

## Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be “used for any purpose other than private study, scholarship, or research.” If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “fair use” that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

**Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation**

Printing note: If you do not wish to print this page, then select “Pages from: first page # to: last page #” on the print dialog screen

The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

## **ABSTRACT**

### **VOICE ONSET TIME VARIATION IN STOP CONSONANT TO VOWEL TRANSITIONS**

**by**

**Abhishek Gunvant Parikh**

Reduced duration, increased consistency, and improved intelligibility are goals of reducing the motor complexity of speech for individuals with cerebral palsy having dysarthria. In this study, measurement and analysis were made to compare an individual with spastic Cerebral Palsy (CP) having dysarthria to an individual with athetoid CP having dysarthria as well as to a non-dysarthric individual. Each participant's normal speech, whispering, and speech using an artificial larynx was evaluated, utilizing the source-filter theory methodology. The plausibility of dysarthric speech duration reduction by minimizing vocalization is tested by stop consonant "P" to vowel transitions. The data suggest that speech duration is dependent on voice onset time (VOT) variation among the participants. This study could serve as a basis to encourage further research analyzing neuromotor and physiological articulatory control, which could lead to interventional treatment for individuals having dysarthria.

**VOICE ONSET TIME VARIATION IN STOP CONSONANT TO VOWEL  
TRANSITIONS**

by  
**Abhishek Guvant Parikh**

**A Thesis  
Submitted to the Faculty of  
New Jersey Institute of Technology  
In Partial Fulfillment of the Requirements for the Degree of  
Master of Science in Biomedical Engineering**

**Department of Biomedical Engineering**

**May 2004**

Blank Page

**APPROVAL PAGE**

**VOICE ONSET TIME VARIATION IN STOP CONSONANT TO VOWEL  
TRANSITIONS**

**Abhishek Guvant Parikh**

---

Richard Foulds, PhD, Thesis Advisor Date  
Associate Professor of Biomedical Engineering, NJIT

---

David Kristol, PhD, Committee Member Date  
Professor of Biomedical Engineering, NJIT

---

Dr. Beverly Baird, Ed.D, OTR, FAOTA, Committee Member Date  
Consultant to the Occupational Therapist Dept., Matheny School and  
Hospital; Retired Professor of Occupational Therapy, NYU

---

William Hunter, PhD, Committee Member Date  
Chair of Biomedical Engineering, NJIT

## **BIOGRAPHICAL SKETCH**

**Author:** Abhishek Gunvant Parikh

**Degree:** Master of Science

**Date:** January 2004

### **Undergraduate and Graduate Education:**

- Master of Science in Biomedical Engineering,  
New Jersey Institute of Technology, Newark, NJ, 2004
- Bachelor of Science in Engineering Science: Concentration in Pre-Medicine/Biomedical Engineering,  
New Jersey Institute of Technology, Newark, NJ, 2000

**Major:** Engineering Science

To my beloved family:

Words cannot express nor provide the gratitude deserved by my family, who has provided continuous encouragement and unconditional support in my academic endeavors and in my life.

To Matheny School and Hospital's students:

Whose hopes and aspirations are truly inspirational, having true spirit, no matter what the obstacles in life may be.

I hope to be able to give something back to these students, starting from these studies. Improving the quality of life of others is in itself the driving force for my work, and I hope to be able to do more in the future.



## ACKNOWLEDGEMENT

Foremost, I am indebted to Dr. Richard Foulds, who while serving as my thesis advisor, has gone beyond what is required to aid, guide, mentor, support and encourage me in my research by always making himself available. Thanks are given especially to Dr. Beverly Bain who has gone out of her way to guide and assist me in my Thesis, and Dr. Kenneth Robey both of whom have taught me so much about research. Dr. Stanley Reisman who has steered me in the right direction in my graduate career, Susan Buin for helping to choose dysarthric participants and allowing me to utilize speech pathology equipment countless times, Dr. Susan Roeloffs for taking the time to review my proposal, and Dr. Gary Eddey who has inspired me not only academically but professionally. I would like to give special thanks to everyone actively participating in my committee.

Many of my fellow graduate students in the Biomedical Engineering Department as well as the Adult Services Department at Matheny School and Hospital are deserving of recognition for their support, as well as Matheny management for allowing me to work with flexibility around my academic pursuits. I would also like to thank Dr. David Kristol, not only for his assistance but also for his support over the years.

## TABLE OF CONTENTS

<b>Chapter</b>	<b>Page</b>
1 INTRODUCTION.....	1
2 PURPOSE AND SCOPE .....	8
3 EXPERIMENTAL DESIGN.....	13
4 DATA ANALYSIS: TEMPORAL ANALYSIS.....	29
5 RESULTS.....	33
6 CONCLUSION.. ..	42
APPENDIX A INFORMED CONSENT.....	49
APPENDIX B SIEMANS SERVOX ELECTRO LARYNX DEVICE.....	54
REFERENCES .....	55

## LIST OF FIGURES

<b>Figure</b>		<b>Page</b>
2.1	Source-Filter Model.....	10
2.2	Voice onset time.....	12
3.1	Waveform of sentence measurement.....	15
3.2	Waveform of individual words.....	16
3.3	Waveform of voice onset time.....	16
3.4	Waveform of occlusion time.....	17
3.5	Waveform of stop consonant to vowel.....	17
3.6	Waveform of vowel durations.....	18
3.7	Sony DCR - PC5 camcorder.....	22
3.8	Shure SA10A microphone.....	23
3.9	Sony MDR-V600 headphones.....	24
3.10	Sony DCR – PC5 firewire connection to PC.....	26
5.1	Durational measurements for natural speech.....	33
5.2	Durational measurements for whispering.....	34
5.3	Durational measurements for using the artificial larynx.....	35
5.4	Min/ Max/ Mean of natural voicing (Nat).....	36
5.5	Min/ Max / Mean of whispering (W).....	37
5.6	Min/ Max/ Mean of artificial larynx (AL).....	38
5.7	Durational changes of C-T (control talker).....	39
5.8	Durational changes of D-T-A (athetoid talker).....	40

5.9	Durational changes of D-T-S (spastic talker).....	41
6.1	Waveform of D-T-S for natural speech.....	42
6.2	Waveform of C-T for natural speech.....	45
6.3	Waveform of D-T-A for natural speech.....	45
6.4	Waveform of D-T-S for whispering.....	46
6.5	Waveform of D-T-S for using the artificial larynx.....	46

## LIST OF DEFINITIONS

Artificial Larynx (AL) - (see Appendix B)

Dysarthria - difficulty in articulating words leading to imperfection of speech due to disturbances in muscular control resulting from damage or disease of the central or peripheral nervous system (CancerWeb, 2003, Health-Dictionary, 2004, Taber's, 1985).

Filter - a hardware device, material or software program that provides a frequency-dependent transmission of energy (sound). Commonly a filter is used to exclude, suppress, or minimize energy at certain frequencies while passing the energy at other frequencies. A low pass filter passes the frequencies below a certain cut-off frequency; a high-pass filter passes the frequency above a certain cut-off frequency; and a band pass filter passes the energy between a lower and upper cut-off frequencies. The device used in spectrophotometric analysis to isolate a segment of the spectrum. (CancerWeb, 2003, Merriam-Webster Online, 2003, Taber's 1985).

Formant - a resonance of the vocal tract. A formant is specified by its center frequency (commonly called formant frequency) and bandwidth. Formants are denoted by integers that increase with the relative frequency location of the formants. F1 is the lowest-frequency formant; F2 is the next highest, and so on (The American Heritage College Dictionary, 1993, Merriam-Webster Online, 2003).

Formant Transition - a change in formant pattern typically associated with a phonetic boundary; for example, the CV formant transition refers to formant pattern changes associated with the consonant-vowel transition (Kent, 1979, 1999).

Fundamental Frequency (F0) - the lowest frequency (first harmonic) of a periodic signal. In speech, the fundamental frequency refers to the first harmonic of voice. Fundamental frequency is the reciprocal of the fundamental period. Ideally, fundamental frequency is used to refer to a physical measure of the lowest periodic component of vocal fold vibration. Pitch should be on a continuum from low to high (Kent, 1979, 1999).

Source-filter theory - a theory of an acoustic production of speech that states that the energy from a sound source is modified by a filter or a set of filters (Foulds, 2002).

Spastic CP - a disability resulting from damage to the brain before, during, or shortly after birth and outwardly manifested by speech disturbances and muscular incoordination.

Spastic Paralysis - paralysis with tonic spasm of the affected muscles and with increased tendon reflexes (Tortura, Gabrowski, 1993, Merriam-Webster Online, 2003).

Spasticity - is an increased resistance to externally imposed motion, which also increases with increasing speed of motion, furthermore which is associated with features of increased reflex excitability. Causes may include CP, spinal cord and brain injury. Spasticity in the clinical sense seems mainly to be related to changes in the threshold of the muscles for eliciting behavior like that of a spring (Tortura, Gabrowski, 1993, Lance, 1970, Taber's, 1985).

Spectrogram - a pattern of sound analysis containing information on intensity, frequency and time. The typical spectrogram provides a three-dimensional display of time on the horizontal axis, frequency on the vertical axis, and intensity on the gray scale. A spectrogram can be printed as copy or displayed on a video monitor (The American Heritage College Dictionary, 1993, Merriam-Webster Online, 2003, Taber's, 1985).

Spectrum - a graph showing the distribution of signal energy as a function of frequency; a plot of intensity by frequency. (The American Heritage College Dictionary, 1993, Taber's, 1985).

Voice Onset Time (VOT) - time from the start of the stop-consonant to the beginning of voicing (see Figure 2.2).

Wave Summation/ Temporal Summation - The algebraic addition of the excitatory and inhibitory effects of many stimuli applied to a nerve cell body. The increased strength of muscle contraction that results when stimuli follow in rapid succession (Tortura, Gabrowski, 1993).

Wavelength - the distance that a periodic wave travels in one complete cycle.  $\text{Wavelength} = \text{speed of sound} / \text{frequency}$ . (Health-Dictionary, 2004, Merriam-Webster Online, 2003, Taber's, 1985).

# CHAPTER 1

## INTRODUCTION

It is important to increase the communicative abilities of individuals with dysarthria, as verbal communication is vital to function effectively in our society. People with dysarthria may speak slowly, or appear breathless, and are frequently misunderstood by the listener who becomes disinterested due to the time spent in the communication and the lack of understanding of what has been said. A person who communicates using poorly articulated speech, and has difficulty expressing certain sounds of words, is prone to being less intelligible to others.

Previous studies, confirmed by this study, suggest that some CP talkers having dysarthria can reduce their utterance durations by whispering, and by using an artificial larynx, though it is questionable if this improves intelligibility or the consistency of their speech (Foulds, 1980, 2002). Therefore, within this pilot study, segments of speech were analyzed to determine if some variables, such as phoneme duration or occlusion time were responsible for the increased duration in speech. There was a possibility that some phoneme-to-phoneme transitions were more difficult than others; therefore, stop consonant to vowel transitions were also analyzed for durational differences.

In order to understand dysarthria, knowledge of causality is essential. Cerebral Palsy is a medical term used to describe a group of chronic conditions affecting body movements and muscle coordination. It is caused by damage to one or more specific areas of the brain, usually occurring during fetal development, or during infancy, and can also occur before, during or shortly following birth. “Cerebral” refers to the brain and

“Palsy” to a disorder of movement or posture. Children with CP may not be able to talk, walk, eat or play in the same ways as most other children (Cerebral-Palsy, 2003).

Cerebral Palsy may affect the muscles of one part of the body, or one side of the body, and sometimes the entire body. Hyperactive reflex movements and muscle tightness (spasticity) occur with varying severity. Some babies born with severe CP may have a floppy or very stiff body. Birth defects such as a small jawbone, an irregularly shaped spine, or small head may also occur along with CP. Some of these symptoms appear at birth, others may not be obvious until the nervous system matures, and others may become more pronounced with age (WebMD, 2003).

Some people with CP have little or no control over their mouths and tongues. Others have only a slight limp or an uncoordinated movement of other parts of their body, such as arms and legs. People with severe forms of CP are more likely to have other problems, such as seizures or mental retardation (WebMD, 2003). Depending on which areas of the brain have been damaged, one or more of the following may occur:

- impairment of speech, sight, or hearing
- muscle tightness or spasm
- involuntary movement
- disturbance in gait and mobility
- abnormal sensation and perception
- seizures
- learning disabilities



The three main types of CP are: 1. spastic (difficult and stiff movement), 2. athetoid (rhythmic, uncontrolled and involuntary movement), and 3. Ataxic (disturbed sense of depth perception and balance); however, one person may have a combination of these.

Spastic CP is the most common type of CP, accounting for nearly 80 percent of all CP cases. The most widely accepted definition for spasticity states that there is an increased resistance to externally imposed motion, which also increases with increasing speed of motion, furthermore which is associated with features of increased reflex excitability (NorthWestern University Medical School, 2003). In the clinical sense, spasticity seems mainly to be related to changes in the threshold of the muscles for eliciting behavior like that of a spring. Spasticity is measured by reflex threshold, length tension or torque/angle relation (stiffness, spring constant), or velocity dependence (viscosity). It should be noted that there is a difference between cerebral and spinal spasticity, as the mechanisms are potentially different. Cerebral spasticity is related to abnormal descending excitation, while mechanisms for spinal cord spasticity are less clear other than that being of an interneuronal mechanism for flexion withdrawal, and extension spasms (Northwestern University Medical School, 2003). Individuals with this type of CP may have one or more tight muscle groups that limit movement. They may have abnormal muscle tone, abnormal co-contraction, and hypersensitive reflexes. With spastic CP these individuals have stiff and jerky movements, resulting in having difficulty speaking fluidly.

Because CP influences the way children develop, it is known as a developmental disability. In the United States today, more people have CP than any other developmental disability, including Downs Syndrome, Epilepsy, and Autism. About two

children out of every thousand born in this country have some type of CP. Approximately five thousand children up to the age range of toddlers plus approximately 1300 preschoolers are diagnosed with CP yearly. About five hundred thousand individuals in the United States have some degree of CP, yet since there is no system for monitoring CP's onset it is not known whether it is declining, stagnant, or increasing. With the large incidence in the United States alone, the opportunity of improving the quality of affected individuals is substantial. CP is a lifelong disability and usually, the movement and other problems associated with CP affect what a child is able to learn and do to varying degrees throughout their life reinforcing the value of further scientific investigation(s) as the benefit to individuals treated early could possibly assist them throughout their lifespan (Cerebral-Palsy, 2003).

