared with the performance of the fricative / affricate consonants. Labial and velar stop consonants in word-initial position have a better recognition performance than alveolar stop consonants. In addition, the problems with vowel distinctions, particularly vowel duration, should also be considered in the word-list design. Vowels that are at extreme positions in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, should be chosen to let the vowels be as distinctive as possible and to avoid lax - tense confusions. Because JS' front vowels have better distinctions than the back ones, the front vowels should have higher priority than the back ones when the new word list is designed. However, this difference is relatively minor.

4.3.4 Improvement of Speech Recognition Accuracy

From the findings described above for JS, a new word list with 52 words was designed for him to dictate to the computer. Table 4.1 shows that the recognition accuracy for the new word list increased from 37% (R.T.(%)) based on the 70 diagnosis words to 75% (New_L(%)) based on the new 52-word list (or 44% for R.T.-35(%) and 74% for New_L-35(%) based on the same vocabulary size, 35). An improvement of 38%

(Improve(%)) or 30% (Improve-35(%)) has been achieved. After designing this new list and following Part II of the flow chart in Figure 3.1, some of the words which were still unsatisfactory were modified. The final list is in Appendix 4.3.

Two sentences, "I am brilliant." and "No one here will run away." were dictated by following the alphabet spelling, e.g., "air" is for "a" and "bad" is for "b". The dictation speed for these two sentences was compared with his usual typing method, using the nose. Table 4.4 shows the results of the test. His average typing time for sentence I is 1 minute and 20 seconds compared with a dictating time of 48 seconds; for sentence II, the typing time is 1 minute and 52 seconds compared with a dictating time of 1 minute and 40 seconds. These data show that JS' typing speed is slower than the speech dictation; therefore the speech input is a useful tool for him to control the computer. He can watch simultaneously the computer monitor when he controls the computer by using the speech recognizer and is free from the painful typing method in which he uses his nose. Moreover, he can use his speech command to replace the function of the mouse and operate the computer efficiently since his awkward and involuntary movements make use of the mouse control difficult.

4.4 Subject 4 (DW)

4.4.1 Background and Symptoms

DW, who is 45, has a master's degree. His mother had difficuity in childbirth and her lung was collapsed for ten minutes. After emergency help, DW had brain damage, but his twin brother was healthy. After birth, DW had evidence of spastic and athetosis cerebral palsy. DW's arm and leg muscles move involuntarily. His jaw muscle control is impaired and is spastic to cause his upper teeth to grind heavily with his lower ones, so that his teeth are ground down. He can only use his index fingers to type or program. His speech sounds very disordered to the unfamiliar listener. The intelligibility of his speech is worse than that of JS, but his speech is less throaty than that of JS. His speech impairment is indicative of poor respiratory control, exhibiting a forced, throaty voice quality and a large range of jaw and head movements. Some of his words are cut in half by unexpected movements of the larynx or respiratory system. His speech is particularly time variant. His speech pattern and rate change seriously from one utterance to another.

4.4.2 Analysis and Discussion

(i) Intelligibility Test (I.T.)

Table 4.1 shows that DW's intelligibility for the I.T. is 57%. The Chi-Square test value of I.T. is 3.6 (< 9.49 based on $df = 4$ and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the listening responses between each of the five judges for DW's speech are not significantly different. Figure 4.7-a shows the summary of results from the intelligibility test based on contrast analysis of 19 phonetic groups. The five most frequent errors are Stop - Affricate (53%), Initial Voicing Contrast (29%), Initial Cluster - Singleton (27%), Alveolar - Palatal Place (23%), and Stop - Fricative (21%). Figure 4.7-b shows the summary of results from the intelligibility test based on contrast analysis of 40 phonetic groups. The top seven most frequent errors are Affricate - Stop Consonant (53%), Init. Voiceless - Voiced Contrast (43%), Alveolar - Other Fricative Place (40%), Fricative - Affricate Consonant (27%), Initial Cluster - Singleton (27%), Final Voiceless - Voiced Contrast (26%), and Alveolar - Palatal Place (23%). Because DW's speech is seriously impaired (I.T.: 57%), the error distributions show very serious confusion errors or special articulatory problems. The main errors concentrate on the alveolar and palatal articulatory position control problems, especially for fricatives / affricates. Initial cluster and initial voiceless - voiced control are problems too. The unvoiced consonants are especially confused with voiced ones in word-initial position. Further, DW's vowel control also has some problems although the impairments of the vowel contrasts, from 5% to 20% in Figure 4.7-b, are not so serious as the high confusion error groups mentioned above.

From Table 4.2, the I.T. results for DW indicate that the average accuracy for the 70 words is 57%. Further, groups SONOR_IN (74%), WTOT_OB (83%), and OBS_IN (45%) show that DW has the most severe problem for words with an obstruent consonant in word-initial position. The words without any obstruent consonant (83%) have the best I.T. performance. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (32%) and the CLUS_IN (33%) show a more serious position control than the STOP_IN (65%). More detailed study of the FR_AF_IN group shows that the /ch/_IN (7%) has the lowest accuracy in this FR_AF_IN group. /s/_IN (23%) and /sh/_IN (33%) also have low scores. Table 4.3 shows the detailed I.T. analysis for the STOP_IN group which includes LABIAL (73%), ALVEOLAR (80%), and VELAR (40%) stop consonant groups. It appears that the

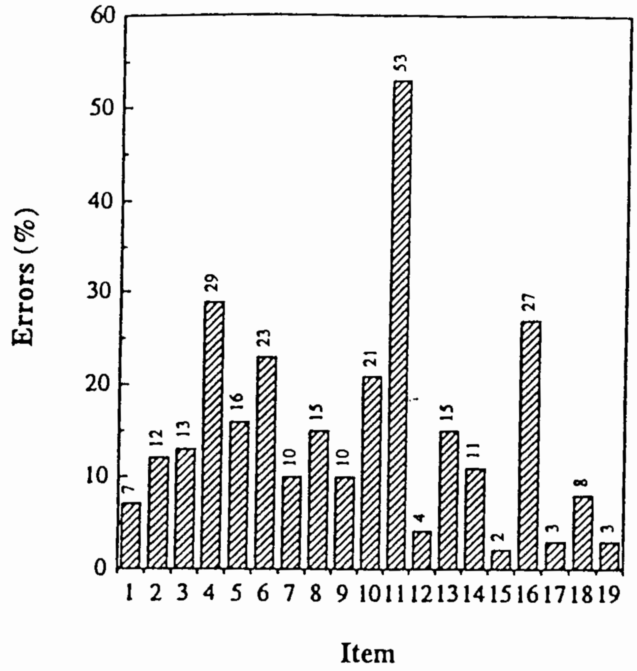


Figure 4.7-a DW's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution shows the five most frequent errors are Stop - Affricate (53%), Initial Voicing Contrast (29%), Initial Cluster - Singleton (27%), Alveolar - Palatal Place (23%), and Stop - Fricative (21%).

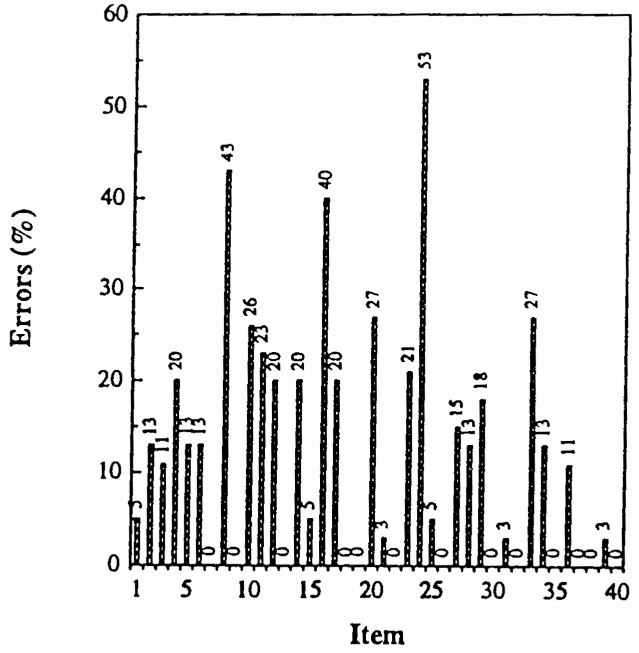


Figure 4.7-b DW's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution shows the top seven most frequent errors are Affricate - Stop Consonant (53%), Init. Voiceless - Voicing Contrast (43%), Alveolar - Other Fricative Place (40%), Fricative - Affricate Consonant (27%), Initial Cluster - Singleton (27%), Final Voiceless - Voicing Contrast (26%), and Alveolar - Palatal Place (23%).

words with labial and alveolar stop consonants in word-initial position have better I.T. performance than velar consonants, and also better than the I.T. average of the 70-word list (57%).

In summary, this subject has particular difficulties with obstruent consonants in word-initial position. The alveolar and palatal position control in word-initial position has particular impairments, especially for fricatives and affricates. That is, the words with /ch/, /s/, and /sh/ in word-initial position, and initial clusters have low accuracy in comparison with the average intelligibility of the entire word list (57%). Further, initial and final voiced - voiceless contrast errors show DW has some problems with glottal control or with adjustment of vowel duration before voiced and voiceless consonants. The words with stop consonants in word-initial position have very good performance compared to the I.T. average of the 70-word list. Finally, the vowel duration and position errors show that DW cannot make good distinctions in the vowels. However, his vowel control impairments are not as serious as the other problems mentioned above. Finally, the words without any obstruent consonant have the best I.T. performance.

(ii) Transcription (T.T.)

From Table 4.1, the Chi-Square test value of T.T. is 0.62 (< 9.49 based on df = 4 and α = 0.05), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the transcription responses between two judges for DW's speech are not significantly different. From Table 4.9, the T.T. results for DW indicate the average accuracy judged by two experienced judges for the obstruent (21%) and sonorant (74%) consonants, identified correctly in every feature, in word-initial position. Vowel, voicing, and obstruent place are represented in the form of confusion matrices. The vowel confusions, which only focus on one feature identification, show that this speaker makes mild distinction errors between front vowels (76%), back vowels (76%), and diphthongs (75%). The confusions associated with the voiced - voiceless contrast, which addresses only one feature identification, show that this speaker does not have a serious glottal control problem for voiced consonants (94%) in word-initial position. However, there is a very serious confusion for the voiceless consonants (37%). The confusions for place of articulation of obstruents indicate that palatal position control is the most serious one (8%) compared to labial (88%), alveolar (72%), and velar (63%) control. More detailed study shows that most of these errors are made by fricative or affricate consonants. This is supported by the similar conclusions from the I.T. Both tests indicate severe alveolar

Obstruent Accuracy: 21%

Sonorant Accuracy: 74%

Vowels

T. \ R.	Front	Back	Middle	Diph.	[]
Front	76%	16%		7%	1%
Back	6%	76%	3%	15%	
Diph.	17%	8%		75%	

Voicing

T. \ R.	+	-	[ ]
+	94%	4%	2%
-	63%	37%	

Place for Obstruents

T. \ R.	Lab.	Dent.	Alv.	Pal.	Vel.	[ ]
Lab.	88%		8%			4%
Alv.	3%	13%	72%		3%	9%
Pal.			33%	8%	58%	
Vel.			13%	13%	63%	13%

Table 4.9 DW's transcription data for vowel, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position. The accuracies of obstruent and sonorant consonants are percent correct identification of the obstruent and sonorant consonants in word-initial position. The target items are listed in the left column and the response items are displayed across the top.

and palatal articulatory position control problems, especially for fricative / affricate consonants.

Overall, obstruent accuracy (21%) is much worse than sonorant accuracy (74%) for consonants in word-initial position. The data for obstruents in word-initial position show that this subject has much difficulty in pronouncing fricative and affricate consonants. This observation is consistent with the results from Figures 4.7-a and 4.7-b, which also show a high percentage of errors for alveolar - palatal place and fricatives / affricates.

(iii) Recognition Test (R.T.)

From Table 4.2, the R.T. results for DW indicate that the average accuracy for the 70 words is 30% compared with the I.T. of 57%. Further, groups SONOR_IN (34%), WTOT_OB (37%), and OBS_IN (26%) show that DW has a somewhat worse articulatory consistency control problem for words with an obstruent consonant in word-initial position. This is similar to the finding of the I.T. described above. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (23%) and CLUS_IN (23%) groups appear to show slightly more serious consistency control problems than the STOP_IN (29%). More detailed study of the FR_AF_IN shows that /ch/ in word-initial position (7%) has the lowest accuracy in this group. /s/ in word-initial position (20%) also is worse than the average R.T. accuracy (30%). The STOP_IN accuracy (29%) is similar to the average R.T. accuracy (30%). Table 4.3 shows the detailed R.T. analysis for the STOP_IN group which includes LABIAL (35%), ALVEOLAR (33%), and VELAR (15%) stop consonant groups. It appears that the words with labial and alveolar stop consonants in word-initial position have better recognition performance than velar consonants, and also better than the R.T. average of the 70-word list (30%). This is consistent with the findings in the I.T. On the other hand, the R.T. rejection rate of the 70-word list is 43% (Table 4.1) compared with the normal subject's 0.3%. It shows that DW's pronunciation of the words is very inconsistent.

In summary, this subject has particular consistency control difficulties with /ch/ in word-initial position (7%), /s/ in word-initial position (20%), and initial clusters (23%) in comparison with the average R.T. of the entire word list (30%). In addition, the words without any obstruent consonant (37%) have the best R.T. performance in comparison with the SONOR_IN and OBS_IN groups. The words with stop consonants in word-initial position have a fair performance compared to the R.T. average of the 70-word list.

(iv) Acoustic Analysis (A.A.)

Figures 4.8-a and 4.8-c show DW's spectrograms corresponding to "sip" and "chair". Figures 4.8-b and 4.8-d show MJ's, the normal subject's, spectrograms corresponding to "sip" and "chair". Figure 4.8-a at 200 ms shows that DW's glottis has begun to vibrate before the release of /s/ (390 to 560 ms) and keeps vibrating for the whole period of /s/. This might be the reason that three of the five judges confused DW's "sip" with "zip" in the I.T. Further, DW's /s/ in "sip" is more like an aspirated consonant /h/ rather than /s/ because it lacks the strong high-frequency noise ,and because the formants of this consonant are continuous with the following vowel, /I/. Figure 4.8-b shows that MJ's spectrogram of "sip" has strong high frequency energy around 5.5 kHz for /s/, produced by the turbulence flow passing through the tongue tip constriction,

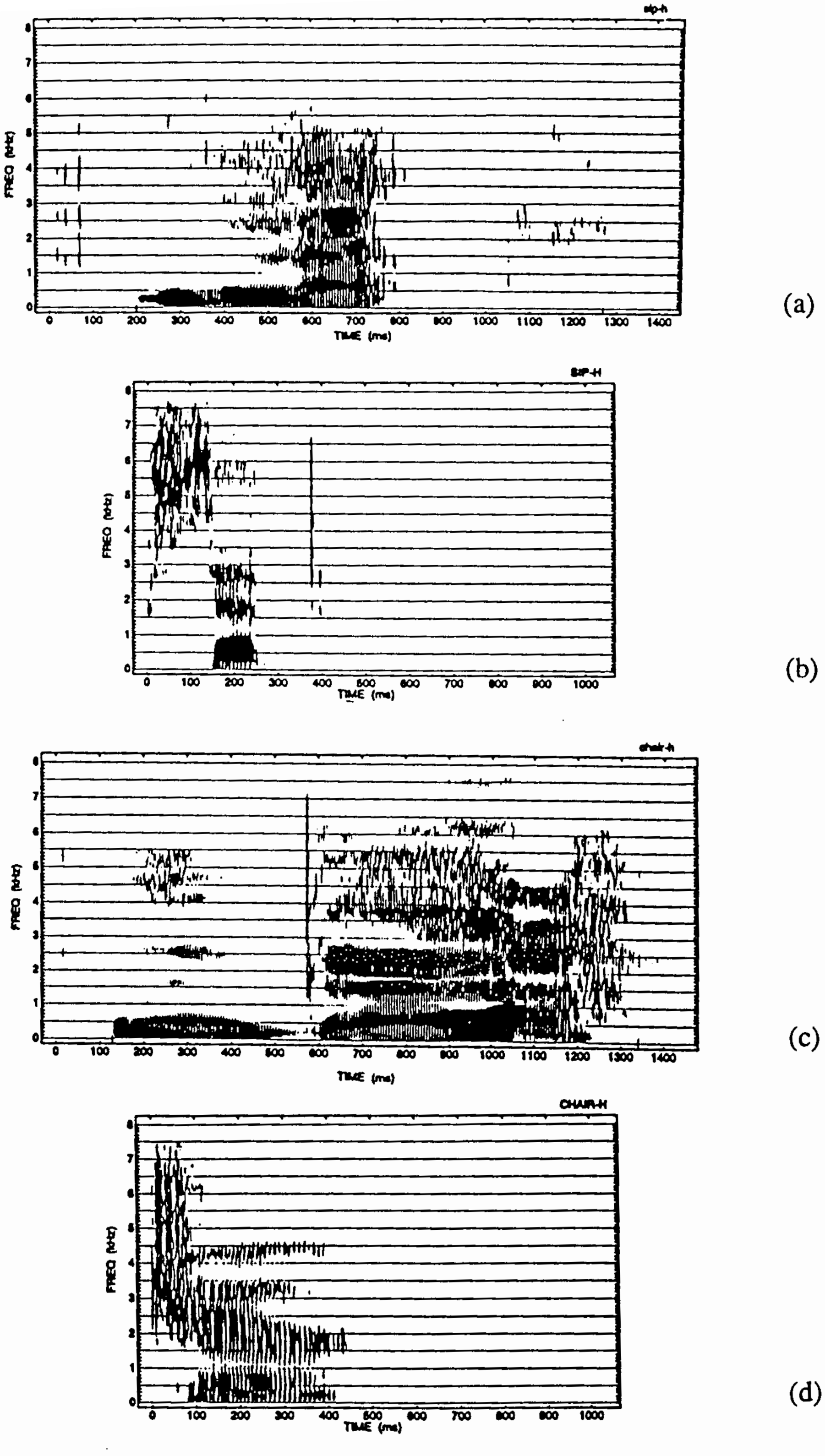


Figure 4.8: DW's spectrograms of "Sip" (4.8-a) and "Chair" (4.8-c) compared with MJ's spectrograms of "Sip" (4.8-b) and "Chair" (4.8-d).

without any glottal vibration at all. Further, DW has some noise from 0 to 80 ms. This happens very often in DW's speech.

Figure 4.8-c shows that the frication part of /ch/ is missing and there are only the release and aspiration parts (570 to 600 ms) for DW's "chair". It also shows noise with voiced vibration from 120 ms to 570 ms. Therefore, the glottis vibrates before the release of the affricate /ch/ of "chair" in the same way as the pre-voicing of "sip". This pre-voicing problem is often found in DW's speech. Figure 4.8-d shows MJ's spectrogram of "chair". The release of /ch/ starts around 0 ms and is followed by the frication part of /ch/ with strong high-frequency noise. For these reasons, four of the five judges heard /t/ rather than /ch/ in word-initial position for DW's utterance of "chair" and chose the response "tear".

Figures 4.8-a and 4.8-c show that there is mild aspiration noise in the vowel especially above 2.5 kHz, indicating that the subject's speech is breathy. However, DW's speech is not so breathy as that of JS.

The average duration measurement based on the spectrograms for the 70 diagnosis words is 1195 ms compared with MJ's 478 ms. The prolongation ratio is 2.5; that is, the duration of DW's words is 2.5 times that of the subject who does not have dysarthria, MJ, on the average.

4.4.3 Findings and Design Criteria

In conclusion, the four tests described above (I.T., T.T., R.T., and A.A.) show that DW has control problems in alveolar and palatal obstruents in word-initial position (for the fricatives / affricates especially), consonant clusters in word-initial position, and voicing contrast in word-initial and -final positions. All of the vowels (front vowels, back vowels, and diphthongs) are equally distinctive but have some duration and position control impairments. Both the I.T. and R.T. show that the words without any obstruent consonants and the words with sonorant consonants in word-initial position have a better performance than the words with obstruents in word-initial position. This observation is important for the design of a new word list. The words with stop consonants in word-initial position also have a fair performance. Labial and alveolar stop consonants in word-initial position have a better I.T. and R.T. performance than velar stop consonants.

Moreover, the problems with vowel distinctions, particularly vowel duration, should also be considered in the word-list design. Vowels that are at extreme positions in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, should be chosen to let the vowels be as distinctive as possible and to avoid lax - tense confusions. Further, because DW's vowels show similar accuracy for the front ones, back ones, and diphthongs, these three types of vowels have the same priority when the new word list is designed.

From the A.A., Figures 4.8-a and 4.8-c both show that DW often speaks with a heavy voiced breathing noise or a slurp noise coming before the words or utterances, e.g., 0 to 80 ms for Figure 4.8-a and 120 to 570 ms in Figure 4.8-c. These involuntary noises could cause the speech recognizer to become confused or to recognize the noises as input utterances. Furthermore, DW not only has very serious consistency control problems (43% rejection rate for the 70 diagnosis words), but also variability in timing for his speech. These two reasons indicate that there may be severe limitations in the accuracy that can be achieved with a new word list.

4.4.4 Improvement of Speech Recognition Accuracy

From the findings described above for DW, a new word list with 52 words was designed for him to dictate to the computer. Table 4.1 shows that the recognition accuracy for the new word list has been improved from 30% (R.T.(%)) based on the 70 diagnosis words to 53% (New_L(%)) based on the new 52-word list (or 36% for R.T.-35(%) and 57% for New_L-35(%) based on the same vocabulary size, 35). An improvement of 23% (Improve(%)) or 21% (Improve-35(%)) has been achieved. The final list is in Appendix 4.4.

Unfortunately, even though there is a huge improvement (23%) for the recognition accuracy, the 53% accuracy for the 52-word list is still not good enough to dictate all the new words to the computer because the accuracy is still too low. More detailed study shows that the main reason for getting such low recognition accuracy for this new word list comes from the high rejection rate (36%) compared to the misrecognition rate (11%). The word distinction for this new list is almost as high as possible. There is not too much space for improvement for the new word-list design unless the high rejection rate problem can be solved. Additionally, his speech varies with time, especially for the words or utterances with more than one syllable. The time variation problem is another important difficulty in DW's dictation of this new word list. The heavy voiced breathy noises or slurp noises coming before the utterances also cause the speech recognition rate to be reduced.

Therefore, DW just uses some words with good stability, time-consistency, and no pre-breathy (or slurp) noises from the 52 words to dictate to the computer. Moreover, a special function, ADAPT, in the DragonLab offered by the Dragon Writer-1000 was used to adapt DW's language model which was built by using five of the ten tokens for each of the 52 words. The main function of ADAPT is to adapt a model from a set of new tokens. Two sentences, "I am brilliant." and "No one here will run away." were dictated by following the alphabet spelling, e.g., "air" is for "a" and "bad" is for "b". In the dictation test, only the words used in dictating Sentence I (or II) were activated, in order to reduce the active word list size and the possibility of misrecognition. Since his speech had serious time variation, one extra token of each active word was trained to adapt the language model before starting the test. After finishing the adapting-model procedure, the formal dictation tests were recorded. DW's dictation speed for these two sentences was compared with his usual typing speed, using his index fingers. Table 4.4 shows the results of the test based on the average of three tries. His average typing time for sentence I is 24 seconds compared with a dictating time of 56 seconds; for sentence II, the typing time is 40 seconds compared with a dictating time of 1 minute and 37 seconds. These data show that DW's typing speed is two times faster than his speech dictation. However, the speech input can still be a useful tool for him to control the computer. He can simultaneously use typing and speech dictation methods to reduce the overload of using his index fingers and hands. For example, he can use his index fingers for typing characters and his speech dictation, which is based on a very limited word list, for changing different word fonts and sizes or choosing software function commands. However, in order to reduce the speech variability in timing, he needs to repeat the adapting-model procedure by using one extra token for each of the active words every time before he dictates to the computer. In addition to using the IBM PC, DW has tried to dictate the computer commands on his Macintosh by using Voice-Navigator II. By using the new words or utterances, DW has successfully controlled his Macintosh and has replaced some of the basic functions of the computer mouse to pick the command menu or edit a file, although the word list that he can use is very limited. Since his hands have involuntary movements and tremor, the use of the mouse or trackball is not as easy as it is for normal people. Therefore, the speech recognizer using the newly designed words or utterances has become an important tool for DW.

4.5 Subject 5 (ES)

4.5.1 Background and Symptoms

ES, who is 61, has a bachelor's degree. She has spastic cerebral palsy from birth. Her muscle movements are stiff, move sluggishly through a limited range, and are weak. The muscles have increased tone with heightened deep tendon reflexes. The behavior of fingers behavior is constringent. However, she can still ambulate by herself. She can use her right index finger, which is constringent, to type on the keyboard.

Her cognitive and linguistic abilities are intact, and her speech is the best of all eight subjects. However, her speech is still slow and seems to emerge with difficulty. She has an air flow and vital capacity control problem. Her speech becomes weak and decays in amplitude after continuous talk. Therefore, in separate utterances, her speech is quite clear and intelligible (I.T. = 97%), but not in continuous communication.

4.5.2 Analysis and Discussion

(i) Intelligibility Test (I.T.)

