ACOUSTIC SIGNALS AS VISUAL BIOFEEDBACK

IN THE SPEECH TRAINING OF

HEARING IMPAIRED CHILDREN

by

Elizabeth E. Crawford

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Audiology

in the Department of Communication Disorders

University of Canterbury

March 2007

Acknowledgements

Firstly I would like to thank my supervisor, Emily Lin, for her guidance, encouragement and many hours spent helping me collect data and write-up this thesis. Thank you also to the Advisors of Deaf Children at van Asch Deaf Education Centre, who helped recruit the participants for this study. Thank you to the Oticon Foundation New Zealand for their contribution which enabled me to present this research abroad at the ASHA conference. Finally, thank you to my husband, Mike, and to my family, for their support over the last two years during my thesis and studies.

Contents

List of Figures

List of Tables

Abstract

This study investigated the effectiveness of utilizing acoustic measures as an objective tool in monitoring speech errors and providing visual feedback to enhance speech training and aural rehabilitation of children with hearing impairment. The first part of the study included a comprehensive description of the acoustic characteristics related to the speech deficits of a hearing impaired child. Results of a series of t-tests performed on the experimental measures showed that vowel length and the loci of formant frequencies were most relevant in differentiating between correctly and incorrectly produced vowels, while voice onset time along with measures of Moment 1 (mean) and Moment 3 (skewness) obtained from speech moment analysis, were related to consonant accuracy. These findings, especially the finding of an abnormal sound frequency distribution shown in the hearing impaired child's consonant production, suggest a link between perceptual deficits and speech production errors and provide clues to the type of compensatory feedback needed for aural rehabilitation.

The second part of the study involved a multiple baseline design across behaviours with replication across three hearing impaired children to assess the efficacy of treatment with acoustic signals as visual feedback. Participants' speech articulations following traditional speech training and training using spectrographic and RMS displays as visual feedback (referred to as "visual treatment") were compared, with traditional non-visual treatment followed by visual treatment on one or two targets in a time-staggered fashion. Although no statistically significant difference on the experimental measures was found between the two training approaches based on perceptual assessment, some objective acoustic measures revealed more subtle changes toward normal speech patterns with visual treatment as compared to a traditional approach. Further acoustic-perceptual studies with a larger sample size and longer experimental period are needed to better understand the general and long-term effectiveness of visual treatment.

1 Introduction

A close relationship has traditionally been assumed between speech perception and the development of speech production skills (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). In the case of hearing-impaired children, limitations of their auditory system can prevent them from perceiving differences in sounds, resulting in speech production that is delayed or disordered (Ruffin-Simon, 1983). The degree of intelligibility or clarity of the speech produced by children with sensorineural hearing impairment varies considerably, ranging from totally unintelligible to near normal. Although several studies have documented the typical speech errors of hearing-impaired children in comparison with normal-hearing children, the majority of these studies have been confined to perceptual analysis of phonetic and phonologic errors (Hudgins & Numbers, 1942; Smith, 1982; Dunn & Newton, 1986; Culbertson & Kricos, 2002).

In recent years, the development of digital technology has greatly improved the efficiency in studying speech acoustics, which can provide more precise details than perceptual analysis regarding the timing and the frequency distribution of acoustic energy for individual speech segments. While perceptual studies present with great face validity, their results tend to be considerably variable due to the subjective nature of the analysis. Acoustic analysis, on the other hand, allows for more precise and objective tracking of subtle speech sound changes in guiding speech training and monitoring progress (Uchanski & Geers, 2003). In addition, the acoustic characteristics associated with the articulation errors produced by hearing impaired children may reveal how the acoustic information lacking in their auditory feedback affects the way they establish the association between a sound classification and its production scheme and thus help identify the missing differentiating components that may be supplemented through other types of feedback.

Special techniques of speech training have been taught to hearing impaired children in order to develop correct speech production. Such training techniques often include the use of visual, tactile, and kinesthetic cues to compensate for the lack of access to auditory cues (Dodd, 1976; Ling, 1976). With advances in technology in the past few decades, devices providing visual cues in the form of real-time visual displays of acoustic as well as kinematic or physiological signals are now easily accessible. However, there is a paucity of literature on the treatment efficacy of these applications. More research is needed before many of these devices can be accepted by professionals for routine clinical use.

1.1 Thesis Overview

This thesis includes two acoustic-perceptual studies to identify methods useful for improving speech training of hearing impaired children. Study One was designed to identify particular acoustic characteristics of a hearing-impaired child's speech relating to whether a production is perceived as correct or incorrect. Based on the theoretical understanding of the acoustic-articulatory relationship in vowel and consonant productions, the hypothesis was that the acoustic analysis would be useful for revealing the underlying constraints responsible for the speech deficits of the hearing-impaired. Study Two was designed to examine whether the use of spectrograms as visual biofeedback were more effective in speech training of hearing impaired children compared to a traditional training approach. The hypothesis was that the provision of objective, real-time visual feedback, including acoustic displays suggestive of manner of articulation (pitch, airflow, intensity and spectral pattern) as well as placement of articulation, would have a more positive effect than traditional therapy approaches on improving speech production in children with hearing impairment. The purpose of this project was to provide empirical evidence for clinicians to consider in deciding how and whether to employ acoustic signals as visual biofeedback in delivering speech training to hearing impaired children.

1.2 Background

This section will provide a review of the theoretical background on the relationship between speech production and acoustics of speech as well as previous findings regarding the acoustic characteristics of English phonemes. To establish a link between the fundamentals of speech acoustics and the speech training of hearing impaired children, this literature review will also include current understanding of the human hearing mechanism and hearing loss, and previous findings regarding the effect of hearing loss on speech perception, speech errors of the hearing impaired, traditional speech training, and visual feedback for the hearing impaired.

1.2.1 Speech Production and Acoustics of Speech

The speech production process begins from the moment when the air flows from the lungs and travels up the trachea, a tube consisting of rings of cartilage. From the trachea, the air flows up through the larynx toward the vocal tract. The larynx consists of the epiglottis, true vocal folds, and false vocal folds. When the air flows through the true vocal folds, the vocal folds may remain open (abducted) resulting in a voiceless speech sound, or may close (adduct) resulting in phonation of a voiced speech sound. During phonation of voiced sounds, the vocal folds open and close rapidly and relatively periodically. The semi-periodic vibration of the vocal folds provides the basic harmonic structure of the speech sound (Fry, 1979). The first harmonic, which has the lowest frequency, is called the fundamental frequency. The air stream flowing through or modulated by the vocal folds then passes through the vocal tract, which forms the resonance tube in the speech production system and serves as a filter for all speech sounds as well as the source of noise for consonants. The vocal tract consists mainly of the pharynx, nasal cavity, and oral cavity. The shape of the vocal tract is varied by the soft plate (velum), tongue, lips, and jaw, which are referred to as the articulators. The speech sound waves moving through the vocal tract are shaped by the articulators, causing the distribution of the sound energy across frequencies to be modified by the resonance of the vocal tract configuration. The formants, which are the loci of frequencies where acoustic energy is concentrated, are created by the sound filtering effect of the vocal tract. The two spectral peaks with the lowest frequencies, namely, Formant 1 (F1) and Formant 2 (F2), are the most important formants for vowel recognition (Yost, 2000).