The type and severity of dysarthria depends on which area of the nervous system is affected (ASHA, 2003). Depending on the extent and location of damage to the nervous system a person with dysarthria may experience many of the following symptoms:

- “slurred” speech
- significantly greater energy expenditure during speech articulation
- speaking softly or barely able to whisper
- slow rate of speech
- rapid rate of speech with a “mumbling” quality
- limited tongue, lip, and jaw movement
- abnormal intonation (rhythm) when speaking
- changes in vocal quality (“nasal” speech or sounding “stuffy”)

- hoarseness
- breathiness
- drooling or poor control of saliva
- chewing and swallowing difficulty

#### Causes of Dysarthria:

Dysarthria is caused by many different conditions that involve the nervous system, including:

- Cerebral Palsy
- Stroke
- Brain Injury
- Multiple Sclerosis
- Tumors
- Parkinson's disease
- Lou Gehrig's disease (ALS)
- Huntington's disease

Dysarthria is generally apparent in daily conversation where there is difficulty expressing certain sounds or words. Though participants in this study have a speech disorder from birth or early childhood, it can be noted that this condition may also be caused by taking excess medications such as narcotics, phenytoin, or carbamazepine. Alcohol intoxication also causes dysarthria. Some former severe alcoholics who have developed brain damage (Korsakoff syndrome) due to drinking may have continued problems with language, even after years of sobriety. Degenerative neurological disorders affecting the cerebellum or brainstem can also cause dysarthria such as a

cerebral vascular accident (stroke) that affects brainstem or cerebellar regions. Dysarthria is occasionally confused with aphasia. It is important to distinguish between a difficulty in articulation of words versus a problem with the production of language, as these have different causes (ASHA, 2003).

Individuals with CP having dysarthria were chosen for this study specifically because a majority of them have this speech disorder, caused by a weakness in the muscles that produce speech, resulting in slurring for mild cases, to inability to vocalize in severely extreme cases. Furthermore the reason for choosing spastic CP individuals with dysarthria, stems from the fact that spasticity in muscles from this type of disorder could possibly be the causality for extended duration of such individuals' speech (CerebralPalsyfyi, 2004).

### **Current Treatment**

Treatment depends on the cause, type, and severity of the symptoms. A speech language pathologist (SLP) works with the individual to improve communication abilities. Goals may include slowing the rate of speech, improving breath support so the person can speak more loudly, muscle strengthening exercises, increasing mouth, tongue, and lip movement, or improving articulation so that speech is clearer. The SLP can also help the person's caregivers or family learn to adapt the environment so that they can understand the person better and can teach compensatory strategies that will enhance communication. In cases of severe dysarthria, it may be impossible for the person to speak intelligibly and an alternative means of communication may be needed. These range from using simple gestures or alphabet boards to more sophisticated electronic or

computer-based equipment known as augmentative communication devices (ASHA, 2003).

With the development of speech recognition systems and an increased awareness of the needs of people with dysarthric speech, there is a clear opportunity to develop systems for such users to enable them to gain greater independence and control of their lives. Modern automatic speech recognition is based on training statistical models of the acoustic manifestation of speech units (words, phones or context-dependent phones), using pre-recorded speech databases (Gold and Morgan, 2003). Good performance may be achieved when the speech to be recognized is sufficiently similar to that used in training. However, automatic speech recognition performance remains brittle when compared to the human ability to deal with abnormal or distorted speech: word error rates in such conditions are typically an order of magnitude worse for the machine (Lippmann, 2003).

Consumer automatic speech recognition systems have been used for people with mild and moderate dysarthria as a means of inputting text, but there is a lack of consensus over whether these systems are appropriate for people with severe dysarthria (Boves, Rosengren , 2003).

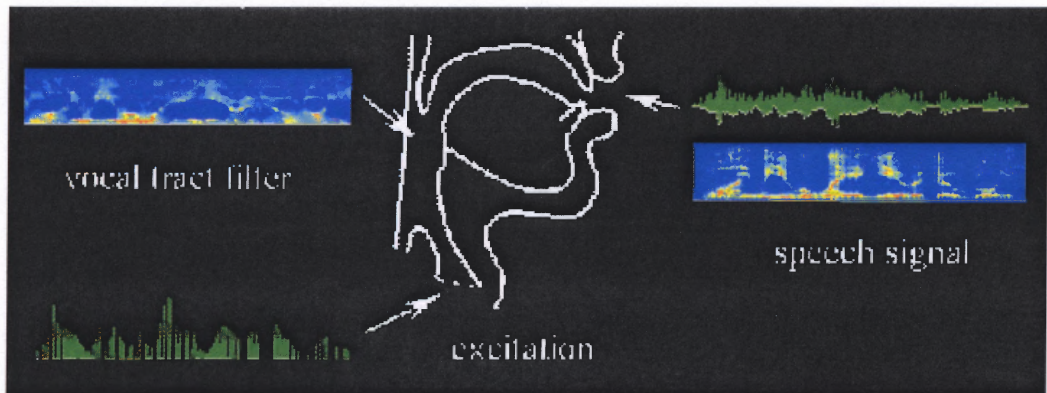
## CHAPTER 2

### PURPOSE AND SCOPE

Dysarthria is the most common acquired speech disorder affecting 170 per 100,000 of the population, this equates to roughly 590,000 individuals in the US, and approximately 11 million worldwide. In its severest form dysarthric speech is unintelligible to others and may take the form of producing vocal utterances, rather than words recognizable to unfamiliar communication partners. The combination of speech and general physical disability can make it particularly problematic for them to interact in their environment and limits independence. This is the primary problem with spastic CP dysarthric individuals (ASHA, 2003).

Research evidence is limited (Foulds, 2002, Kent, 1979, 1999). Most reports are single case studies or are anecdotal and there are difficulties in comparing results between reports due to uncertain definitions of severity of dysarthria and a lack of detail about methods employed (Kent, Netshell, Abbs 1979). This study differs in two ways; one, the three participants were clinically diagnosed by speech-language pathologist (SLP) as being athetoid CP having dysarthria, or diagnosed as being spastic CP having dysarthria, or the control subject who is clinically non dysarthric. The second difference, is that the standardized methodology used in the pilot study was outlined in detail. Hawley, suggests that intensive training is the key to improving speech performance, especially while whispering and with the artificial larynx so that the participants feel comfortable speaking in these methods prior to actual recordings (Hawley, 2003). As a result, in this study, prior training of AL use and whispering was performed.

This study attempted to investigate possible uncontrolled spastic reflex movements effecting dysarthric speech by employing a source-filter theory model. First proposed by Johannes Müller in the 19th century, source-filter theory accounts for the acoustic properties of what are called "voiced" speech sounds (sounds during whose articulation the vocal chords vibrate). For "unvoiced" sounds (e.g., *Shh*), the source is air forced through a constriction in the vocal tract (Columbia, 2004). In other words, this theory states the acoustic production of speech that is the result of the energy from a sound source which is modified by a filter or a set of filters (Foulds, 2002). For example, for vowels the vibrating vocal folds are usually the source of sound energy, and the vocal tract resonances (formants) are the filters: part of the training to be able to control the turbulence in vocalization was achieved by employing an external artificial larynx with a vibration set to the fundamental frequency of normal voice. Whispering reduces the motoric complexity of speech, particularly the vocalization aspect that is the source, in the source-filter theory model. When whispering, the vocal folds are not completely closed, and the source does not stimulate the same excitation as compared to natural speech; therefore, the minimum threshold for muscle excitation is not reached. The artificial larynx (AL) speech differs as it results in reducing motoric complexity of speech by acting as the source of vocalization for the participant. As a result, there is no excitation of the vocal folds, but the vocalization is produced from the turbulence created from the vibrations of the AL. In other words, the utterance was just mouthed, not vocalized by the individual. The jaw, lips, and tongue would collectively comprise the filter aspect of the source-filter theory model (see Figure 2.1).



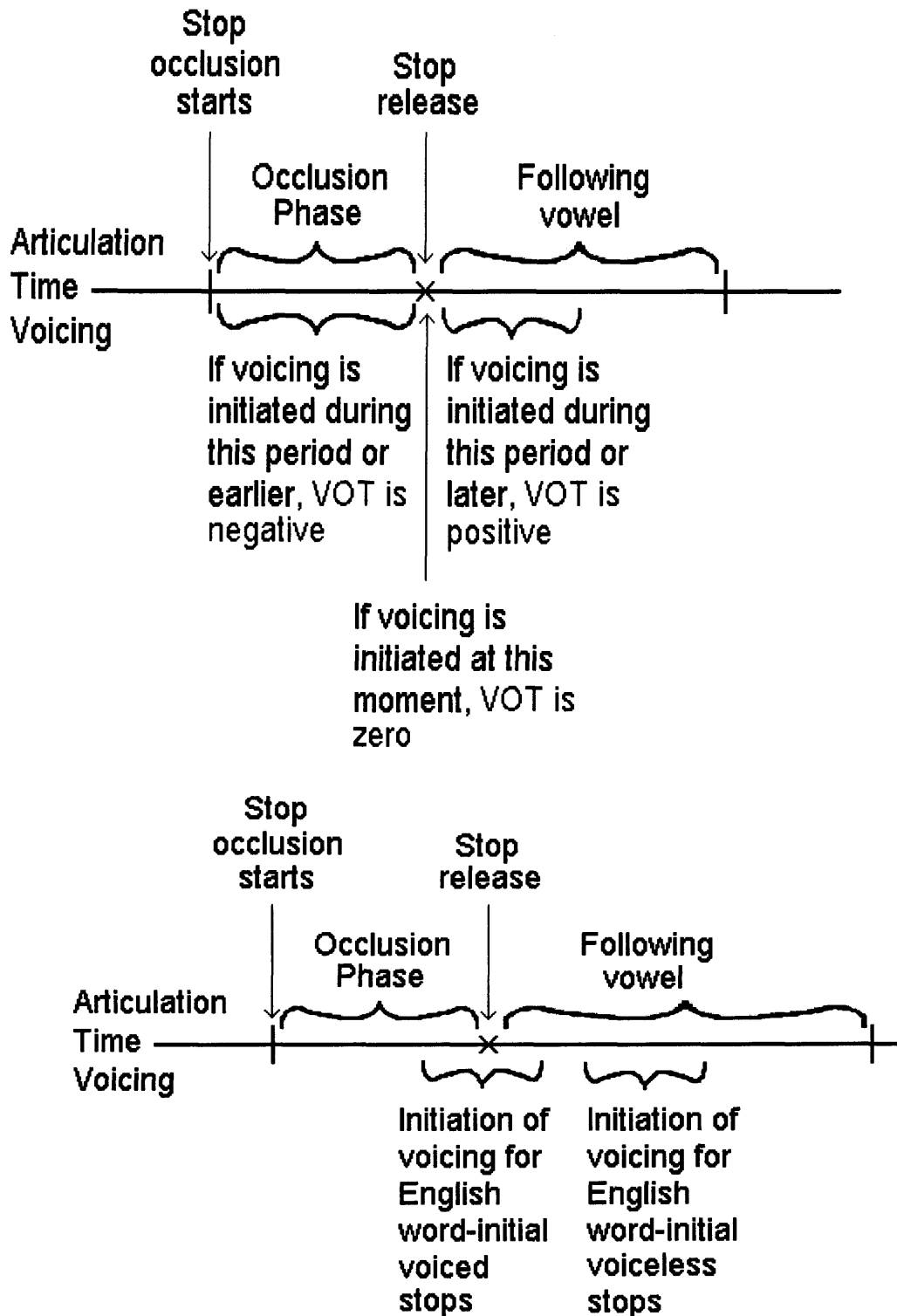
**Figure 2.1** Representation of the Source – Filter Model shows the source as excitation, the vocal tract filter which dampens certain frequencies and intensifies others, and resulting speech signal, which was generated and captured for analysis with Matlab/Colea, a full speech analyzer software (Columbia, 2004).

From a review of past research (Kent et al. 1999), as well as review of past related studies, there is a need to quantify Foulds' hypothesis. which states that when motoric complexity of speech/minimizing vocalization, it may result in reducing the duration of speech, to first determine if this led to a reduction in speech duration of a Spastic CP dysarthric individual. If the duration is reduced, what specifically is the durational determinant, the phonemes or the occlusion time, if any or both? Second, voice onset time, the time from the stop consonant/plosive release to the beginning of voicing of the following vowel, is another possible cause of durational increase in dysarthric speech (see Figure 2.2).

This study compared durational aspects of speech between an individual with spastic CP having dysarthria to an individual with athetoid CP having dysarthria, to those of a non-dysarthric control talker. The data can also be the basis of future studies, with a larger participant set, investigating inter-relationships of specific parameters such as comparisons between different individuals having dysarthria, as well as for studies



employing an electroglottograph allowing detailed physiological motor analysis of the dysarthric speech (University of Virginia, 2004).



**Figure 2.2** Graphical depiction of voice onset time describing different phases of articulation and voicing (Macquarie University, 2004).

## **CHAPTER 3**

### **EXPERIMENTAL DESIGN**

This study was designed to measure the effects of reducing the motor complexity of speech on the coordination of speech articulation (tongue/lips/jaw) in spastic CP individuals having dysarthria.

This was accomplished by employing a source-filter model, based on a theory of an production of speech that states that the energy from a sound source is modified by a filter or a set of filters, therefore separating the source (respiration and vocal cord control) from the filter (lips/tongue/jaw) for durational analysis. (see Figure 2.1).

Speech production involves the above components among others. Specifically the organs are the lungs, trachea or windpipe, the larynx, the throat or pharyngeal cavity, the buccal or oral cavity (mouth), and the nasal cavity or nose. Usually the oral and pharyngeal cavities are grouped together and referred to as the vocal tract, while the nasal cavity is referred to as the nasal tract. In accord, the vocal tract begins at the output of the larynx (glottis or vocal cords), and ends at the lips. The nasal tract begins at the velum and terminates at the nostrils. When the velum, which is a mechanism at the back of the mouth acting like a trap-door, is lowered, the nose is acoustically coupled with the mouth to produce the nasal sounds of speech. After air enters the lungs via the normal breathing mechanism, it is expelled from the lungs through the trachea, as the tensed vocal cords within the larynx are caused to vibrate by the air flow. The air is chopped into quasi-periodic pulses whose frequency is controlled by the tension in the vocal cords, in passing out through the throat, the mouth, and possibly the nose.

Sounds responsible for speech are relatively unstructured when generated. Vibrations of the vocal cords create the “buzz-like” sound that has relatively simple spectral properties with harmonics at frequencies corresponding to integer multiples of the fundamental frequency. Friction noise generated by turbulence within the oral cavity, has a relatively broad frequency distribution with a somewhat high pass characteristic.