Table 4.1 shows that ES' intelligibility for I.T. is 97%. The Chi-Square test value of I.T. is 4.1 (< 9.49 based on $df = 4$ and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the listening responses between each of the five judges for ES's speech are not significantly different. Figure 4.9-a shows the summary of results from the intelligibility test based on contrast analysis of 19 phonetic groups. The two most frequent errors are Fricative - Affricate (8%) and Alveolar - Palatal Place (3%). Figure 4.9-b shows the summary of results from the intelligibility test based on contrast analysis of 40 phonetic groups. The top two most frequent errors are Affricate - Fricative Consonant (13%) and Palatal - Alveolar Place (10%). Because ES' speech is very mildly impaired, the error distributions do not show very serious confusion errors or special articulatory problems. The main errors are in the palatal articulatory position control (fricatives and affricates).

From Table 4.2, the I.T. results for ES indicate that the average accuracy for the 70 words is 97%. Further, groups SONOR_IN (99%), WTOT_OB (97%), and

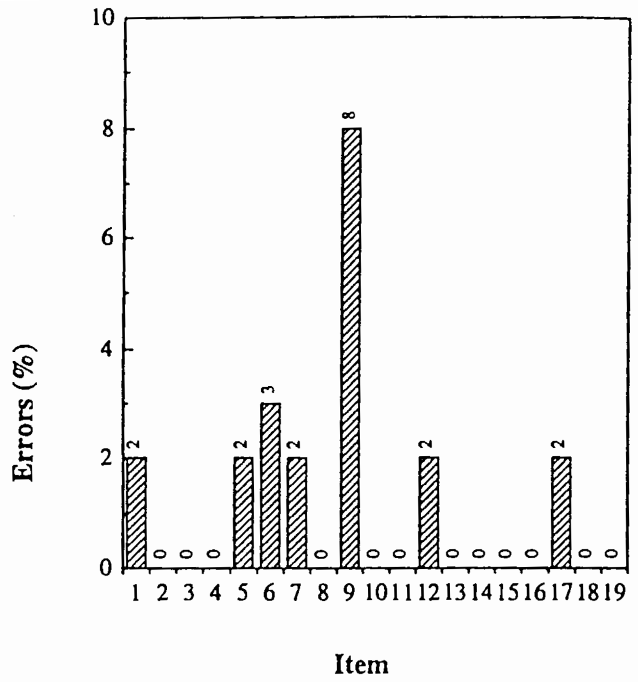


Figure 4.9-a ES' intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution shows the two most frequent errors are Fricative - Affricate (8%) and Alveolar - Palatal Place (3%).

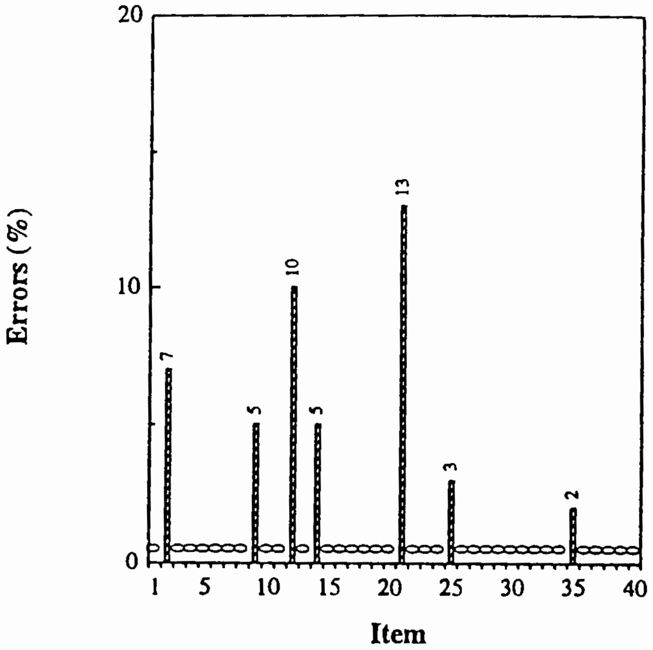


Figure 4.9-b ES' intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution shows the top two most frequent errors are Affricate - Fricative Consonant (13%) and Palatal - Alveolar Place (10%).

OBS_IN (96%) suggest that ES has a slightly worse articulatory position control problem for words with an obstruent consonant in word-initial position. The words with sonorant consonants in word-initial position (99%) have the best I.T. performance. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (94%) group shows a slightly more serious position control problem than the CLUS_IN (97%) and the STOP_IN (100%). More detailed study of the FR_AF_IN group shows that the /ch/_IN (73%) has the lowest accuracy in this FR_AF_IN group. /f/_IN (95%) and /sh/_IN (93%) are also below the I.T. average (97%). However, all of these consonant errors are very mild compared with the other subjects. Table 4.3 shows the detailed I.T. analysis for the STOP_IN group which includes LABIAL (100%), ALVEOLAR (100%), and VELAR (100%) stop consonant groups. It appears that all of the labial, alveolar, and velar stop consonants in word-initial position have the same good performance and have better I.T. performance than the I.T. average of the 70-word list (97%).

In summary, ES' speech is not very impaired. Her I.T. accuracy is the highest of the eight subjects. This subject only has very mild difficulties with obstruent consonants in word-initial position. ES' main articulatory errors concentrate on the palatal consonants (fricatives / affricates). The words with /ch/ in word-initial position have especially low accuracy in comparison with the average intelligibility of the entire word list. The words with a sonorant consonant in word-initial position have the best I.T. performance.

(ii) Transcription Test (T.T.)

From Table 4.1, the Chi-Square test value of T.T. is 0.07 (< 9.49 based on $df = 4$ and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the transcription responses between two judges for ES' speech are not significantly different. From Table 4.10, the T.T. results for ES indicate the average accuracy judged by two experienced judges for the obstruent (99%) and sonorant (96%) consonants, identified correctly in every feature, in word-initial position. Vowel, voicing, and obstruent place are presented in the form of matrices. The vowel confusions, which only focus on one feature identification, show that this speaker makes good distinctions between front vowels (100%), back vowels (100%), and diphthongs (100%). The confusions associated with the voiced - voiceless contrast, which addresses only one feature identification, show that this speaker does not have a glottal control problem for voiced-voiceless consonants in word-initial position. The confusions for place of

Obstruent Accuracy: 99%

Sonorant Accuracy: 96%

Vowels

T. \ R.	Front	Back	Middle	Diph.	[]
Front	100%				
Back		100%			
Diph.				100%	

Voicing

T. \ R.	+	-	[ ]
+	100%		
-		100%	

Place for Obstruents

T. \ R.	Lab.	Dent.	Alv.	Pal.	Vel.	[ ]
Lab.	100%					
Alv.			100%			
Pal.			8%	92%		
Vel.					100%	

Table 4.10 ES' transcription data for vowel, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position. The accuracies of obstruent and sonorant consonants are percent correct identification of the obstruent and sonorant consonants in word-initial position. The target items are listed in the left column and the response items are displayed across the top.

articulation for obstruents, which are based only one feature identification, indicate that palatal position control is the most serious one (92%) compared to labial (100%), alveolar (100%), and velar (100%) control. However, the palatal position control error (all made by the fricative / affricate consonants) is still mild.

Overall, obstruent accuracy (99%) is better than sonorant accuracy (96%) for consonants in word-initial position. The data for obstruents in word-initial position show that this subject does not have much difficulty with consonants that require pressure buildup behind a constriction in the vocal tract. However, the palatal confusions indicate a slight palatal position control error (Table 4.10). This observation is consistent with the results from Figures 4.9-a and 4.9-b, which also show some errors for alveolar - palatal place and fricative - affricate consonants.

(iii) Recognition Test (R.T.)

From Table 4.2, the R.T. results for ES indicate that the average accuracy for the 70 words is 80% compared with the I.T. of 97%. Further, groups SONOR_IN (88%), WTOT_OB (86%), and OBS_IN (72%) show that ES has the worst articulatory consistency control problem for words with an obstruent consonant in word-initial position. This is consistent with the finding of the I.T. described above. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (65%) group shows a more serious consistency control problem than the CLUS_IN (70%) and the STOP_IN (81%). More detailed study of the FR_AF_IN shows that /s/ in word-initial position (62%) has the lowest accuracy in this group. /ch/_IN (67%), /sh/_IN (67%), and /f/_IN (75%) all are worse than the average R.T. accuracy (80%). The STOP_IN accuracy (81%) is roughly equal to the average R.T. accuracy (80%). Table 4.3 shows the detailed R.T. analysis for the STOP_IN group which includes LABIAL (70%), ALVEOLAR (100%), and VELAR (90%) stop consonant groups. It appears that the words with alveolar and velar stop consonants in word-initial position have better recognition performance than labial consonants, and also better than the R.T. average of the 70-word list (80%). This is roughly consistent with the findings in the I.T. On the other hand, the R.T. rejection rate of the 70-word list is 10% (Table 4.1) compared with the normal subject's 0.3%. This rejection rate is lower than that of the other abnormal subjects, and shows that ES' speech has the most mild control problem.

In summary, this subject has particular consistency control difficulties with alveolar and palatal positions (especially for fricatives / affricates in word-initial position, e.g., /s/, /sh/, and /ch/) and with the initial clusters in comparison with the average R.T. of the entire word list. Further, the words with stop consonants in word-initial position also have a fair performance compared to the 70-word list. In addition, the words with sonorant consonants in word-initial position (88%) have the best R.T. performance, similar to the finding of the I.T.

(iv) Acoustic Analysis (A.A.)

Figures 4.10-a and 4.10-c show ES' spectrograms corresponding to "chop" and "cheer". Figures 4.10-b and 4.10-d show MJ's, the normal subject's, spectrograms corresponding to "chop" and "cheer". Figure 4.10-a shows an abrupt release (40 to 60 ms) and then the frication with high frequency energy of /ch/. Figure 4.10-b shows that

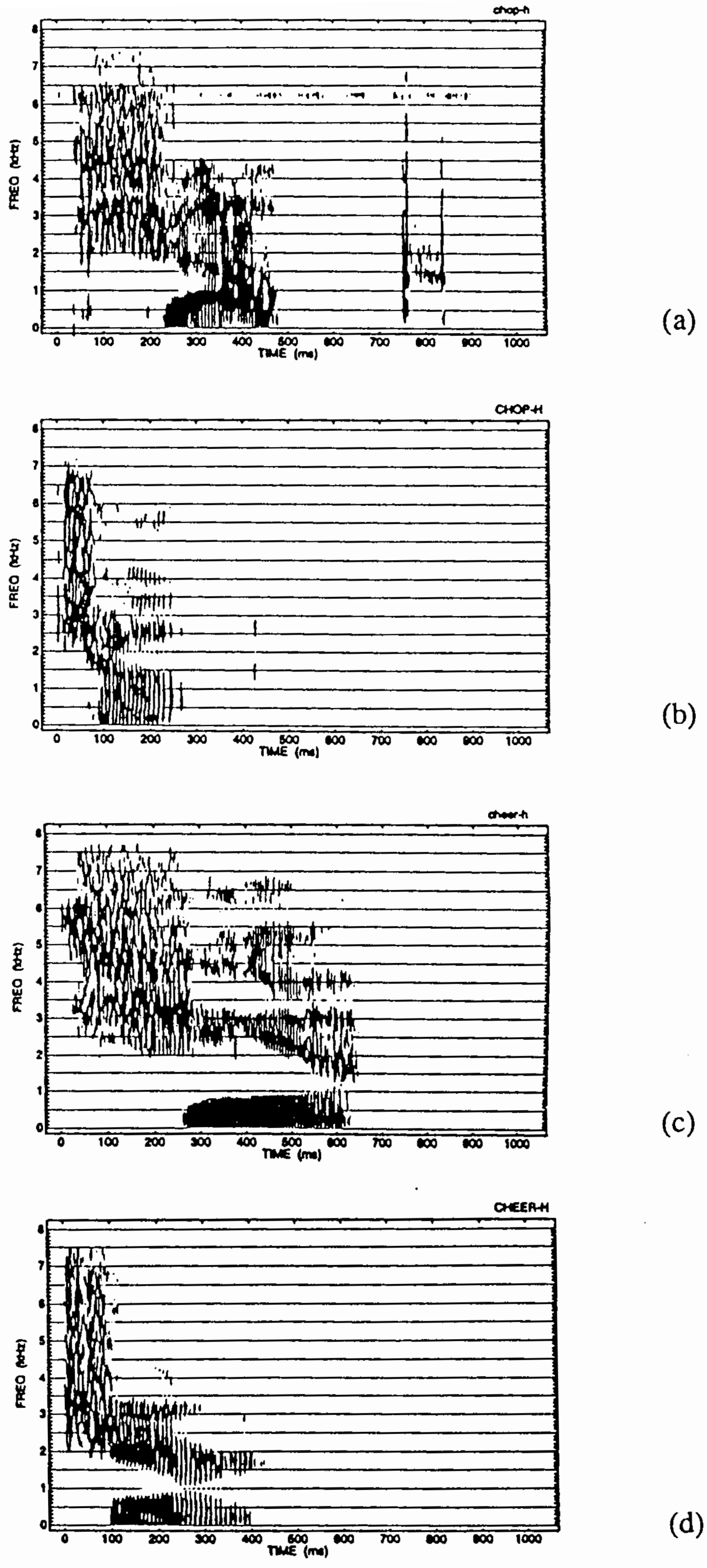


Figure 4.10: ES' spectrograms of "Chop" (4.10-a) and "Cheer" (4.10-c) compared with MJ's spectrograms of "Chop" (4.10-b) and "Cheer" (4.10-d).

MJ's spectrogram of "chop" has an abrupt release (0 to 10 ms) and a frication with strong high-frequency noise (10 to 70 ms). Comparison of these two figures shows similar characteristics for /ch/ in "chop". Therefore, the I.T. accuracy for this word is 100%. On the other hand, Figure 4.10-c shows that there is frication with high-frequency energy but lack of an abrupt release part for the /ch/ of "cheer". Figure 4.10-d shows MJ's spectrogram of "cheer" has both of the abrupt release and the frication of /ch/. It appears that ES' /ch/ of "cheer" looks like /sh/ not /ch/. This is the reason why three of the five judges confused ES' "cheer" with "sheer" in the closed intelligibility test, I.T. However, this situation is not as serious as the problems with other subjects' speech.

Figures 4.10-a and 4.10-c show that there is mild aspiration noise in the vowel especially above 2.5 kHz, indicating that the subject's speech is breathy. However, ES' speech is not so breathy as other subjects.

The average duration measurement based on the spectrograms for these 70 diagnosis words is 763 ms compared with MJ's 478 ms. The prolongation ratio is 1.6; that is, the duration of ES' words is 1.6 times that of the subject who does not have dysarthria, MJ, on the average.

4.5.3 Findings and Design Criteria

In conclusion, the four tests described above (I.T., T.T., R.T., and A.A.) show that ES has only mild control problems in producing fricatives / affricates in word-initial position (especially in the alveolar and palatal positions) and initial clusters. She also has a very slight error rate for the distinctions in back and front vowels (7% in Figure 4.9-b). Both the I.T. and the R.T. show that the words without any obstruent consonants and the words with sonorant consonants in word-initial position have a better performance than the words with obstruents in word-initial position. This observation is important for the design of a new word list. Furthermore, the words with stop consonants in word-initial position also have a fair performance compared with the FR_AF_IN group. The words with alveolar and velar stop consonants in word-initial position have a better recognition performance than the ones with labial stop consonants. The vowel distinction problem should also be considered in the word-list design. Vowels that are at extreme positions in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, should be chosen to let the vowels be as distinctive as possible. Although the back vowels have slight confusion with the front vowels in Figure 4.9-b, the front and back vowels are still considered as the same priority since the T.T. shows these two types of vowels do not have any difference in performance. Overall, ES does not have seriously impaired speech.

4.5.4 Improvement of Speech Recognition Accuracy

From the findings described above for ES, a new word list with 52 words was designed for her to dictate to the computer. Table 4.1 shows that the recognition accuracy for the new word list has been improved from 80% (R.T.(%)) based on the 70 diagnosis words to 93% (New_L(%)) based on the new 52-word list (or 88% for R.T.-35(%) and 91% for New_L-35(%) based on the same vocabulary size, 35). An improvement of 13% (Improve(%)) or 3% (Improve-35(%)) has been achieved. After designing this new list and following Part II of the flow chart in Figure 3.1, some of the words which were still unsatisfactory were modified. The final list is in Appendix 4.5.

Two sentences, "I am brilliant." and "No one here will run away." were dictated by following the alphabet spelling, e.g., "air" is for "a" and "bad" is for "b". The dictation speed for these two sentences was compared with her usual typing method, using her right index finger. Table 4.4 shows the results of the test. Her average typing time for sentence I is 31 seconds compared with a dictating time of 31 seconds; for sentence II, the typing time is 1 minute and 9 seconds compared with a dictating time of 1 minute and 50 seconds. These data show that ES' dictation speed is close to but not better than the typing speed. In spite of this result, the speech input can still be a useful tool for her to control the computer. She can watch simultaneously the computer monitor when she uses the computer by using the speech recognizer and is free from the painful typing method by using only one constringent finger. In addition, she can use her speech command to replace the function of the mouse and operate efficiently the computer in the future since her awkward and involuntary movements make control of the mouse difficult.

4.6 Subject 6 (JR)

4.6.1 Background and Symptoms

JR, who is 22, is studying in an undergraduate school. At birth, JR had evidence of athetosis and spastic cerebral palsy. However, her symptoms showed that the evidence for spastic is more apparent. Her neuromotor condition is more characteristic of spastic cerebral palsy: the muscles are stiff and the movements awkward. The muscles have increased tone with heightened deep tendon reflexes, followed by contraction of the fingers and rotation of the wrist. Moreover, the involuntary movements of the articulatory and pharyngeal muscles indicate that she should be characterized as dysarthria and dysphagia (Brain, 1969). She uses her right thumb most of the time and her left index finger sometimes 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 the face and the involuntary movements of the 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 interruption of breathing. The slow, rasping, and labored speech comes out with a large range of jaw movement, and each word is prolonged. JR's speech sounds weak to the unfamiliar listener and less throaty than that of JS' speech. Her cognitive and linguistic abilities are intact.

4.6.2 Analysis and Discussion

(i) Intelligibility Test (I.T.)

Table 4.1 shows that JR's intelligibility for I.T. is 64%. The Chi-Square test value of I.T. is 13.4 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the listening responses between each of the five judges for JR's speech are significantly different. JR is the only subject whose Chi-Square test value shows significantly different between the judges' responses. Figure 4.11-a shows the summary of results from the intelligibility test based on contrast analysis of 19 phonetic groups. The five most frequent errors are Alveolar - Palatal Place (45%), /r/ - /w/ (33%), Final Voicing Contrast (26%), Initial Cluster - Singleton (25%), and Fricative

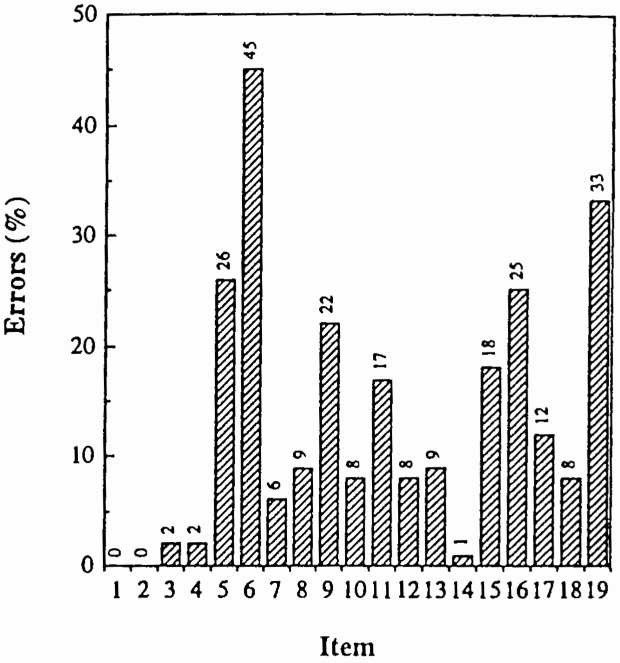


Figure 4.11-a JR's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution shows the five most frequent errors are Alveolar - Palatal Place (45%), /r/ - /w/ (33%), Final Voicing Contrast (26%), Initial Cluster - Singleton (25%), and Fricative - Affricate (22%).

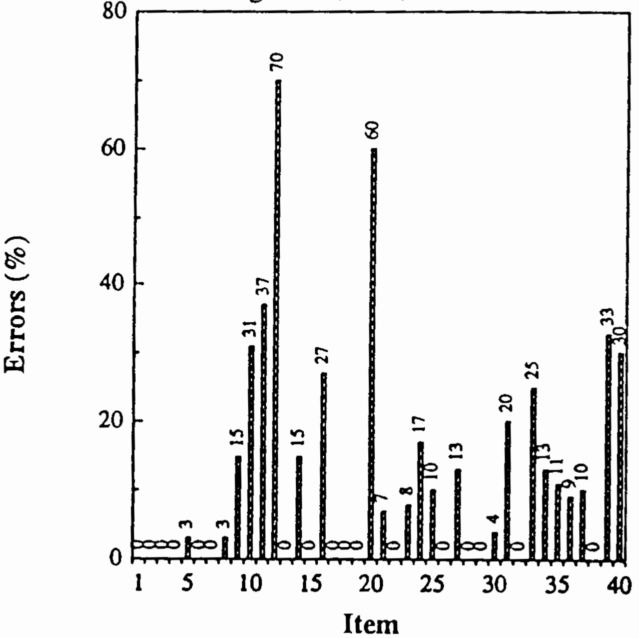


Figure 4.11-b JR's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution shows the top eight most frequent errors are Palatal - Alveolar Place (70%), Fricative - Affricate Consonant (60%), Affricate - Stop Consonant (53%), Alveolar - Palatal Place (37%), /r/ - /w/ (33%), Final Voiceless - Voicing Contrast (31%), /w/ -/r/ (30%), Alveolar - Other Fricative Place (27%), and Initial Cluster - Singleton (25%).

- Affricate (22%). Figure 4.11-b shows the summary of results from the intelligibility test based on contrast analysis of 40 phonetic groups. The top eight most frequent errors are Palatal - Alveolar Place (70%), Fricative - Affricate Consonant (60%), Affricate - Stop Consonant (53%), Alveolar - Palatal Place (37%), /r/ - /w/ (33%), Final Voiceless - Voiced Contrast (31%), /w/ -/r/ (30%), Alveolar - Other Fricative Place (27%), and Initial Cluster - Singleton (25%). Because JR's speech is seriously impaired (I.T.: 64%), the error distributions show large confusion errors and special articulatory problems. The main errors concentrate on the fricative / affricate control problems (especially for the alveolar and palatal positions). Initial cluster, final voiceless - voiced, and /r/ - /w/ control are serious too. The percentage error for Final Voiceless - Voiced Contrast indicates that this subject has problems with glottal control or with adjustment of vowel duration before voiced and voiceless consonants. Further, JR might also have some problems with tongue-blade control because of the above mentioned problems and the high percentage errors of /r/ - /w/. The percentage error for /r/ - /w/ (33% in Figure 4.11-b) indicates that this subject has particular difficulty shaping the tongue blade to produce a retroflex consonant /r/.

From Table 4.2, the I.T. results for JR indicate that the average accuracy for the 70 words is 64%. Comparing groups SONOR_IN (67%), WTOT_OB (97%), and OBS_IN (62%), JR evidences a worse articulatory position control problem for words with an obstruent consonant in word-initial position. The words without any obstruent consonant (97%) have the best I.T. performance. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The CLUS_IN (47%) group has the worst performance in these three sub-groups. The FR_AF_IN (53%) group shows a more serious position control problem than the STOP_IN (76%). More detailed study of the FR_AF_IN group shows that the /sh/_IN (27%) and /s/_IN (38%) are the most serious two in the FR_AF_IN group. Table 4.3 shows the detailed I.T. analysis for the STOP_IN group which includes LABIAL (70%), ALVEOLAR (87%), and VELAR (80%) stop consonant groups. It appears that the words with alveolar and velar stop consonants in word-initial position have better I.T. performance than labial consonants, and also better than the I.T. average of the 70-word list (64%).

In summary, this subject has particular difficulties with obstruent consonants in word-initial position. JR's tongue-blade control is impaired. The alveolar and palatal position control in word-initial position is particularly impaired, especially for fricatives and affricates. The words with /s/ in word-initial position, /sh/ in word-initial position, and initial clusters have low accuracy in comparison with the average intelligibility of the entire word list. Further, the errors of final voiceless - voiced contrast show that JR has some problems with glottal control or with adjustment of vowel duration before voiced and voiceless consonants. The words with stop consonants in word-initial position have good performance compared to the I.T. average of the 70-word list. Finally, the words without any obstruent consonant have the best I.T. performance.

(ii) Transcription Test (T.T.)