1.2.2 Acoustic Characteristics of English Phonemes

Phonemes, the fundamental units of phonology, are the sounds which differentiate one word from another. English contains about 42 phonemes, which can be classified into four general categories according to how they are produced: vowels, diphthongs, semivowels and consonants (Weiss, Gordon & Lillywhite, 1980).

Vowels are produced by laryngeally modulated air stream flowing through a fixed open vocal tract. The particular vowel sound identity depends on the position of the articulators, in particular the tongue. For example, when the tongue is highly arched in the back of the mouth, a high back vowel is produced. When the tongue is arched in the centre of the mouth, central vowels are produced. The formant frequencies of a vowel are determined by the configuration of vocal tract, namely the vocal tract length, tongue positioning, and lip spreading or rounding. The formant frequencies of vowels vary between men, women, and children, due to the differences in the size and shape of their vocal tract (Fry, 1979). When analysing vowels, the ‘point vowels’ /i, a, u/, which represents the three corner points in a F1-F2 frequency plot, are commonly used to denote the extent of vowel differentiation. Use of the point vowels dates back to Joos (as cited in Vebrugge & Strange, 1976), who noted that they represented the extreme positions of an individual’s articulatory vowel space and the extremes of formant frequency values in an individual’s acoustic vowel space (Verbrugge & Strange, 1976).

Diphthongs are voiced sounds that are produced by a continuous change in vocal tract position from one vowel towards another within the same syllable, resulting in a smooth switch from one set of formants in the direction of another (Fry, 1979). Semivowels, /y/ and /w/, are characterized by a gliding formant frequency transition at a rate which is considerably quicker than that of vowel-to-consonant transitions. The acoustical characteristics of these sounds depend strongly on the context in which they occur (Fry, 1979).

Consonants make up the bulk of English sounds and are the most variable phonemes of English. While vowel sounds produce the bulk of the energy of speech, consonants provide most important acoustic clues for speech intelligibility (Ling, 2001). In contrast to vowels, consonants require greater constriction of the vocal tract and faster, more precise adjustments of the articulators. Consonants are produced when the tongue or the lips change the nature of the breath stream in some way. They can be classified in terms of manner of articulation, place of articulation, and voicing (Ling, 1989). The manner of articulation, including the main classification of “fricatives”, “affricates”, “stops”, “nasals”, and “liquids” in English, is based on the path of the airflow and the degree to which it is altered by vocal tract constrictions. ‘*Fricative’* sounds are generated by a constriction at some point in the vocal tract which causes turbulence in the airflow. The turbulent air slowly released from the narrow passage between the two closely approximated articulators creates a sound like a hiss. *‘Plosive’* or ‘*stop’* sounds are created by an initial blockage of the vocal tract by the lips and nasal cavity, allowing the air pressure to build up, followed by a sudden release. ‘*Affricates’* are a combination of a plosive and a fricative. For example, ‘t! ’ is produced by the stop /t/ followed by the fricative /! / (Grant, Walden & Seitz, 1998). ‘*Nasals’* are produced by closing the vocal tract and lowering the soft palate so that the air travels out through the nasopharynx. ‘Liquids’ are produced by creating a partial closure in the mouth with the tongue, resulting in a resonant, vowel-like consonant (Weiss et al., 1980).

Place of articulation refers to the point of greatest constriction in the vocal tract. The places of articulation of relevance in English include the lips (bilabial), lips and teeth (labio-dental), teeth (dental), upper gums (alveolar), hard plate (palatal), soft plate (velar), and glottis (glottal) (Ling, 2001). Table 1 summarises the classification of each of the consonants of English.

Table 1. Classification of Consonants by Manner, Place and Voicing (Grant et al., 1998).

VOICING					
Voiced				Unvoiced	
b,d,g,m,n,v,ð,z,ʒ,dʒ				p,t,k,f,θ,s,ʃ,tʃ	
MANNER OF ARTICULATION					
Stop		Nasal	Fricative		Affricate
b,p,g,k,d,t		m,n	v,f,ð,θ,z,s,ʒ,ʃ		dʒ,tʃ
PLACE OF ARTICULATION					
Bilabial	Lingua-Velar	Lingua-Alveolar	Lingua-Dental	Lingua-Palatal	Labio-Dental
b,p,m	g,k	d,t,n,s,z	ð,θ	ʒ,ʃ,dʒ,tʃ	v,f

The articulatory differences in production of different phonemes contribute to the variety of differences on the frequency and intensity patterns between speech sounds. The identification of speech sounds requires the ability to detect and discriminate the subtle changes in frequency, intensity, and duration (Ling & Ling, 1978). There are also the various suprasegmental elements of speech, including rate, rhythm, and intonation, that may be related to the acoustic changes of speech sounds. The process of speech perception, therefore, depends on the ability to detect the various segmental and suprasegmental elements that comprise speech. Without a normal auditory system, the development of auditory, speech, and language skills may be seriously affected (Culbertson & Kricos, 2002).

1.2.3 The Hearing Mechanism and Hearing Loss

The human auditory system comprises the peripheral auditory system and the central auditory system. The peripheral auditory system consists of the outer, middle and inner ear, and the cochlear nerve. The central auditory system includes the brainstem nuclei, thalamic nuclei and the auditory cortex. Sound waves collected by the pinna travel through the ear canal to the tympanic membrane, which transmits the vibrations to the ossicles of the middle ear. The movement of one of the ossicles, known as the stapes footplate, through its attachment to the oval window of the cochlear, creates a wave within the fluids of the cochlea. The wave sets the basilar membrane and tectorial membrane into motion, triggering signals that can be picked up by sensory cells and sent as nerve impulses to the brainstem and brain (Martin & Clark, 2003).