Though sound sources used in speech are simple in characteristic, the speech signal itself is structured in both frequency and time and is complex. This structure derives from the response characteristics of the vocal tract with resonances (poles) and anti-resonances (zeroes) located at frequencies determined by a variety of factors, but primarily the length and cross-sectional area of the vocal tract above the location of the sound source. The signal is further structured in time by the motion of articulators (tongue/lips/jaw) which constantly effects changes in the vocal tract response characteristics (University of Delaware, 2004).

A hypothesis (Foulds, 2002) which considers that spasticity in some way interferes with the changes in the source from unvoiced to voiced with a direct relationship to durational increase in dysarthric speech, was tested.

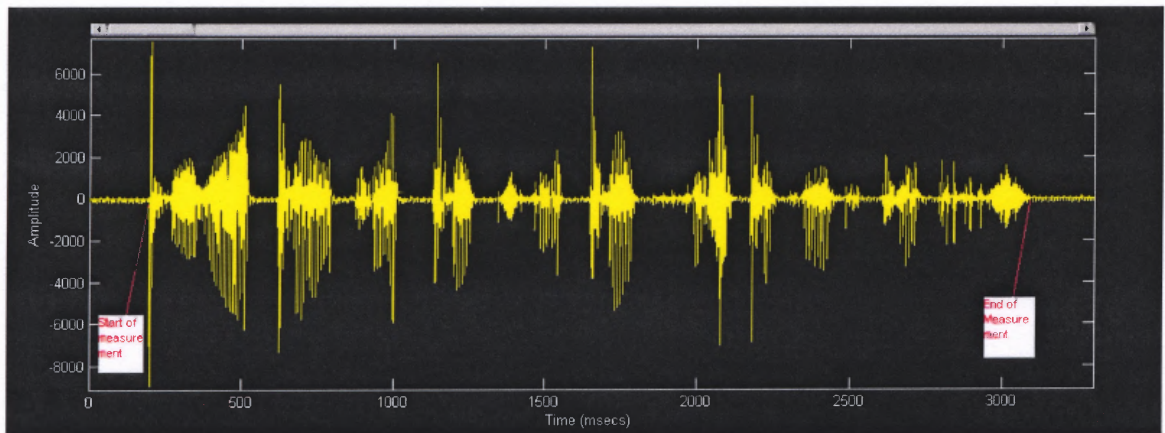
Accordingly, three types of speech from each participant were used for analysis of speech, natural speech, whispering, and speech via the use of an external artificial larynx (APPENDIX C). Natural speech produced speech as described above. Whispering altered speech by producing speech through aspiration without voicing, because the vocal folds were not vibrating, they were slightly closed leaving just enough constriction to cause some frication (is similar to adding an “h” sound to all articulations). Vocalization for the

participants is produced externally by the AL; therefore, alleviating the vocal folds from having to vibrate.

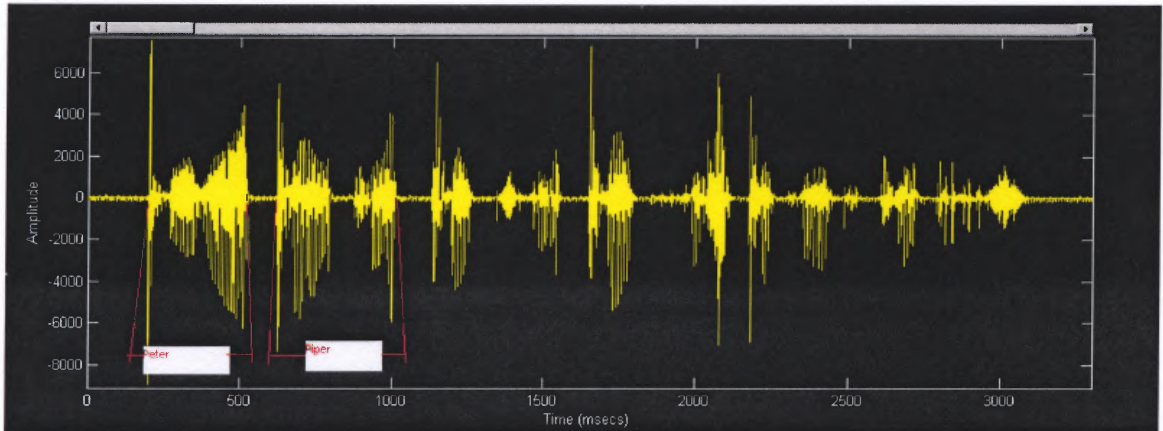
The stop consonant (the letter “P” is used in this study) to vowel transitional change (P-Vowel), was chosen for this study, as the stop plosives of P provide the longest voice onset time to the vowel in comparison to other consonants (O’ Shaughnessy, 1987). In accord, the following sentence was chosen due to the large number of measurable P consonant to vowel transitions:

- Peter piper picked a peck of pickled peppers. Therefore: Pi (short i), PI (long i), Pe (short e), PE (long e) are present and different durational measurements are examined.

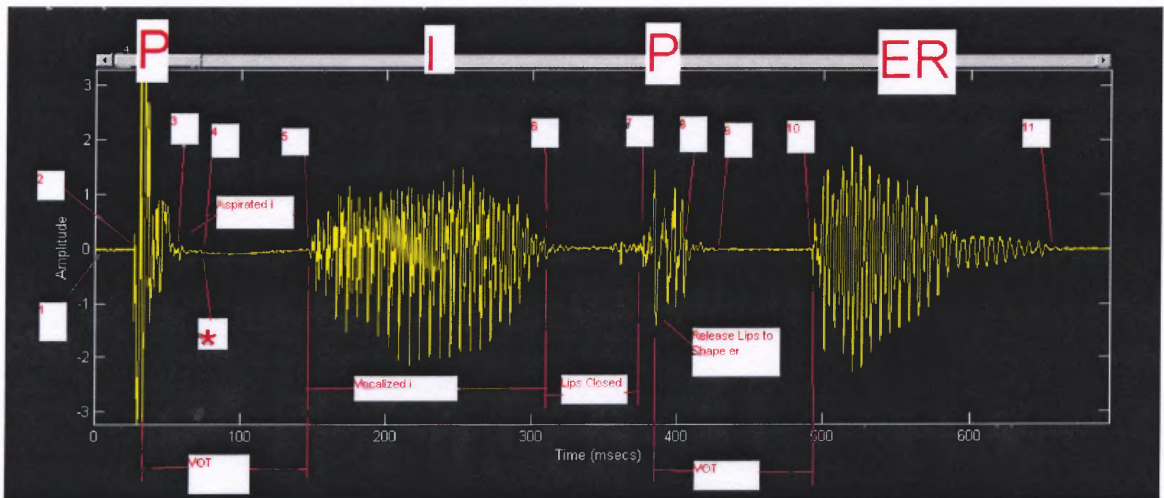
Durational measurements include: sentence duration, duration of individual words in the sentence, voice onset time, occlusion time, time from stop consonant release to end of vowel, and vowel duration.



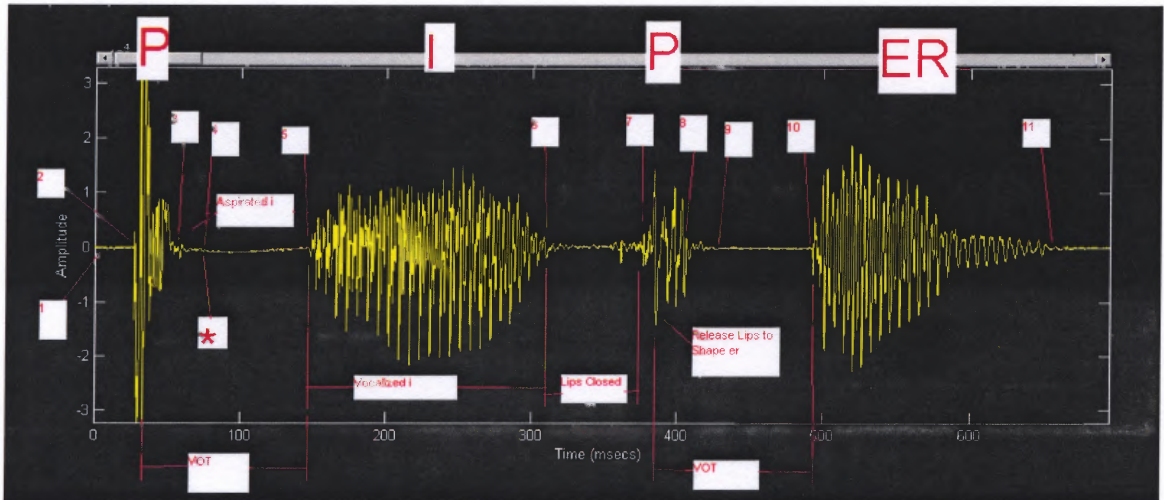
**Figure 3.1** This is the waveform (using Colea, a full speech analyzer software from University of Texas) of Peter Piper Picked a Peck of Pickled Peppers, and was used to measure total sentence duration. The measurement is made from before the plosive release of “P” to the end of Pickles.



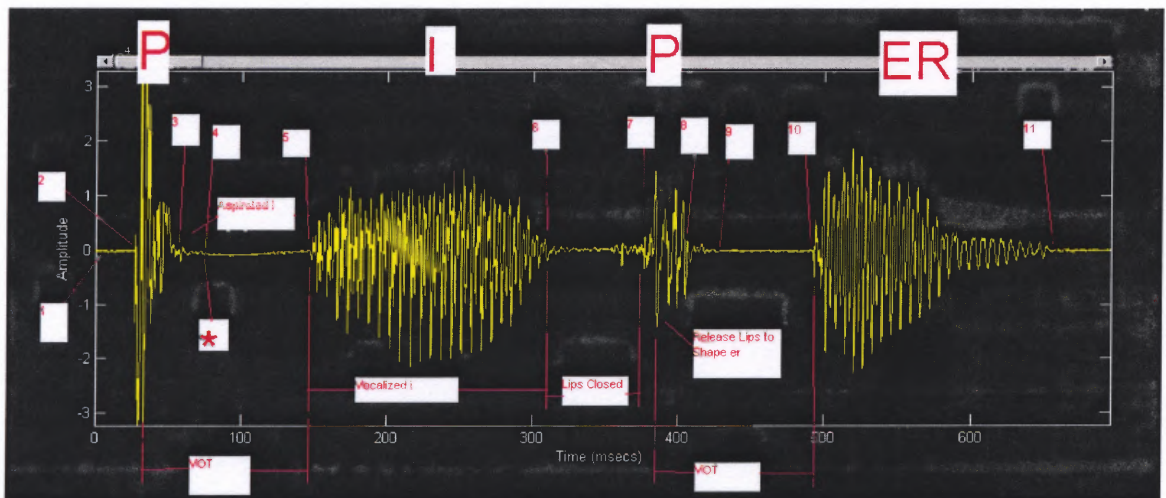
**Figure 3.2** This is the waveform of the entire sentence: Peter Piper Picked a Peck of Pickled Peppers, and depicts how the word durations were individually measured using the example of the first two words.



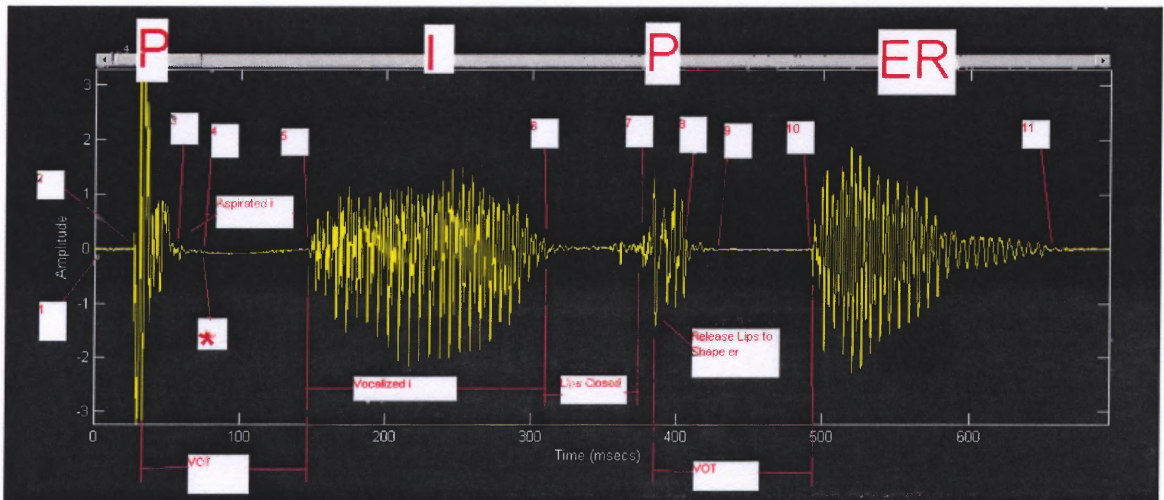
**Figure 3.3** This is the waveform of the word Piper from the sentence; Peter Piper Picked a Peck of Pickled Peppers. Shown as an example of Voice Onset Time (VOT) measurement, and it is represented from point 2 (release of stop consonant “P”) to point 5 (before the start of vocalization of the following vowel).



**Figure 3.4** This is the waveform of the word Piper from the sentence; Peter Piper Picked a Peck of Pickled Peppers. Shown as an example of Occlusion Time measurement, and it is represented from point 6 (closing of the lips at the end of vowel voicing) to point 7 (the point before the stop consonant “P” release).



**Figure 3.5** This is the waveform of the word Piper from the sentence; Peter Piper Picked a Peck of Pickled Peppers. Shown as an example of the measurement from stop consonant to end of vowel, and it is represented from point 2 (release) to point 7 (before the release of the next stop consonant).



**Figure 3.6** This is the waveform of the word Piper from the sentence; Peter Piper Picked a Peck of Pickled Peppers. Shown as an example of vowel duration measurement, and it is represented from point 5 (start of vowel voicing) to point 6 (end of vowel voicing).

In measuring the duration of spastic dysarthric speech in individuals with CP, three participants were selected: one, individual with spastic CP having dysarthria (D-T-S) for the case of spasticity; a second, individual with athetoid CP having dysarthria (D-T-A) but who is not clinically spastic; a third, a non-dysarthric, non-CP participant served as the control talker (C-T) with no known other speech impediment. The D-T-A speech is compared with the D-T-S speech and to the speech of the C-T.

The duration of the utterance is perceptually longer with D-T-S than with the other two participants, D-T-A and C-T. The duration of “Peter Piper Picked a Peck of Pickled Peppers” was measured in the following ways to better understand the causality of durational increase, especially in D-T-S.