From Table 4.1, the Chi-Square test value of T.T. is 0.34 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the transcription responses between two judges for JR's speech are not significantly different. From Table 4.11, the T.T. results for JR indicate the average accuracy judged by two experienced judges for the obstruent (34%) and sonorant (56%) consonants, identified correctly in every feature, in word-initial position. Vowel, voicing, and obstruent place are presented in the form of matrices. The vowel confusions, which only focus on one feature identification, show the following results: front vowels (85%), back vowels (62%), and diphthongs (83%). The data show that the back vowels have the worst position control. Further, both front and back vowels have confusion errors with diphthongs: front vowel - diphthong (10%) and back vowel - diphthong (29%). The confusions associated with the voiced - voiceless contrast, which addresses only one feature identification, show that this speaker does not have a serious glottal control problem for voiceless consonants (92%) in word-initial position, but the glottal control problem for voiced consonants (69%) in word-initial position causes a serious confusion. This is just inverse to the I.T. test conclusion about the voicing control in word-initial position (Init. Voiced - Voiceless Contrast 0% in Figures 4.11-b). With more detailed study, it was found that the I.T. only studied the voiced - voiceless confusion for the obstruent consonants, but the T.T. studied both the obstruent and sonorant consonants for the voiced - voiceless confusion. JR pronounces some of the voiced sonorants as voiceless consonants. The confusions for place of articulation for obstruents, which are based only one feature identification, indicate that palatal (0%) and alveolar (3%) position control both are impaired compared to labial (96%) and velar (100%) control. It shows that JR has a serious position control problem for alveolar and palatal regions (especially for fricatives / affricates). Figures 4.11-a and 4.11-b both supported the same problems for the alveolar and palatal position control errors. JR has the inclination to pronounce

Obstruent Accuracy: 34%

Sonorant Accuracy: 56%

Vowels

T. \ R.	Front	Back	Middle	Diph.	[]
Front	85%	5%		10%	
Back	9%	62%		29%	
Diph.	17%			83%	

Voicing

T. \ R.	+	-	[ ]
+	69%	19%	12%
-	5%	92%	3%

Place for Obstruents

T. \ R.	Lab.	Dent.	Alv.	Pal.	Vel.	[ ]
Lab.	96%				4%	
Alv.			3%	3%	75%	19%
Pal.		8%		0%	83%	8%
Vel.					100%	

Table 4.11 JR's transcription data for vowel, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position. The accuracies of obstruent and sonorant consonants are percent correct identification of the obstruent and sonorant consonants in word-initial position. The target items are listed in the left column and the response items are displayed across the top.

alveolar and palatal obstruents as velars. A possible reason is given below in the A.A. section.

Overall, obstruent accuracy (34%) is worse than sonorant accuracy (56%) for consonants in word-initial position. The data for obstruents in word-initial position show that this subject has much difficulty in pronouncing alveolar and palatal consonants (especially for fricatives or affricates), a result that is supported by Figures 4.11-a and 4.11-b. Further, JR stops her glottal vibration in some of her voiced sonorants in word-initial position.

(iii) Recognition Test (R.T.)

From Table 4.2, the R.T. results for JR indicate that the average accuracy for the 70 words is 59% compared with the I.T. of 64%. Comparison of groups SONOR_IN (74%), WTOT_OB (71%), and OBS_IN (49%) shows that JR evidences a worse articulatory consistency control problem for words with an obstruent consonant in word-initial position than for words with sonorant-initial consonants. This is consistent with the finding of the I.T. described above. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (39%) group shows a more serious consistency control problem than the CLUS_IN (47%) and the STOP_IN (65%). More detailed study of the FR_AF_IN shows that /s/ in word-initial position (31%) has the lowest accuracy in this group. /f/_IN (45%) and /sh/_IN (47%) both are worse than the average R.T. accuracy (59%). The STOP_IN accuracy (65%) is better than the average R.T. accuracy (59%). Table 4.3 shows the detailed R.T. analysis for the STOP_IN group which includes LABIAL (68%), ALVEOLAR (80%), and VELAR (50%) stop consonant groups. It appears that the words with labial and alveolar stop consonants in word-initial position have better recognition performance than velar consonants, and also better than the R.T. average of the 70-word list (59%). This is not very consistent with the findings in the I.T. On the other hand, the R.T. rejection rate of the 70-word list is 17% (Table 4.1) compared with the normal subject's 0.3%. It shows that JR's speech has some control problems with inconsistency.

In summary, this subject has particular consistency control difficulties with fricatives in word-initial position, e.g., /f/, /s/, /sh/, and initial clusters in comparison with the average R.T. of the entire word list (59%). Further, the words with a sonorant consonant in word-initial position (74%) have the best R.T. performance. This subject is one of the two subjects, JR and GG, whose best performance among the groups, SONOR_IN, WTOT_OB, and OBS_IN, is not the same for the I.T. and the R.T., i.e., WTOT_OB (97%) has the best performance of these three groups for the I.T. but SONOR_IN (74%) has the best R.T. performance for the three groups.

(iv) Acoustic Analysis (A.A.)

The T.T., mentioned above, has shown that JR's obstruents in word-initial position tend to be pronounced as velar consonants, e.g., velar stop and fricative consonants. The place analysis of obstruents for the T.T. has not shown very serious palatal - alveolar confusion (only 3% alveolar - palatal consonant confusion). However, the I.T. shows the high accuracy errors for alveolar - palatal error and fricative - affricate error. Figures 4.12-a, 4.12-b, 4.12-c, and 4.12-d show JR's spectrograms corresponding to "shoot", "sip", "ship", and "hat". These spectrograms can help to explain the conflict mentioned above. Figures 4.12-a to 4.12-c all show fricative consonants with a formant around 2K to 2.5K Hz. From theory, the acoustic constriction for this fricative consonant could be in the velar position because the formant frequency is so low. Since there is no apparent release part in the beginning of this consonant, this fricative is possibly like a voiceless velar fricative, /x/. The T.T. shows that one judge transcribed the /sh/ of "shoot", Figure 4.12-a, as /kx/ and the other judge as /x/. On the other hand, three of the five judges in the I.T. chose "suit" as the response of "shoot". Because there is no word with /x/ in word-initial position in the four-choose-one word list, these judges in the I.T. chose only the closest word, "suit", as the response. This is one of the reasons why the I.T. shows a high percentage error of the Alveolar - Palatal Place cluster (45% in Figure 4.11-a) but the T.T. shows a high percentage error of the Palatal - Velar confusion (83% in Table 4.11). In the spectrogram of "sip", Figure 4.12-b, there is still a low frequency formant at 2.1K Hz from 20 to 240 ms. The T.T. shows that one judge transcribed the /s/ of "sip", Figure 4.12-b, as /x/ and the other judge as /h/. However, three of the five judges in the I.T. chose "tip" as the response of "sip" and one of the five judges chose "ship" as the response, for the same reason as "shoot". Further, with more detailed study of the T.T., it appears some of these fricative consonants were transcribed as stop consonants, followed by the velar fricative /x/, e.g., /kx/. Thus, this fricative noise sometimes could be possibly considered as a stop consonant. For example, the word "ship" shown in Figure 4.12-c is transcribed as /k/ by one judge and /h/ by the other judge in the T.T., but three of the five judges in the I.T. chose "chip" as the response and one of the five judges chose "tip" as the response. Thus, the I.T. shows a high percentage error for Fricative - Affricate cluster in Figure 4.11-a, but The T.T. shows a high percentage error for Palatal - Velar confusion and Palatal - [] confusion, Table 4.11. Further, Figure 4.12-d shows a regular /h/ of JR's "hat". From 20 to 100 ms, there is aspiration noise with formants which are the trajectory of the following vowel's formants. This is a "good" /h/ compared with other fricative consonants for JR's speech. Both the I.T. and T.T. show that all of five judges in the I.T. have correct responses and both of the two judges in the T.T. transcribed the aspiration consonant of "hat" as /h/.

In summary, from the A.A., one learns that this subject has a particular trend for using the velar position for the obstruents. This velar consonant with frication noise is sometimes transcribed as /x/, /k/, /h/, or even /kx/ and /ch/ because there is not an apparent release in this "new" velar consonant.

The average duration measurement based on the spectrograms for these 70 diagnosis words is 812 ms compared with MJ's 478 ms. The prolongation ratio is 1.7; that

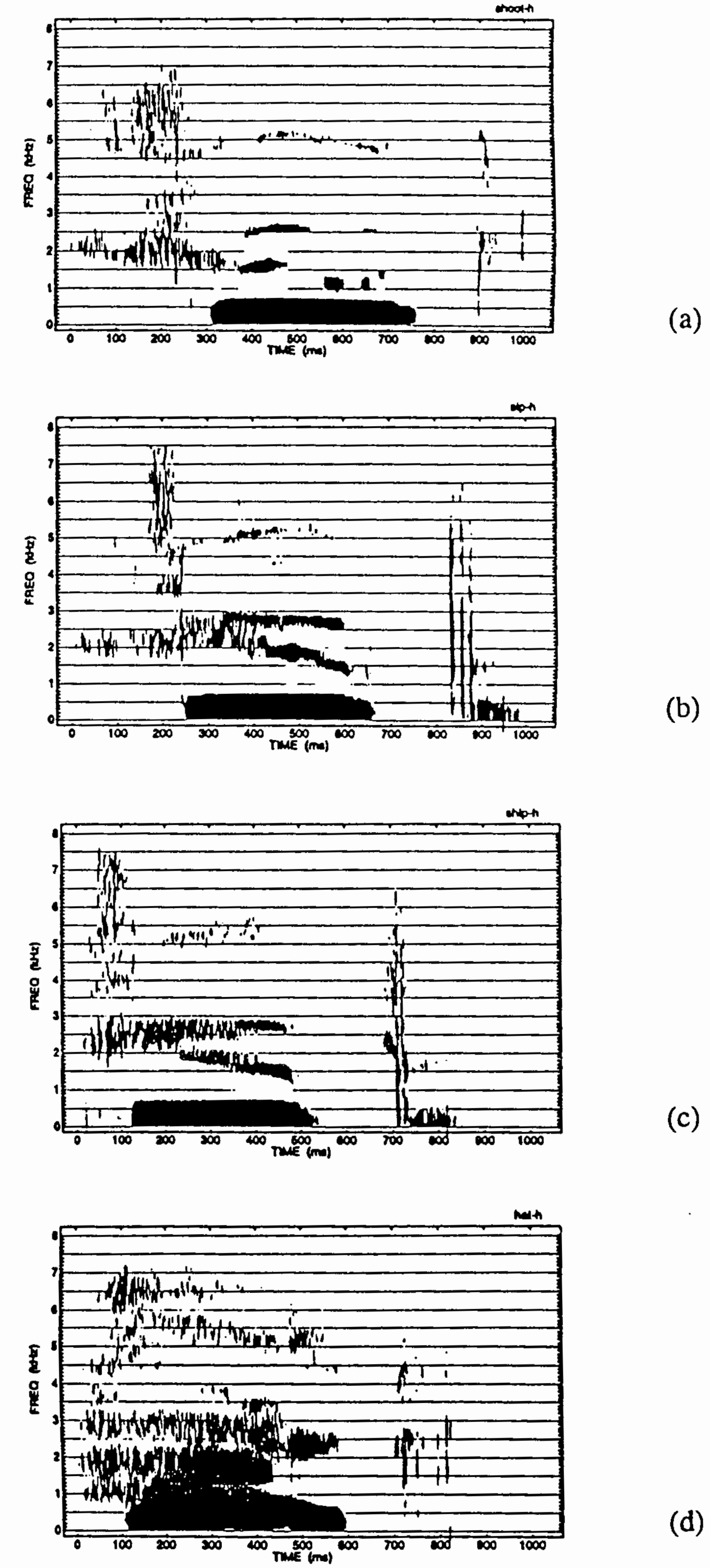


Figure 4.12: JR's spectrograms of "Shoot" (4.12-a), "Sip" (4.12-b), "Ship" (4.12-c), and "Hat" (4.12-d).

is, the duration of JR's words is 1.7 times that of the subject who does not have dysarthria, MJ, on the average.

4.6.3 Findings and Design Criteria

In conclusion, the four tests described above (I.T., T.T., R.T., and A.A.) show that JR has control problems in alveolar and palatal obstruents in word-initial position (especially for the fricatives / affricates), tongue blade control, and glottal control in word-initial and -final positions. She is inclined to pronounce alveolar and palatal obstruents as velars. Both the I.T. and the R.T. show that the words without any obstruent consonants and the words with sonorant consonants in word-initial position have a better performance than the words with obstruents in word-initial position. This observation is important for the design of a new word list. The words with stop consonants in word-initial position also have a fair performance. Labial and alveolar stop consonants in word-initial position have a better recognition performance than velar stop consonants. The vowel distinction problems are also considered in the word-list design. Vowels that are at extreme positions in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, should be chosen to let the vowels be as distinctive as possible and to avoid lax - tense confusions. Further, because JR's front vowels and diphthongs are distinguished better than the back vowels, the back vowels should have a lower priority than the front vowels and diphthongs when the new word list is designed.

4.6.4 Improvement of Speech Recognition Accuracy

From the findings described above for JR, a new word list with 52 words was designed for her to dictate to the computer. Table 4.1 shows that the recognition accuracy for the new word list has been improved from 59% (R.T.(%)) based on the 70 diagnosis words to 73% (New_L(%)) based on the new 52-word list (or 64% for R.T.-35(%) and 78% for New_L-35(%) based on the same vocabulary size, 35). An improvement of 14% (Improve(%)) or 14% (Improve-35(%)) has been achieved. After designing this new list and following Part II of the flow chart in Figure 3.1, some of the words which were still unsatisfactory were modified. The final list is in Appendix 4.6.

Two sentences, "I am brilliant." and "No one here will run away." were dictated by following the alphabet spelling, e.g., "air" is for "a" and "bad" is for "b". The dictation speed for these two sentences was compared with her usual typing method, using her right thumb. Table 4.4 shows the results from the typing test. Her average typing time for sentence I is 33 seconds compared with a dictating time of 34 seconds; for sentence II, the typing time is 59 seconds compared with a dictating time of 1 minute and 33 seconds. These data show that JR's typing speed is faster than her speech dictation. Nevertheless, the speech input can be a potentially useful tool for her to control the computer. She can watch simultaneously the computer monitor when she uses the computer by using the speech recognizer and is free from the painful typing method by using her thumb. Especially, she can use her speech command to replace the function of mouse since her involuntary movements and constringent fingers are hard to use the mouse or joystick.

4.7 Subject 7 (CH)

4.7.1 Background and Symptoms

CH, who is 62, has a 5th grade education. At birth, CH had apparent spastic cerebral palsy. Her neuromotor condition is like JR's: the muscles are stiff, the movements awkward with heightened deep tendon reflexes, followed by contraction of the fingers and rotation of the wrist. She can use only her right index finger for typing. However, her speech was intact originally until ten years ago when she had surgery to remove the acoustic neuroma. After this operation, the left side of her face, the left side of her tongue, her left ear, and her 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 control of aspiration. CH's speech sounds very weak to the unfamiliar listener and more throaty than other subjects. Her speech volume is too weak even to be sampled by the speech recognizer. Some of the utterances come out with heavy breathy and explosive noise. With the effort of speech, her face grimaces, as though the sound is produced against considerable resistance.

Her cognitive and linguistic abilities are intact. However, she needs a reading assistant for pronunciation in the I.T. and R.T. because she can not pronounce spontaneously from reading the words on cards and monitors.

4.7.2 Analysis and Discussion

(i) Intelligibility Test (I.T.)

Table 4.1 shows that CH's intelligibility for the I.T. is 61%. The Chi-Square test value of I.T. is 3.2 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the listening responses between each of the five judges for CH's speech are not significantly different. Figure 4.13-a shows the summary of results from the intelligibility test based on contrast analysis of 19 phonetic groups. The five most frequent errors are Alveolar - Palatal Place (25%), Initial Glottal Consonant - Null (25%), Stop - Affricate (23%), /r/ - /w/ (23%), and Final Voicing Contrast (22%). Figure 4.13-b shows the summary of results from the intelligibility tes: based on contrast analysis of 40 phonetic groups. The top nine most frequent errors are Final Voiced - Voiceless Contrast (60%), Initial Null - Glottal Consonant (40%), Final Cluster - Singleton (40%), Alveolar - Palatal Place (30%), Fricative - Affricate Consonant (27%), Velar - Other Consonant Place (25%), Long - Short Vowel Duration (23%), Affricate - Stop Consonant (23%), and /r/ - /w/ (23%). Because CH's speech is seriously impaired (I.T.: 61%), the error distributions show significant confusion errors or special articulatory problems. The main errors concentrate on alveolar and palatal (especially for fricatives / affricates), velar, and larynx control problems, e.g., null - glottal and final voiceless - voiced confusion. On the basis of these data, it appears that CH has problems with tongue-blade control. The percentage error for /r/ - /w/ (23% in Figure 4.13-b) indicates that this subject has particular difficulty shaping the tongue blade to produce a retroflex consonant /r/. Because of CH's glottis paralysis, she often has heavy glottal explosive noise associated with her speech. The errors of final voiceless - voiced contrast show that CH has some problems with glottal control or with adjustment of vowel duration before voiced and voiceless consonants. Initial and final cluster control is a problem too. Further, the vowel duration errors show that CH cannot make good distinctions between long and short vowels.

From Table 4.2, the I.T. results for CH indicate that the average accuracy for the 70 words is 61%. Further, groups SONOR_IN (66%), WTOT_OB (80%), and OBS_IN (58%) show that CH has the most severe problem for words with an obstruent consonant in word-initial position. The words without any obstruent consonant (80%) have the best I.T. performance. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The CLUS_IN (50%) has the worst I.T. performance in these

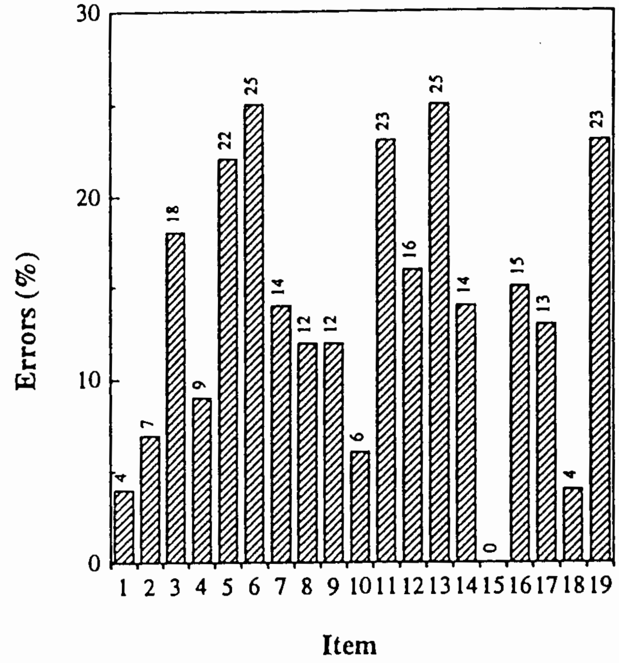


Figure 4.13-a CH's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution shows the five most frequent errors are Alveolar - Palatal Place (25%), Initial Glottal Consonant - Null (25%), Stop - Affricate (23%), /r/ - /w/ (23%), and Final Voicing Contrast (22%).

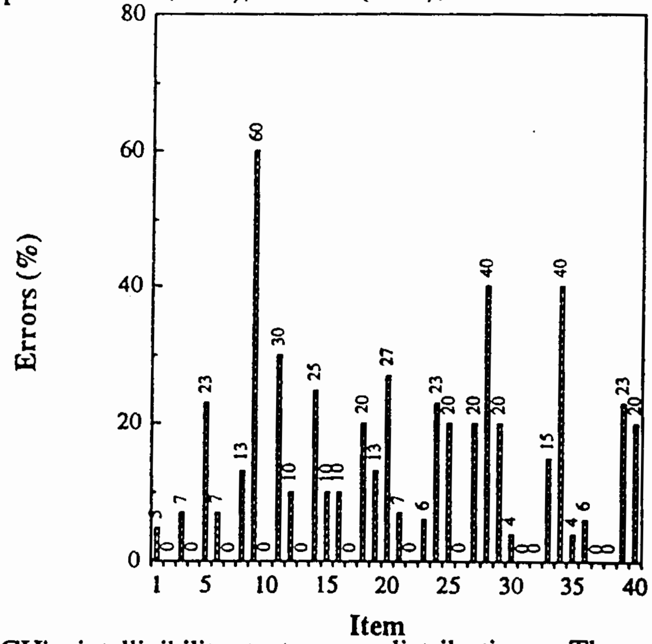


Figure 4.13-b CH's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution shows the top nine most frequent errors are Final Voicing - Voiceless Contrast (60%), Initial Null - Glottal Consonant (40%), Final Cluster - Singleton (40%), Alveolar - Palatal Place (30%), Fricative - Affricate Consonant (27%), Velar - Other Consonant Place (25%), Long - Short Vowel Duration (23%), Affricate - Stop Consonant (23%), and /r/ - /w/ (23%).

three sub-groups. The FR_AF_IN (55%) group shows a more serious position control problem than the STOP_IN (64%). More detailed study of the FR_AF_IN group shows that the /ch/_IN (47%) has a slightly lower than /s/_IN (52%) and /f/_IN (55%); all of these are lower than the average accuracy of the 70-word list (61%). Table 4.3 shows the detailed I.T. analysis for the STOP_IN group which includes LABIAL (60%), ALVEOLAR (53%), and VELAR (80%) stop consonant groups. It appears that the labial and velar stop consonants in word-initial position have better I.T. performance than alveolars.

In summary, this subject has difficulties with obstruent consonants and with larynx control in word-initial position. CH has particular difficulties with alveolar and palatal (especially for fricatives / affricates), and velar articulatory places in word-initial position in comparison with the average intelligibility of the entire word list. The labial fricative /f/ also has some impairment, presumably because of her lip paralysis. CH's tongue-blade and glottis control are impaired. The errors of final voiceless - voiced contrast show that CH has some problems with glottal control or with adjustment of vowel duration before voiced and voiceless consonants. Additionally, the clusters in word-initial and -final positions also are seriously impaired. The words with stop consonants in word-initial position have good performance compared to the I.T. average of the 70-word list. On the other hand, CH also has some vowel duration control problems. Finally, the words without any obstruent consonant have the best I.T. performance.

(ii) Transcription Test (T.T.)

From Table 4.1, the Chi-Square test value of T.T. is 8.84 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the transcription responses between two judges for CH's speech are not significantly different, but are very close to the critical value, 9.49, which cuts off an upper-tail area of 0.05 (Sincich, 1987) [p. 715]. From Table 4.12, the T.T. results for CH indicate the average accuracy judged by two experienced judges for the obstruent (16%) and sonorant (54%) consonants, identified correctly in every feature, in word-initial position. Vowel, voicing, and obstruent place are presented in the form of matrices. The vowel confusions, which only focus on one feature identification, show that this speaker makes errors in front vowels (83%), back vowels (82%), and diphthongs (67%). The data show that the diphthong vowels have the worst position control. Both the back vowels and diphthongs are confused with the front vowels (back - front vowels (12%) and diphthong - front

Obstruent Accuracy: 16%

Sonorant Accuracy: 54%

Vowels

T. \ R.	Front	Back	Middle	Diph.	[]
Front	83%	12%	1%	4%	
Back	12%	82%		6%	
Diph.	33%			67%	

Voicing

T. \ R.	+	-	[ ]
+	75%		25%
-	17%	53%	31%

Place for Obstruents

T. \ R.	Lab.	Dent.	Alv.	Pal.	Vel.	[ ]
Lab.	29%		4%			67%
Alv.	9%		28%	19%	16%	28%
Pal.	17%		8%	25%	25%	25%
Vel.	13%				0%	87%

Table 4.12 CH's transcription data for vowel, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position. The accuracies of obstruent and sonorant consonants are percent correct identification of the obstruent and sonorant consonants in word-initial position. The target items are listed in the left column and the response items are displayed across the top.

vowels (33%)). The confusions associated with the voiced - voiceless contrast, which addresses only one feature identification, show that this speaker has a glottal control problem for both the voiced (75%) and voiceless (53%) consonants in word-initial position. The confusions for place of articulation for obstruents, which are based only one feature identification, indicate that velar (0%), palatal (25%), alveolar (28%), and labial (29%) position control all are impaired compared to the other subjects. These data show that CH has a serious position control problem for obstruent consonants. In addition, many of the obstruent consonants in word-initial position disappeared in CH's speech. Figures 4.13-a and 4.13-b both support the same problems for the labial, alveolar, palatal, and velar position control errors.

Overall, obstruent accuracy (16%) is worse than sonorant accuracy (54%) for consonants in word-initial position. The data for obstruents in word-initial position show that this subject has much difficulty in pronouncing labial, alveolar, palatal and velar consonants - a finding that is supported by Figures 4.13-a and 4.13-b. The main reason may come from the paralysis of the left part of her neuron system.

(iii) Recognition Test (R.T.)

From Table 4.2, the R.T. results for CH indicate that the average accuracy for the 70 words is 26% compared with the I.T. of 61%. Further, groups SONOR_IN (49%), WTOT_OB (60%), and OBS_IN (17%) show that CH has the worst articulatory consistency control problem for words with an obstruent consonant in word-initial position. This is consistent with the finding of the I.T. described above. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (13%) group shows a more serious consistency control problem than the CLUS_IN (23%) and the STOP_IN (24%). With more detailed study of FR_AF_IN, it shows that /f/ in word-initial position (0%) has the lowest accuracy in this group. /s/_IN (15%), /sh/_IN (20%), and /ch/_IN (13%) all are worse than the average R.T. accuracy (26%). The STOP_IN also is a little lower than the average R.T. accuracy. Table 4.3 shows the detailed R.T. analysis for the STOP_IN group which includes LABIAL (28%), ALVEOLAR (27%), and VELAR (15%) stop consonant groups. It appears that the words with labial and alveolar stop consonants in word-initial position have better recognition performance than velar consonants, and also better than the R.T. average of the 70-word list (26%). This is not consistent with the findings for the I.T. in STOP_IN group, which shows that the words with alveolar stop consonants in word-initial position have the worst I.T. performance. Further, the R.T. rejection rate of the 70-word list is 33% (Table 4.1) compared with the normal subject's 0.3%. It shows that CH's speech is very inconsistent.