A disruption to any anatomical or physiological component of the process can result in a hearing loss. Hearing loss is usually divided into three types: conductive, sensorineural and mixed. A fourth category, central auditory processing disorders, refers to lesions in the central auditory nervous system. A conductive hearing loss occurs due to an interference with the transmission of sound vibrations through the ear canal, the middle ear, or both. The interference may be due to any pathologic condition involving either the outer ear (pinna), the ear canal, the tympanic membrane, or the middle ear structures. The most common pathologic conditions are fluid in the middle ear space, or abnormalities of the tympanic membrane or ossicular chain (Martin & Clark, 2003). A mixed hearing loss occurs when both a conductive and sensorineural hearing loss are present in the same ear. In a sensorineural hearing, the outer and middle ear systems function normally but the pathology of the cochlea or neural pathways creates a distortion of the sound. Treatment of a sensorineural hearing loss usually involves the use of an amplification system, or for severe-profound hearing losses, a cochlear implant (Kidd, 2002).

The type of hearing loss and the degree of its severity is determined by a hearing test. The most commonly used hearing test is pure tone audiometry. It measures the hearing threshold across the frequency range of speech, usually 250 Hz to 8000 Hz. A hearing threshold, measured in decibels (dB), is defined as "the lowest (sound level) at which the listener can identify the presence of the signal at least 50% of the time" (Harrell 2002, p .71). Each threshold is noted on an audiogram (Figure 1). The normal range of hearing is controversial, but generally considered to be 15 dB. The severity of a hearing loss is described as mild, moderate, severe or profound. Table 2 shows the commonly used scale for classification of degree of the hearing loss.

Table 2. Recommended scale for classification of degree of hearing loss (Goodman, 1965).

dB HL	Degree of Hearing Loss
-10 to 15	Normal hearing
16 to 25	Slight hearing loss
26 to 40	Mild hearing loss
41 to 55	Moderate hearing loss
56 to 70	Moderately-severe hearing loss
71 to 90	Severe hearing loss
> 90	Profound hearing loss

An important section of the audiogram that has significant implications for hearing impaired children developing speech is the range known as the 'speech banana'. The speech banana, as shown in Figure 1, is the frequency range of speech sounds in an audiogram and is thus an important area for determining a hearing impaired child's ability to perceive different speech sounds.

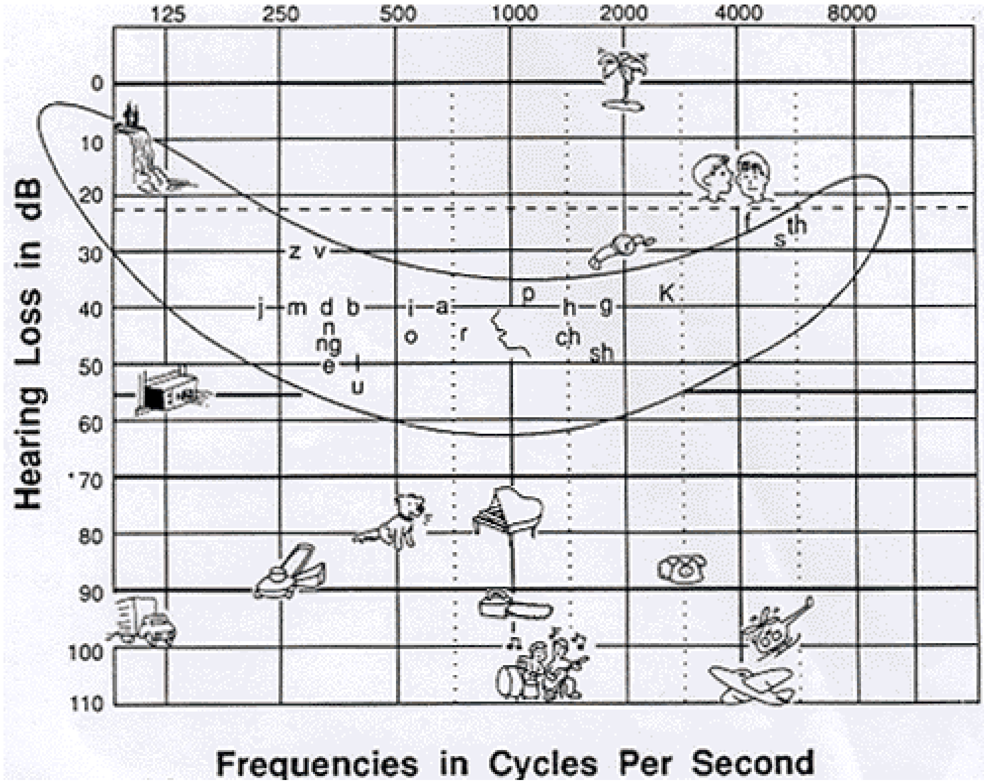

Figure 1. Example of an audiogram showing the speech banana.

1.2.4 Effect of Hearing Loss on Speech Perception

The time from birth to five years is regarded as a critical period for the acquisition of speech and language (Culbertson & Kricos, 2002). According to Moore, Perazzo and Braun (1995), a fetus hears speech at approximately 28 weeks gestation and by six months of age a child has learnt to discriminate the sounds of his/her own language (Downs and Yoshinaga-Itano, 1999). The presence of hearing loss during this period alters a child's ability to perceive the different speech sounds which are necessary for auditory feedback (Carney & Moeller 1998). Impaired hearing limits the perception of normal intensity and frequency ranges of the speech signal, resulting in difficulty discriminating between speech sounds (Culbertson & Kricos, 2002).

Auditory feedback is important because it allows a child to compare his/her own productions with correct models. Stoel-Gammon and Otomo (1986) showed a link between the random articulatory movements of infant babbling and the ability to self-monitor through auditory feedback. They found that a baby's hearing loss compromised its ability to hear its own vocalizations and limited its expansion of new babbling sounds. These studies provide evidence for a link between speech perception and speech production.

1.2.5 Speech Errors of the Hearing Impaired

The speech production of hearing-impaired children varies considerably and has been studied extensively over many years. One of the most cited studies is by Hudgins and Numbers (1942), who identified the main speech production errors in hearing-impaired children as deletion of initial and final consonants, consonant cluster errors, voicing and nasality errors, consonant substitutions, and vowel distortions. Many similar studies agreed on the complex and varied nature of speech errors present in hearing-impaired speech (Smith, 1982; Dunn & Newton, 1986; Culbertson & Kricos, 2002). There is a well-documented relationship between the severity of hearing loss and intelligibility of speech (Boothroyd, 1984; Smith, 1982; Perkell et al., 2000). Boothroyd (1978) studied the auditory perception and speech production of 122 primary-school aged hearing impaired children with sensorineural hearing loss. The children all communicated verbally and had a pure-tone average (PTA = average of hearing thresholds at 500, 1000 and 2000 Hz) in the range of 55-123 dB HL. They wore hearing aids during the testing. Patterns were identified between PTA levels and speech perception and production errors. Those with a PTA of 75 dB HL confused consonant place; PTA of 85 dB HL confused initial consonant voicing; PTA of 90 dB HL had difficulty with consonant continuance; PTA of 100 dB HL had difficulty with vowel place; PTA of 105 dB HL and above had difficulty with the suprasegmental aspects of speech such as syllabic pattern and intonation. Results indicated that the development of speech production could, to some extent, be predicted by the child's degree of hearing loss. Figure 2 illustrates the relationship between the intelligibility of a hearing impaired child's speech and their hearing threshold.