- Duration of whole sentence.
- Duration of individual words.



The difference between the sentence duration and the individual word duration accounts for time between measurable starts and stops of the words. This determines the duration due to the consonant occlusion time.

- Duration of voice onset time (VOT) measured.

The VOT is analyzed for any reduction in duration of speech as stated above. The VOT is the time from the release of the stop consonant plosive, P, to the beginning of voicing of the following vowel.

To summarize, the parameters durational measurements of speech include: one, duration of utterance; measured overall, this means the entire sentence; two, comprising words; three, the occlusion time before the stop consonant release, and VOT (Figure 2.2).

Measures of the articulatory accuracy and timing, and intelligibility are suitable for future study.

The three subjects selected for the study fit the following criteria:

- Control talker (C-T)
- Dysarthric talker athetoid (D-T-A)
- Dysarthric talker spastic (D-T-S)

The selection criteria for the C-T participant specify that sex was matched to D-T-S, and does not have any known speech disorder. The selection criteria for the D-T-A are the participant was sex matched to D-T-S, and clinically diagnosed with dysarthric athetoid CP. The D-T-S participant was clinically diagnosed with spastic CP having dysarthria. Though both dysarthric participants have a speech disorder, and have the capability of vocalizing, they do not necessarily have to be intelligible.

The recruitment of participants is as follows:

- C-T participants as volunteers from Matheny School and Hospital, Staff/Faculty at NJIT, and/or NJIT students are acceptable. It should be noted the sex of the C-T must match that of the D-T-S.
- D-T-A and D-T-S: Recommended by Matheny School and Hospital speech pathologist, Susan Buin (Director of Speech Pathology), or Richard Foulds PhD (Associate Professor of Biomedical Engineering, NJIT) as having sufficient vocalization abilities, significant dysarthria, a sufficient cognitive level to participate, as well as having interest in voluntarily participating in this study.

Exclusion criteria for potential dysarthric participants include individuals with high quality of speech, as there is no need for involvement. On the other hand the Inclusion criteria for potential dysarthric participants include individuals who are unable to produce high quality speech and have difficulty expressing certain sounds of words that may be unintelligible to the untrained listener.

## Methodology

Examination of overall sentence duration of D-T-A and D-T-S speech to evaluate any differences to that of C-T speech and in differences in natural speech, whispering, and in AL speech was made.

Possible sources of durational increases were; VOT, which is the delay from the stop consonant release to the onset of voicing, breaths taken during speech, occlusion time before the stop consonant release, and general phoneme length.

Before any tests were conducted, questions that the participants may have were answered thoroughly and to the best of the researcher's ability, furthermore consent forms (Appendix A) were signed by the participants and/or guardians. The participants were informed that if they wished to cease participation in this study; they could do so at any point in time.

The participants practiced whispering with and without using the non-invasive Artificial Larynx (AL) (Appendix B), which helped alleviate any user-device usability difficulties prior to the data collection, and was the basis of the analysis made to determine if the D-T-S and D-T-A duration of speech approaches that of the C-T duration of speech.

During communication with a dysarthric individual, a relaxed, quiet environment where external stimuli are kept to a minimum is crucial. In accord, recording sessions were conducted in a quiet but familiar and naturally comfortable environment using a digital audio video recorder (see Figure 3.7).



**Figure 3.7** Pictured here is the Sony DCR-PC5 digital audio and video recorder which was used to capture the participants' speech production. The data was saved using Panasonic MiniDV digital tape, which is to the right of the camera.

A headset microphone (see Figure 3.8), was used in capturing audio in synchronization with the video. A set of digital auranomic, circum-aural, wired (for maximization of clarity) stereo headphones (see Figure 3.9) utilized for audio feedback of participant's vocalization were used in synchronization with the video input and vocalized audio output.



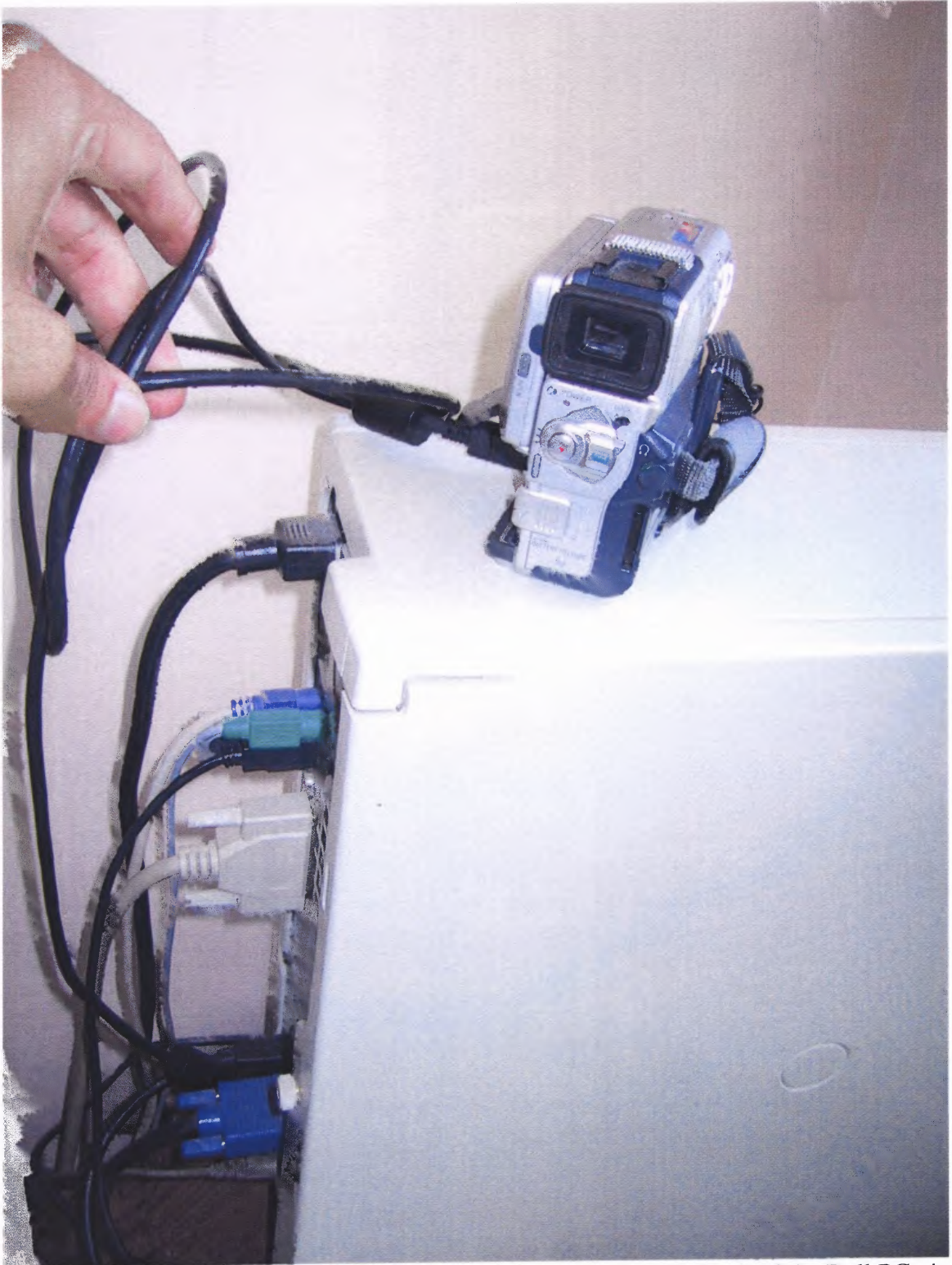
**Figure 3.8** Pictured here is the Shure SM10A headset microphone used to capture the audio from the participants' speech. It was connected to the Sony DCR-PC5.



**Figure 3.9** Pictured here are the stereo headphones (Model: Sony MDR-V600) used for auditory feedback of the participants' speech, in real time, while speaking.

The use of these headphones was found to be extremely helpful during preliminary testing with all subjects to provide amplified feedback of their whispered and AL speech. It is known that individuals who lose their hearing exhibit a gradual degradation in speech production due to the loss of auditory feedback. This occurs since speech production depends upon hearing to produce correct movement patterns of the vocal articulators. During this study, it was found that both the whispered speech of the two subjects with dysarthria and the AL speech of all three subjects was barely audible to the talkers. Amplified feedback assisted them in correctly articulating the words.

The D-T-A or D-T-S was seated, typically in a wheelchair (C-T typically in a standard chair), next to the experimenter. The experimenter led the talker through a series of recordings that includes the set of words comprising the sentence: Peter Piper Picked a Peck of Pickled Peppers. Each word of the sentence was stated first by the experimenter and then repeated by the talker. This assisted all talkers in pronunciation of words and was essential for subjects with limited hearing, eyesight or literacy. The sentences of both the dysarthric talker and the C-T digitally recorded with the Sony DCR-PC5 audio/video recorder were uploaded to a Dell Dimension P3 500Mhz personal computer (see Figure 3.10) via a firewire cable connected to a firewire internal PCI card installed in the Dell personal computer.



**Figure 3.10** Pictured here is the Sony DCR-PC5 connected to the back of the Dell PC via the firewire cable to the internal firewire PCI card.



This high speed firewire protocol is essential in high quality data transfer required in video processing. Sony EZEditor video editing software is used to upload the data and save the video file as type Microsoft Video File. Adobe Audition Version 1.0 was used to extract the audio from the audio/video file, which was then saved as a sound file of type “.wav.” This allowed Colea, a full speech analyzer software running on Matlab Version 6.5 as a “toolbox,” to open the audio files of “.wav” format. Words in each sentence were then marked using the waveform display in Colea to examine segmental duration (intelligibility and consistency can be further analyzed from this data in future studies.).

Each subject was asked to speak the sentence (Peter piper....) three times under each of the following conditions:

1. Natural voice /Vocalized speech (Nat)
2. Whispering (W)
3. Using an Artificial Larynx (AL)

Another set of recordings was taken of C-T saying the same sentences using the described experimental conditions.

### Study recordings of data collection:

Measure durations of all sentences, words and occlusion time, VOT, and vowel length.

1. Analyze the average duration of natural speech, whispering, and with the AL.
2. AL (non dysarthric speech, C-T) duration = 110% natural (non dysarthric speech, C-T) duration.

### Questions:

Does the:

1. Average duration of the C-T  
Natural = 90% AL
2. Average duration of the D-T (within the same subject):  
Natural (Nat) > Whisper (W) > AL or AL < W < Nat
3. Average duration of the D-T (within the same subject):  
Nat > Control Nat
4. Average duration of the D-T (within the same subject):  
W > Control W
5. Average duration of the D-T (within the same subject):  
AL ~ Control AL
6. Ultimately: AL speech duration of D-T approach duration of C-T. Or does motor control or filter improve when source control is minimized? And how does this possibly result in speech change?

## CHAPTER 4

### DATA ANALYSIS: TEMPORAL ANALYSIS

Digital video sentence data was separated into audio and video signals for evaluation of spastic onset and interference. Vocalization onset and interference, as well as physical filter correlation with the audio output was analyzed. Full speech analyzer software, Matlab V 6.5/Colea, is utilized for signal analysis of recorded data.

Acoustic analysis of recordings in the time-domain (via waveform or energy-envelope of speech) was made. Any conformation that the dysarthric talker's natural speech was longer than their whispered speech, or if whispered speech was longer than compared to the duration of AL speech, was noted. This study compared and contrasted three individuals having three different types of speech, though statistical validity can further support the findings in this study by having a larger participation population set than how many are involved in this study.

Although the frequency domain can be examined in future studies to see where dysarthric speech is longer and different from non-dysarthric speech, waveform examination was used in this study. Waveform analysis was used to investigate whether the artificial larynx speech became more like non-dysarthric speech, or if there is another reason such as VOT variation in transitions from stop consonants to vowels. This helped define the D-T-A and D-T-S difficulties and relationships.

Analysis, utilizing the data obtained, to determine the relationships, differences, and properties between the D-T-A, D-T-S and C-T.

Predictions:

- Increased duration within words.

- Overall duration of Spastic sentence was longer than Athetoid and Control.
- Overall duration of each word was longer in Spastic than Athetoid and Control.

Difference in Spastic Speech- Increased Duration:

- Increased duration was not due to occlusion time.
- Increased duration was not due to breathing.

This raises a question as to where the additional duration comes from. It was suspected that the duration from the plosive to vowel voicing may be the reason, as it takes time for the onset of vocalization to occur. If the vocalization variable was minimized or removed, and an alternate source of vocalization was utilized, the delay of voice onset time would, thus, be removed as a direct result, producing speech that was perceptually shorter in duration for dysarthric individuals with spasticity and CP.

- Voice onset time (VOT) between unvoiced stops and vowels – the most dramatic example was used: The consonant “P” as the stop consonant, used to transition into a vowel. It is the most dramatic because unvoiced stops have the longest VOT.
- Vowel (V) transition to Consonant (C) is *ballistic* and requires one muscle to contract while the other relaxes. Such movements are often faster than controlled movements and are seen in closures of the vocal tract such as lip closures or the tongue movement toward the palate.
- C transition to V is *controlled* and therefore more complex, as it requires both antagonist muscles to contract– Closure to complex shape to make the right formants are therefore slower than ballistic movements.

Measuring VOT in the following cases: Pe (Peck, Peppers), PE (Peter), Pi (Picked, Pickled), PI (Piper) for the three participants C-T, D-T-A, and D-T-S, it was predicted that the dysarthric spastic talker would have had the longest durational measurement.

Utilizing these duration measurements:

Peter, Piper, Picked, Peck, Pickled, And Peppers:

- For Control: D1c, D2c, D3c, D4c, D5c, D6c.
- For Athetoid: D1a, D2a, D3a, D4a, D5a, D6a.
- For Spastic: D1s, D2s, D3s, D4s, D5s, and D6s.

$D1s = D1c + \text{extra VOT}_{\text{spastic}}$

Compared to the increased duration also due to articulatory/coordination delays in spastic and athetoid cases, most of the increased duration, relatively, was predicted to be due to extra VOT, stop consonant, “P,” to vowel transition(P→V). A possible explanation was that it was possible that spasticity increases the duration in transitioning from voiced Consonants to voiced vowel.

The control for unvoiced to voiced transitions when whispering, where speech was aspirated and there was no voicing brought about a question as to the cause of the extra duration, and whether it was due to motor control (brain planning). Reductions in whispering durations from that of natural D-T-S speech using: Peter, Piper, Picked, Peck, Pickled, and Peppers, was predicted be primarily from VOT (O’ Shaughnessy, 1987).