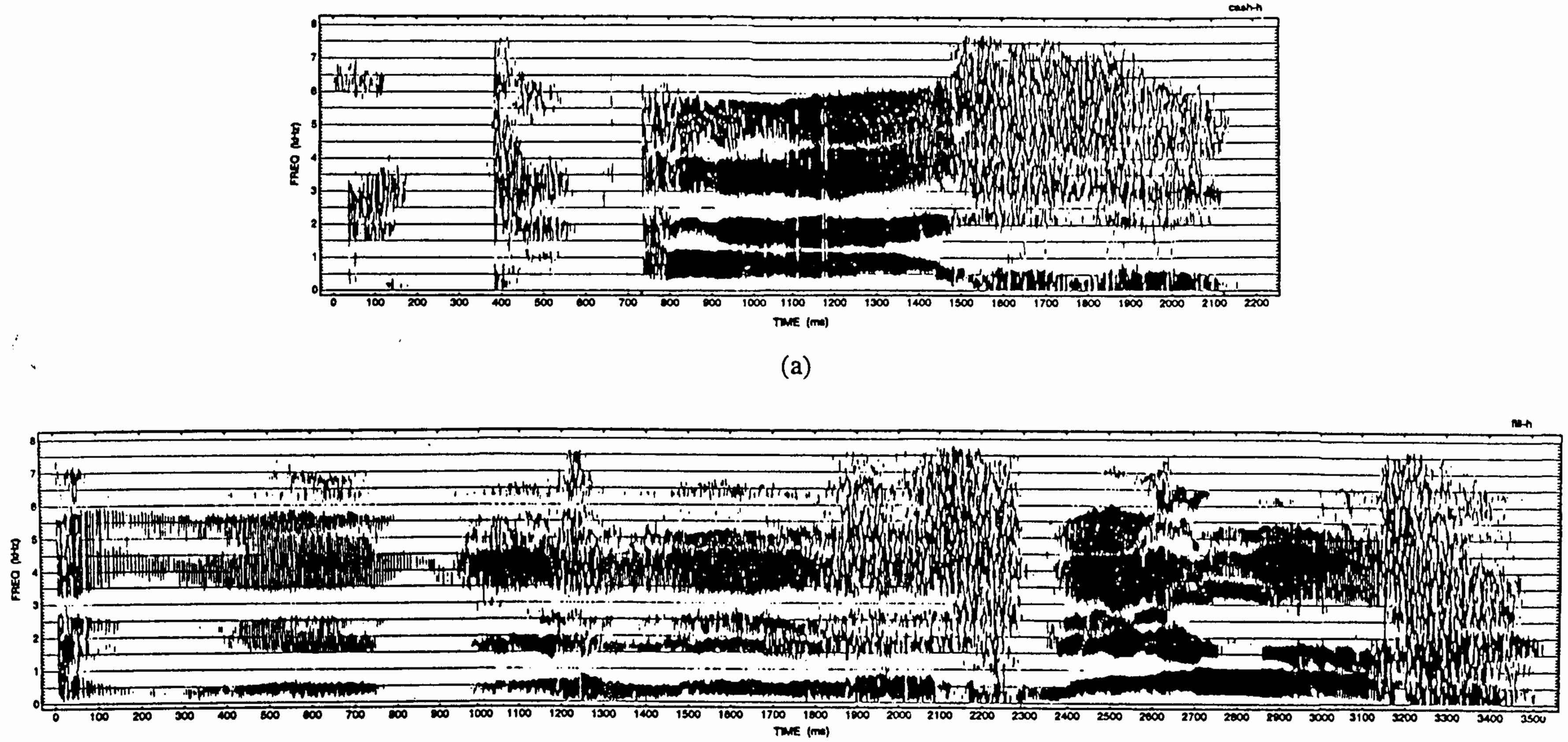
In summary, this subject has particular consistency control difficulties with the FR_AF_IN (13%) in comparison with the average R.T. of the entire word list (26%). Moreover, the words without any obstruent consonant in word-initial position (60%) have the best R.T. performance. Further, the recognition accuracy of the words with stop consonants in word-initial position (24%) is roughly equal to the average recognition accuracy of the 70-word list (26%).

(iv) Acoustic Analysis (A.A.)

Figures 4.14-a, 4.14-b, and 4.14-c show CH's spectrograms corresponding to "cash", "fill", and "had". Figure 4.14-a shows there is heavy and deep breathing noise (0 to 570 ms) to build up enough air pressure to pronounce the /k/ (720 to 780 ms) of "cash". However, even when she has built up intraoral pressure, she still does not pronounce the correct /k/. There are no apparent release and aspiration parts of /k/. Instead, CH mixes the /k/ with the vowel /æ/. This problem caused two judges in the T.T. to consider that CH's /k/ in "cash" had disappeared. In the I.T., the five judges have the four-choose-one list for reference. Even though they missed the first consonant /k/, they could still choose the correct answer from the following acoustic information of "cash", i.e., /æ/ and /sh/. This is one of the reasons of why T.T. shows a very serious velar control problem (100% errors) but the I.T. only shows a mild error, Velar - Other Consonant Place (25%) in Figure 4.13-b. In addition, because of CH's heavy breathing noise, it sometimes caused the speech recognizer to become confused and to recognize the noise as part of the utterance. This noise becomes especially serious for words with velar stop consonants in word-initial position.

Figure 4.14-b shows the spectrogram of "fill". It shows a very serious and long pre-voicing glottalized signal before the frication of /f/. This long pre-voicing noise is very common in CH's speech. This happened especially in front of fricatives, affricates, /h/, and some of the stop consonants. The frication and aspiration of /f/ started from 1890 to 2280 ms with glottal vibration. On the other hand, at the end of the /l/ of "fill", there is wide-band aspiration noise (3150 to 3530 ms). This aspiration noise happens often in CH's utterances whenever the end of a word is not an obstruent consonant. She would try to release all of the extra air from the articulatory system at the end of the word. This strong release at the end of the word sometimes caused the speech recognizer to consider this aspiration as another word, i.e., the speech recognizer gets two responses instead of one.

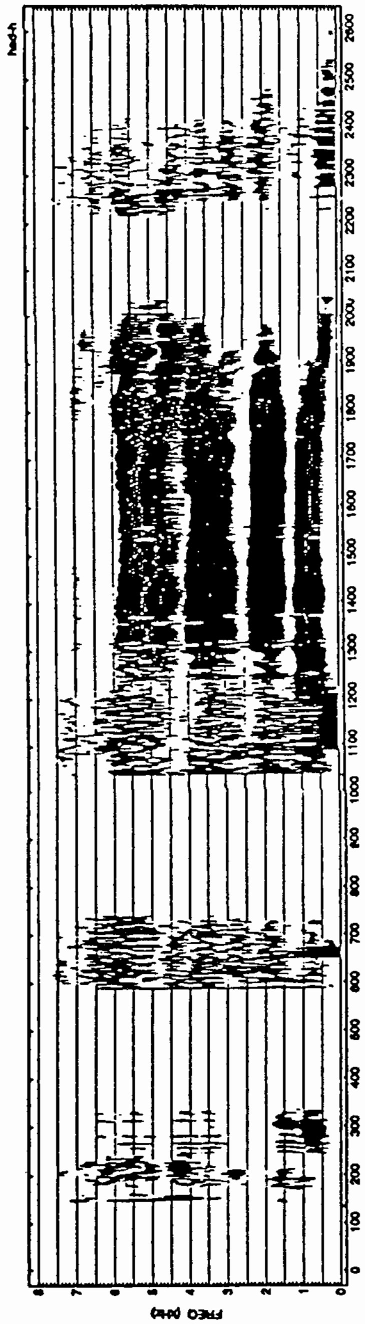
Figure 4.14-c shows that the glottal explosive noise (120 to 320 ms and 590 to 720 ms) happened before the /h/ of "had". She needed to put much more effort than normal subjects into the pronunciation of /h/ since the surgery, which was done ten years ago, has damaged the neuron system of her vocal fold. The real aspiration consonant /h/ with glottalization started from 1030 to 1220 ms. At the end of the word, she released all of the energy and air to pronounce /d/. This release caused the strong aspiration part for /d/ on the spectrogram. The low-frequency energy at 300 Hz (Figure 4.14-c) in /d/



(a)

(b)

Figure 4.14: CH's spectrograms of "Cash" (4.14-a), "Fill" (4.14-b), and "Had" (4.14-c).


had-h
FREQ (kHz)
0
1
2
3
4
5
6
7
8
TIME (ms)
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600

(c)

probably comes from airflow hitting the microphone because CH released too much air at the end of the word. From these observations, it is evident that CH has particularly serious problems with the words with /h/ in word-initial position.

Figures 4.14-a, 4.14-b, and 4.14-c show that there is mild aspiration noise in the vowel especially above 2.5 kHz, indicating that the subject's speech is breathy. However, CH's speech is not as breathy as that of JS.

The average duration measurement based on the spectrograms for these 70 diagnosis words is 1452 ms compared with MJ's 478 ms. The prolongation ratio is 3.0; that is, the duration of CH's words is 3.0 times that of the subject who does not have dysarthria, MJ, on the average. She has the longest speech prolongation of any of these subjects.

4.7.3 Findings and Design Criteria

Because of the damage caused by her left ear tumor surgery, CH has control problems with aspiration and obstruent consonants in word-initial position, e.g., /h/, /ch/, /sh/, and /s/. The status for /h/ in word-initial position is the most serious one. She can't easily pronounce most of the words with /h/ at the beginning. In the recognition test, most of the words with /h/ at the beginning were skipped if she could not easily say the word. Moreover, her fricative / affricate consonants in word-initial position are especially bad, including the labial, alveolar, and palatal positions. This problem might be a consequence of the paresis of the left part of her tongue and lips. The affricate consonants, e.g., /ch/, in word-initial position are much worse than the fricative consonants, e.g., /sh/. In the fricatives, /s/ also is difficult for her to say. Some of the words with /s/ in word-initial position have too low a volume to initialize the Dragon Writer-1000. The same problem sometimes happens on the words with /r/ in word-initial position too, but it is not as serious as /s/. In the fricative group, the words with /f/ in word-initial position have the worst consistency control problem.

From the four tests described above (I.T., T.T., R.T., and A.A.), one learns that CH has serious control problems with the consonant /h/, fricatives, and affricates in word-initial position. CH's tongue-blade and glottis control are impaired. Her internal and final clusters also are impaired. Both the I.T. and the R.T. show that the words without any obstruent consonants and the words with sonorant consonants in word-initial position have a better performance than the words with obstruents in word-initial position. This is important information for guiding the design of a new word list. The accuracy of the words with stop consonants in word-initial position is roughly equal to the average recognition accuracy of the 70-word list. Labial and alveolar stop consonants in word-initial position have a better recognition performance than velar stop consonants. Most of the words with velar stop consonants in word-initial position have heavy breathing noises before the utterance. The vowel duration and distinction problems are considered also in the word-list design since CH has vowel duration and place control errors. Vowels that are at extreme positions in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, should be chosen to let the vowels be as distinctive as possible and to avoid lax - tense confusions. Further, because CH's front vowels and back vowels have better distinctions than the diphthongs, the diphthongs should have a lower priority than the front and back vowels when the new word list is designed.

In addition, except for the findings mentioned, the extra pre-voicing, glottalized noise, or heavy breath before the utterance (Figures 4.14-a to 4.14-c) could make the speech recognizer consider them as another utterance or as being part of the utterance. Therefore, the recognizer's performance would decrease and the rejection rate would increase. On the other hand, the strong release at the end of a word which has no obstruents at the end could mislead the recognizer, which considers this as another input command or utterance and gives a wrong response. Therefore, these two extra factors should also be included in the design criteria.

4.7.4 Improvement of Speech Recognition Accuracy

From the findings described above for CH, a new word list with 52 words was designed for her to dictate to the computer. Table 4.1 shows that the recognition accuracy for the new word list has been improved from 26% (R.T.(%)) based on the 70 diagnosis words to 59% (New_L(%)) based on the new 52-word list (or 34% for R.T.-35(%) and 61% for New_L-35(%) based on the same vocabulary size, 35). An improvement of 33% (Improve(%)) or 27% (Improve-35(%)) has been achieved. After designing this new list and following Part II of the flow chart in Figure 3.1, some of the words which were still unsatisfactory were modified. The final list is in Appendix 4.7.

Two sentences, "I am brilliant." and "No one here will run away." were dictated by following the alphabet spelling, e.g., "air" is for "a" and "bad" is for "b". The dictation speed for these two sentences was compared with her usual typing method, using her right index finger. Table 4.4 shows the test results. Her average typing time for sentence I is 1 minute and 5 seconds compared with a dictating time of 1 minute and 21 seconds. For sentence II, the typing speed is 1 minute and 8 seconds compared with a dictating time of 4 minutes and 16 seconds. She can't control her speech and breathe consistently for a long time. Thus, her speech performance would become worse for the dictation with long period. These data show that CH's typing speed is faster than her speech dictation; the speech input is only useful as a supplementary tool for her to control the computer. She can choose the computer commands by her speech and type with her right index finger.

4.8 Subject 8 (GG)

4.8.1 Background and Symptoms

GG, who is 24 years old, is a student at a university. At birth, GG had evidence of cerebral palsy. Her neuromotor condition is characteristic of spastic cerebral palsy: the muscles are stiff and the movements awkward. The muscles have increased tone with heightened deep tendon reflexes. GG's speech sounds very weak to the unfamiliar listener and less throaty than that of JS' speech but the intelligibility of her speech is much better. Her speech and muscle movement are similar to JR's. In addition, GG's speech accuracy is the 3rd highest of the eight subjects in I.T. She can only use pencil grasped by her left or right fingers to type or program on the computer.

Her cognitive and linguistic abilities are intact. However, she needs a reading assistant for pronunciation because she can not pronounce spontaneously from reading the words on cards or monitors.

4.8.2 Analysis and Discussion

(i) Intelligibility Test (I.T.)

Table 4.1 shows that GG's intelligibility for the I.T. is 89%. The Chi-Square test value of I.T. is 1.4 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715].

The Chi-Square test value indicates that the listening responses between each of the five judges for GG's speech are not significantly different. Figure 4.15-a shows the summary of results from the intelligibility test based on contrast analysis of 19 phonetic groups. The percentage error for a specific contrast was computed as the total number of errors made across the judges for that contrast divided by the total number of trials in which that specific contrast was available as a response for the judges. The three most frequent errors are Stop - Affricate (20%), Other Fricative Place (16%), and Vowel Duration (9%). Figure 4.15-b shows the summary of results from the intelligibility test based on contrast analysis of 40 phonetic groups. The top five most frequent errors are Final Cluster - Singleton (33%), Affricate - Stop Consonant (20%), Glottal - Other Fricative Place (13%), Alveolar - Other Fricative Place (13%), and Long - Short Vowel Duration (13%). The Affricate - Fricative Consonant (10%) along with Affricate - Stop Consonant indicate that this speaker has some problems with palatal control, especially for affricates. From these data, it also indicates that fricatives have some control problems. Furthermore, the percentage errors for Glottal - Other Fricative Place and Init. Voiceless - Voiced Contrast (10%) indicates that this subject has very mild problems with larynx control. Because GG's speech is not seriously impaired (I.T.: 89%), the error distributions do not show very serious confusion errors or special articulatory problems except for the control of fricative / affricate consonants and the final cluster - singleton consonants. Moreover, the vowel duration error shows that GG can not make good distinctions between the long and short vowels but the problem is not very serious.

From Table 4.2, the I.T. results for GG indicate that the average accuracy for the 70 words is 89%. Groups SONOR_IN (82%), WTOT_OB (89%), and OBS_IN (92%) show that GG has slightly better articulatory position control for words with an obstruent consonant in word-initial position. She is the only subject whose OBS_IN group has the best articulatory position control. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (89%) group is worse than the CLUS_IN (100%) and the STOP_IN (96%). More detailed study of the FR_AF_IN group shows that the /ch/_IN (53%) has the lowest accuracy in this FR_AF_IN group. Table 4.3 shows the detailed I.T. analysis for the STOP_IN group which includes LABIAL (93%), ALVEOLAR (100%), and VELAR (100%) stop consonant groups. It appears that the words with velar and alveolar consonants in word-initial position have better I.T. performance than labial consonants, and also better than the I.T. average of the 70-word list (89%).

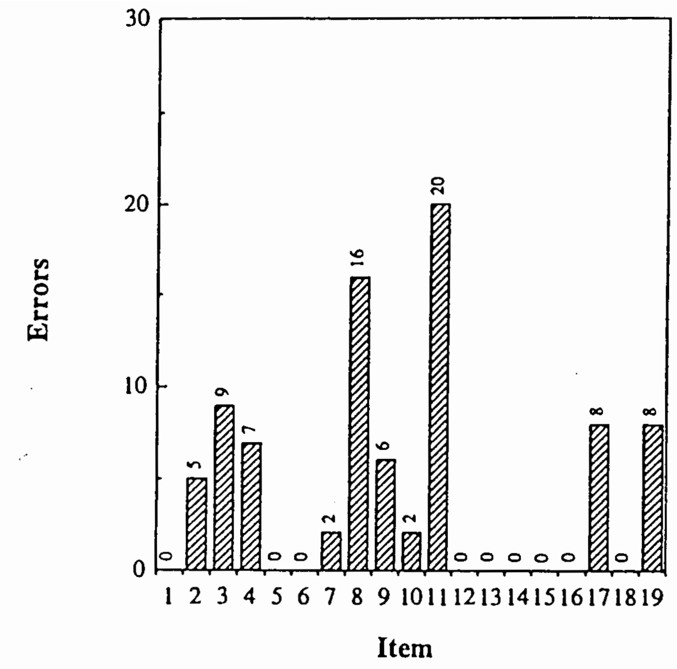


Figure 4.15-a GG's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 19 are described in Table 3.1. The error distribution shows the three most frequent errors are Stop - Affricate (20%), Other Fricative Place (16%), and Vowel-Duration (9%).

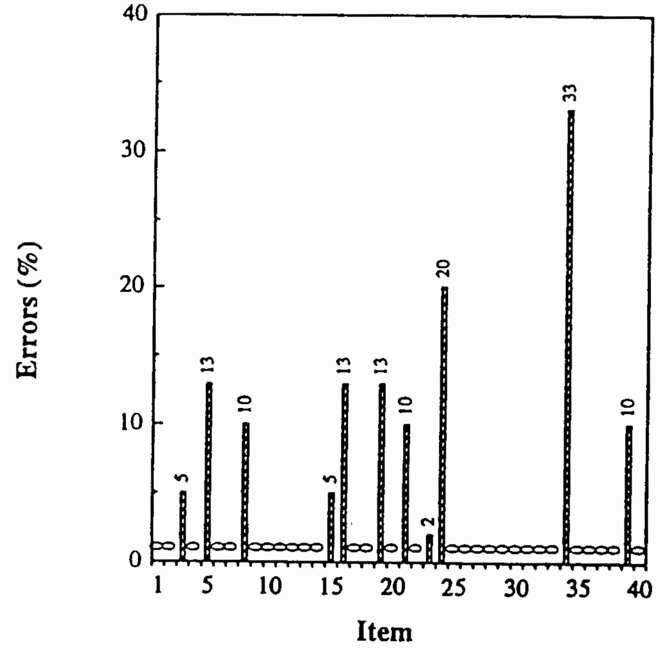


Figure 4.15-b GG's intelligibility test error distribution. The acoustic-phonetic error contrasts from item 1 to 40 are described in Table 3.2. The error distribution shows the top five most frequent errors are Final Cluster - Singleton (33%), Affricate - Stop Consonant (20%), Glottal - Other Fricative Place (13%), Alveolar - Other Fricative Place (13%), and Long - Short Vowel Duration (13%).

In summary, this subject is different to other subjects and does not have particular place control difficulties for the obstruent consonants in word-initial position except the affricates in word-initial position and final clusters. The words with /ch/ (53%) in word-initial position have especially low I.T. accuracy in comparison with the average intelligibility of the entire word list (89%) using a reference list (Table 4.2). GG also has mild fricative, larynx and vowel duration control problems. However, except the affricates in word-initial position and the final clusters, the other articulatory errors are only mild.

(ii) Transcription Test (T.T.)

From Table 4.1, the Chi-Square test value of T.T. is 0.01 (< 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. The Chi-Square test value indicates that the transcription responses between two judges for GG's speech are not significantly different. From Table 4.13, the T.T. results for GG indicate the average accuracy judged by two experienced judges for the obstruent (80%) and sonorant (76%) consonants, identified correctly in every feature, in word-initial position. This is consistent with the findings in the I.T. Vowel, voicing, and obstruent place are presented in the form of confusion matrices. The vowel confusions, which only focus on one feature identification, show that this speaker makes good distinctions between front vowels (98%), back vowels (100%), and diphthongs (92%). The confusions associated with the voiced - voiceless contrast, which addresses only one feature identification, show that this speaker does not have a glottal control problem for voiced-voiceless consonants in word-initial position. There is a very mild confusion for the voiceless consonants (5%). That is consistent with the Test 1 results for the error of the Init. Voiceless - Voiced Contrast (10%), Figure 4.15-b. The confusions for place of articulation for obstruents, which are based only on one feature identification, indicate that palatal position control (75%) is the most serious one compared to labial (83%), alveolar (94%), and velar (100%) control. However, the palatal position control errors (75%), which all are made by the fricatives / affricates, are still not very serious. A detailed examination of the errors for obstruents shows that all of the alveolar confusion errors come also from fricative consonants but not from stop consonants.

Overall, obstruent accuracy (80%) is better than sonorant accuracy (76%) for consonants in word-initial position, like JF's speech. Moreover, the alveolar and palatal confusions indicate a slight alveolar and palatal position control error (especially for fricatives and affricates).

Obstruent Accuracy: 80%

Sonorant Accuracy: 76%

Vowels

T.\R.	Front	Back	Middle	Diph.	[]
Front	98%	2%			
Back		100%			
Diph.	8%			92%	

Voicing

T.\R.	+	-	[ ]
+	98%		2%
-	5%	95%	

Place for Obstruents

T.\R.	Lab.	Dent.	Alv.	Pal.	Vel.	[ ]
Lab.	83%		13%		4%	
Alv.	3%		94%	3%		
Pal.			25%	75%		
Vel.					100%	

Table 4.13 GG's transcription data for vowel, voicing, and place for obstruent confusion and obstruent - sonorant comparison in word-initial position. The accuracies of obstruent and sonorant consonants are percent correct identification of the obstruent and sonorant consonants in word-initial position. The target items are listed in the left column and the response items are displayed across the top.

(iii) Recognition Test (R.T.)

From Table 4.2, the R.T. results for GG indicate that the average accuracy for the 70 words is 63% compared with the I.T. of 89%. Groups SONOR_IN (71%), WTOT_OB (86%), and OBS_IN (55%) show that GG has the worst articulatory consistency control problem for words with an obstruent consonant in word-initial position. This is not consistent with the finding of the I.T. described above. The OBS_IN group can be split into three groups: CLUS_IN, STOP_IN, and FR_AF_IN. The FR_AF_IN (50%) and the CLUS_IN (50%) groups appear to show a more serious consistency control problem than the STOP_IN (63%). More detailed study of FR_AF_IN shows that /f/ in word-initial position (35%) has the lowest accuracy in this group. /s/_IN (52%) and /sh/_IN (40%) both are worse than the average R.T. accuracy (63%). In addition, the STOP_IN group has 63% accuracy, which is roughly equal to the average R.T. accuracy (63%). Table 4.3 shows the detailed R.T. analysis for the STOP_IN group which includes LABIAL (53%), ALVEOLAR (73%), and VELAR

(75%) stop consonant groups. It appears that the words with alveolar and velar stop consonants in word-initial position have better recognition performance than labial consonants, and also better than the R.T. average of the 70-word list (63%). This is consistent with the findings for I.T. in STOP_IN group, which shows that the words with labial stop consonants in word-initial position have the worst I.T. performance. On the other hand, the R.T. rejection rate of the 70-word list is 17% (Table 4.1) compared with the normal subject's 0.3%. It shows that GG's speech has a mild problem with inconsistency in pronunciation.

In summary, this subject has particular consistency control difficulties with fricatives and affricates in word-initial position and the initial clusters in comparison with the average R.T. of the entire word list. Moreover, the words without any obstruent consonant in word-initial position (86%) have the best R.T. performance. This is very different from the results of the I.T. The labial stop consonants have the worst performance in STOP_IN group.

(iv) Acoustic Analysis (A.A.)