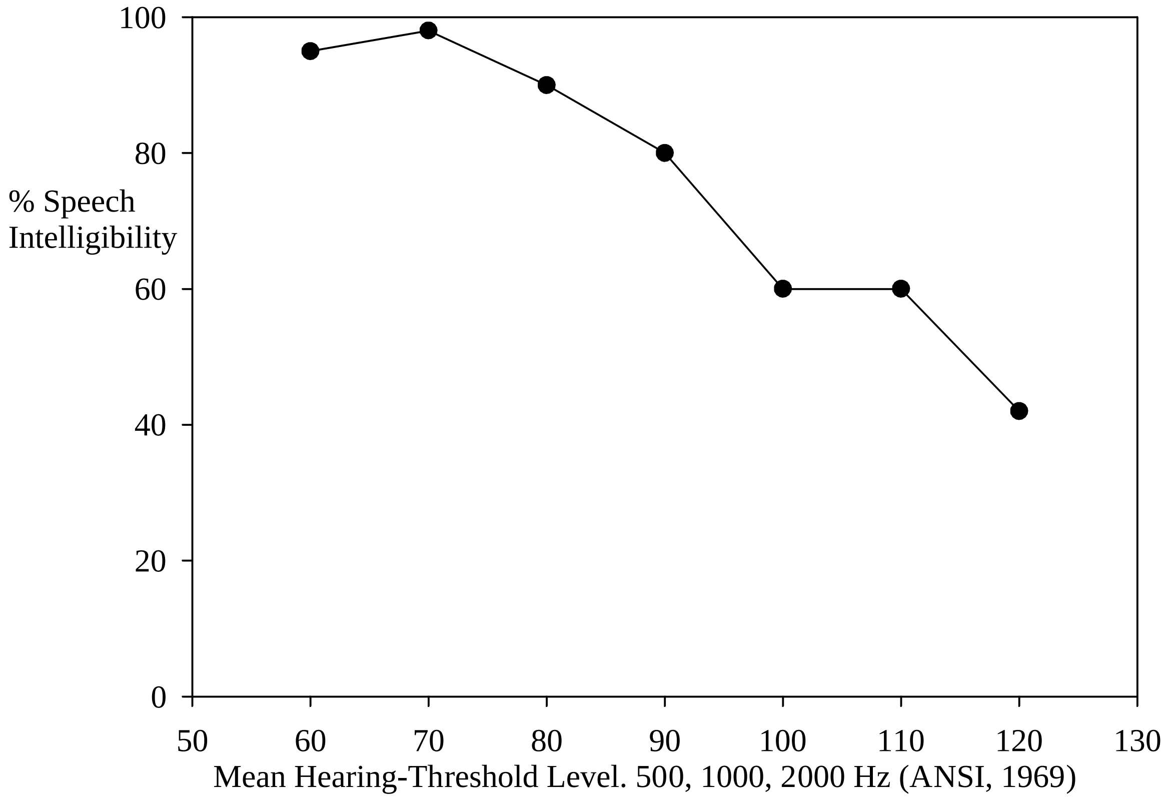

Figure 2. Acoustic speech intelligibility of 122 hearing-impaired children as a function of mean hearing level for 500, 1000 and 2000 Hz in the better ear (Boothroyd, 1984)

The relationship between hearing ability and speech intelligibility supports the acoustic theory of speech production (Kuhl, 1981; Stevens, 2002), which claims that the acoustic patterns of a speech signal are processed and organized into an internal map, which can be distorted if the acoustic patterns have not been adequately received during the input process. With a compromised input process, such as that associated with a hearing loss, the incorrect mapping will result in distorted or deleted speech sounds in speech production (Stevens, 2002). For instance, children with mild-moderate losses tend to develop intelligible speech, but have production errors mainly involving affricates, fricatives and blends (Elfenbein, Hardin-Jones & Davis, 1994). In contrast, children with severe to profound hearing losses have significantly reduced intelligibility due to difficulty with consonant, vowel and diphthong production, as well abnormal voice production (Culbertson & Kricos, 2002). Dunn and Newton (1986) described the speech of people with severe to profound hearing impairments as including both suprasegmental and segmental errors. Suprasegmental errors typically seen were slow speech rate, slow articulatory transitions, poor breath control, inappropriate stress patterns, and poor resonance. Segmental errors involved substitutions, omissions and distortions of vowels, consonants and diphthongs (Osberger & McGarr, as cite in Dodd, 1976). Smith (1982) summarized the speech of children with severe to profound hearing impairment as consisting of "stacks of errors which are complex and interrelated" (p. 27). The reason why children with severe-profound hearing losses, as compared with children with mild-moderate hearing impairment, have more complex errors and more reduced speech intelligibility is most likely related to the fact that their hearing ranges fail to include a greater portion of the speech spectrum. As a result of their impaired access to the speech spectrum, they may only develop limited spontaneous speech and must rely on the visibility of phonetic features to learn speech. Consequently, the speech that these children develop varies, but often shows unusual phonetic deviations due to the fact that their speech sound repertoire has been developed through vision and residual hearing (Dodd, 1976).

The majority of studies describing the speech production of hearing-impaired children have been confined to perceptual analysis of phonetic and phonologic errors. However, the development of technology in more recent years has enabled researchers to study the acoustic characteristics associated with speech production (Uchanski & Geers, 2003). The majority of the acoustic studies of the speech in the hearing-impaired population have focused on temporal parameters, such as consonant and vowel durations (Forrest & Elbert, 1990), as well as voice onset time (VOT) (Monsen, 1974). Hudgins and Numbers (1942) reported abnormally long durations for vowels and consonants in the hearing-impaired children. Uchanski & Geers (2003) also found that children with cochlear implants produced vowels and words of longer durations compared with normal-hearing children and suggested that the longer durations were related to vowel and consonant substitutions and distortions, commonly found in this population. Monsen (1974) found VOT to increase for both /t/ and /d/ in hearing-impaired children and suggested that this VOT lengthening was due to a lack of brief, timely execution in speech production. According to Monsen (1974), the lengthening of the VOT, an important acoustic feature for distinguishing between voiced and voiceless consonants, significantly affected the speech intelligibility scores of hearing-impaired children. These preliminary findings have shown that the temporal measurements of the acoustic signals are useful in describing the speech characteristics of the hearing-impaired children in relation to their underlying production difficulties.