Predictions in durational measurements are as follows for the whispering trials:

- Durations of C-T and D-T-A did not change much.
- Durations of D-T-S were reduced.

Via waveform analysis (whispering) for E (Peter,), e (Peck, Peppers), I (Piper), i (Picked, Pickled).:

- The time from stop consonant to “formation” (no voicing) of vowel, in all three cases; C-T, D-T-A, and D-T-S, should have been similar.

Predictions in durational measurements were as follows for the artificial larynx (AL) trials:

- Time should have been similar in all three cases; C-T, D-T-A, and D-T-S.
- Durations of all words will be similar in all three cases; C-T, D-T-A, and D-T-S.

For the sentence, identification of phonemes in waveform analysis yielding duration measurements were used to inspect if all phonemes and all pauses shared the same scaling from:

For durational measurements:

- D-T-S Natural (N) to C-T (N)
- D-T-A (N) to C-T (N)
- D-T-S (N) to D-T-S Artificial Larynx (AL)
- D-T-A (N) to D-T-A (AL)

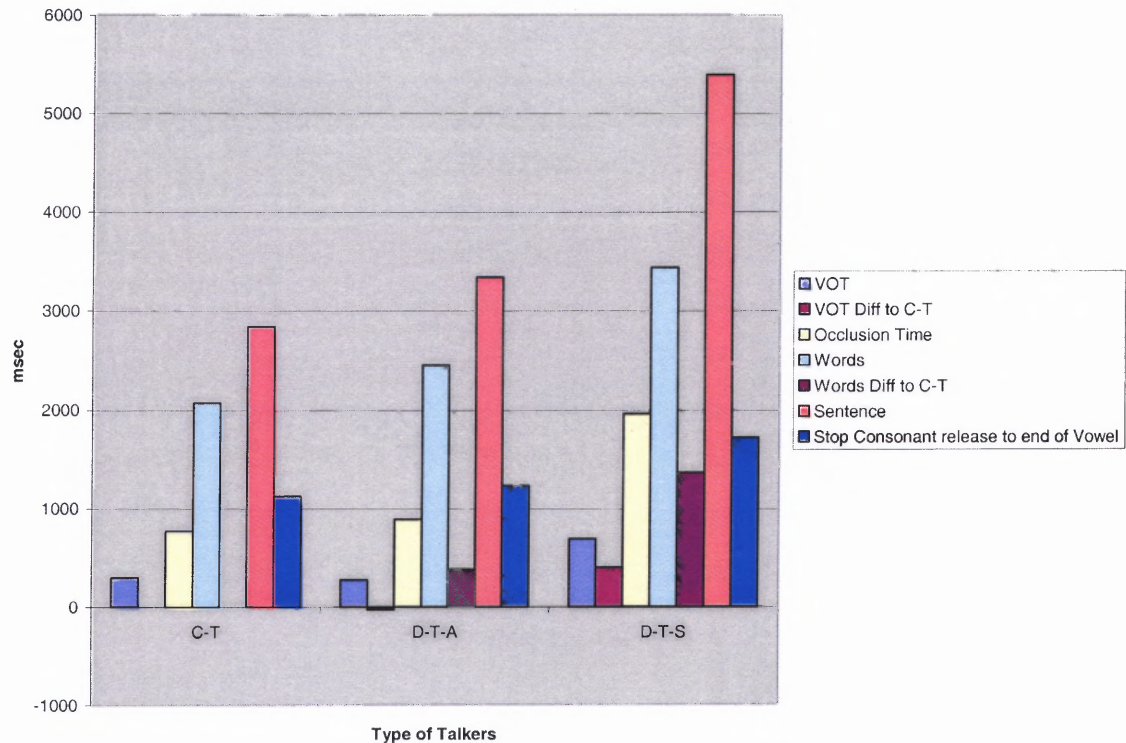
It was predicted that D-T-A (AL) and D-T-S (AL) should have been similar to C-T(N).

Non study-relevant statistical data collected for D-T and C-T groups may also be applicable for future analysis, e.g. inter-relationships within the dysarthric participants.

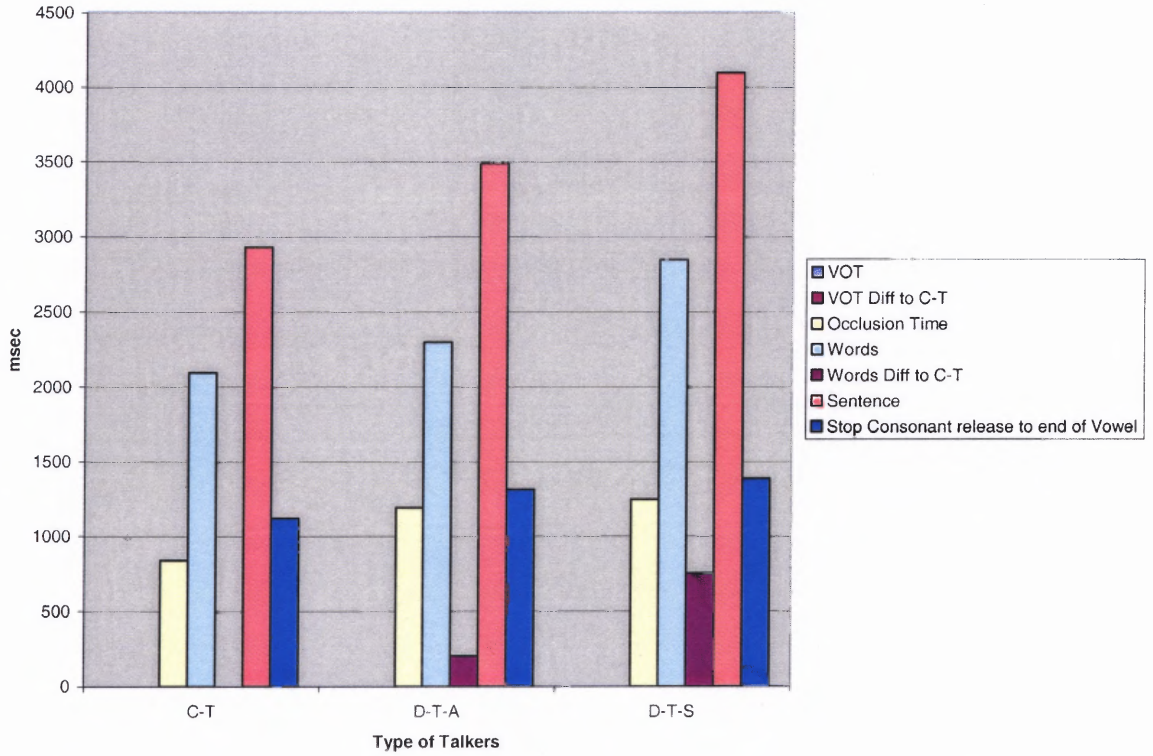
## CHAPTER 5

### RESULTS

Average sentence duration of D-T-S (5394.1 msec) was found to be longer than that of D-T-A (3341.0 msec), which was in turn longer than C-T (2842.9 msec) in natural speech (see Figure 5.1). Specifically, natural spastic dysarthric speech was 189.73% longer in duration than that of natural control speech and 161.45% longer than that of natural athetoid speech (see figures 5.1, 5.2, 5.3, 5.4, 5.5, 5.6).