From observation of each spectrogram of the 70-word list, there is a strong extra formant at 4.5 kHz for most of the obstruent consonants. From the listening study and theoretical predictions based on vocal-tract acoustics, this extra formant is in a frequency range normally expected for /s/ and /sh/. This extra spectral peak is very apparent particularly for her stop consonants. This frication noise in the closure period of her stop consonants suggests that she has problems in making a complete closure. Because she can not close completely her lips, tongue tip, or tongue body, the air passes through the constriction formed by the incomplete closure and makes the turbulence noise. Acoustic theory predicts that the front tube length to produce this frication noise would be about 2 cm (c / (4*L) = 4500 Hz where c = 35400 cm / sec and L represents the length of the front tube). This fricative noise sometimes masks or substitutes for the abrupt onset for the affricate consonant /ch/, resulting in a fricative /sh/ or /s/. Figures 4.16-a, 4.16-b, and 4.16-c show spectrograms corresponding to "cheer", "sip", and "wax". Figure 4.16-a shows the example in which /ch/ is confused with /sh/ because the onset of /ch/ has been masked or substituted by the fricative noise. This is the error type of the affricate - fricative confusion, as in Figure 4.15-a. Figure 4.16-b is an example of a word for which there was a voiceless - voiced error. There is a voicing cue in the low-frequency range of /s/. This cue results in the Test 1 confusion of the target "sip" for response "zip". The

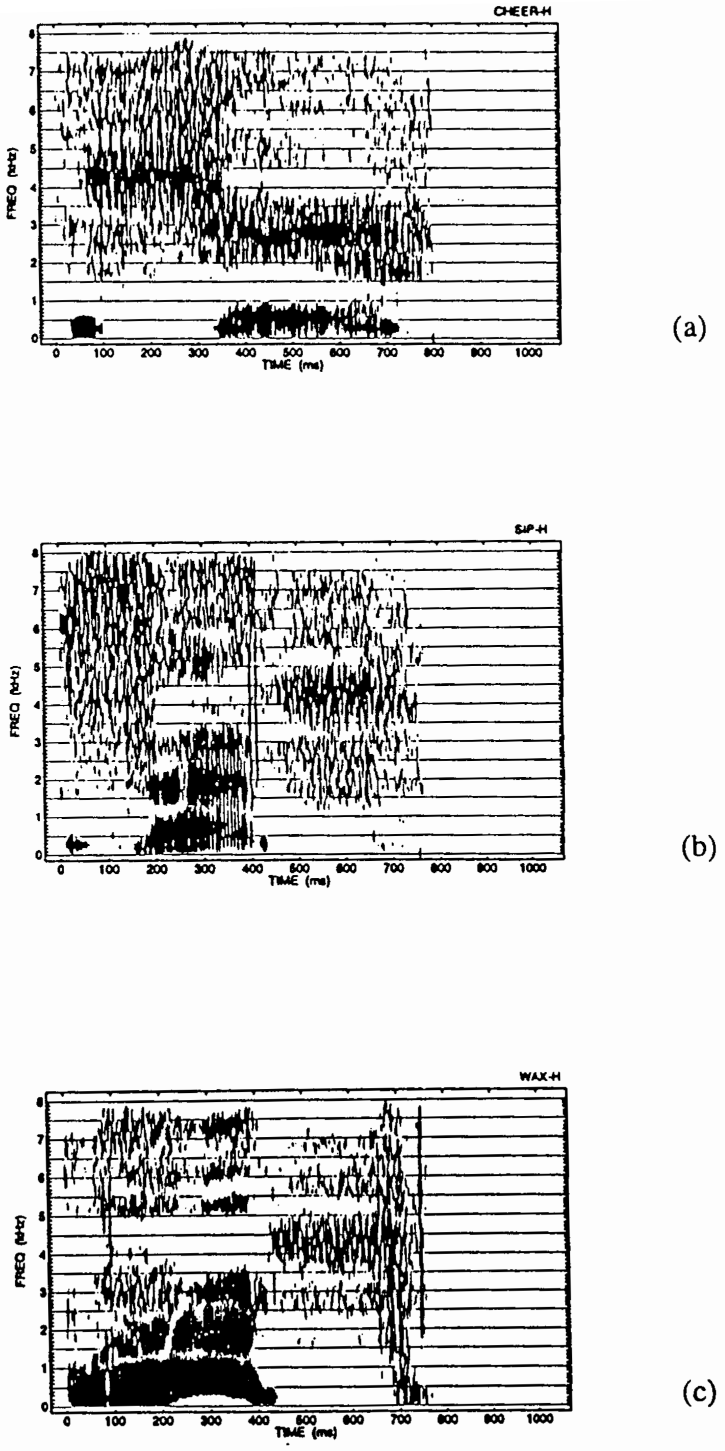


Figure 4.16: GG's spectrograms of "Cheer" (4.16-a), "Sip" (4.16-b), and "Wax" (4.16-c).

same problem also happens in "cheer" whose pre-voicing cue shows up at 30 ms. Further, Figure 4.16-b also shows the fricative noise happens again in the final /p/ of "sip". Because of the fricative noise, the final /p/ of this word does not have a clear release. Figure 4.16-c shows an example in which the cluster is confused with final singleton. The fricative noise shows up from 430 to 700 ms because of the apparent lack of a complete closure for the /k/ of "wax". At the end of this fricative noise, a lot of air is released and causes too much air blow on microphone. Therefore, her "wax" does not have the final cluster /ks/ but a fricative noise. This is one reason for the high error percentage of Final Cluster - Singleton in Figure 4.15-b. Further, Figures 4.16-a, 4.16-b, and 4.16-c show that there is substantial aspiration noise in the vowels above 2.5 kHz, indicating that the subject's speech is breathy but is not as serious as that of JS' speech.

The average duration measurement based on the spectrograms for these 70 diagnosis words is 770 ms compared with MJ's 478 ms. The prolongation ratio is 1.6; that is, the duration of GG's words is 1.6 times that of the subject who does not have dysarthria, MJ, on the average.

4.8.3 Findings and Design Criteria

In conclusion, the four tests described above (I.T., T.T., R.T., and A.A.) show that GG has some control problems in producing fricatives / affricates in word-initial position and consonant clusters in word-initial position. Frication noise often appears in the closure period of her stop consonants. It appears that she has the difficulty forming a closure for stop consonants. Because she can not close completely her lips, tongue tip and blade, or tongue body, the air passes through the constriction and makes turbulence noise. The R.T. shows that the words without any obstruent consonants (86% for WTOT_OB in R.T.) and the words with sonorant consonants in word-initial position (71% for SONOR_IN in R.T.) have a better recognition performance than the words with obstruents in word-initial position (55% for OBS_IN in R.T.). However, the words with obstruents in word-initial position have slightly better intelligibility performance (92% for OBS_IN in I.T.) than the accuracy of the words without any obstruent consonants (89% for WTOT_OB in I.T.) and the words with sonorant consonants in word-initial position (82% for SONOR_IN in I.T.). She is the only one subject whose I.T. and R.T. show different results in the comparison of sonorants and obstruents in word-initial position. This observation is important for the design of a new word list. Furthermore, the words with stop consonants in word-initial position also have a fair performance compared to FR_AF_IN group. The words with alveolar and velar stop consonants in word-initial position have a better recognition performance than the ones with labial stop consonants. The vowel duration and distinction problems should also be considered in the word-list design since GG still has some mild vowel duration and position control problems. Vowels that are at extreme positions in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, should be chosen to let the vowels be as distinctive as possible and to avoid lax - tense confusions. Further, because GG's vowels all have good distinctions. All of these vowels have the same priority when the new word list is designed. Overall, GG does not have very seriously impaired speech.

4.8.4 Improvement of Speech Recognition Accuracy

Because she moved to other state, GG stopped joining this research project. However, the results from the studies of her speech are reported here to provide additional evidence of the kinds of speech problems experienced by these individuals.

4.9 Summary

Although these eight subjects have different symptoms and suffered nervous system damage in various parts of the brain, the new methods introduced in this project have been shown to be efficient for finding and classifying the speech disorder problems for these speakers with dysarthria, and for facilitating the use of speech recognition technology. Using the four methods, I.T., T.T., R.T., and A.A., as described above, the basic speech characteristics of each of these speakers with dysarthria can be obtained and can provide a basis for designing the new personal word list. There are some common difficulties in certain articulations across these subjects, e.g., fricatives and affricates in word-initial position. These common features can be used to help describe the detailed malfunction of these subjects' articulatory system in an overall viewpoint. This information can be of great help for improving the design of the initial word list to achieve the best recognition accuracy. Moreover, other acoustic information and the user's preferred words or other nonvocal sounds can be included in the test vocabulary. The following common features exhibited by these subjects could be used as general criteria for designing the new personal list:

(i) Words without any obstruent at all have the best I.T. and R.T. performance for most of these subjects. Additionally, words with sonorants in word-initial position are the second choice for designing the new word list.

(ii) Avoid using initial clusters and fricatives / affricates in word-initial position since all of the subjects' I.T. and R.T. for these two groups are lower than the average I.T. and R.T. accuracy for the 70-word list. The consistency performance of initial-clusters and /s/ in word-initial position are both especially poor for all of these eight subjects.

(iii) Even though some of the subjects do not have a serious voicing contrast control problem in word-initial position, it is preferable not to use voiced - voiceless pairs in the word pool, e.g., "pat" and "bat". Such pairs could still potentially confuse the speech recognizer. Most of the subjects have glottal control problems to a certain extent.

(iv) The STOP_IN group for most subjects has a fair I.T. and R.T. performance compared with the average accuracy of the 70-word list. Therefore, stop consonants in word-initial position could still be potential candidates for the new word-list design, even though they are obstruents.

(v) The A.A. and T.T. show that most of these subjects' vowels errors are less prevalent than consonants. It appears that vowels deserve less consideration than consonants. However, the vowel duration and distinction problems still need to be considered. The vowels would be chosen in the quadrilateral, e.g., /i/, /æ/, /a/, and /u/, to make them as distinctive as possible and to eliminate lax - tense confusion.

(vi) Practical experience shows that the more syllables there are in a word the worse the speech consistency for these subjects. However, the lack of acoustic information would also reduce the recognition accuracy if the utterance duration of the designed word is too short. Therefore, selection of the number of syllables in a word is an important issue. One to two syllables in each word is a good range for designing the word list.

(vii) Alphabet spelling should be one of the basic requirements for applying this word list to writing a text-editor package, e.g., PE2 or Microsoft Word. Therefore, a good and ease of use alphabet spelling system should be included in the design of the words along with the speech characteristics for each specific subject, e.g., "air" represents "a", and "be" is for "b".

Except for the common criteria mentioned above, detailed rules for each specific subject should be included by considering the results from the four analysis tools. Therefore, the design of the initial word list should take into consideration: user-preferable interface, alphabet spelling system, and the impaired speech characteristics for each subject. After extensive efforts, a final word list for each subject has been derived from trial-and-error and from iteration tests, following Part II in Figure 3.1. Even though all of the analyses and tests mentioned above have been done, there is still a lack of adequate data to cover all of the phonetic problems for each dysarthric speaker. Thus, a modification of the initially designed word list is still necessary. The optimal final list should be designed after the actual recognition tests are done.

Chapter 5 Conclusions and Recommendations

5.1 Introduction

Overall, the intelligibility test, transcription test, recognition test, and acoustic analysis results contribute to the clinical assessment of dysarthria. The perceptual assessment can explore the potential articulatory position-control impairments for the specific dysarthric speaker. The speech recognition test can be developed and used to study the consistency of dysarthric speech. The acoustic method can reveal parameters that assist in determining the factors contributing to reduced speech intelligibility and lack of consistency in production and serve as a guide to improving the clinician's understanding of what the patient is doing to produce impaired speech (Simmons, 1982).

There are several implications of observations from these four tests in relation to assessment and treatment of motor speech disorders. Although each subject has his / her own symptoms and speech characteristics, some common features are found in these subjects. The following discussion will give a summary of the findings from these eight subjects, organized according to the primary and complementary goals mentioned in Section 1.3: (i) Primary Goal: determine how best to use speech recognition techniques for augmenting the communication abilities of dysarthric computer users; (ii) Complementary Goal 1: identify features of dysarthric speech which affect speech recognition performance; (iii) Complementary Goal 2: supply different viewpoints that can lead to a better understanding of the speech patterns of the dysarthrias. Most of the findings will be compared with the relevant results from other research and clinical assessment papers. From these findings, a standard route has evolved for designing the new word list for input to the computer. A recognition accuracy comparison between the 70-word diagnosis list and two new lists of 35 and 52 words for each of the eight subjects is discussed. From study of the recognition improvement, it is possible to determine the gain that has been achieved for these subjects. In addition to these findings, we consider questions of ease of computer use in practical situations and these questions add further dimensions to the project. Finally, the contributions of this project are summarized and future steps and some recommendations are also discussed.

5.2 Findings and Conclusions

From the four tests mentioned in Chapters 3 and 4, estimates of the basic speech characteristics for each subject have been made. Some of the common features across these subjects and the findings from this project are discussed below in the order of the steps mentioned in Section 1.3. Since only eight subjects participate in this project, it is hard to generalize the overall performance for the dysarthric speech. The following discussion will only focus on the findings from these subjects and some comparisons will be made with data reported in the literature. The subjects in all of the figures in this chapter will be displayed in the order of increasing scores for the I.T.

5.2.1 **Primary Goal: determine how best to use speech recognition techniques for augmenting the communication abilities of dysarthric computer users.**

The characteristics of dysarthric speech found in the assessment process can help to develop criteria for designing a new word list which has a higher speech recognition performance for the subject with dysarthria. Further, the integration of the new speech technology, i.e., speech recognizer and speech synthesizer, can improve the communication abilities of dysarthric computer users. The detailed outcomes are listed as follows:

(1) Show the recognition accuracy improvement with and without making a new list from assessment of these subjects with dysarthria.

The conclusion from preliminary work that preceded this thesis showed that optimal design of a speech recognition interface for persons with dysarthria requires an understanding of each subject's speech characteristics and preferences (Goodenough-Trepagnier and Rosen, 1991 and Goodenough-Trepagnier et al., 1992). A comparison of I.T. and recognition accuracy improvement (based on a comparison between the 35-new word list and the 35-word diagnostic word list which is randomly picked from the 70-word diagnosis list) across all of these eight subjects is shown in Table 4.1 and Figure 5.1. The new word list designed for each subject has led to an improvement in speech recognition accuracy which ranges from 3% to 30% depending on the subjects. By using a paired t-test (a two-tailed test at the 5% significant level with 6 df for the null hypothesis Ho: [the recognition accuracy has not significantly improved in the comparison of the diagnostic-word list with the new word list], the t value for this test is 3.46 (> $t_{0.025}$ (=

2.447)), refer to Sincich (1987) [p. 468]. Thus, the hypothesis is rejected. The results in Table 4.1 and Figure 5.1 have shown significant improvement in the recognition accuracy by using the new word lists designed for these subjects. The data show that the less speech impairment there is the smaller improvement in recognition accuracy. Subjects ES, JF, and MW show only a small improvement since their speech performance is close to that of normal subjects. The mild malfunctions of their articulatory system will not greatly influence their production. However, there is a lot of room for improvement for the speakers with severe dysarthria, e.g., CH, JS, and DW. Since the control of the articulatory systems for these speakers with dysarthria has been severely impaired, their speech performance can be improved efficiently and rapidly if the correct malfunctions, i.e., the phonetic combinations having the most inconsistent performance, are found and removed from the inventory of words. However, the consistency and time variation problems will become more and more serious as the severity of the dysarthria increases. The accuracy improvement for the speech recognizer will be limited. The amount of accuracy improvement will depend on the severity of the specific speaker's impairment. For example, because the serious impairment causes poor speech consistency for DW's speech, the potential for improving his speech recognition performance is limited. Consequently, the improvement curve in Figure 5.1 has a peak at JS, and there is not a continued increase in improvement from JS to DW.

From the physiological view, the more serious the speech impairment is, the less the articulatory system can be used consistently. When there are many malfunctions of these subjects' articulatory systems, the impairments will not only influence the speech position accuracy control but also the consistency control. Therefore, the recognition accuracy for these subjects can be improved by using the new word list, at least up to a point.

(2) Improve the performance of the speech recognizer for dysarthric users based on a new 52-word list.

From the information and assessment results from each subject, an initial word list for each specific individual has been designed. By passing through the flow chart in Figure 3.1, a final word list with 52-word list has been developed for the use of each individual in their practical computer work (Appendices 4.1 - 4.7). From the analysis across all eight subjects, some of the specific findings and common features have been

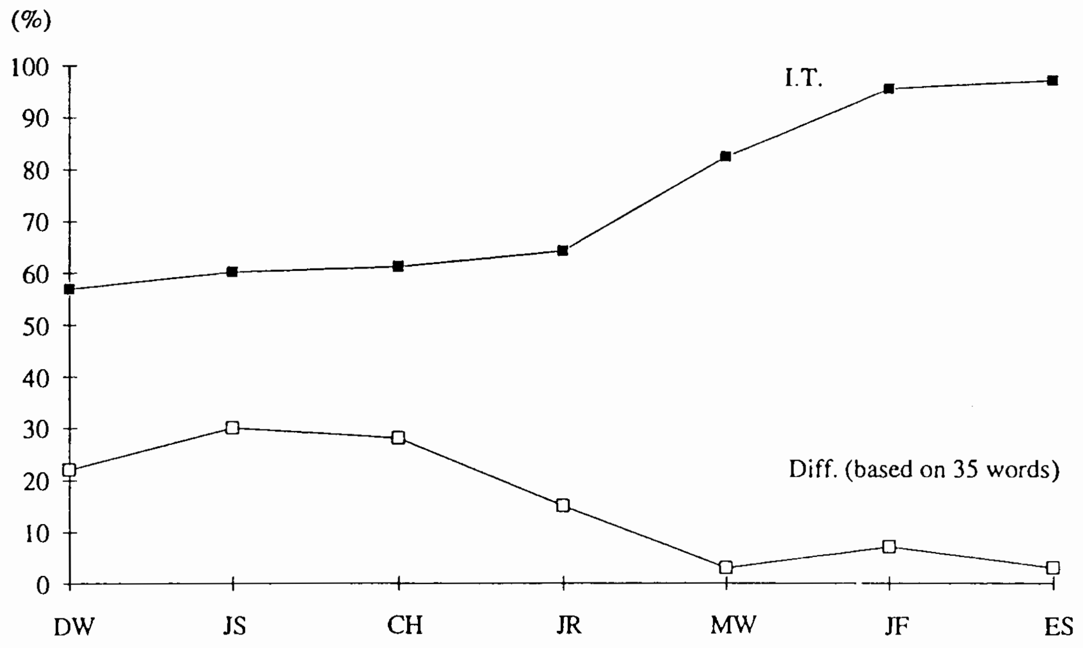


Figure 5.1 Recognition accuracy improvement (labeled as "Diff.(based on 35 words)") for the 35 word-vocabulary size, based on a comparison between the 35-new word list and the 35-diagnostic word list, relative to the intelligibility test scores. The subjects are ordered according to increasing scores for the I.T.

discussed above. The detailed recognition performance improvements for each subject will be covered in the following.

Figure 5.2 shows the recognition accuracy improvement, based on the comparison between the new 52-word list and the 70-word diagnostic list, compared to the intelligibility performance. The new method introduced in this project has led to an improvement in speech recognition accuracy from 9% to 38%. The data distribution in Figure 5.2 is similar to the one in Figure 5.1. The data in both Figures 5.1 and 5.2 indicate that the less speech impairment there is, the less recognition accuracy performance improvement can be achieved. For example, Figure 5.2 demonstrates that subjects ES, JF, and MW show only a small improvement since their speech performance is close to that of normal subjects. However, there is a lot of room for improvement for the severely impaired speakers with dysarthria, e.g., CH, JS, and DW. For the same reason as in Figure 5.1, the potential for improving DW's speech recognition performance will be limited; he cannot even use all of the 52 words designed for him. Consequently, the improvement curve in Figure 5.2 also has a peak at JS, and there is no continued increase in improvement from JS to DW. The effect of speech consistency will be discussed again in the section of "complementary goal 2".

In summary, the new method introduced in this project leads to a successful recognition improvement for these subjects' recognition performance from 9% to 38%. However, there is a limit to the practical application of this method for some of the subjects who have severely impaired speech. The speech consistency and time variation problems will be the factors that destroy the recognition performance.

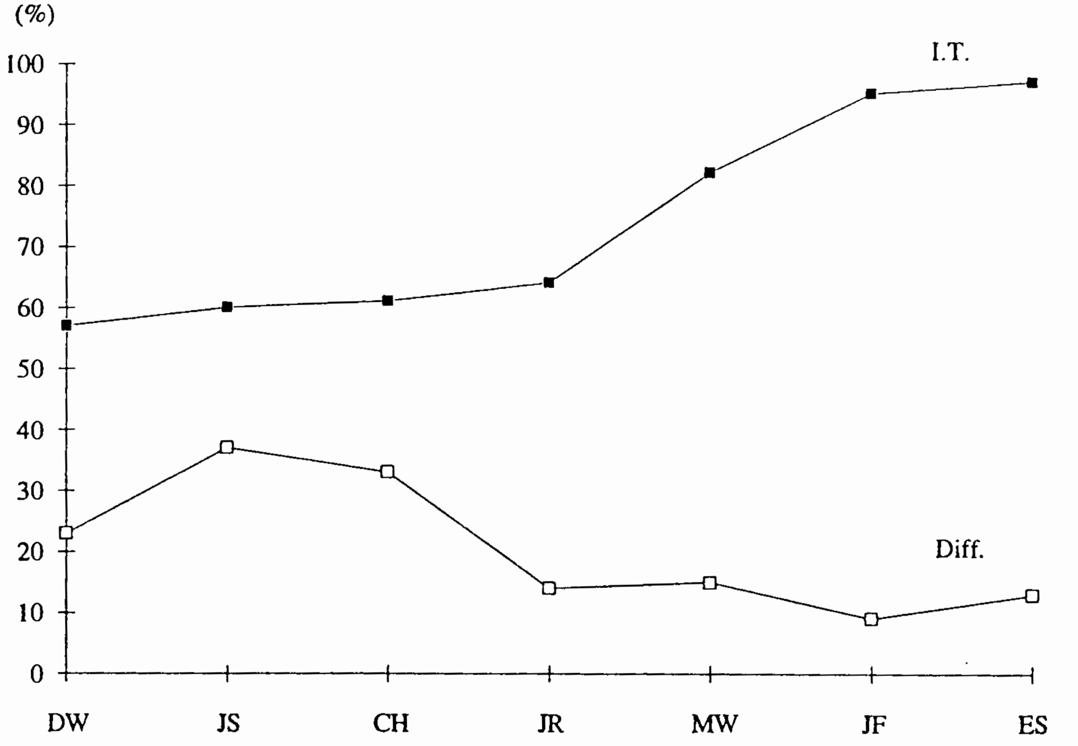


Figure 5.2 The recognition accuracy improvement (labeled as "Diff."), based on a comparison between the 52-new word list and the 70-word diagnostic list, relative to the intelligibility test. The subjects are ordered according to increasing scores for the I.T.

(3) Show the possibility of using the new speech technology integration to supply new efficient input and output channels for the computer.

(a) Designing a new word list to improve the computer operation speed for these subjects is possible.

Table 4.4 has shown the possibility of designing a new word list that can be used for dictating to the computer for text-editing. All of these subjects except GG, who has moved to another state, have finished the comparison of typing and dictation speed on the IBM PC. Some of them, MW and DW, also have practiced using a speech recognition system (Voice Navigator) on the Macintosh.

Table 4.4 has shown the comparison of typing and speech dictation speed for each subject, except GG. All of the subjects except GG successfully dictated the two sentences, "I am brilliant.", and "No one here will run away.", including "." at the end of both sentences. For some of the subjects, the dictation speed is very close to their typing speed, i.e., MW, ES, and JR. JS' dictation speed is especially faster than his usual typing method, using his nose. These results have shown that the speech recognizer can be a useful input channel for these subjects, even though they have speech impairments. However, there is another important advantage for these subjects in using this new input design. When they are using the speech recognizers as the input of the computer, they need not incline their bodies forward, put their hands on the T-board, or hold the hand which is used to type by using the other hand in order for making their movements stable. They can just watch the computer monitor and talk to the computer as easily as unimpaired persons. They can correct the mistakes immediately when the computer displays wrong responses. By using this new technology, they may operate computers more easily, since they need not type one character and then raise their heads to check the computer monitor. This method has successfully reduced the overload of using the impaired limbs or fingers to type on the computer keyboard.

(b) The integration of the speech recognizer and speech synthesizer can supply new computer input and output channels for these subjects with dysarthria.

By using a speech synthesizer, e.g., Speech - Plus 1000, these subjects can directly communicate with other persons and hear the commands they just gave. They can talk to the computer first and then the computer sends the dictating command to the text-to-speech synthesizer. Then, the listeners can understand the subject's commands or utterances. In some of the cases where the subject is too unintelligible for humans to understand, the machine is still able to recognize the utterances that are produced consistently and make the appropriate word choice.

Furthermore, the speech recognizer can be used not only to dictate one single character but can also produce a whole sentence (or paragraph) and computer command functions (e.g., open a file, copy a section, or change the font size) using a single spoken command. These subjects have tried some of these dictation practices to control the computer, e.g., to call a computer macro function or replace the function of the computer mouse by using the speech recognizers. This project demonstrates successfully a potentially useful input channel for persons with speech and / or movement disability.