While temporal parameters have received most of the attention in the literature, there have been few studies measuring the acoustic energy of speech sounds, especially consonants. Various frequency regions of the speech signal have been examined, with the majority of studies focussing on the formant characteristics of vowels produced by hearing-impaired speakers (Monsen, 1976; Angelocci, Kopp & Holbrook, 1964; Uchanski & Geers, 2003). The first two formant frequencies of vowels have been found to provide the greatest information on a vowel's phonetic quality and identity, with Formant 1 reflecting tongue height and Formant 2 reflecting the forward and backward movement of the tongue (Monsen, 1976). Angelocci et al. (1964), in comparing the frequency values of the first three formants of vowels between hearing-impaired and normal-hearing children, found vowels produced by the hearing-impaired children to be more centralised and concluded that the deaf did not have “clearly defined articulatory vowel target areas” (p.169) to allow them to achieve correct placement of the articulators in vowel production. In a similar study, Monsen (1976) also found a reduced vowel space for hearing-impaired children and suggested that the reduced vowel space was due to the hearing-impaired children’s inability to hear the frequency range where F2 frequency was normally situated. Uchanski & Geers (2003) found that cochlear implant users, who generally have better auditory access to high frequencies as compared to hearing aid users, produced a much higher F2 for /i/ than hearing aid users. These studies have provided valuable information on how the spectral characteristics of vowels could be related to the development of speech deficits in hearing-impaired children.

Significantly fewer studies have examined the spectral properties of consonants in the hearing-impaired population. Spectral moment analysis, a technique that quantifies the acoustic energy of consonants using statistical measures, has been developed to provide a description of the spectral concentration and energy range of the speech signal (Forrest, Weismer, Milenkovic & Dougall, 1988; Uchanski & Geers, 2003). Forrest and Elbert (1990) used spectral moment analysis to determine whether consonant errors of children with phonological disorders contained particular acoustic patterns and confirmed that a lack of difference on measures of spectral moments between /t/ and /d/ resulted in poor phonemic contrast between these two consonants. Uchanski and Geers (2003) used spectral moment analysis to examine the fricatives /s/ and /! / and found that children with normal hearing, as compared to children with cochlear implants, produced /s/ with a higher spectral moment 1 value (mean) and exhibited more negative spectral skew and more peaked spectra for /s/ than for /! /. Since spectral moment 1 was considered to reflect the degree of frontness of vocal tract constriction, Uchanski and Geers (2003) concluded that cochlear implant users who exhibited lower spectral moment 1 values for /s/ were not placing their tongue forward sufficiently to produce a correct /s/ sound. These spectral moment studies provide an interesting basis for further exploration of the spectral analysis of consonants in hearing-impaired children, to clarify the complex nature of their speech errors and help guide speech training.

1.2.6 Traditional Speech Training

Various approaches for speech training of children with hearing impairment have been used over the years. All have typically maximized the use of residual hearing through amplification, along with traditional hierarchical speech sound training using visual and tactile cuing, and aural training (Ling, 1976). These traditional training approaches rely on the clinician to provide a variety of stimuli and appropriate feedback (Ertmer, Stark & Karlan, 1996). In general, training has tended to follow Liberman et al.'s (1967) motor theory of speech perception, which claims that speech is perceived as a specific pattern of articulatory gestures which are represented in the brain as motor commands. Thus by correcting the placement of articulation we improve the child's perception of the speech sound. However, the effectiveness of these therapy approaches is often restricted due to the lack of visual cues available for many speech sounds produced within the oral cavity. Another problem is the delayed and subjective nature of the clinician's feedback, which may affect the child's ability to associate the clinician's feedback with correct articulation placement. Consequently, traditional speech training based on the motor theory of speech perception, often results in improved articulation for visible consonants but less improvement for consonants produced within the oral cavity (Ertmer et al., 1996).

An alternative to the motor theory of speech perception is the acoustic theory (Kuhl, 1981; Miller, 1977; Stevens, 2002). The acoustic theory claims that the speech signal is processed into a sequence of acoustic segments and patterns which form an internal map of speech sounds. Hence if the acoustic patterns are not analysed accurately, an incorrect map of speech is created which will then affect the production of those speech sounds (Stevens, 2002). For example, the acoustic theory would suggest that the incorrect production of high frequency speech sounds typically seen in children with high frequency hearing losses is due to their lack of access to the correct speech segments. This causes an incorrect map to be created and as a result, these speech sounds are substituted or deleted in their speech production. Speech training based on the acoustic theory of speech perception is now available with the development of technology that can provide immediate objective feedback.

1.2.7 Visual Feedback for the Hearing Impaired

The development of visual feedback devices is an attempt to compensate for the lack of auditory access to acoustic patterns of speech sounds. Visual feedback devices can record and display both acoustic and physiologic information to the child and provide them with an awareness of acoustic patterns that are unavailable through their hearing. By displaying this type of information the child can compare his/her own speech production with a correct visual template from the clinician, allowing immediate objective feedback for the child (Dagenais, Critz-Crosby, Fletcher & McCutheon, 1994). There has been a substantial increase in the development of these technological devices, however more research on their effectiveness is required before they are accepted by clinicians as an effective clinical tool.

Electropalatography (EPG) appears to be the most widely researched of these devices and is claimed to be a safe and convenient technique for use in the assessment, diagnosis and treatment of children and adults with articulation disorders (Dagenais et al., 1994). EPG uses electrodes on the tongue to record and provide real-time visual feedback of tongue-palate contacts. It can be used to monitor and physically practise new articulatory placements. Dagenais et al. (1994) compared traditional aural-oral therapy and electropalatography therapy for the development of consonants in adults with hearing impairment. They found dramatic advantages in the use of EPG for particular speech sounds, such as velar stops, and concluded that EPG provided a viable alternative to traditional training. Crawford (1995) examined the use of EPG to train correct velar production with two ten-year-old children with profound hearing loss. Previous traditional therapy for these targets had shown no improvement, however after two months of EPG training both children showed significant improvement in velar production. Similarly, Pantelemidou, Herman and Thomas (2003) examined the use of electropalatography to train velar consonants in children with cochlear implants. They found a significant improvement in production both immediately after the training period and five weeks after training had ceased, indicating maintenance of the learnt targets.