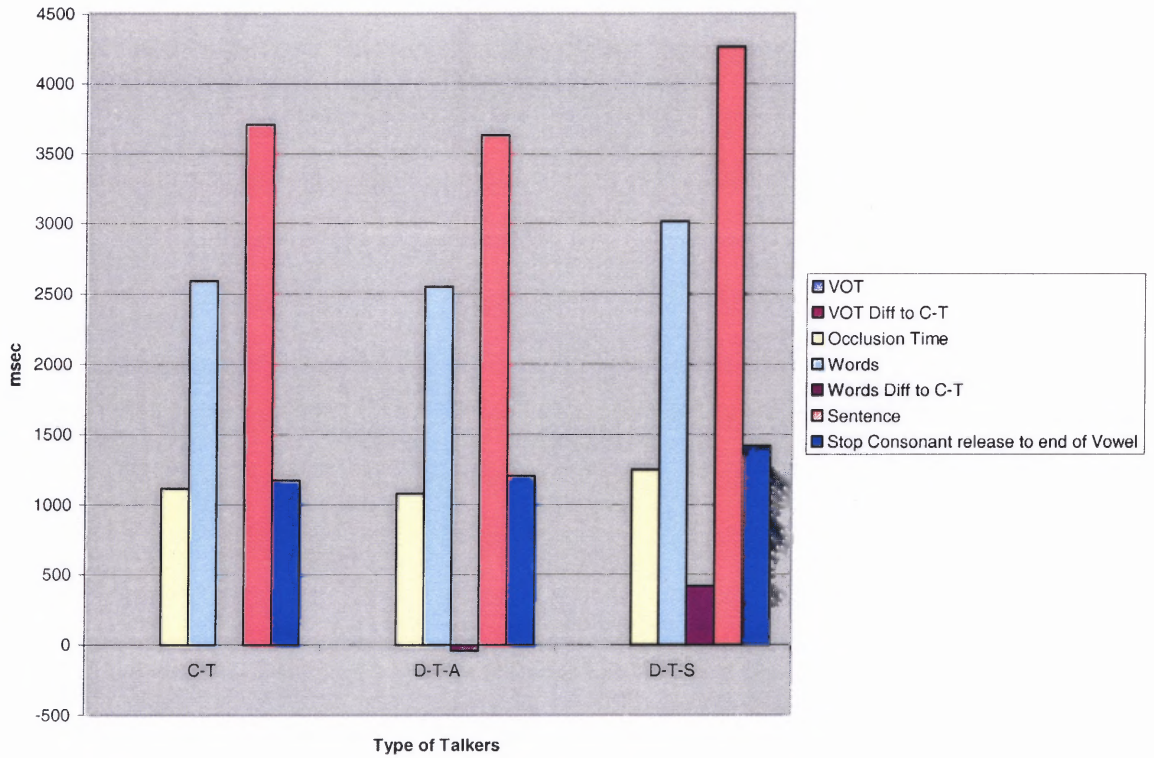


**Figure 5.1** The following graph depicts the durational measurements of natural speech for the three types of talkers.

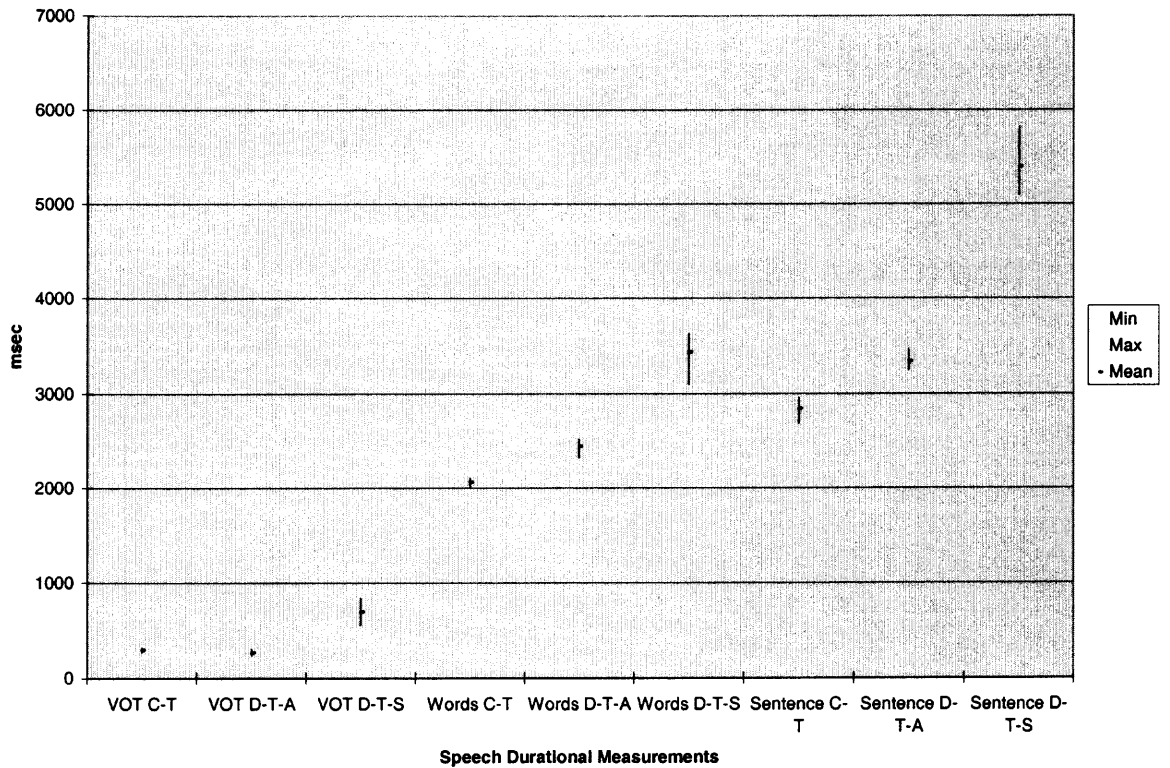


**Figure 5.2** The following graph depicts the durational measurements of whispering for the three types of talkers.

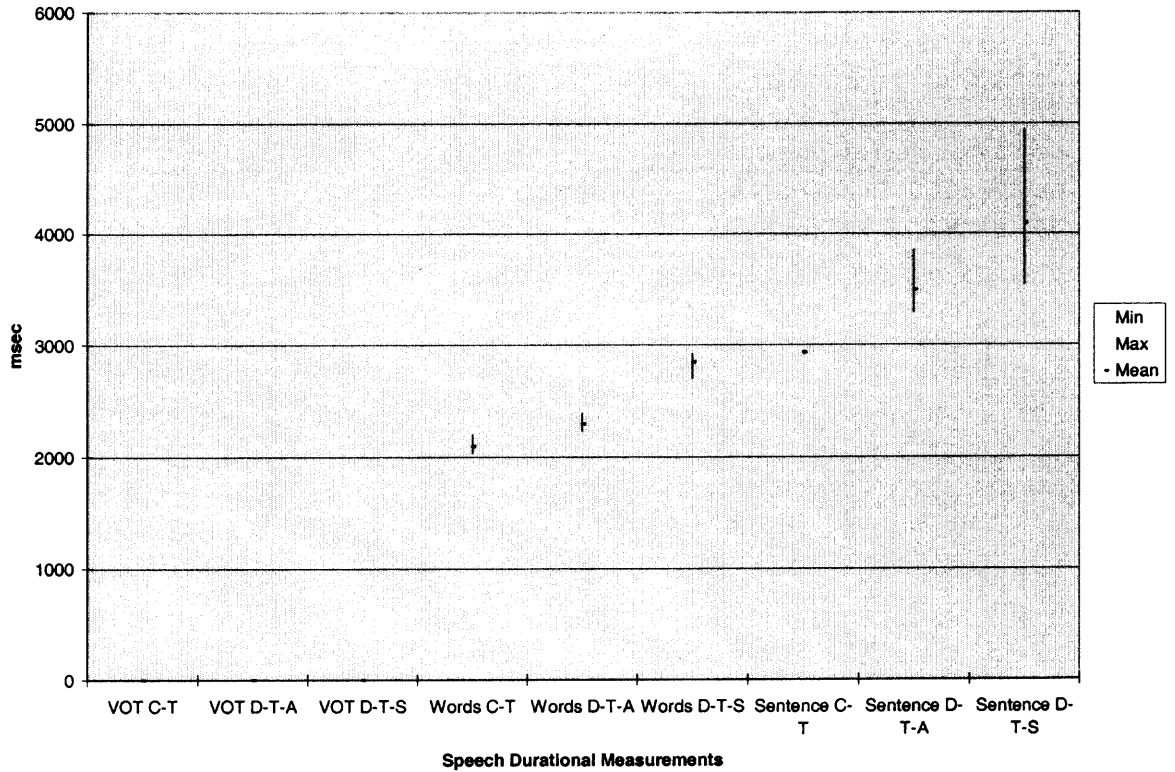




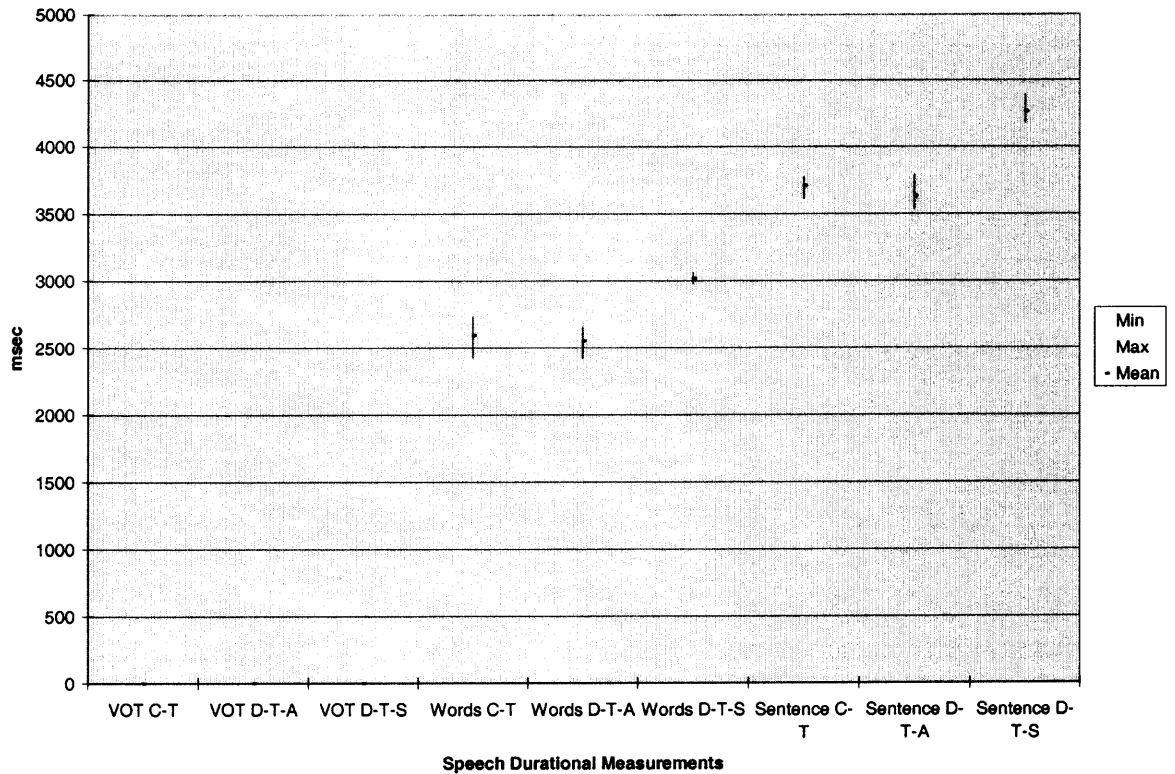
**Figure 5.3** The following graph depicts the durational measurements using the AL for the three types of talkers.



**Figure 5.4** The above figure shows the minimum, mean, and maximum, for natural speech (by C-T, D-T-A, and D-T-S) for durations of voice onset time (VOT), Words, and the Sentence



**Figure 5.5** The above figure shows the minimum, mean, and maximum, for speech when whispering (by C-T, D-T-A, and D-T-S) for durations of voice onset time (VOT), Words, and the Sentence.



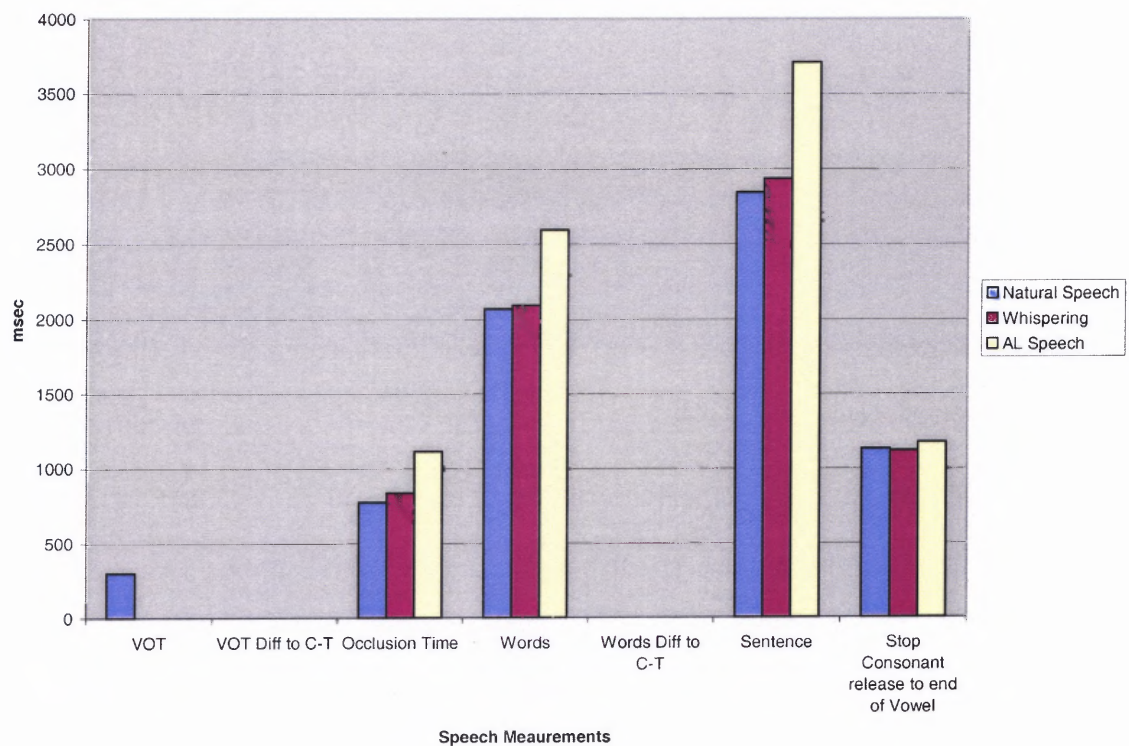
**Figure 5.6** The above figure shows the minimum, mean, and maximum, for AL speech (by C-T, D\_T-A, and D-T-S) for durations of voice onset time (VOT), Words, and the Sentence.

By choosing the most dramatic case of P → C (Stop consonant, “P,” to vowel transition) it was shown that additional VOT was the primary cause of the increased duration within the words. Even though articulatory coordination did contribute to durational increases, it was relatively minimal. Video review of speech in participants revealed that the increased duration was not due to pauses between words due to complexities brought about from breathing during speech.

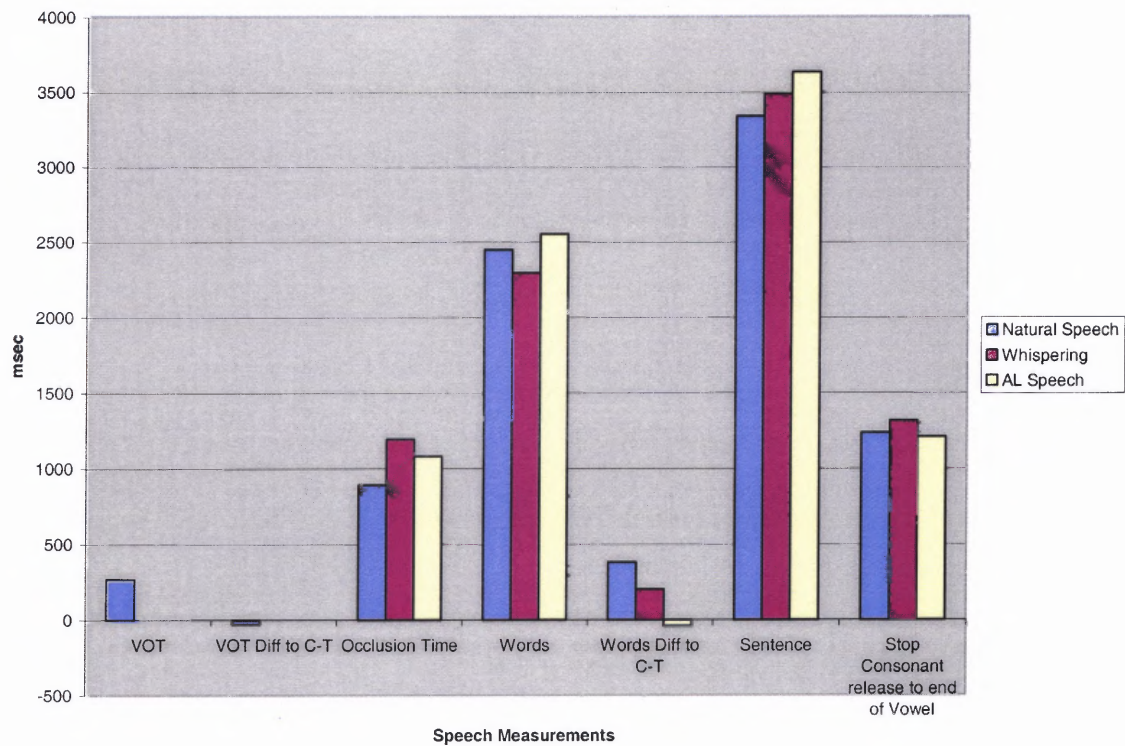
By using whispering (pure aspiration) or AL (pure voicing), it was shown that the extra VOT was not due to the articulatory complexity (coordination) of moving from the stop to vowel articulation patterns, but possibly to the onset of natural voicing, which was noted from the increased similarity in causality of minimization of vocalization directly

relating to the durational decrease in dysarthric speech to being even closer in duration to that of C-T's duration (see Figure 5.3). The time from stop constant to the "formation" (no voicing) of the vowel, in all three types of participants/cases were closer in durational similarity for whispering, and even more so for the AL scenario. For example in the case of D-T-S, durational difference of VOT to that of C-T, decreased from 398.3 msec for natural speech, to near zero (287.3 msec for the plosive) msec for whispering, to near zero (96.7msec for the plosive) msec for the AL (see Figures 5.7, 5.8, 5.9).

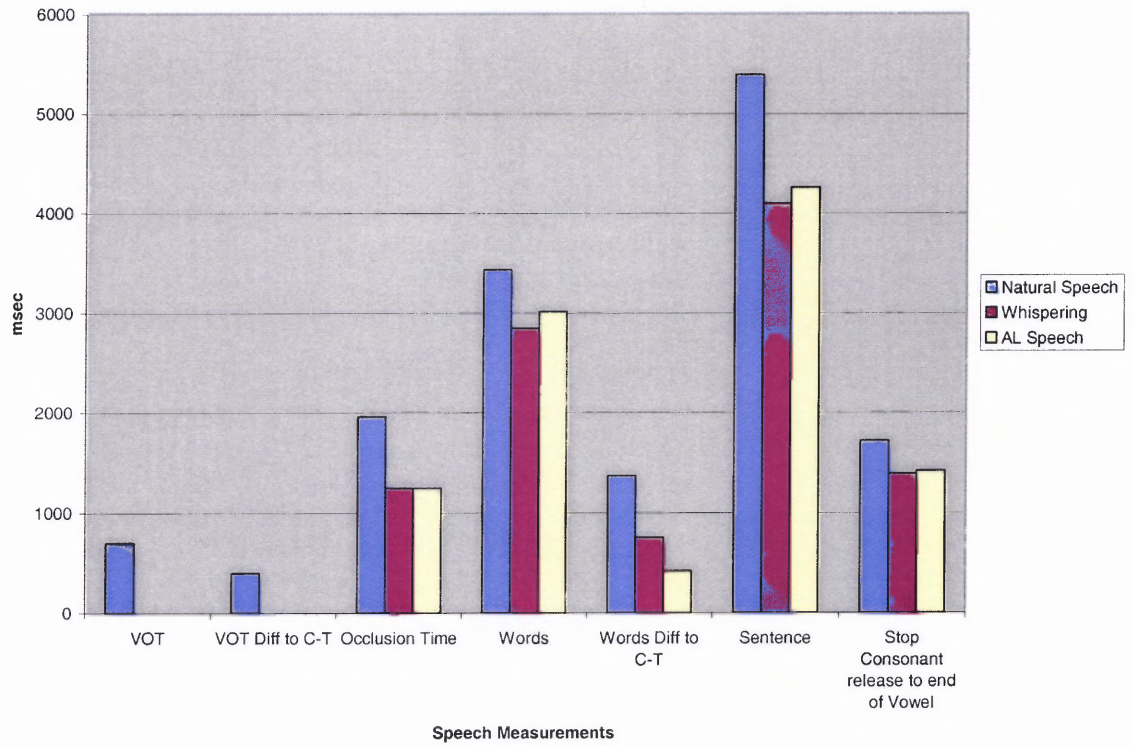
It has been shown that VOT for unvoiced constants to vowel is approximately 25-30 msec. In other words the time from the beginning of the P constant through transitioning to the following vowel should have shown approximately a 25 msec VOT duration in a Control Talker (O' Shaughnessy, 1987).