5.2.2 Complementary Goal 1: identify features of dysarthric speech which affect speech recognition performance.

Assessment of the dysarthric speech can provide information about the main malfunctions for each subject's speech (including articulatory-position and -consistency control problems) and the relative factors which influence the recognition accuracy for each subject with impaired speech. The detailed outcomes are listed as follows:

(1) Obstruent consonants caused a worse problem than sonorants in word-initial position for these subjects.

he intelligibility test and recognition test both show that the words with sonorants in word-initial position have better performance than the words with obstruents across all of these subjects with one exception. The exception is GG, whose I.T. score for OBS_IN is better than that for SONOR_IN (Table 4.2). Furthermore, the words without any obstruent have the best I.T. and R.T. performances for most of the subjects. Six of the eight subjects have the best I.T. performance in the WTOT_OB group, one of them has the best I.T. performance in the SONOR_IN group, and the last one has the best I.T. performance in OBS_IN group. In addition, six of the eight subjects have the best R.T. performance in WTOT_OB group, two of them have the best R.T. performance in the SONOR_IN group, and none has the best R.T. performance in the OBS_IN group. Table 5.1 shows a summary of the test results for I.T., T.T., and R.T. The scores in Table 5.1 are presented

Table 5.1 Summary of the test accuracy for I.T., T.T., and R.T. The columns give the results for the percentage scores, which are averaged across all of the eight subjects, with all of the 70 words, without any obstruent at all, sonorant in word-initial position, and obstruent in word-initial position.

Test \ Group	ALL	WTOT_OBS	SONOR_IN	OBS_IN
I.T.	76	89	79	73
T.T.			72	56
R.T.	55	70	64	49

in the form of average accuracy across all of the eight subjects. This table shows that the WTOT_OBS (I.T.: 89%, R.T.: 70%) and SONOR_IN (I.T.: 79%, R.T.: 64%) groups both have higher scores than the ALL group (I.T.: 76%, R.T.: 55%) and have better performance than the OBS_IN group (I.T.: 73%, R.T.: 49%) in I.T. and R.T. The WTOT_OBS particularly has the best performance in these four groups. In the comparison of T.T., Table 5.1 shows that sonorants (72%) in word-initial position have better accuracy than obstruents (56%) in word-initial position. Overall, the performance for words with sonorants in word-initial position and the words without any obstruent is better than the performance for words with obstruents in word-initial position. This is a very important finding for this project and for the new word-list design. This project uses this finding as the first criterion for designing new word lists.

The transcription test shows that obstruent consonants in word-initial position have worse accuracy than sonorants for MW, JS, DW, JR, and CH, but not for JF, ES, and GG. This finding indicates that there are five subjects who have difficulty producing consonants which need a buildup of pressure in the vocal tract. Further study of this phenomenon indicates that the three subjects whose obstruents in word-initial position have better performance than sonorants are the top 3 subjects in the I.T. scores. These three subjects have good articulatory control for consonants in general, and, apparently, also good control for sounds requiring a buildup of pressure in the vocal tract. However, the rest of these five subjects, whose I.T. scores are lower, have some difficulties producing the obstruents since their fine control of the articulators is too abnormal. Obstruent production requires either complete closure of the vocal tract (stop consonants), or a small constriction (fricatives), or both (affricates). These kinds of obstruents need better articulatory control than the sonorants in order to maintain a stably small constriction area or a complete closure. Any position mistake, tremor, or abnormality for the constriction in the vocal tract can cause a misarticulation or failure to build up enough intraoral pressure for obstruents.

(2) When comparing these subjects' speech data, anterior lingual places (particularly for tongue tip or blade position) of articulation for fricatives / affricates caused significant errors for the I.T. and were inconsistent for the R.T. in word-initial position.

The control required to produce a small constriction area is particularly difficult for the tongue tip (or blade) and lips. Any small perturbation, muscle tremors, or any significant change in intraoral pressure will cause the constriction area, e.g., between tongue tip (or blade) and upper wall of vocal tract (including alveolar and palatal positions), to be unstable. This is one of the reasons why all of these eight subjects have anterior place control problems (especially for the fricative and affricate confusion), as determined from the I.T., T.T., and R.T. From the T.T. data, two of the eight subjects, MW and JS, have alveolar position control problems. Four of the eight subjects, JF, DW, ES, and GG, have palatal position control problems. Two of the subjects, JR and CH, have serious problems for both the alveolar and palatal positions. Three of the subjects, JS, CH, and GG, have labial control problems. Most of the mistakes for the alveolar and palatal positions are caused by the fricatives or affricates. The fricatives and affricates, especially, need more complex and better position control than the stop consonants because the constriction area must be kept small enough to build up the turbulent flow through the vocal tract but not so small as to close the constriction. This observation may help to explain the following point: the words with anterior fricatives / affricates in word-initial position have worse performance in both I.T. and R.T. across all of the eight subjects. Available descriptions of the articulatory problems of children with cerebral palsy (Byrne, 1959) suggest that anterior lingual, continuant fricative classes of English phonemes, e.g., /t/, /sh/, and /s/, are most likely to be abnormal in cerebral-palsy speech. The Platt et al. (1980a) paper presented some findings on specific phonetic features of adult cerebral-palsy speech. It was reported that anterior lingual places of articulation for the production of fricative and affricate manners of consonants in the word-initial position were often inaccurate.

Byrne (1959) found that speech elements involving the complex movement of the tongue tip were the most frequently misarticulated by children with cerebral palsy. The misarticulation of the anterior portion of the tongue is the main reason for mispronunciation of alveolar and palatal consonants, especially for the fricative and affricate consonants. The data of Hixon and Hardy (1964) also suggested that there may be a number of cerebral palsy children who have difficulty with sounds requiring movement of the anterior portion of the tongue.

Study of the cönfusion errors of lateral consonants, /r/ and /l/, and the glide consonant, /w/, for each subject's speech shows that most of these subjects have some degree of lateral consonant confusion errors, especially JS, JR, and CH. The Platt et al. (1980a) also reported that anterior lingual places of articulation for the production of /r/ in the word-initial position were often inaccurate. Therefore, combining the alveolar and palatal position control problems and the lateral consonant confusions indicates most of these subjects have tongue tip or blade control problems to various degrees.

(3) **The errors for initial clusters were high for the I.T. and the production of these clusters was inconsistent, as determined from R.T. for all of these subjects.**

Study of the initial clusters for both I.T. and R.T. leads to the following two observations: (i) All of the subjects' initial clusters have worse R.T. performance than the average accuracy of the 70-word list. A similar conclusion is found from the I.T. data, except for GG (Table 4.2). (ii) In a comparison of the STOP_IN, CLUS_IN, and FR_AF_IN groups (Table 4.2), four subjects' worst I.T. performance (JF, MW, JR, and CH) is in CLUS_IN and the other four subjects' (JS, DW, ES, and GG) is in FR_AF_IN. Further, three subjects' worst R.T. performance (JF, MW, and JS) is in CLUS_IN, three subjects' (ES, JR, and CH) is in FR_AF_IN, and two subjects' (DW and GG) is equal in both CLUS_IN and FR_AF_IN. None of them has the worst I.T. and R.T. in STOP_IN. From these results, it appears that the initial clusters are similar in difficulty to fricatives / affricates in word-initial position. The reason may come from the articulatory components involved in the initial clusters defined in the 70-word list: /bl/, /sl/, /sp/, and /st/. These initial clusters all include at least one alveolar or lateral consonant. From the findings noted above, these subjects have tongue tip or blade control problems, and therefore they tend to misarticulate the initial clusters. However, it is also recognized that the production of clusters requires a complex coordination and sequencing of articulator movements - a skill which appears to be deficient for these speakers. Raghavendra et al. (1994) also reported that the words were misrecognized as other words due to the distorted consonant clusters, e.g., "stava" ➔ "talat" in Swedish.

(4) In comparison with other classes of obstruents, all of these subjects' stop consonants performance in word-initial position is fair.

Detailed study of the words with stop consonants in word-initial position indicates that all of the subjects' STOP_INs have a better I.T. performance than the average accuracy of the 70-word list and an R.T. performance similar to the average accuracy of the 70-word list. Furthermore, almost all of these subjects' STOP_INs have a better performance than CLUS_INs and FR_AF_INs for both I.T. and R.T. In short, these data for STOP_IN show that the words with stop consonants in word-initial position have a quite reasonable performance compared with the overall 70-word list accuracy average. It seems that a complete closure of the constriction area for stop consonants is more easily controlled than a slight opening of the constriction area for fricatives and affricates. Platt et al. (1980a) suggested, "cerebral-palsied adults are generally capable of accurate production of the manner categories of stop, nasal and glide, and the place categories of bilabial and velar" [p. 37].

(5) These subjects' errors in voicing occurred more frequently for voiceless consonants than for voiced cognates in word-initial position.

The results of the analysis of these subjects' speech show that errors in voicing occurred more frequently for the voiceless consonants than for their voiced cognates in word-initial position (e.g., /p/ is voiceless and its cognate /b/ is voiced). Four of the eight subjects (JS, DW, CH, and GG) have more errors in voicing for the voiceless consonants in word-initial position than for the voiced ones (in T.T.), but only one subject, JR, has the inverse result. The other three subjects (JF, MW, and ES), whose I.T. scores are good compared with other subjects, have no distinctive difference in the voiced - voiceless confusion. This problem will cause some of the words with voiceless consonants in word-initial position to be confused with their voiced cognates, e.g., "pad" and "bad". Byrne (1959) reported the same results for the cerebral-palsied children. It appears, then, that the speakers have difficulty in abducting the glottis to produce voiceless obstruent consonants.

(6) For all of these subjects, vowel front - back errors were less prevalent than obstruent consonant errors.

All of the subjects' results found in this research indicate that front and back vowel errors were less prevalent than consonant errors. Two subjects in particular, JF and ES, have 100% accuracy for all of the front - back - diphthong vowel comparisons in T.T. Further, all of the subjects' front - back - diphthong confusion matrixes in T.T. show better accuracy than the performance of obstruents. For example, DW's lowest score in the vowel matrix is 75% but the lowest score for place of articulation in the obstruent matrix is 8% (Table 4.9). In addition, for all subjects, the front - back contrast errors are less serious than the consonant contrast errors in the I.T. These speakers with dysarthria have better tongue-body control for front - back vowels than tongue-tip (or blade) control for consonants. In Coleman and Meyers' (1991) paper reporting on the recognition performance for speakers with dysarthria, they noted, "For the dysarthric group, vowels in an h-d environment were recognized significantly more than the consonants followed by a neutral vowel." [p. 34]. They also indicated that the dysarthric group had significantly fewer correct recognitions for consonants than for vowels. This finding is compatible with the conclusion of Johns and Darley (1970): "most characteristics of the speech of the dysarthric subjects in this study was the consistently imprecise production of consonants" [p. 579]

(7) Some of the subjects show excessive breathiness in their speech.

Subjects, JF, JS, DW, ES, CH, and GG all show a problem with excessive breathiness, characterized by high-frequency noise, in the vowel, especially above 2.5 kHz. There appears to be too much air flow through the glottis, resulting in whispered and hoarse phonation. Berry and Eisenson (1956) noted the whispered and breathy phonation of speakers with cerebral palsy. Hardy (1961) observed that children with cerebral palsy used more air volume per syllable than normal children. He also mentioned that these subjects' articulatory rate is often slow, motor control problems at the larynx may cause abnormal stress, and fluctuating velar control introduces variable nasality. Underlying these upper airway control problems may be a basic respiratory asynchrony which produces a poorly driving force for articulatory valving. The breathiness may have been indicative of defective moderation of air; that is, laryngeal airway resistance during vowel phonation was probably inadequate, resulting in excessive air flow through the larynx (Smitheran and Hixon, 1981). Evidently the vocal folds are positioned in a way that leaves an opening in the glottis during phonation.

(8) Utterance prolongation influences the speech recognition performance of these subjects.

Acoustic analysis was used to measure the average duration (A.D.), which is a dimensionless scale, for each subject's utterances (based on the 70-word list used in the I.T.), as shown in Table 4.1. The relation between the recognition rejection rate (Rej.%) and the average duration, A.D., for each specific subject is shown in Figure 5.3. Because

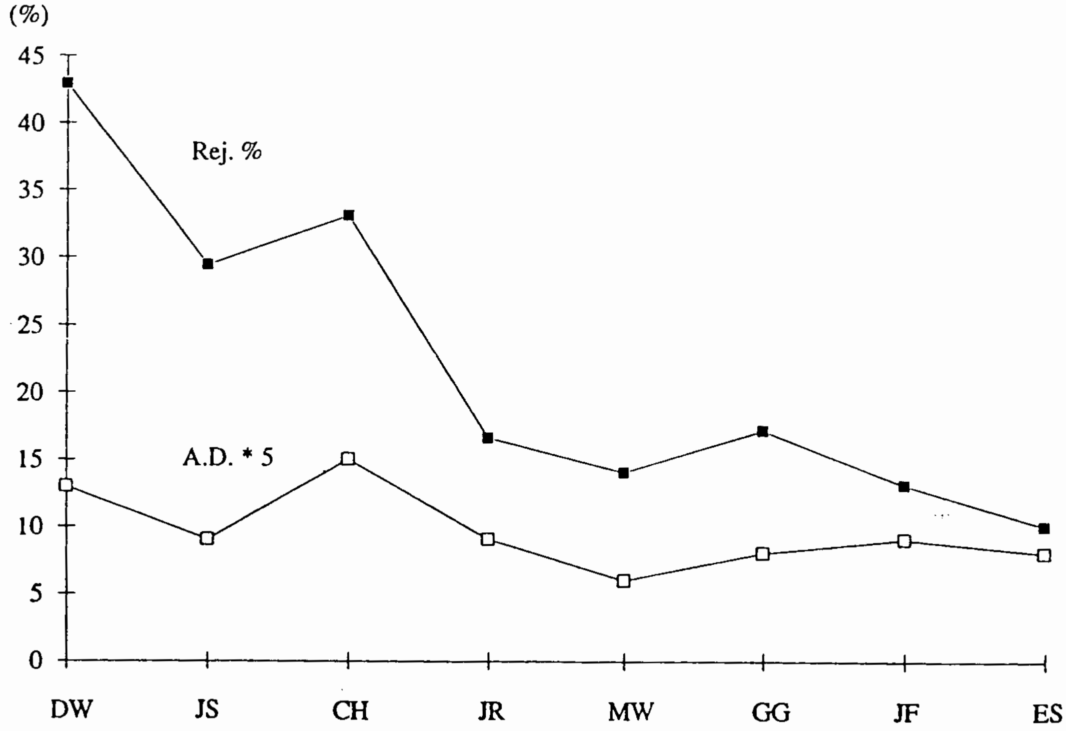


Figure 5.3 Comparison between the recognition rejection rate (Rej.%) and the average duration of the 70-word diagnostic list used in intelligibility test. The subjects are ordered according to increasing scores for the I.T. The A.D. values for each subject have been multiplied by five (A.D. * 5) in order to show the apparent relation between A.D. and Rej.% on the same scale.

the values of A.D. are too small to show the apparent relation between A.D. and Rej.%, all of the values of A.D. are multiplied by five. These data also show a correlation between A.D. and Rej.%, with a Pearson correlation coefficient of 0.79 with $0.005 < p < 0.01$, refer to Sincich (1987) [p. 576]. The figure shows that the local slopes for these two curves are very similar except for the last two data points, JF and ES. The average utterance duration based on the 70-word list used in the I.T. for CH is particularly large in comparison with other subjects' A.D.

CH took extra time to adjust her articulation to the right position and to build up lung pressure since her glottis and the left part of her articulatory system are paralyzed. Her speech sometimes sounded like a stutter. However, her speech position accuracy control is not as serious as DW's. Thus, her recognition rejection rate is still lower than that for DW's speech. In short, the high correlation coefficient with low p value and local curve slope similarity in Figure 5.3 demonstrate that the utterance duration can be one of the potentially important factors which influence speech consistency. These speakers with dysarthria need a longer time to adjust their articulation. However, in the adjustment procedure, when they produce a longer utterance, they have more chance to make mistakes or local fluctuations, e.g., tremors. These uncertain factors will make their speech sound variable and unstable. On the other hand, the overall prolongation of utterances will cause pauses within words (e.g., the pause in the /t/ of "sheet") longer than those of the utterances without prolongation. This unusually long period of pause will sometimes cause the speech recognizer to confuse one word as two words. Ferrier (1994) said pauses within words are also problematic for the commercial speech recognizer, Dragon Dictate, which she was using for speakers with dysarthria. She found, "If the speaker has too long of a pause between consonants, the machine may recognize the utterance as two words" [p. 13]. She also said, "We found the technology is more accurate if the user mispronounces a word consistently but speaks at a moderate-to-fast rate rather than if the user says the word nearly correctly but speaks too slowly because of this pausing problem" [p. 13].

(9) Use only 1 - 2 syllables in each word designed for each individual.

Practical experience shows that the more phonetic syllables there are in one word the worse is the speech consistency for these subjects. However, the lack of acoustic information would reduce the recognition accuracy if the utterance duration of the designed word is too short. Therefore, selection of the utterance length, particularly the number of syllables, is an important issue. One to two syllables in each word would be a good range for designing the word list. Hixon and Hardy (1964) also suggested, "children with cerebral palsy frequently have difficulty in producing a relatively large number of syllables on one expiration. Moreover, it is also commonly believed that speakers with cerebral palsy have difficulty in "coordinating" the processes of articulation, phonation, and respiration" [p. 300]. Hardy (1961) observed that children with cerebral palsy utilized more air volume per syllable than normal speakers. Together with a reduced vital capacity, the inefficient valving of the breath stream results in respiratory support for speech that is often inadequate. The problems with velopharyngeal closure and with laryngeal and oral articulation dysfunction may also contribute to insufficient intraoral air pressure for supporting the proper articulation. In summary, the reduced vital capacity, inefficient valving control, difficult and abnormal articulation control cause the subjects with cerebral palsy to have articulatory difficulty for words with several syllables.

5.2.3 Complementary Goal 2: supply different viewpoints that can lead to a better understanding of the speech patterns of the dysarthrias.

Assessment of the dysarthric speech includes the study of the articulatory-position control problem (using I.T. and T.T.), utterance-consistency control problem (using R.T.), and acoustic interpretation (using A.A.). In particular, a sub-goal is to develop and evaluate methods for quantitative assessment of dysarthric speech (R.T. and A.A.) as well as methods involving subjective observation (I.T. and T.T.) The detailed outcomes are listed as follows:

(1) Acoustic analysis and recognition test provide two diagnostic tools that are quantitative and can supplement the traditional perception methods.

The results from the perception methods (intelligibility test and transcription test), the recognition test, and the acoustic analysis, can supplement each other in the assessment of the speech of persons with dysarthria. Ludlow and Bassich (1982) made a comparison between an acoustic assessment and a perceptual assessment. They demonstrated that both assessment systems are valid for differentiating between dysarthric and normal speech. That paper also illustrated how acoustic analysis of patients' speech can be helpful in planning suitable treatment approaches. Therefore, the new speech technology can supply more referrals to the clinicians and can assist in making the correct assessments. Hammen (1994) also mentioned, "Modern advances in technology are the axis of current diagnostic work in dysarthria" [p. 12]. The two methods, R.T. and A.A., utilizing the new speech science technology can supply two more efficient diagnostic tools to study disordered speech. In particular, these two methods can provide quantitative assessments and are easy to replicate. Thus, they can overcome the disadvantage of I.T. and T.T., which are difficult to standardize over time and across different settings and judges. Table 4.1 has shown that the overall inter-judge consistency between judges in I.T is significant (overall average of I.T.-X2 = 4.6 < 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. However, JR's I.T.-X2 value shows that the inter-judge consistency for JR's speech is significantly different. It indicates that different judges could give different responses. The same study was also made for T.T. Table 4.1 has shown that the overall inter-judge consistency between each judge in T.T is significant (overall average of T.T.-X2 = 1.99 < 9.49 based on df = 4 and $\alpha = 0.05$), refer to Sincich (1987) [p. 715]. However, JS and CH's Chi-Square values (5.55 and 8.84) are high compared to the rest of subjects' values. Therefore, although I.T. and T.T. have significant inter-judge consistency, the difficulty of standardizing across different judges is still one of the disadvantages for I.T. and T.T. However, the R.T. and A.A. can efficiently overcome this difficulty. Furthermore, both R.T. and A.A. can be applied to study the speech consistency of specific individuals, but it is very hard to evaluate consistency for I.T. and T.T.

(2) There is a strong correlation between the I.T. (position accuracy control) and the R.T. (consistency control) from these subjects' speech data.

A comparison of I.T. and R.T. across all of these eight subjects is displayed in Figure 5.4. This figure shows that the I.T. and R.T. change together (Pearson correlation coefficient = 0.9 and p value < 0.005), refer to Sincich (1987) [p. 576]. The finding suggests that with better position accuracy control (giving a higher I.T. score) there is better consistency control (yielding a higher R.T. score). Although, in fact, the strong overall correlation coefficient scores do not mean that the detailed outcomes of these two tests can exactly predict each other (e.g., JS' "steak" has only 20% accuracy for I.T. but 100% for R.T.), the strong overall correlation between I.T. and R.T. still shows the potential relation between the computer recognizer test and the human listener test for assessment of dysarthric speech. Sy and Horowitz (1993) built up a statistical causal model to relate a speech impairment rating with speech recognition accuracy. They found that this model is very appealing to support the relation between listeners' subjective judgments of impaired speech and the performance of a laboratory version of a speech recognizer. Therefore, if there is a significant correlation between the intelligibility test (for articulatory error patterns of dysarthric speech) and the response of a speech recognizer, then the evaluation of the dysarthric speech could be deduced from the speech recognition test. On the other hand, the R.T. curve in Figure 5.4 has an exception in the data, for CH, which shows abnormally low recognition accuracy performance. Because CH's speech impairment comes in parts from surgery damage, her speech performance is a little different from the other subjects, as mentioned in Chapter 4. It is not surprising, then, that her data deviate from the pattern for other speakers.

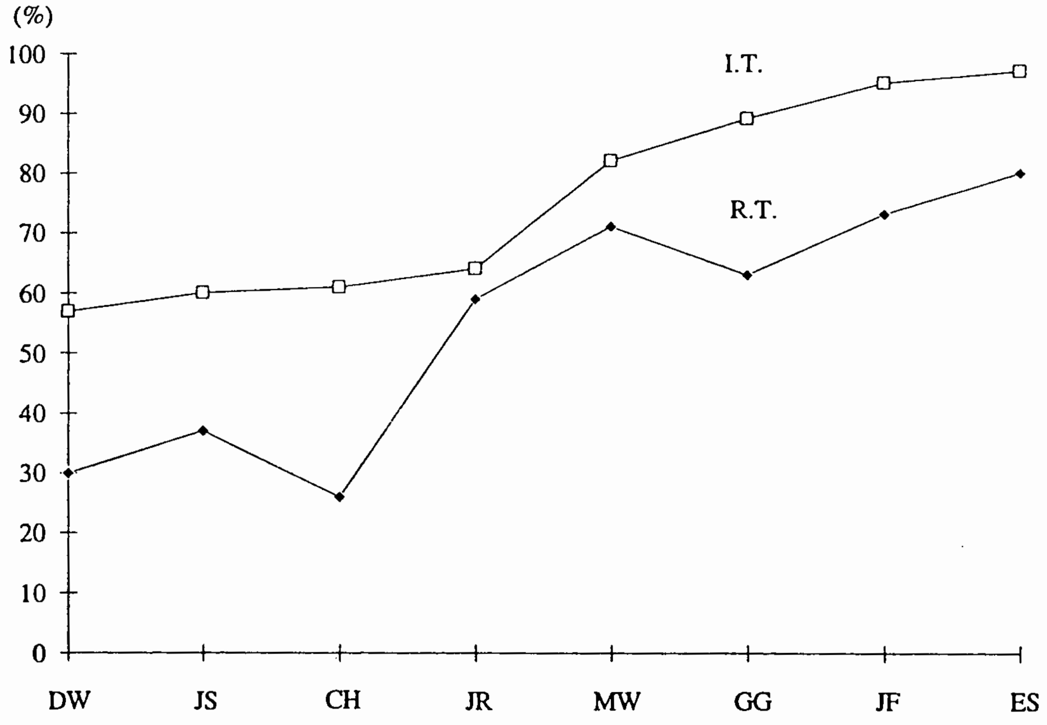


Figure 5.4 Comparison of the different test results across subjects. The subjects are ordered according to increasing scores for the I.T.