Visual feedback in speech training has also been reported via microcomputer-based applications, such as the IBM SpeechViewer. The SpeechViewer provides games which allow children to practice speech components including pitch, amplitude, duration and voicing. Following some initial training by the clinician, the child is able to independently practice his/her own speech production using the software (Shuster, Ruscello & Smith, 1992). Pratt, Heintelmann and Deming (1993) examined the effectiveness of the SpeechViewer in the treatment of vowels in six preschool children. They found the treatment to be at least partially effective with all of the preschoolers showing some treatment effect, however they experienced difficulties and inconsistencies with the SpeechViewer feedback system. The researchers recommended that the SpeechViewer be used as a supplementary treatment tool rather than as a primary means of improving speech production. Similarly, Ryalls, Le Dorze, Boulanger and Laroche (1995) used objective measures to compare treatment using the SpeechViewer with traditional therapy to lower the fundamental frequency (Fo) of two adolescent girls with hearing losses. Their results showed that the SpeechViewer was effective in lowering the Fo to a normal level in one participant, however the SpeechViewer produced a negative result by raising the Fo further in the second participant. Following these unexpected results, the researchers stressed the importance of using objective measures in research, as they otherwise would not have noticed the second participant's negative treatment effect. They concluded that visual feedback, such as the SpeechViewer can give good results when used to lower the Fo in some people with hearing impairment.

Another visual feedback tool which has become increasingly popular more recently is the spectrographic display of speech. Real-time spectrograms provide a visual representation of the frequency, intensity, and time domains of an acoustic signal (Ertmer & Maki, 2000). Unlike many other visual feedback devices that provide feedback on a single dimension of speech, spectrographic displays can provide many segmental and suprasegmental speech features simultaneously. The spectrographic displays are capable of providing visual contrasts between a correct model and an incorrect production (Ertmer et al., 1996). The immediate feedback generated by these devices provides an increased awareness of the relationship between speech movements and the acoustic consequences of these movements (Ertmer & Maki, 2000). Maki and Streff (1978) examined the effects of speech training with spectrograms for adults with severe-profound hearing loss. After 40 hours of individual training participants who received training with spectrograms made greater improvement for /i/ and /ai/ than those in traditional therapy, indicating that training with spectrograms was more effective than non-instrumental training methods. Stark (1971) investigated the usefulness of spectrograms in vowel production training in an 8-year old child with a hearing loss. After two weeks of intensive training with spectrograms the child produced several vowels and diphthongs in combination with consonants to form syllables that were perceived as words by trained listeners. Shuster et al., (1992) investigated the use of a spectrogram to train the correct production of [ɾ] in an adult with hearing impairment when previous traditional therapy had been ineffective. Within three sessions of using the spectrogram as an aid to traditional therapy, the participant was able to produce the sound correctly. In fact the participant reported that "having the visual feedback enabled [him] to find the correct position for [his] tongue in a way that verbal description of positioning by the clinician could not" (p. 32). In a later study by Ertmer et al. (1996), spectrographic displays were used to train vowel production in two children with profound hearing losses. Both participants used the spectrographic displays to judge their own productions during the training sessions. One participant showed some improvement in one target, while the other showed significant improvement in all targets. Generalisation of practiced vowel targets to untrained words was also observed in both participants. The authors suggested that the limited success of one participant may have been due to her strong bias toward self-judging her productions as "correct". They therefore emphasized the importance of "frequent instruction, clarification, and ongoing practice to facilitate self-evaluation" (p .14). In a more recent study, Ertmer and Maki (2000) compared traditional therapy and therapy using a spectrographic display for four teenage children with hearing impairment and found improvements in consonant production in a relatively short time for both therapy methods. Only one participant had significant differences in improvement between approaches at the end of training, indicating that spectrographic training was more successful for her than traditional training. Greater results were found in the generalization scores. Two participants showed considerably higher generalization scores for their spectrographic-trained targets compared to their traditionally trained targets, indicating that the spectrographic displays facilitated greater generalization to untrained targets than the traditional training approach. The authors concluded that spectrographic displays can be at least as effective as traditional therapy for establishing certain consonants. However they cautioned that the success of the approach likely depends on the distinctive spectrographic features of the chosen speech targets as well as the children's ability to understand the visual representations of correct and incorrect productions. They suggested that spectrographic training may be indicated for training speech sounds that are relatively invisible, or when teaching different manners of production.

1.2.8 Summary

As detailed above, there have been numerous studies on the characteristics of speech production and alternative speech training methods for children with hearing impairment. The speech production of this population is very complex, therefore further research is needed so that we may better understand the characteristics of their speech and thus provide more effective aural habilitation and speech training. The following two studies described in this thesis aim to add to add to this understanding and provide a foundation for further research in the area.

2 Study One

The purpose of this study was to identify the acoustic features characterizing the speech deficits of a hearing-impaired child and examine how the acoustic properties were related to the perceived accuracy of the child's speech production. The hypothesis was that certain acoustic features would be useful in differentiating between speech sounds perceived to be "correctly produced" and those perceived to be "incorrectly produced".

2.1 Method

A perceptual and acoustic analysis was performed on one hearing impaired child's speech productions as sampled from the acoustic recordings made during repeated administrations of a standardized articulation test over the course of a three-and-a-half-month experimental period.

2.1.1 Participant

Based on a convenience sampling method, a 12-year old female sibling of a student attending the university in which the study took place was recruited and an informed consent form signed by the child's legal guardian was obtained. This child will be referred to as Participant A. Participant A reportedly had a bilateral severe sloping to profound sensorineural hearing loss caused by ototoxicity from age three years. Hearing testing prior to the study using pure-tone audiometry revealed a pure-tone average (PTA; an average of the hearing thresholds for pure tones at 1, 2 and 4 kHz) of 110 dB HL in both ears. Based on the experimenter's assessment, Participant A's speech was minimally intelligible to an unfamiliar listener and was characterised by numerous vowel and consonant errors. Participant A had previously received several years of speech and language therapy, which had ceased approximately one year prior to the study. She attended a school for the deaf that used total communication (New Zealand sign language with verbal speech). Participant A showed no other sign of physical or learning handicaps according to the report of her main caregivers and the observation of the experimenters. Participant A wore bilateral hearing aids full-time and throughout the recording sessions.

2.1.2 Participant's Task

To yield data for the purpose of Study One, the participant was asked to name the pictures or read the word list included in the Goldman-Fristoe Test of Articulation (see Appendix 1). For pictures the participant was unable to name, she was prompted by the experimenter or asked to repeat the word after the experimenter.