**Figure 5.7** The following graph depicts the durational changes for C-T.



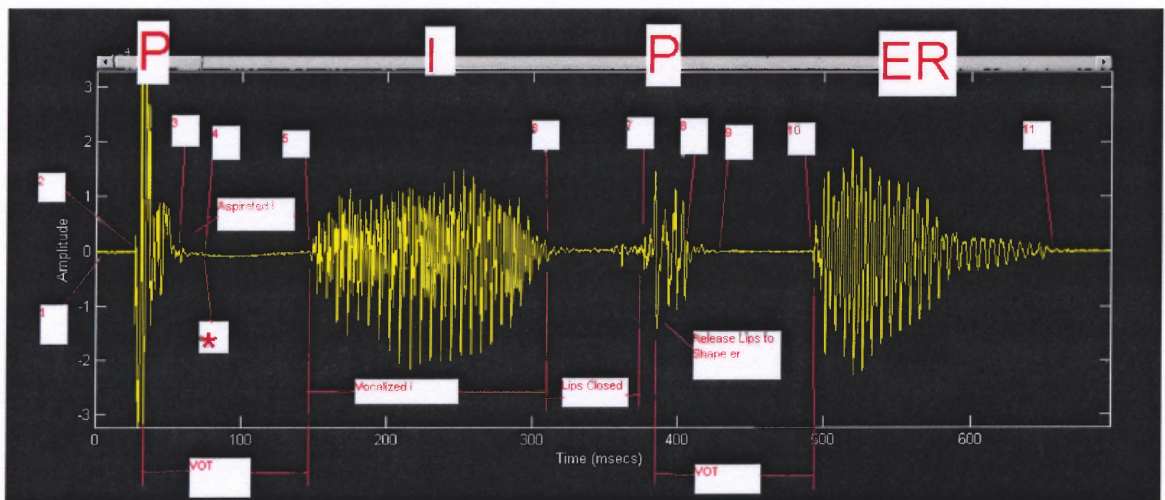
**Figure 5.8** The following graph depicts the durational changes for D-T-A.



**Figure 5.9** The following graph depicts the durational changes for D-T-S.

## CONCLUSION

Piper from Peter Piper Picked a Peck of Pickled Peppers durational measurements and segmental analysis were found to be in agreement with the predictions, confirming that the gaps between phonemes are correlated to the time between consonant to vowel and vowel to consonant transitions and was further supported by Matlab acoustic analysis as these are occlusion times (see Figure 6.1).



**Figure 6.1** Waveform of an individual with spastic Cerebral Palsy having dysarthria speaking naturally.

P1, I, P2, ER.

P1 to I (vowel, V1) is a Consonant to Vowel Transition.

I to P2 is a Vowel to Consonant Transition.

P2 to ER (vowel, V2) is a Consonant to Vowel Transition.



1. P1 consists of closure of the lips.
2. P1 closure of the lips - aspiration of high amplitude as air rushes out.
3. P1-V1 transition is shaping of the articulators. Tongue (pre-shaping), lips and jaw (both must wait for the release before they move to form "I.").
4. P1- V1 Transition represents VOT (in natural speech only).
5. represents the duration of the vowel.
6. P2 represents closure of the lips once again.
7. P2 represents release of the closed lips.
8. P2 – V2 Transition is reshaping of the vocal folds.
9. P2 – V2 Transition represents VOT.
10. V2 Begins.
11. End of vocalization.

Resulting measurements from the three voicing methods yielded a similar duration for the vowel lengths for each respective talker, independent of type of talker. In other words the duration of the vocalized sounds was similar.

The implication of the utterance signifies that the motor-control needed to produce the utterance was predetermined neurologically, and therefore the duration of these vocalizations were not altered.

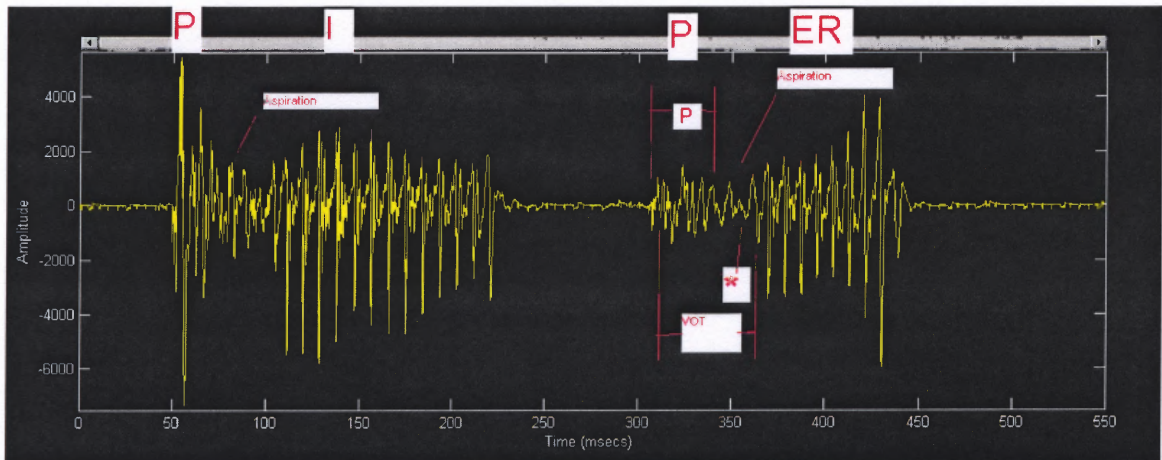
When the vocalization was delayed following the consonant, the delay time before the vocalization started was not subtracted from the duration of the vowel, but the duration of the vowel started upon vocalization.

The question as to whether or not VOT was as long in C-T and D-T-A, in natural speech, as in D-T-S, was answered from noting three points of interest; Stop Release (Release), Aspiration, and Vocalization represented in the speech waveform. For C-T,

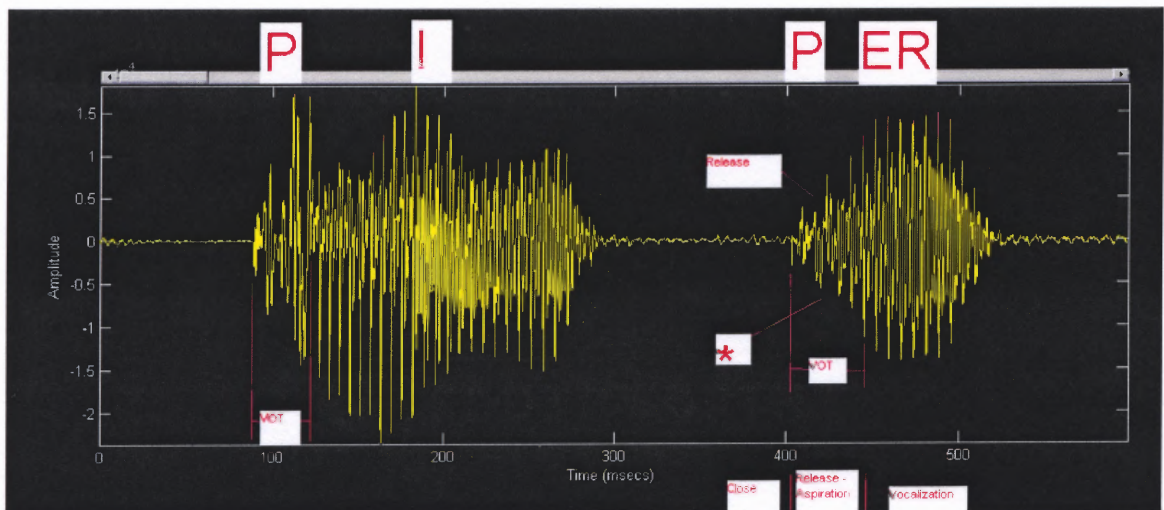
immediately following Release, aspiration started, then ended at what's labeled "\*", where vocalization had begun ( see Figure 6.2). For the spastic case, D-T-S, following Release aspiration ended at "\*", similar to that of the control. However, there was a durational gap before vocalization had begun.

At "\*" was where the vocal folds were at closure for articulatory voicing. For VOT (measured from stop release to the onset of vocalization), for both C-T and D-T-A resulted in vocalization after the combined time of release plus aspiration, as vocalization started immediately after aspiration (see Figures 6.2, 6.3). Spastic VOT, D-T-S, had the same durational VOT time as C-T and of D-T-A, though there was an additional VOT time measured from when aspiration ended to when the vocal folds started to produce excitation. The following measurements were exclusively measured in the word: Piper. Additional VOT delays were noticed observing D-T-S's natural speech VOT and measured to be approximately 117.8 msec, from point 4 to 5 time and point 9 to 10 time, approximately 80.0 msec. On the other hand, VOT was not observed at these points in either whispering or in AL speech. For whispering and AL speech, in comparison to natural D-T-S speech from point 2 to 4, the total time in whispering and in AL was approximately the time from point 2 to 4 in natural speech, approximately 41.7 msec. From point 4 to 5, VOT was not observed in either whispering or in AL. From point 5 to 6 the length of vowel duration was approximately the same at 150.0 msec in all three types of speech. From point 6 to 7, lip closure duration was approximately the same at 51.7 msec. From point 7 to 8, release and burst, the duration was approximately the same at 61.7 msec. From point 9 to 10, it was noted that this segment was non-existent in both

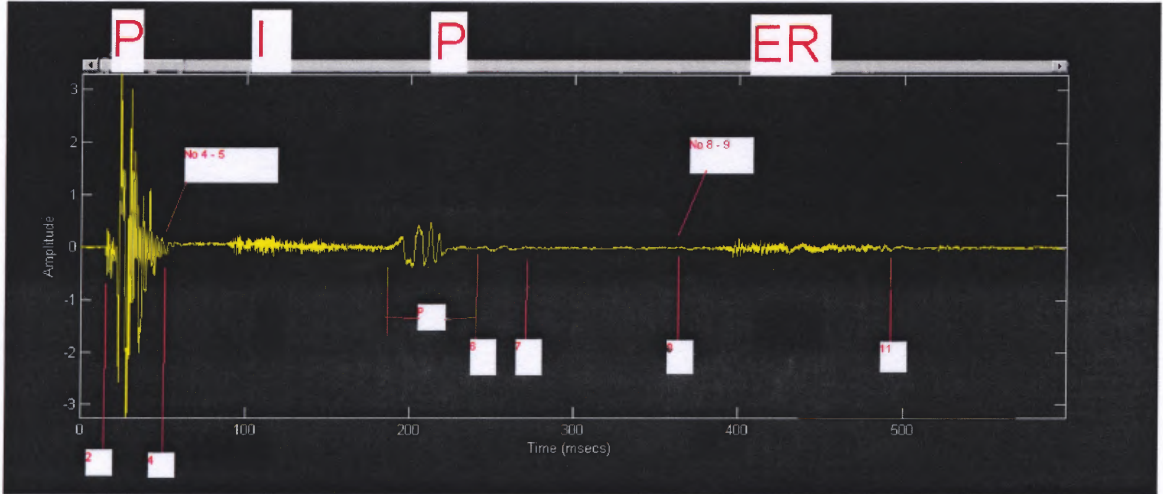
whispering and in AL speech. Finally, for point 10- 11, the length of the vowel was similar, approximately 130.0 msec (see Figures 6.1, 6.2, 6.3, 6.4, 6.5).



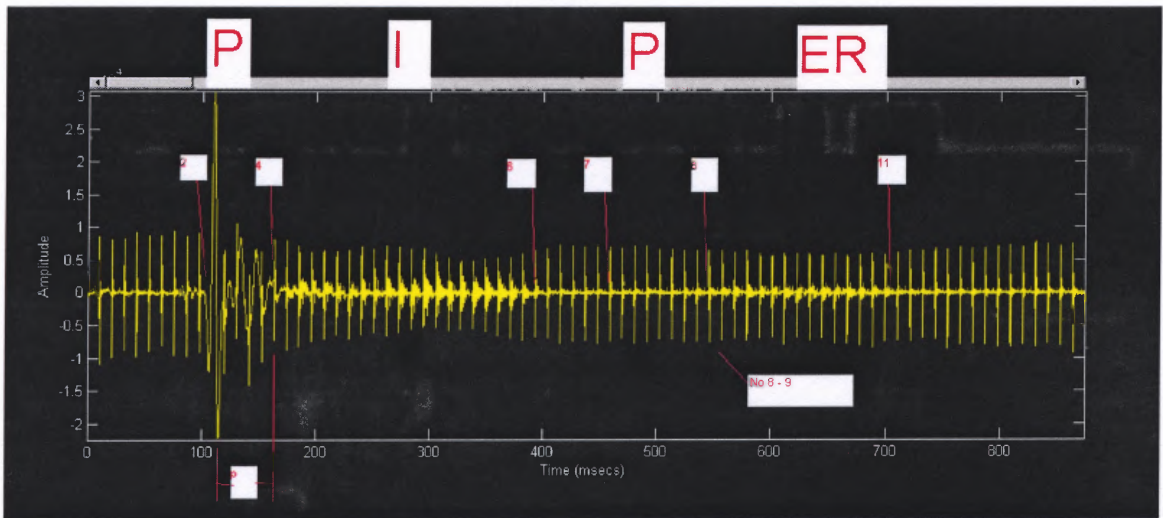
**Figure 6.2** Waveform of an individual with no known speech disorder speaking naturally.



**Figure 6.3** Waveform of an individual with athetoid Cerebral Palsy having dysarthria speaking naturally.



**Figure 6.4** Waveform of an individual with spastic Cerebral Palsy having dysarthria speaking by whispering.



**Figure 6.5** Waveform of an individual with spastic Cerebral Palsy having dysarthria speaking using the artificial larynx.

The working hypothesis, based on Dr. Foulds' hypothesis stating durational reduction via minimized vocalization, was that the time duration difference are due

primarily to VOT in Natural speech as per the measurements in this pilot study. It thus was observed that when transitioning from a stop plosive “P” to a vowel, dysarthric speech appeared closer to that of natural C-T duration, when there was minimal or no VOT influence. In other words the extra duration seemed to be due to the effect of spasticity on the excitation of the vocal cords, required for voicing. The data from this study suggests that spasticity may be one of the reasons for this delay in VOT for D-T-S, where the muscle excitation threshold was too low, as compared to that of the controls C-T and D-T-A. It seems that; eliciting a response, the closed vocal folds, as they were excited, reacted in the form of a spastic reflex to being opened by the aspirated air, and was observed as the extra delay prior to vocalization onset in the speech waveform. This suggested, therefore, that spasticity directly increased the VOT duration.

In D-T-S, within the extra VOT section of speech utterance, the vocal cords are closed, otherwise if they were open, in the time waveform, there would appear to be a continuation of the aspiration that had begun after the stop consonant plosive release, up to the beginning of vowel vocalization. Thus, there is no waveform amplitude present in that portion of VOT, suggesting spasticity affects the vocal folds, by holding them closed for the extended time period creating the delay to the onset of vocalization.