Further study of the I.T. performance and the recognition rejection rate in Figure 5.5 (based on the 70-word diagnosis list) shows a clear tendency: the lower the I.T. scores, the higher the recognition rejection rates. The Pearson correlation coefficient between I.T. and recognition rejection rate is -0.83 with $0.005 < p < 0.01$, refer to Sincich (1987) [p. 576]. The recognition rejection rate is a measure of the variability in production of the words. A possible interpretation is the following: since the poorest I.T. scores represent the more serious speech impairments, the malfunction of the specific subject's articulatory system causes more variability, including articulatory in consistency and timing variability. This interpretation can be used to explain why DW has such serious timing variability and speech consistency control problems. Moreover, because of the high recognition rejection rate for the severe speakers with dysarthria, not all of the subjects can benefit from the method used in this project. The benefits which can be obtained from this method will depend on how seriously impaired the specific subject is. However, this

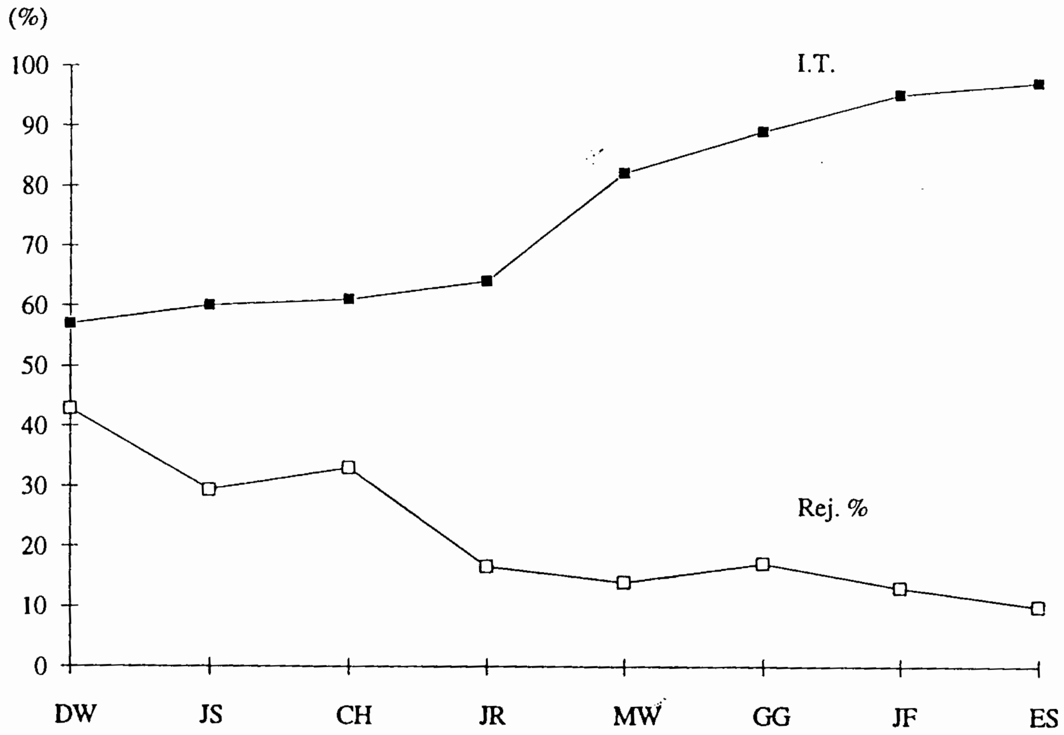


Figure 5.5 Comparison between the intelligibility test performance and the recognition rejection rate (Rej.%) for all eight subjects. The subjects are ordered according to increasing scores for the I.T.

project does indicate the possibility of using a new input channel to help most persons with dysarthria.

In summary, by combining all or part of the intelligibility test, transcription test, recognition test, and acoustic analysis, an efficient procedure and tool to configure a speech input system can be established, thereby assisting dysarthric individuals to achieve the highest possible speech recognition accuracy with a commercially available speech recognizer.

5.3 Summary and Project Contributions

The findings and the final improvements for each subject mentioned above show that integration of speech recognition and synthesis can successfully provide new communication channels for these subjects with dysarthria. This has been the goal for this project. However, although a primary goal of this project is to help to develop new input computer channels - speech recognizer and synthesis - for dysarthric computer users, it also provides the information about speech characteristics of dysarthric speakers, and has helped us to understand in depth the characteristics of their speech impairments. Thus our findings have not only provided help to specific speakers with dysarthria to control the computers by using particular words or sounds (which may be unclear to listeners but nevertheless are pronounced consistently), but have also supplied a new approach to make assessment for the speech characteristics of dysarthria. The four methods, I.T., T.T., R.T., and A.A., were integrated and implemented in this project.

5.4 Future Investigations and Alternatives

From this project and the findings mentioned above, many issues about dysarthric speech have been explored and studied. Although a great deal of recognition improvement for each subject has been achieved, there are still many difficult problems which need to be understood. An outcome of the testing and analysis is an initial word list for the specific subject to use as input to the computer. This word list is obtained from a large word pool so that the highest recognition accuracy is achieved. In the short term, developing an automatic word assessment system will be necessary to reduce the time of analyzing data from the human listeners. In this project, a short program that helps to make the speech assessment has been written. It has successfully reduced the operation time. However, the connection between the intelligibility test program and the recognition test should be developed in order to make the whole assessment procedure more automatic. After this step, a connection between the word-pool data base and the analysis outputs from I.T. and R.T. must be made in order to achieve the ideal, automatic word-design system.

The traditional speech pathology method for speech assessment is time consuming and is not objective. The recognition data have shown there is a strong relation between I.T. and R.T. In the future, it may be possible to use the recognition test to replace the intelligibility test. The computers can be used to replace the human listeners as judges. The method is more objective than the traditional method involving perception tests. The computer is used to collect the abnormal speech data, the data are analyzed off line, and a final assessment of the subjects' speech is obtained. Then, by supplementary the computer-based results with acoustic data analysis, the basic speech characteristics for the specific individual can be obtained. This step involves developing an efficient method for acoustic analysis of the dysarthric speech. For example, it will be necessary to detect acoustic landmarks to measure utterance durations, and to track formants.

A comparison of the findings of I.T. with T.T. indicates that some of the conclusions from I.T. and T.T. are conflicting. For example, CH's I.T. shows that her alveolar - palatal confusion problems are more serious than velar - other consonant ones (Alveolar - Palatal Place error (30% error) and Velar - Other Consonant Place (25% error) in Figure 4.13-b). However, her T.T. shows that her velar obstruents have the most serious errors (Velar Position (0% accuracy), Palatal Position (25% accuracy), and Alveolar Position (28% accuracy) in Table. 4.8). The findings from I.T. show CH's alveolar and palatal position controls are the main errors, but the findings from T.T. show CH's main malfunction is in the velar position. These two findings are conflicting; alveolar position belongs to the anterior portion of the oral cavity but the velar position belongs to the more posterior portion. The main difference may come from the judgment data base. I.T. has a four-choose-one list as a reference. Therefore, even if the judges lose the first consonants of the words, they can still choose the right word from the following acoustic information, i.e., vowels and final consonants. However, T.T. is judged based on what the human listeners hear for the consonants and vowels, without the reference list. The judges in T.T. only focus on each separate phoneme and do not focus on the lexical status of the utterance. Therefore, the I.T. is not as sensitive as T.T. in reflecting all of the articulatory malfunctions. The designed confusion pairs for I.T. can not include all of the possible confusion pairs. Thus, the findings of I.T. are less reliable than T.T. in studying word-initial consonants. However, the I.T. is convenient for applying to practical clinical work because the judges do not require prior training and the data analysis of I.T. can be performed by computers.

In addition, except for the reliability problem for the I.T., the error-pair contrast distributions are not uniform. Some of the error contrasts have too many trials and some of the contrasts have too few trials. The four-choose-one list has some bias in errors due to the fact that the phonetic contrast groups are not uniformly distributed, i.e., some phonetic contrast groups have too few tokens to provide a reliable estimate of the errors. For example, the /r/ - /l/ contrast group has 35 tokens but the /l/ - /r/ contrast group only has 10 tokens. The error scores are less reliable for the /l/ - /r/ group due to the limited number of tokens. In the future, this 70-word list and the reference list need to be improved to represent more reliably the types of abnormalities for particular speakers with dysarthria. These biased distributions can cause inaccurate conclusions from the analysis. For example, all of the pairs in affricate - stop contrasts only focus on affricates confused with stop consonants but there are no pairs studying the possibility of stop consonants confused with affricates. The diagnostic conclusions may give the analyst a wrong idea about a specific subject: "This subject has no stop consonant (goal) - affricate (response) confusion problem". The analyst will consider this subject as normal in this error contrast; however, in fact, this kind of trial does not exist at all. Thus, this diagnostic list should be redesigned to give more balanced estimates of the distribution of errors.

This project not only studies the speech characteristics of speakers with dysarthria, but it also aims to design a final word list for practical applications. Therefore, a user-preference dictation system is an important factor in the applications in order to make the users easily operate the computer by using this new system. For example, in order to let the subjects remember easily the specific word list and to fit the requirement of an alphabet character dictation, an alphabet dictation system has been designed for each subject. However, for the purpose of reducing the misrecognition rate, proper grouping of these 52 words into different classes depending on the task will be another issue which needs to be considered. Further grouping of the 52-word list will reduce efficiently the misrecognition rate since only a subset of the 52 words is activated in each group, e.g., only 10 words might be activated in each specific group instead of activating all 52 words. In addition, by using the classifications, the same words can be used in different groups to represent different meanings, e.g., "air" represents "a" in the alphabet group but it also can be used to represents "1" in the numerical group. Hence, the number of words needed for these subjects to dictate can be reduced and the recognition accuracy can be increased by using the grouping method.

From the analysis of I.T., T.T., R.T., A.A., and real experience, it is evident that the consonants in word-initial position have more serious impairments than the vowels. Thus, most of the efforts were put on word-initial consonants. However, the vowels still supply very important information and have a strong influence on speech recognition performance. In the future, the study of vowel defects will be of great help in the design of the word lists. A complete study of all of the relevant phonetic features should be included in this project in order to understand more about these subjects' speech abnormalities.

It is also evident that most of these subjects have the special difficulty of pronunciation in word-initial position, e.g., some utterances begin like a stutter or require a great deal of effort to initialize. In addition, the wheelchair noise coming from the involuntary body movements, deep breathing noise, and saliva noise causes the speech recognizer to misrecognize these noises as a command or word. This is a very common type of error when these subjects use the speech recognizers. Raghavendra et al. (1994) also reported that stuttering-like repetitions and involuntary sounds led computers to misrecognize as a word (or words). Although an optimal word-list can efficiently solve most of the articulatory errors (e.g., use the sonorant consonants or vowels in the word-initial position), it still can not eliminate the influence of the extraneous noises on speech recognizer. Thus, in the future, a new front-end algorithm of the speech recognizer should be designed to reduce these environmental or involuntary noises, e.g., rejecting the utterances whose duration is less than 100 ms.

Speech recognition presents an interesting possibility as an alternative communication system. Baker (1989) has reported that with practice, single word entry rates of up to 60 words per minute can be obtained for normal speakers by using the speech recognizer system. If the speech recognition accuracy and the dictation rate can be improved for dysarthric computer users, speech recognition can offer a new, rapid, and efficient communication system.

Finally, as Lass (1994) said, "The more people understand about dysarthria and other communication disorders, the less likely they will be to judge a book by its cover. People with dysarthria are not disordered people; they are people first who have a disorder" [p. 13]. Many persons with the speech disorder still strive to use their voices effectively to communicate with the other persons. A device that utilizes the user's speech for communication, computer operation, or writing still seems worth investigating and studying for these persons who are affected with dysarthria. This project has shown that the speech recognition technique is available to address these unique communication disorders.

Glossary:

A.A.: acoustic analysis.

A.D.: average duration.

Affricate: a stop consonant and its immediately following release through the articulatory position for a continuant nonsyllabic consonant, e.g., /t/ + /sh/ become as /ch/ as "church".

Automatic Speech Recognition (ASR): a speech input interface for the computer.

Cerebral Palsy: referring to disorders of movement resulting from damage to the brain at some time during the period of growth and development.

Chi-Square test: a statistics method used to identify the probability of the hypothesis to be correct.

Clusters: two or more successive consonants in a word, e.g., /kl/ and /st/ in "cluster".

Dysarthric Speech: a group of speech disorders resulting from disturbances in the muscular control of the speech mechanism due to neuromuscular disease.

Fricative: characterized by frictional passage of the expired voiced or voiceless breath against a narrowing at some point in the vocal tract, e.g., /sh/ as "shoe".

Hidden Markov Model (HMM): a powerful statistical method of characterizing the spectral properties of the frames of a pattern.

I.T.: intelligibility test.

Obstruent: a phone with building up apparent pressure drop across the constriction in the vocal tract, e.g., /t/, /d/, and /ch/.

R.T.: recognition test.

Sonorant: a phone with source at glottis and without building up apparent pressure drop across the constriction in the vocal tract, e.g., /m/, /i/, and /r/.

T.T.: transcription test.

Voiced Consonant: a consonant with the glottal source vibration, e.g., /b/, /d/, /g/.

Voiceless Consonant: a consonant without the glottal source vibration, e.g., /t/, /p/, and /k/.

Tolerance: a index number between 1 and 100 to define the level of variability in the subject's speech to the speech recognizer, e.g., a value of 100 makes the Dragon Writer-1000 very tolerant of variability in the subject's speech (but also makes it more prone to mistakenly recognize extraneous noise as real commands) and a value of 1 makes the Dragon Writer-1000 reject almost all noise but also most valid utterances.

Appendix:

Appendix 3.1 The four-choose-one lists for the 70-diagnostic words are listed in the same form as the original list adapted from Kent et al. (1989). The first column gives the target words. The four items in each row give the possible responses for the listeners.

1	bad	bed	bat	pad
2	sip	ship	tip	zip
3	spit	pit	sit	it
4	knot	dot	nod	nut
5	sigh	shy	tie	thigh
6	sheet	seat	feet	eat
7	sticks	six	ticks	stick
8	knew	know	knee	gnaw
9	leak	lick	league	reek
10	chair	share	tear	air
11	nice	knife	night	dice
12	write	ride	light	white
13	side	sign	sight	sigh
14	pat	bat	pot	pad
15	hand	and	sand	fanned
16	ate	hate	aid	fate
17	witch	wish	rich	wit
18	much	mush	mut	muck
19	sew	show	toe	foe
20	feed	food	feet	fee
21	him	hem	ham	hum
22	at	hat	fat	add
23	air	hair	fair	are
24	pit	pet	pat	bit
25	read	lead	weed	rid
26	sell	tell	shell	fell
27	blend	bend	lend	end
28	shoot	suit	sheet	shot
29	see	she	he	tea
30	slip	sip	lip	sleep
31	steak	snake	take	sake
32	blow	low	bow	bloat
33	beat	boot	bit	meat
34	sin	shin	in	tin
35	rock	walk	lock	rocks
36	geese	goose	guess	gas
37	chop	chap	shop	top
38	ship	sheep	chip	tip
39	feet	fit	heat	fat
40	coat	goat	code	tote
41	dug	tug	duck	bug
42	cash	gash	catch	cat
43	fill	hill	pill	full
44	hat	fat	pat	that
45	hold	old	fold	cold
46	heat	eat	feet	hate

47	bill	mill	dill	gill
48	ache	aches	ape	ate
49	lip	leap	lit	rip
50	reap	rip	leap	weep
51	rise	wise	lies	eyes
52	row	woe	low	owe
53	wax	wack	lax	racks
54	dock	docks	mock	knock
55	cheer	sheer	sear	tear
56	hash	hatch	ash	dash
57	tile	dial	pile	mile
58	bunch	munch	punch	bun
59	ease	is	cheese	peas
60	seed	see	seeds	feed
61	sink	sing	pink	ink
62	harm	arm	charm	farm
63	cake	cakes	take	ache
64	meat	me	meats	neat
65	had	add	pad	hid
66	hail	ail	sail	tail
67	hall	all	tall	ball
68	fork	four	forks	cork
69	rake	ray	rakes	lake
70	leak	lee	leaks	Luke

Appendix 3.2 Word pairs from the intelligibility test listed by 19 phonetic contrasts (adapted from Kent et al. (1989)).

(1.) Front - Back Vowel Contrasts
Pairs: 11

knew	pat	him	shoot	beat	geese	feed	air
knee	pot	hum	sheet	boot	goose	food	are
chop	fill	leak					
chap	full	Luke					

(2.) High - Low Vowel Contrasts
Pairs: 12

knew	knew	him	him	shoot	geese	geese	pit
know	gnaw	hem	ham	shot	gas	guess	pet
pit	feet	heat	had				
pat	fat	hate	hid				

(3.) Vowel Duration Contrasts
Pairs: 11

beat	slip	leak	knot	read	ship	feet	lip
bit	sleep	lick	nut	rid	sheep	fit	leap
ease	reap	bad					
is	rip	bed					

(4.) Initial Voicing Contrasts
Pairs: 9

pat	bad	pit	sip	coat	dug	cash	tile
bat	pad	bit	zip	goat	tug	gash	dial
bunch							
punch							

(5.) Final Voicing Contrasts
Pairs: 11

feed	bad	leak	knot	write	side	coat	dug
feet	bat	league	nod	ride	sight	code	duck
ate	at	pat					
aid	add	pad					

(6.) Alveolar - Palatal Place Contrasts

Pairs: 8

sip	shoot	sigh	sell	sin	sew	see	sheet
ship	suit	shy	shell	shin	show	she	seat

(7.) Consonant Place Contrasts

Pairs: 10

dug	tile	cake	meat	bill	bill	ache	ache
bug	pile	take	neat	dill	gill	ape	ate
lip	coat						
lit	tote						

(8.) Other Fricative Place Contrasts

Pairs: 17

sheet	sigh	fill	hand	sew	see	nice	hat
feet	thigh	hill	sand	foe	he	knife	fat
sell	feet	hat	hold	hail	harm	seed	hand
fell	heat	that	fold	sail	farm	feed	fanned
heat							
feet							

(9.) Fricative - Affricate Contrasts

Pairs: 10

chair	witch	much	ship	chop	cash	cheer	cheer
share	wish	mush	chip	shop	catch	sheer	sear
hash	harm						
hatch	charm						

(10.) Stop - Fricative Contrasts

Pairs: 19

sip	sigh	sell	sin	sew	nice	sea	cash
tip	tie	tell	tin	toe	night	tea	cat
hat	fill	hold	ship	had	hail	hall	fork
pat	pill	cold	tip	pad	tail	tall	cork
hash	hall	sink					
dash	ball	pink					

(11.) Stop - Affricate Contrasts

Pairs: 6

chair	much	chop	witch	much	cheer
tear	mut	top	wit	muck	tear

(12.) Stop -Nasal Consonant Contrasts
Pairs: 10

beat	knot	side	nice	steak	bill	dock	dock
meat	dot	sign	dice	snake	mill	mock	knock

bunch	tile
munch	mile

(13.) Initial Glottal Consonant - Null Contrasts
Pairs: 11

air	ate	at	hand	hold	heat	hash	harm
hair	hate	hat	and	old	eat	ash	arm

had	hail	hall
add	ail	all

(14.) Initial Consonant - Null Contrasts
Pairs: 14

air	ate	at	sin	sheet	chair	spit	blend
fair	fate	fat	in	eat	air	it	end

ease	ease	sink	cake	rise	row
peas	cheese	ink	ache	eyes	owe

(15.) Final Consonant - Null Contrasts
Pairs: 9

feed	side	blow	fork	rake	leak	meat	bunch
fee	sigh	bloat	four	ray	lee	me	bun

seed
see

(16.) Initial Cluster - Singleton Contrasts
Pairs: 12

slip	slip	spit	spit	blend	blend	sticks	sticks
sip	lip	pit	sit	bend	lend	six	ticks

steak	steak	blow	blow
take	sake	low	bow

(17.) Final Cluster - Singleton Contrasts
Pairs: 12

sticks	rock	seed	sink	cake	meat	fork	rake
stick	rocks	seeds	sing	cakes	meats	forks	rakes

leak	ache	wax	dock
leaks	aches	wack	docks

(18.) /r/ - /l/ Contrasts
Pairs: 8

read	write	leak	rock	rake	lip	reap	rise
lead	light	reek	lock	lake	rip	leap	lies

row	wax
low	lax

(19.) /r/ - /w/ Contrasts
Pairs: 8

read	write	witch	rock	reap	rise	row	wax
weed	white	rich	walk	weep	wise	woe	racks

Appendix 3.3 Word pairs from the intelligibility test listed by 40 phonetic contrasts. These pairs show the target (before "-") and the actual response (after "-") that were scored as an error.

(1.) Front - Back Vowel Contrasts
Pairs: 8

pat	him	beat	geese	feed	air	fill	leak
pot	hum	boot	goose	food	are	full	Luke

(2.) Back - Front Vowel Contrasts
Pairs: 3

knew	shoot	chop
knee	sheet	chap

(3.) High - Low Vowel Contrasts
Pairs: 11

knew	knew	him	him	shoot	geese	geese	pit
know	gnaw	hem	ham	shot	gas	guess	pet

pit	feet	heat
pat	fat	hate

(4.) Low - High Vowel Contrasts
Pairs: 1

had
hid

(5.) Long - Short Vowel Duration Contrasts
Pairs: 8

beat	leak	knot	read	feet	ease	reap	bad
bit	lick	nut	rid	fit	is	rip	bed

(6.) Short - Long Vowel Duration Contrasts
Pairs: 3

slip	ship	lip
sleep	sheep	leap

(7.) Initial Voicing - Voiceless Contrasts
Pairs: 3

bad	dug	bunch
pad	tug	punch

(8.) Initial Voiceless - Voicing Contrasts

Pairs: 6

pat	pit	sip	coat	cash	tile
bat	bit	zip	goat	gash	dial

(9.) Final Voicing - Voiceless Contrasts

Pairs: 4

feed	bad	side	dug
feet	bat	sight	duck

(10.) Final Voiceless - Voicing Contrasts

Pairs: 7

leak	knot	write	coat	ate	at	pat
league	nod	ride	code	aid	add	pad

(11.) Alveolar - Palatal Place Contrasts

Pairs: 6

sip	sigh	sell	sin	sew	see
ship	shy	shell	shin	show	she

(12.) Palatal - Alveolar Place Contrasts

Pairs: 2

shoot	sheet
suit	seat

(13.) Alveolar - Other Consonant Place Contrasts

Pairs: 2

dug	tile
bug	pile

(14.) Velar - Other Consonant Place Contrasts

Pairs: 4

cake	ache	ache	coat
take	ape	ate	tote

(15.) Consonant Place Contrasts

Pairs: 4

meat	bill	bill	lip
neat	dill	gill	lit

(16.) Alveolar - Other Fricative Place Contrasts
Pairs: 6

sigh	sew	see	nice	sell	seed
thigh	foe	he	knife	fell	feed

(17.) Palatal - Other Fricative Place Contrasts
Pairs: 1

sheet
feet

(18.) Labial - Other Fricative Place Contrasts
Pairs: 2

fill	feet
hill	heat

(19.) Glottal - Other Fricative Place Contrasts
Pairs: 8

hand	hat	heat	hat	hold	hail	harm	hand
sand	fat	feat	that	fold	sail	farm	fanned

(20.) Fricative - Affricate Contrasts
Pairs: 3

ship	cash	hash
chip	catch	hatch

(21.) Affricate - Fricative Contrasts
Pairs: 6

chair	witch	much	chop	cheer	cheer
share	wish	mush	shop	sheer	sear

(22.) Glottal Fricative - Affricate Contrasts
Pairs: 1

harm
charm

(23.) Fricative - Stop Contrasts
Pairs: 19

sip	sigh	sell	sin	sew	nice	sea	cash
tip	tie	tell	tin	toe	night	tea	cat

hat	fill	hold	ship	had	hail	hall	fork
pat	pill	cold	tip	pad	tail	tall	cork

hash	hall	sink
dash	ball	pink

(24.) Affricate - Stop Contrasts
Pairs: 6

chair	much	chop	witch	much	cheer
tear	mut	top	wit	muck	tear

(25.) Stop -Nasal Consonant Contrasts
Pairs: 8

beat	side	steak	bill	dock	dock	bunch	tile
meat	sign	snake	mill	mock	knock	munch	mile

(26.) Nasal - Stop Consonant Contrasts
Pairs: 2

knot	nice
dot	dice

(27.) Initial Glottal Consonant - Null Contrasts
Pairs: 8

hand	hold	heat	hash	harm	had	hail	hall
and	old	eat	ash	arm	add	ail	all

(28.) Null - Initial Glottal Consonant Contrasts
Pairs: 3

air	ate	at
hair	hate	hat

(29.) Initial Consonant - Null Contrasts
Pairs: 9

sin	sheet	chair	spit	blend	sink	cake	rise
in	eat	air	it	end	ink	ache	eyes

row
owe

(30.) Initial Null - Consonant Contrasts
Pairs: 5

air	ate	at	ease	ease
fair	fate	fat	peas	cheese

(31.) Final Consonant - Null Contrasts

Pairs: 8

feed	side	fork	rake	leak	meat	bunch	seed
fee	sigh	four	ray	lee	me	bun	see

(32.) Final Null - Consonant Contrasts

Pairs: 1

blow
bloat

(33.) Initial Cluster - Singleton Contrasts

Pairs: 12

slip	slip	spit	spit	blend	blend	sticks	sticks
sip	lip	pit	sit	bend	lend	six	ticks
steak	steak	blow	blow				
take	sake	low	bow				

(34.) Final Cluster - Singleton Contrasts

Pairs: 3

sticks	sink	wax
stick	sing	wack

(35.) Final Singleton - Cluster Contrasts

Pairs: 9

rock	seed	cake	meat	fork	rake	leak	ache
rocks	seeds	cakes	meats	forks	rakes	leaks	aches
dock							
docks							

(36.) /r/ - /l/ Contrasts

Pairs: 7

read	write	rock	rake	reap	rise	row
lead	light	lock	lake	leap	lies	low

(37.) /l/ - /r/ Contrasts

Pairs: 2

leak	lip
reek	rip

(38.) /w/ - /l/ Contrasts

Pairs: 1

wax
lax

(39.) /r/ - /w/ Contrasts

Pairs: 6

read	write	rock	reap	rise	row
weed	white	walk	weep	wise	woe

(40.) /w/ - /r/ Contrasts

Pairs: 2

witch	wax
rich	racks

Appendix 3.4 The word lists for the accuracy study of each group in Tables 3.4 and 3.5.