2.1.3 Instrumentation

For signal recording, a simultaneous acoustic and electroglottographic (EGG) recording system was set up. Descriptions regarding the EGG recording device, which is non-invasive involving only two small electrode plates placed over the participant's neck, will be omitted because the EGG data will only be used in an area beyond the scope of this thesis and reported in a separate project. The acoustic recording system included a headset microphone (AKG C420, Austria) and a mixer (Eurorack MX602A, Behringer) for amplification of the microphone signal. The output of the microphone amplifier was connected to one of the eight channels of a 12-bit A/D converter (National Instrument DAQCard-AI-16E-4, USA) via a SCB-68 68-pin shielded connector box equipped with a low-passed filter (cutoff frequency = 20 KHz). The A/D converter was housed by a laptop computer (Compaq 650 MHz Pentium 4, Taiwan) for direct digitization. The sampling rate was set at 44.1 kHz. A locally developed algorithm written in Matlab was used for signal digitization. For signal playback and measurement, a time-frequency analysis software (TF32; copyright: Paul Milenkovic, 2000) was used to derive temporal measures and the PRAAT software (copyright: 2005 Paul Boersma and David Weenink) to perform spectral analysis.

2.1.4 Procedures

The participant was seated in a quiet room. During an acoustic recording session, the headphone microphone was positioned at a distance of approximately 5 cm off-axis from the participant's mouth. The participant was asked to perform the participant's task while the recording system was in place. Over the full length of the experimental period, a total of five acoustic recordings for administration of the Goldman-Fristoe Test were obtained at different points of time (i.e., initial day, 5^{th} week immediately after a 20-minute training session, 6^{th} week immediately before and after a training session, and 15^{th} week immediately after a training session). Details regarding the training sessions will be described in Study Two.

2.1.5 Measurement

The dependent variables measured from the recorded speech samples included the segmental length of each of the consonants and the corner vowels /i, a, u/, the frequency of the first two formants for each of the corner vowels, calculation of the vowel space, and four measures from spectral moment analysis for each of the plosives, fricatives, affricates, and nasals.

<u>Segmental Lengths</u>. Three temporal properties of the acoustic signals were measured, including vowel duration for each of the three corner vowels /i/, /a/, and /u/ which appear in the

first syllable of a word, voice onset time for voiced and voiceless plosives in the word-initial position, and consonant duration for fricatives, affricates and nasals in the word-initial position. Voice Onset Time (VOT) is defined as the length of time between the release of a plosive burst and the onset of periodic vocal fold vibration for a following vowel (Monsen, 1974).

Formant Frequencies and Vowel Space. The spectral properties of the vowels were represented through measures of Formant 1 (F1) and Formant 2 (F2) frequencies in the three corner vowels, /i/, /a/, and /u/. These three vowels were chosen because they represented the extreme articulatory vowel positions and thus yielded the extreme values for F1 and F2 (Liberman et al., 1967). The vowel space, a measure of the amount of space delimited by the three corner vowels on a Formant 1-Formant 2 plot, was also extracted from the average values for each of the three corner vowels. The vowel space was calculated because it provided an indication of the adequacy of vowel articulation and differentiation (Monsen, 1976), with a smaller vowel space indicating a closer proximity of vowels and less between-vowel differentiation. The formula used to calculate the acoustic vowel space, as described by Liu, Tsao and Kuhl (2005), was:

$$\text{Vowel triangle area } (Hz^2) = ABS\{[F1i*(F2a-F2u) + F1a*(F2u-F2i) + F1u*(F2i-F2a)]/2\},$$

where “ABS” is absolute value, “F1i” symbolizes the F1 value of vowel /i/, “F2a” symbolizes the F2 value of vowel /a/, etc.

Spectral Moments. The spectral properties of the consonants were obtained through Spectral Moments Analysis, which calculates the location of the central frequency and the distribution of energy of a speech signal. The first moment (M1) is the mean, representing the centre of gravity or concentration of the spectrum. The second moment (M2) is the standard deviation, representing the spread of energy. The third moment (M3) is the skewness, representing the asymmetry or spectral tilt of the energy distribution. A symmetrical distribution is indicted by a skewness of zero. A negatively skewed distribution indicates greater energy in the higher frequencies, while a positive skewed distribution indicates a concentration of energy in the lower frequencies. The forth moment (M4) is the kurtosis, representing the degree of peakedness or the extent the frequency distribution departs from the shape of a normal distribution. Positive kurtosis values indicate a high peakedness, reflecting a clearly defined spectrum, while negative values indicate a flatter distribution lacking defined peaks (Jongman, Wayland & Wong, 2000). Measures of spectral moment were used in this study for their potential to provide information on articulatory placement and the position of vocal tract constriction. For example, when distinguishing between /s/ and /! /, a reduced M1 value for /s/, which normally has a high M1 value due to a forward articulatory movement and a narrow vocal tract constriction, will make it less distinctive from /! / (Tjaden & Turner, 1997).

2.1.6 Data Analysis

Speech samples taken from the five Goldman-Fristoe recordings were submitted to perceptual and acoustic analyses.

Perceptual analyses. The TF32 software was used to play back the acoustic signals through a computer sound card. The experimenter listened to each item of the recorded words and conducted a phonemic transcription task. From the phonemic transcription, the speech clinician tabulated the numbers of vowel and consonant errors and the types of error processes.

Acoustic analyses. For temporal analysis, the acoustic signal was displayed in the form of both time waveform and spectrogram using the TF32 software. The experimenter used the cursors on the computer to mark the time of onset and finish of a single sound. One cursor was placed at the onset of the selected segment and the other cursor was placed at the end of the segment. The values of the interval length displayed on the screen were recorded. Durations were measured regardless of any sound errors in the intended word. Vowel onset was determined by high-amplitude, periodic signals in the time waveform display. Based on the spectrographic display, plosive onset was determined by the presence of a burst and release, fricative and affricate onset by the beginning of high-frequency noise, and nasal onset by presence of the nasal murmur.

For spectral analysis, the Praat software was used for signal display, segment selection, and derivation of F1 and F2 measures for vowels and spectral moments for consonants. For vowels, the formant analysis was conducted on a 20-30ms steady segment centred at the chosen midpoint of the vowel to reduce the effect of any formant transitions from neighbouring consonants. For consonants, the Spectral Moment Analysis (Hanning Window, window size = 20ms) was performed on each of the selected consonant segments, which were restricted to those single consonants in word-initial position to control for potential position effect. The whole consonant was selected and analysed by placing the cursor at the onset and the end of the selected consonant, as described for duration measurements.

2.1.7 Statistical Analysis

Descriptive statistics were performed to summarize the numerical results of all acoustic measures. A series of correlation procedures were conducted to evaluate measurement reliability.