Studies with a larger participant population set would help support the findings of the study, furthermore significant correlations may be found from future analysis of speech consistency, which would help individuals with dysarthria be able to utilize speech recognition technology. Analysis of vocal fold activity using an electroglottograph could possibly provide further evidence and relationships not seen from acoustic analysis alone.

Potential applications of this work would suggest whispering and/or AL may be a viable intervention to improve the intelligibility of dysarthric individuals with spastic CP, to help improve the quality of communication and independence in daily life.

## **APPENDIX A**

### **INFORMED CONSENT**

The following is the consent form given to the participants to sign after the study was explained as well as what was expected of them when participating,, but before the study is conducted.

#### **CONSENT TO TAKE PART IN A RESEARCH STUDY**

#### **TITLE OF STUDY: VOICE ONSET TIME VARIATION IN STOP CONSONANT TO VOWEL TRANSITIONS**

This consent form is part of an informed consent process for a research study and it will give information that will help me to decide whether I wish to volunteer for this research study. It will help me to understand what the study is about and what will happen in the course of the study.

If I have questions at any time during the research study, I should feel free to ask them and should expect to be given answers that I completely understand.

After all of my questions have been answered, if I still wish to take part in the study, I will be asked to sign this informed consent form.

The study doctor (the principal investigator) or another member of the study team (an investigator) will also be asked to sign this informed consent. I will be given a copy of the signed consent form to keep.

I understand that I am not giving up any of my legal rights by volunteering for this research study or by signing this consent form.

The study doctor is interested in finding out if I can fully understand the information I am being given. I need to fully understand the information before I can give my informed consent to enter into this research study.

I will be told if my illness has affected my ability to give my consent for this study. If this is the case, the study doctor and my guardian will talk about whether I should take part in this research study. My legally authorized person is someone who may make this decision for me if I do not have the ability to make it for myself.

I have the right to say Yes or No to the study doctors talking to my guardian. I also have the right to say Yes or No to taking part in this research study.

If I agree to enter the study, the study doctor will frequently check whether I am willing to stay in the study until I have completed my part in it.

The study doctor's assessment of my ability to give my informed consent applies only to my volunteering for this research study. It is not an assessment of my ability to make decisions for other purposes, such as making financial, legal, and/or medical decisions.

**Why is this study being done?**

The study is intended to help us better understand speech problems (dysarthria) in persons with cerebral palsy, and to explore simple ways to make the speech of persons with dysarthria more understandable. Specifically, the study will test the effects of whispering and/or the use of an electronic “artificial larynx” on the duration and understandability of speech in persons with dysarthria.

**Why have I been asked to take part in this study?**

I have been asked to participate either because (1) I have problems in speaking due to dysarthria, or (2) do not have problems speaking and will serve as a member of a comparison group.

**Who make take part in this study? And who may not?**

Persons in the dysarthria group will be chosen from among people recommended for the study by a speech pathologist and who have dysarthria due to cerebral palsy. They will be capable of vocalizing, but their speech may not be fully intelligible.

Persons in the comparison group will have no known speech disorders and will be matched by age and gender to people in the dysarthria group.

**How long will the study take and how many subjects will participate?**

Four or five individuals from the Matheny School and Hospital will be in the dysarthria group. Four or five additional individuals will be in the comparison group. Each individual’s participation will take about one hour.

**What will I be asked to do if I take part in this research study?**

During my participation, I will be asked to speak words, phrases or sentences. I will be asked to speak them naturally, while whispering, and while using an artificial larynx. An artificial larynx is a battery operated device that is held against the neck that amplifies activity in the larynx (“voice box”) while speaking. I may be asked to wear a set of headphones during some or all of these activities.

**What are the risks and/or discomforts I might experience if I take part in this study?**

The artificial larynx is placed externally against the skin and is battery operated, so there is no risk of electrical shock. There may be some slight vibration from the device that might cause tickling.

**Are there any benefits for me if I choose to take part in this research study?**

I have been told that the benefits of taking part in this study may be:

that I will be contributing to a study that might lead to the development of procedures that will benefit dysarthric speakers

However, I may receive no direct benefit from taking part in this study.

**What are my alternatives if I don’t want to take part in this study?**

There are no alternative treatments available. My only choice is not to take part in this study.



**How will I know if new information is learned that may affect whether I am willing to stay in this research study?**

During the course of the study, I will be updated about any new information that may affect whether I am willing to go on taking part in the study. If new information is learned that may affect me after the study or my follow-up is completed, I will be contacted.

**Who will be allowed to look at my research records from this study?**

In addition to key members of the research team, the following people will be allowed to inspect parts of my medical record and my research records related to this study:

- The Institutional Review Board (a committee that reviews research studies)
- Department of Health and Human Services (DHHS) (regulatory agency that oversees human subject research)
- Office for Human Research Protections (OHRP) (regulatory agency that oversees human subject research)

By taking part of this study, I should understand that the study collects demographic data and data on my health. This data will be recorded by the study doctor/investigator who may store and process my data with electronic data processing systems. The data will be kept as long as the study is being conducted and for six additional years.

My personal identity, that is my name, address, and other identifiers, will be kept confidential. I will have a code number and my actual name will not be used. Only my study doctor will be able to link the code number to my name and will keep this information for six years.

My data may be used in scientific publications. If the findings from the study are published, I will not be identified by name. My identity will be kept confidential. The exception to this rule will be when there is a court order or when a law exists requiring the study doctor to report communicable diseases. In this case, I will be informed of the intent to disclose this information to the state agency. Such a law exists in New Jersey for diseases such as cancer, infectious diseases such as hepatitis, HIV, viruses and many others.

The study doctor/investigator will be allowed to examine the data in order to analyze the information obtained from this study, and for general health research.

If I do not sign this approval form, I will not be able to take part in this research study.

I can change my mind and revoke this approval at any time. If I change my mind, I must revoke my approval in writing. Beginning on the date that I revoke my approval, no new personal health information will be used for research. However, the study doctor/investigator may continue to use the health information that was provided before I withdrew my approval.

I have the right to look at my study data at my study doctor's office and to ask for corrections of any of my data that is wrong.

**Will there be any cost to me to take part in this study?**

There is no cost to individuals who participate in the study..

**Will I be paid to take part in this study?**

Participation is voluntary; there is no payment for involvement.

**What will happen if I do not wish to take part in the study or if I later decide not to stay in the study?**

I understand that I may choose not to be in the study. If I do choose to take part it is voluntary. I may refuse to take part or may change my mind at any time.

If I do not want to enter the study or decide to pull out of the study, my relationship with the study staff will not change, and I may do so without penalty and without loss of benefits to which I am otherwise entitled.

I may also withdraw my consent for the use of my data, but I understand that I must do this in writing.

**Who can I call if I have any questions?**

If I have any questions about taking part in this study, I can call the study doctor:

*Abhishek G. Parikh*  
*Biomedical Engineering Graduate Student, NJIT*  
*Biomedical/Information Systems Engineer, Matheny School and Hospital*  
 (908)234-0011, 1, x445

If I have any questions about my rights as a research subject, I can call:

Susan Roeloffs, MD  
 Chair, Institutional Review Board  
 Matheny School and Hospital  
 (908)234-0011, x785

**What are my rights if I decide to take part in this research study?**

I understand that I have the right to ask questions about any part of the study at any time. I understand that I should not sign this form unless I have had a chance to ask questions and have been given answers to all of my questions.

I have read this entire form, or it has been read to me, and I believe that I understand what has been discussed. All of my questions about this form and this study have been answered.

I agree to take part in this research study.

Subject Name: \_\_\_\_\_

Subject Signature: \_\_\_\_\_ Date: \_\_\_\_\_

**If Appropriate and Approved for this Study:**

Legally Authorized Representative: \_\_\_\_\_

Relationship: \_\_\_\_\_ Date: \_\_\_\_\_

**If Required By the IRB:**

Witness Name: \_\_\_\_\_

Witness Signature: \_\_\_\_\_ Date: \_\_\_\_\_

**Signature of Reader/Translator If the Subject Does Not Read English Well:**

The person who has signed above, \_\_\_\_\_, does not read English well. I read English well and am fluent in \_\_\_\_\_ (*name of the language*), a language that the subject (his/her parent(s)/legal guardian) understands well. I understand the content of this consent form and I have translated for the subject (his/her parent(s)/legal guardian) the entire content of this form. To the best of my knowledge, the subject (his/her parent(s)/legal guardian) understands the content of this form and has had an opportunity to ask questions regarding the consent form and the study, and these questions have been answered (his/her parent(s)/legal guardian).

Reader/Translator Name: \_\_\_\_\_

Reader/Translator Signature: \_\_\_\_\_ Date: \_\_\_\_\_

Witness Name: \_\_\_\_\_

Witness Signature: \_\_\_\_\_ Date: \_\_\_\_\_

**Signature of Investigator or Responsible Individual:**

To the best of my ability, I have explained and discussed the full contents of the study, including all of the information contained in this consent form. All questions of the research subjects and those of his/her parent(s) or legal guardian have been accurately answered.

Investigator/Person Obtaining Consent: \_\_\_\_\_

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

## APPENDIX B

### SIEMANS SERVOX ELECTRO LARYNX DEVICE

This device is what is used to simulate no vocalization by the participant physiologically. All vocalization resonations are generated by a vibrating memberane at it's tip, which is placed externally (non-invasive) against the skin in a location along the vocal tract.

Dunellen Hearing Aid Center - Siemens Servox Electrolarynx

Page 1 of 2

A family tradition of quality care, since 1986

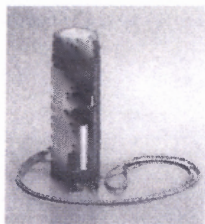
**Dunellen**  
HEARING AID CENTER

*Siemens Servox Electrolarynx  
Artificial Larynx Device*

The Dunellen Hearing Aid Center provides sales and service to laryngectomy patients. We carry the Siemens Servox Inton Electrolarynx, which is regarded as the finest artificial larynx device on the market today.

Thanks to the amazing, precisely engineered Servox speech device, you now have the ability to learn to speak in a new way. Communicate one-to-one, to an audience, even on the phone.

The Servox electrolarynx also offers an oral connector so that you can begin speaking immediately following laryngectomy surgery.



#### Servox features include:

- Unequaled sound quality.
- Weighs only 6.2 ounces.
- Two control buttons for natural sounding pitch variation.
- Dual, automatic battery charging console.
- Masculine/feminine tone adjusting capability.
- Unit is encased in scratch-resistant titanium housing for exceptional durability.
- Battery recharges within 2 hours.
- Extremely wide frequency range.

#### We carry all Servox accessories including:

- |                     |                    |
|---------------------|--------------------|
| • Batteries         | • Oral Connectors  |
| • Dual Chargers     | • Dental Tubes     |
| • Shower Protectors | • Foam Filters     |
| • Stoma Scarves     | • Replacement Caps |
| in assorted colors  | • Canula Brushes   |

ORAL ADAPTERS

**We ship Priority Mail.**

**Credit card orders may be phoned or faxed.  
Call for more information and prices.**

Phone (732) 424-0445 • Fax: (732) 424-1751

Dunellen Hearing Aid Center • 376 North Avenue • Dunellen, NJ 08812  
Call: (732) 424-0445 • Fax (732) 424-1751 E-mail: hearing@inf.net

Dunellen Hearing Home Page  
About Dunellen Hearing Aid Center  
Hearing Aids | Assistive Listening Devices | Servox Artificial Larynx Device  
Comprehensive Hearing Evaluations | Can a Hearing Aid Help Me?

## REFERENCES

- American Speech and Hearing Association,  
<http://professional.asha.org/resources/journals/jslhr/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=25128>, 2003.
- Boves L.(1999), Rosengren E (2000), Perceptual analysis of dysarthric speech in the ENABL project. TMH-QPSR, KTH, 1/2000:13-18 (this volume),  
<http://www.csun.edu/cod/conf/1999/proceedings/session0106.htm>, 2003.
- CancerWEB, <http://cancerweb.ncl.ac.uk/omd/index.html>, 2003.
- Cerebralpalsyfyi, [http://www.cerebralpalsyfyi.com/cerebral\\_palsy\\_glossary.html](http://www.cerebralpalsyfyi.com/cerebral_palsy_glossary.html), 2004.
- Cerebral-palsy.org, <http://cerebral-palsy.org>, 2003.
- Columbia University, [http://www.asel.udel.edu/~bunnell/aip-cd/spch\\_prod.htm](http://www.asel.udel.edu/~bunnell/aip-cd/spch_prod.htm), 2004.
- Foulds, R. (1980), Communication rates for non-speech expression as a function of manual tasks and linguistic constraints. In Proceedings of the International Conference on Rehabilitation Engineering. (Toronto, Canada. June 16-20). Rehabilitation Engineers Society of North America. Toronto, pp. 83-87.
- Foulds, R., "Neuromotor Coordination and its Effect on the Temporal Structure of Dysarthric Speech." Unpublished paper, NJIT, 2002.
- Gold and Morgan (1999), Speech and Audio Signal Processing, Wiley,  
<http://www.dcs.shef.ac.uk/~martin/teaching/speech/speech.pdf>., 2003.
- Hawley, Mark (1999), New technologies supporting people with severe speech disorders, [http://216.239.53.104/search?q=cache:66uIYTn\\_CMAJ:www.dcs.shef.ac.uk/spandh/projects/stardust/presentations/Hawley0206.ppt+Hawley+1999+speech+recognition&hl=en&ie=UTF-8](http://216.239.53.104/search?q=cache:66uIYTn_CMAJ:www.dcs.shef.ac.uk/spandh/projects/stardust/presentations/Hawley0206.ppt+Hawley+1999+speech+recognition&hl=en&ie=UTF-8), 2003.
- Health-Dictionary, [http://www.health-dictionary.com/deafness\\_term\\_details/Dysarthria](http://www.health-dictionary.com/deafness_term_details/Dysarthria), 2004.
- Kent, Ray D.; Netshell, Ronald; Abbs, James H.-University of Wisconsin-Madison (September 1979), Acoustic Characteristics of Dysarthria Associated with Cerebral Disease; Journal of Speech and Hearing Research, pp. 627-648.
- Kent, Ray D., Ph.D., Weismer, Gary, Ph.D.-Communicative Disorders, University of Wisconsin-Madison; Kent, Jane F.-Waisman Center, University of Wisconsin-Madison; Vorperian, Houri K. -University of Wisconsin-Madison; Duffy, Joseph R. -Department of