(1) SONOR_IN:

Total Number: 27

(There are two "leak"s in the 70-word list.)

had hail hall hand harm hat hash
heat him hold knew knot leak-1 leak-2
lip meat much nice rake read reap
rise rock row wax witch write

(2) WTOT_OB:

Total Number: 7

air hail hall harm him knew row

(3) OBS_IN:

Total Number: 38

bad beat bill blend blow bunch cake
cash chair cheer chop coat dock dug
feed feet fill fork geese pat pit
see seed sell sew sheet ship shoot
side sigh sin sink sip slip spit
steak sticks tile

(4) STOP_IN:

Total Number: 15

(i) LABIAL:

Total Number: 8

bad beat bill blend blow bunch pat
pit

(ii) ALVEOLAR:

Total Number: 3

dock dug tile

(iii) VELAR:

Total Number: 4

cake cash coat geese

(5) CLUS_IN:

Total Number: 6

blend blow slip spit steak sticks

(6.) FR_AF_IN:

Total Number: 23

(i) /f/_IN:

Total Number: 4

feet fill feed fork

(ii) /s/_IN:

Total Number: 13

see seed sell sew side sigh sin
sink sip slip spit steak steak

(iii) /sh/_IN:

Total Number: 3

sheet ship shoot

(iv) /ch/_IN:

Total Number: 3

chair cheer chop

Appendix 3.5 The word lists for the accuracy study of each type in transcription test.

(1) Obstruent - Sonorant Accuracy Analysis:

(i) Obstruent Accuracy Analysis:
Total Number: 38

bad	beat	bill	blend	blow	bunch	cake
cash	chair	cheer	chop	coat	dock	dug
feed	feet	fill	fork	geese	pat	pit
see	seed	sell	sew	sheet	ship	shoot
side	sigh	sin	sink	sip	slip	spit
steak	sticks	tile				

(ii) Sonorant Accuracy Analysis:
Total Number: 27
(There are two "leak"s in the 70-word list.)

had	hail	hall	hand	harm	hat	hash
heat	him	hold	knew	knot	leak-1	leak-2
lip	meat	much	nice	rake	read	reap
rise	rock	row	wax	witch	write	

(2) Vowel Accuracy Analysis:

(i) Front Vowel Accuracy Analysis:
Total Number: 47

ache	air	at	ate	bad	beat	bill
blend	cake	cash	chair	cheer	ease	feed
feet	fill	geese	had	hail	hand	hat
hash	heat	him	leak-1	leak-2	lip	meat
pat	pit	rake	read	reap	see	seed
sell	sheet	ship	sin	sink	sip	slip
spit	steak	sticks	wax	witch		

(ii) Back Vowel Accuracy Analysis:
Total Number: 17

blow	bunch	chop	coat	dock	dug	fork	hall
harm	hold	knew	knot	much	rock	row	sew
shoot							

(iii) Diphthong Accuracy Analysis:
Total Number: 6

nice	rise	side	sigh	tile	write

(3) Voicing - Voiceless Accuracy Analysis:

(i) Voiced Consonant Accuracy Analysis:

Total Number: 26

bad beat bill blend blow bunch dock
dug geese
knew knot leak-1 leak-2 lip meat much
nice rake read reap rise rock row
wax witch write

(ii) Voiceless Consonant Accuracy Analysis:

Total Number: 39

cake cash chair cheer chop coat feed
feet fill fork pat pit see seed
sell sew sheet ship shoot side sigh
sin sink sip slip spit steak sticks
tile
had hail hall hand harm hat hash
heat him hold

(4) Obstruent Place Accuracy Analysis:

(i) Labial Place Accuracy Analysis:

Total Number: 12

bad beat bill blend blow bunch feed
feet fill fork pat pit

(ii) Alveolar Place Accuracy Analysis:

Total Number: 16

dock dug see seed sell sew side
sigh sin sink sip slip spit steak
sticks tile

(iii) Palatal Place Accuracy Analysis:

Total Number: 6

(/ch/ is put in this group too.)

chair cheer chop sheet ship shoot

(iv) Velar Place Accuracy Analysis:

Total Number: 4

cake cash coat geese

Appendix 4.1

The following list is used to dictate to the PC for JF.

Root Functions:

"'voice console'":	Mary	"'number key'":	hammer
" 'alphabet spell' ":	normal	" 'command' ":	row
" 'scratch that' ":	/a/	" 'enter' ":	new

Alphabet spelling:

a:	air	b:	bad
c:	cake	d:	dot
e:	earn	f:	fever
g:	gnaw	h:	him
i:	/ai/	j:	jump
k:	knee	l:	long
m:	melon	n:	neon
o:	our	p:	plat
q:	quote	r:	ray
s:	soap	t:	tape
u:	/u/	v:	vain
w:	wine	x:	X
y:	yellow	z:	Zebra

Punctuation:

" . ":	away	" , ":	comma
" _ ":	tile	" ":	well
" 'enter' ":	new	" 'scratch that' ":	/a/

Number Keys:

1 :	air	2 :	bad
3 :	cake	4 :	dot
5 :	earn	6 :	fever
7 :	gnaw	8 :	him
9 :	/ai/	0 :	jump

Command Keys:

" 'escape' ":	only	" 'up-arrow' ":	her
" 'down-arrow'":	linen	" 'left-arrow' ":	_ama_
" 'right-arrow' ":	worm		
" 'begin of line' ":	home	" 'end of line' ":	end

General Keys:

" 'quit list' ":	rumor	" 'manual' ":	menu

Appendix 4.2

The following list is used to dictate to the PC for MW.

Root Functions:

"'voice console'":	Mary	"'number key'":	hammer
" 'alphabet spell' ":	normal	" 'command' ":	row
" 'scratch that' ":	/a/	" 'enter' ":	new

Alphabet spelling:

a:	air	b:	bad
c:	coke	d:	dot
e:	earn	f:	fever
g:	gnaw	h:	him
i:	/ai/	j:	Julia
k:	knee	l:	long
m:	melon	n:	neon
o:	our	p:	pit
q:	quote	r:	ray
s:	seem	t:	tea
u:	/u/	v:	vain
w:	wine	x:	X
y:	yellow	z:	Zebra

Punctuation:

" . ":	away	" , ":	show
" _ ":	tile	" ":	well
" 'enter' ": new		" 'scratch that' ": /a/	

Number Keys:

1 :	air	2 :	bad
3 :	coke	4 :	dot
5 :	earn	6 :	fever
7 :	gnaw	8 :	him
9 :	/ai/	0 :	Julia

Command Keys:

" 'escape' ":	only	" 'up-arrow' ":	her
" 'down-arrow'":	linen	" 'left-arrow' ":	_ama_
" 'right-arrow' ":	worm		
" 'begin of line' ":	home	" 'end of line' ":	end

General Keys:

" 'quit list' ":	rumor	" 'manual' ":	menu

Appendix 4.3

The following list is used to dictate to the PC for JS.

Root Functions:

"'voice console'":	Mary	"'number key'":	hammer
" 'alphabet spell' ":	normal	" 'command' ":	row
" 'scratch that' ":	/a/	" 'enter' ":	new

Alphabet spelling:

a:	air	b:	bad
c:	cake	d:	dot
e:	earn	f:	fever
g:	gnaw	h:	him
i:	/ai/	j:	Julia
k:	knee	l:	lima
m:	melon	n:	neon
o:	our	p:	plat
q:	quote	r:	ray
s:	soap	t:	tea
u:	/u/	v:	vain
w:	wine	x:	X
y:	yellow	z:	Zebra

Punctuation:

" . ":	away	" , ":	comma
" _ ":	tile	" ":	well
" 'enter' ":	new	" 'scratch that' ":	/a/

Number Keys:

1 :	air	2 :	bad
3 :	cake	4 :	dot
5 :	earn	6 :	fever
7 :	gnaw	8 :	him
9 :	/ai/	0 :	Julia

Command Keys:

" 'escape' ":	only	" 'up-arrow' ":	her
" 'down-arrow'":	linen	" 'left-arrow' ":	_ama_
" 'right-arrow' ":	worm		
" 'begin of line' ":	home	" 'end of line' ":	end

General Keys:

" 'quit list' ":	rumor	" 'manual' ":	menu

Appendix 4.4

The following list is used to dictate to the PC for DW.

Root Functions:

"'voice console'":	Mary	"'number key'":	hammer
" 'alphabet spell' ":	more	" 'command' ":	row
" 'scratch that' ":	/a/	" 'enter' ":	new

Alphabet spelling:

a:	air	b:	be
c:	chevy	d:	day
e:	earn	f:	fever
g:	gnaw	h:	him
i:	/ai/	j:	Julia
k:	knee	l:	long
m:	mena	n:	_nana_
o:	"o"	p:	plat
q:	quote	r:	ray
s:	sheet	t:	team
u:	/u/	v:	vain
w:	wine	x:	xenon
y:	yellow	z:	Zebra

Punctuation:

" . ":	our	" , ":	show
" _ ":	tile	" ":	well
" 'enter' ": new		" 'scratch that' ": /a/	

Number Keys:

1 :	air	2 :	be
3 :	chevy	4 :	day
5 :	earn	6 :	fever
7 :	gnaw	8 :	him
9 :	/ai/	0 :	Julia

Command Keys:

" 'escape' ":	only	" 'up-arrow' ":	her
" 'down-arrow'":	linen	" 'left-arrow' ":	_ama_
" 'right-arrow' ":	worm		
" 'begin of line' ":	home	" 'end of line' ":	end

General Keys:

" 'quit list' ":	rumor	" 'manual' ":	menu

Appendix 4.5

The following list is used to dictate to the PC for ES.

Root Functions:

"'voice console'":	Mary	"'number key'":	hammer
" 'alphabet spell' ":	normal	" 'command' ":	row
" 'scratch that' ":	erase	" 'enter' ":	new

Alphabet spelling:

a:	air	b:	bad
c:	cake	d:	dot
e:	earn	f:	fever
g:	gnaw	h:	him
i:	/ai/	j:	Julia
k:	knee	l:	lima
m:	melon	n:	neon
o:	our	p:	plat
q:	quote	r:	ray
s:	sheet	t:	tune
u:	/u/	v:	vain
w:	wine	x:	X
y:	yellow	z:	Zebra

Punctuation:

" . ":	stop	" , ":	show
" _ ":	tile	" ":	space
" 'enter' ": new		" 'scratch that' ": erase	

Number Keys:

1 :	air	2 :	bad
3 :	cake	4 :	dot
5 :	earn	6 :	fever
7 :	gnaw	8 :	him
9 :	/ai/	0 :	Julia

Command Keys:

" 'escape' ":	only	" 'up-arrow' ":	her
" 'down-arrow'":	linen	" 'left-arrow' ":	_ama_
" 'right-arrow' ":	worm		
" 'begin of line' ":	home	" 'end of line' ":	end

General Keys:

" 'quit list' ":	rumor	" 'manual' ":	menu

Appendix 4.6

The following list is used to dictate to the PC for JR.

Root Functions:

"'voice console'":	_mumu_	"'number key'":	hammer
" 'alphabet spell' ":	normal	" 'command' ":	row
" 'scratch that' ":	/a/	" 'enter' ":	new

Alphabet spelling:

a:	air	b:	bad
c:	cake	d:	dot
e:	earn	f:	funny
g:	gnaw	h:	him
i:	/ai/	j:	jump
k:	knee	l:	long
m:	melon	n:	neon
o:	our	p:	pit
q:	quote	r:	ray
s:	soap	t:	type
u:	/u/	v:	very
w:	window	x:	X
y:	yellow	z:	Zebra

Punctuation:

" . ":	away	" , ":	comma
" _ ":	tile	" ":	well
" 'enter' ": new		" 'scratch that' ": /a/	

Number Keys:

1 :	air	2 :	bad
3 :	cake	4 :	dot
5 :	earn	6 :	funny
7 :	gnaw	8 :	him
9 :	/ai/	0 :	jump

Command Keys:

" 'escape' ":	only	" 'up-arrow' ":	her
" 'down-arrow'":	linen	" 'left-arrow' ":	_ama_
" 'right-arrow' ":	worm		
" 'begin of line' ":	home	" 'end of line' ":	end

General Keys:

" 'quit list' ":	rumor	" 'manual' ":	menu

Appendix 4.7

The following list is used to dictate to the PC for CH.

Root Functions:

"'voice console'":	Mary	"'number key'":	bill
" 'alphabet spell' ":	normal	" 'command' ":	row
" 'scratch that' ":	white	" 'enter' ":	new

Alphabet spelling:

a:	air	b:	bad
c:	cake	d:	delay
e:	earn	f:	fever
g:	gnaw	h:	much
i:	ice	j:	Julia
k:	knee	l:	lima
m:	mate	n:	nut
o:	our	p:	plat
q:	meat	r:	ray
s:	seed	t:	tile
u:	/u/	v:	vain
w:	weak	x:	Xenon
y:	yellow	z:	Zebra

Punctuation:

" . ":	away	" , ":	gun
" _ ":	_igi_	" ":	well
" 'enter' ": new		" 'scratch that' ": white	

Number Keys:

1 :	air	2 :	bad
3 :	cake	4 :	delay
5 :	earn	6 :	fever
7 :	gnaw	8 :	much
9 :	ice	0 :	Julia

Command Keys:

" 'escape' ":	only	" 'up-arrow' ":	meat
" 'down-arrow'":	linen	" 'left-arrow' ":	_ama_
" 'right-arrow' ":	worm		
" 'begin of line' ":	nice	" 'end of line' ":	end

General Keys:

" 'quit list' ":	rumor	" 'manual' ":	menu

References

Abbs, J. H., C. J. Hunker, and S. M. Barlow (1982) Differential Speech Motor Subsystem Impairments with Suprabulbar Lesions: Neurophysiological Framework and Supporting Data. In W. R. Berry (Ed.), Clinical Dysarthria, San Diego: College-Hill Press, pp. 21-56.

ASHA (1981) Position Paper for the Ad Hoc Committee on Communication Processes for Nonspeaking Persons, American Speech and Hearing Association, Rockville, MD.

Baker, J. M. (1989) Dragondictate-30K: Speaker - Adaptive Natural Language Speech Recognition with 30,000 Word Active; No Speaker Enrollment Required, Proceedings of the American Voice Input / Output Society, San Jose, CA: AVIOS, pp. 201-206.

Berry, M. F. and J. Eisenson (1956) Speech Disorders: Principles and Practices of Therapy. New York: Appleton-Century-Crofts.

Brain, L. W. and J. N. Walton (1969) Brain's Diseases of the Nervous System, 7th Ed. London: Oxford University Press.

Byrne, M. (1959) Speech and Language Development of Athetoid and Spastic Children, Journal of Speech and Hearing Disorders, 24, pp. 231-240.

Clement, M. and T. Twitchell (1959) Dysarthria in Cerebral Palsy, Journal of Speech and Hearing Disorders, 24, pp. 118-122.

Coleman, C. L. and L. S. Meyers (1991) Computer Recognition of the Speech of Adults with Cerebral Palsy and Dysarthria, AAC Augmentative and Alternative Communication, 7, March, pp. 34-42.

Corcoran, P. J. (1981) Neuromuscular Diseases. In W. Stolov & M. Clowers (Eds.), Handbook of Severe Disabilities, U.S. Department of Education, Rehabilitation Services Administration, pp. 83-100.

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 (1969a) Clusters of Deviant Speech Dimensions in the Dysarthrias, Journal of Speech and Hearing Research, 12, pp. 462-496.

Darley, F., A. Aronson, and J. Brown (1969b) Differential Diagnostic Patterns of Dysarthria, Journal of Speech and Hearing Research, 12, pp. 246-269.

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

Denes, P. B. and E. N. Pinson (1973) The Speech Chain: The Physics and Biology of Spoken Language. New York: Anchor Press.

Denny-Brown, D. (1966) The Cerebral Control of Movement. Springfield, Ill.: Charles C. Thomas Press.

Dixon, N. R. and T. B. Martin (1979) Automatic Speech and Speaker Recognition. New York: IEEE Press.

Dorland's Illustrated Medical Dictionary, 26th Ed. (1981) W.B. Saunders Press.

Easton, J. and D. Halpern (1981) Cerebral Palsy. In W. Stolov & M. Clowers (Eds.), Handbook of Severe Disabilities, U.S. Department of Education, Rehabilitation Services Administration, pp. 137-154.

Erenberg, G. (1984) Cerebral Palsy, Postgraduate Medicine, 75(7), pp. 87-93.

Farmer, A. (1975) Stop Cognate Production in Cerebral Palsied Speakers, paper presented to the 89th Acoustic Society of American Meeting, Austin, Texas.

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

Flanagan, J. L. (1982) Talking with Computers: Synthesis and Recognition of Speech by Machines, IEEE Transactions on Biomedical Engineering, 29, pp. 223-232.

Flanagan, J. L. (1972) Speech Analysis, Synthesis, and Perception. New York: Springer-Verlag.

Goodenough-Trepagnier, C. G. and M. J. Rosen (1991) Towards a Method for Computer Interface Design Using Speech Recognition, Proceedings of 1991 RESNA 14th Annual Conference, Kansas City, MO, pp. 328-329.

Goodenough-Trepagnier, C. G., M. J. Rosen, H. S. Hochheiser, and H. P. Chang (1992) Assessment of Dysarthric Speech for Computer Control Using Speech Recognition: Preliminary Results, Proceedings of 1992 RESNA International Conference, pp. 159-161.

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

Hardy, J. (1961) Intraoral Breath Pressure in Cerebral Palsy, Journal of Speech and Hearing Disorders, 26, pp. 309-319.

Hixon, T. J. and J. C. Hardy (1964) Restricted Motility of the Speech Articulators in Cerebral Palsy, Journal of Speech and Hearing Disorders, 29, pp. 293 - 306.

Johns, D. F. and F. L. Darley (1970) Phonemic Variability in Apraxia of Speech, Journal of Speech and Hearing Research, 13, pp. 556-583.

Kammermeier, M. A. (1969) A Comparison of Phonatory Phenomenε among Groups of Neurologically Impaired Speakers, Ph.D. dissertation, University of Minnesota.

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

Kent, R., R. Netsell, and J. Abbs (1979) Acoustic Characteristics of Dysarthria Associated with Cerebellar Disease, Journal of Speech and Hearing Research, 22, pp. 627-648.

Kent, R. and R. Netsell (1975) A Case Study of an Ataxic Dysarthric: Cineradiographic and Spectrographic Observations, Journal of Speech and Hearing Disorders, 40, pp. 115-134.

Kolb, B. and I. Q. Whishaw (1990) Fundamentals of Human Neuropsychology, 3rd Ed., New York: Freeman Press.

Kraft, G. (1981) Multiple Sclerosis. In W. Stolov & M. Clowers (Eds.), Handbook of Severe Disabilities, U.S. Department of Education, Rehabilitation Services Administration, pp. 111-118.

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

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

Lee, Kai-Fu and R. Reddy (1989) Automatic Speech Recognition: The Development of the SPHINX System. Kluwer Academic Publishers.

Lee, W. C. Jr., S. W. Blackstone, and G. K. Poock (1987) Dysarthric Speech Input to Expert Systems, Electronic Mail and Daily Job Activities, Proceedings of the American Voice Input / Output Society, San Jose, CA: AVIOS, pp. 33-43.

Leith, W. (1954) A Comparison of Judged Speech Characteristics of Athetoid and Spastics, unpublished master's thesis, Purdue University.

Lencione, R. M. (1953) A Study of the Speech Sound Ability and Intelligibility Status of a Group of Educable Cerebral Palsied Children, Ph.D. Dissertation, Northwestern University.

Luchsinger, R. and G. E. Arnold (1965) Voice-Speech-Language: Clinical Communicology - Its Physiology and Pathology. Belmont, Calif.: Wadsworth Publishing Company, Inc.

Ludlow, C. L. and C. J. Bassich (1982) The Results of Acoustic and Perceptual Assessment of Two Types of Dysarthria. In W. R. Berry (Ed.), Clinical Dysarthria, San Diego: College-Hill Press, pp. 121-153.

Neilson, P. and N. J. O'Dwyer (1984) Reproducibility and Variability of Speech Muscle Activity in Athetoid Dysarthria of Cerebral Palsy, Journal of Speech and Hearing Research, 27, pp. 502-517.

Netsell, R. (1986) A Neurobiologic View of Speech Production and the Dysarthrias, San Diego: College-Hill Press.

Nielsen, J. M. (1951) A Textbook of Clinical Neurology, 3rd Ed., New York: Hoeber.

Palmer, M. F. (1952) Speech Therapy in Cerebral Palsy, J. Pediatrics, 40, pp. 514 - 524.

Parker, H. L. (1956) Clinical Studies in Neurology, Springfield, Ill: Charles C Thomas.

Perkins, W. H., and R. D. Kent (1986) Functional Anatomy of Speech, Language, and Hearing. San Diego, C. A.: College-Hill.

Peterson, G., and H. Barney (1952) Control Methods Used in a Study of the Vowels, the Journal of the Acoustical Society of America, 24, pp. 585-594.

Platt, L., G. Andrews, M. Young, and P. Quinn (1980a) Dysarthria of Adult Cerebral Palsy: I. Intelligibility and Articulatory Impairment, Journal of Speech and Hearing Research, 23, pp. 28-40.

Platt, L., G. Andrews, and P. M. Howie (1980b) Dysarthria of Adult Cerebral Palsy: II. Phonemic Analysis of Articulation Errors, Journal of Speech and Hearing Research, 23, pp. 41-55.

Purves-Stewart, J., and C. Worster-Drought (1952) The Diagnosis of Nervous Diseases, 10th Ed., London: Edward Arnold & Co.

Rabiner, L.R. and B. H. Juang (1986) An Introduction to Hidden Markov Models, IEEE ASSP Magazine, January, pp. 4-16.

Rabiner, L.R. and B. H. Juang (1993) Fundamentals of Speech Recognition. New Jersey: PTR Prentice Hall.

Raghavendra, P., E. Rosengren, and S. Hunnicutt (1994) How Does Automatic Speech Recognition Handle Dysarthric Speech, Proceedings of 1994 Fonetik 8th Swedish Phonetics Conference, Lund, Sweden, pp. 112-115.

Rodman, R. D., T. S. Moody, and J. A. Price (1985) Speech Recognizer Performance with Dysarthric Speakers: A Comparison of Two Training Procedures, Speech Technology, pp. 65-71.

Rutherford, B. R. (1944) A Comparative Study of Loudness, Pitch Rate, Rhythm, and Quality of Speech of Children Handicapped by Cerebral Palsy, Journal of Speech and Hearing Disorders, 9, pp. 262-271.

Sataloff, R. T. (1992) The Human Voice, Journal of Scientific American, December, pp. 108-115.

Schmitt, D. G. and J. Tobias (1986) Enhanced Communication for a Severely Disabled Dysarthric Individual Using Voice Recognition and Speech Synthesis, Proceedings of 1986 RESNA 9th Annual Conference, Minneapolis, MN, pp. 304-306.

Simmons, N. N. (1982) Acoustic Analysis of Ataxic Dysarthria: An Approach to Monitoring Treatment. In W. R. Berry (Ed.), Clinical Dysarthria, San Diego: College-Hill Press, pp. 283-294.

Sincich, T. (1987) Statistics by Example. San Francisco, CA: Dellen Publishing Company.

Smitheran, J. and T. Hixon (1981) A Clinical Method for Estimating Laryngeal Airway Resistance During Vowel Production, Journal of Speech and Hearing Disorders, 46, pp. 138-146.

Stevens, K. N. (1972) The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data. In E. E. David and P. B. Denes (Eds.), Human Communication: A Unified View, Inter-University Electronics Series, New York: McGraw-Hill, pp. 51-66.

Stevens, K. N. (1989) On the Quantal Nature of Speech, Journal of Phonetics, 17, pp. 3-45.

Sy, B. K. and D. M. Horowitz (1993) A Statistical Causal Model for the Assessment of Dysarthric Speech and the Utility of Computer-Based Speech Recognition, IEEE Transactions on Biomedical Engineering, 40, pp. 1282 - 1298.

Walshe, F. (1970) Diseases of the Nervous System. 11th ed. New York: Longman.

Wolfe, W. (1950) A Comprehensive Evaluation of Fifty Cases of Cerebral Palsy, Journal of Speech and Hearing Disorders, 15, pp. 234-251.

Yahr, M. D. (Ed.) (1976) The Basal Ganglia. New York: Raven Press.

Yorkston, K. M. (1988) Clinical Management of Dysarthric Speakers. Boston: Little Brown Press.

Yorkston, K. and D. Beukelman (1981a) Assessment of Intelligibility of Dysarthric Speech. Tigard, OR: C.C. Press.

Yorkston, K. and D. Beukelman (1981b) Ataxic Dysarthria: Treatment Sequences Based on Intelligibility and Prosodic Considerations, Journal of Speech and Hearing Disorders, 46, pp. 398-404.

Yorkston, K. and D. Beukelman (1978) A Comparison of Techniques for Measuring Intelligibility of Dysarthric Speech, Journal of Communication Disorders, 11, pp. 499-512.

Zentay, P. J. (1937) Motor Disorders if the Central Nervous System and Their Significance for Speech, Part I: Cerebral and Cerebellar Dysarthrias, Laryngoscope, 47, pp. 147-156.