2.1.8 Reliability

Twenty percent of the data was reanalyzed for test-retest reliability. Results of a series of Pearson's Product Moment Correlation procedures revealed a relatively high measure-remeasure reliability for all acoustic measures, including vowel duration (r = 99.8%), consonant duration (97.6%), F1 (100%), F2 (100%), M1 (98.3%), M2 (95.2%), M3 (69.5%), and M4 (82.1%).

2.2 Results

This section describes results from perceptual and acoustic analyses of Participant A’s speech samples taken from the five recordings of the Goldman-Fristoe Test, detailing information regarding percent correct for each of the three corner vowels /i, a, u/ as well as 17 word-initial consonants.

2.2.1 Perceptual Analyses

The percentage of correct vowel productions was high for all corner vowels (/i/: 80%, /a/: 87%, /u/: 80%). Figure 1 shows the percentage of correct production for each consonant in word-initial position, excluding those in consonant clusters.

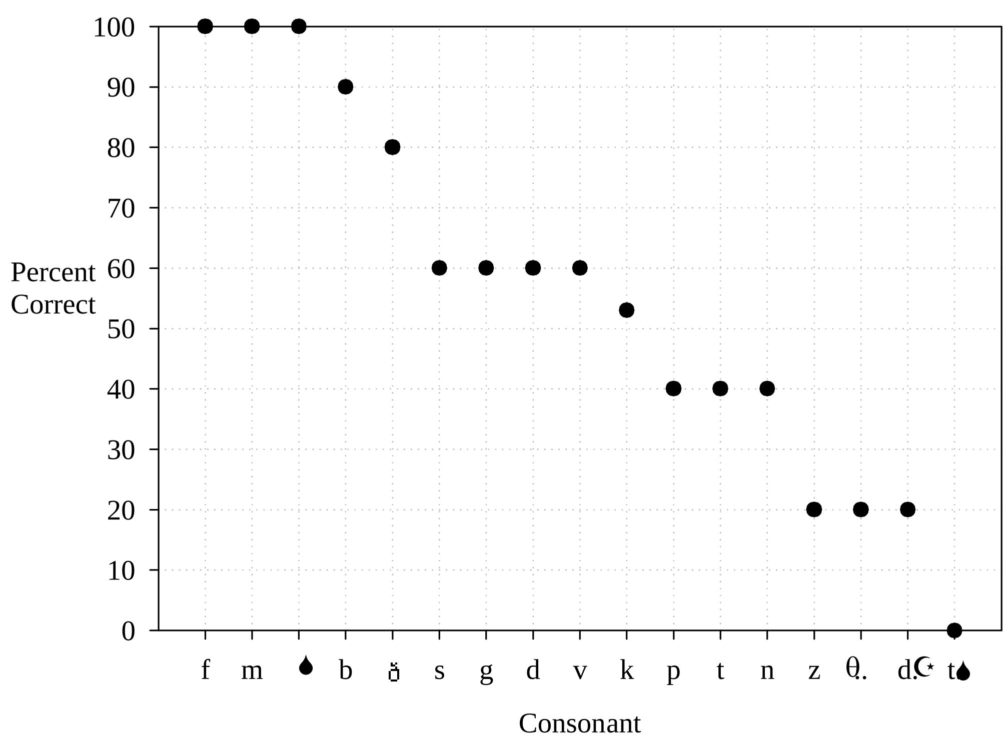

Figure 3. Percentage of total consonants correct in word-initial position for all recordings.

Inspection of Figure 3 revealed that consonants with labial place of articulation (/f/, /m/, /b/) and those involving noise energy covering a broad range of frequencies (/! /) had the highest rate of correct production. Affricates and less visible consonants tended to have lower rates of correct production. Table 3 shows the occurrence rates of the error processes. The most frequently occurring error was consonant cluster reduction, followed by /! / substitution for /s/, /t! / and /d! /. The error processes identified in this study were typical of a child with a severe-profound hearing impairment (Hudgins & Numbers, 1942; Smith, 1982).

Table 3. Percentage of total occurrence of error processes for all recordings.

Error process	No. of opportunities	% occurrence
Consonant cluster reduction	60	62
/! / substitution	105	48
Diphthong errors	35	37
Velar fronting	85	25
Deletion of final consonant	190	21
Weak syllable deletion	120	12
Stopping	210	8

2.2.2 Segmental Lengths

Table 4 shows the average vowel durations for correct and incorrect vowel productions respectively. Inspection of Table 4 revealed that vowel durations were lengthened for all three corner vowels when produced incorrectly.

Table 4. Descriptive statistics for vowel duration and the F1-F2 frequency difference for correct and incorrect vowel productions.

	Duration (in milliseconds)						F2 - F1 difference (in Hz)					
	Correct			Incorrect			Correct			Incorrect		
	n	mean	SD	n	mean	SD	n	mean	SD	n	mean	SD
/a/	5	392	67	2	583	--	5	704	144.1	2	1031	--
/i/	3	143	40	1	181	--	3	1767	347	1	1657	--
/u/	3	342	112	1	394	--	3	1339	677	1	830	--

Table 5 shows the average durations for each consonant in correct and incorrect consonant productions in comparison with norms reported by Umeda (1977). Out of 10 available paired comparisons between correct and incorrect consonant productions, 5 (/t/, /d/, /g/, /v/, /n/) showed longer consonant length while 3 (/k/, /s/, and /z/) showed shorter consonant length in incorrect consonant productions as compared with correct consonant productions. It appeared that the direction of change in consonant length for incorrect consonant productions varied by consonant. The lengths of consonants perceived as correct were not consistent with the normal values as reported in the literature.

Table 5. Duration (in milliseconds) for correct and incorrect production of each consonant in word-initial position in comparison with the normal speaker's data documented in Umeda (1977).

	Correct			Incorrect			
	n	mean	SD	n	mean	SD	Umeda (1977)
/p/	4	40	7	5	36	27	89
/t/	2	17	--	3	117	13	77
/k/	5	146	33	5	118	25	69
/b/	5	15	4	1	16	--	90
/d/	3	23	3	2	188	--	83
/g/	3	37	13	2	41	--	69
/s/	3	139	51	5	129	72	129
/B/	5	122	29	0	--	--	118
/z/	1	130	--	4	85	75	85
/f/	5	211	109	0	--	--	122
/v/	3	41	7	2	68	--	78
/! /	0	--	--	5	17	3	119
/ ð /	5	66	65	0	--	--	--
/t! /	0	--	--	5	46	168	--
/d! /	1	49	--	4	50	58	--
/m/	5	69	18	0	--	--	86
/n/	2	89	--	3	100	35	71

2.2.3 Formant Frequencies and Vowel Space

The average F1 and F2 values of the three corner vowels, /i/, /a/, and /u/, are plotted in Figure 4 for correct and incorrect vowel productions separately.

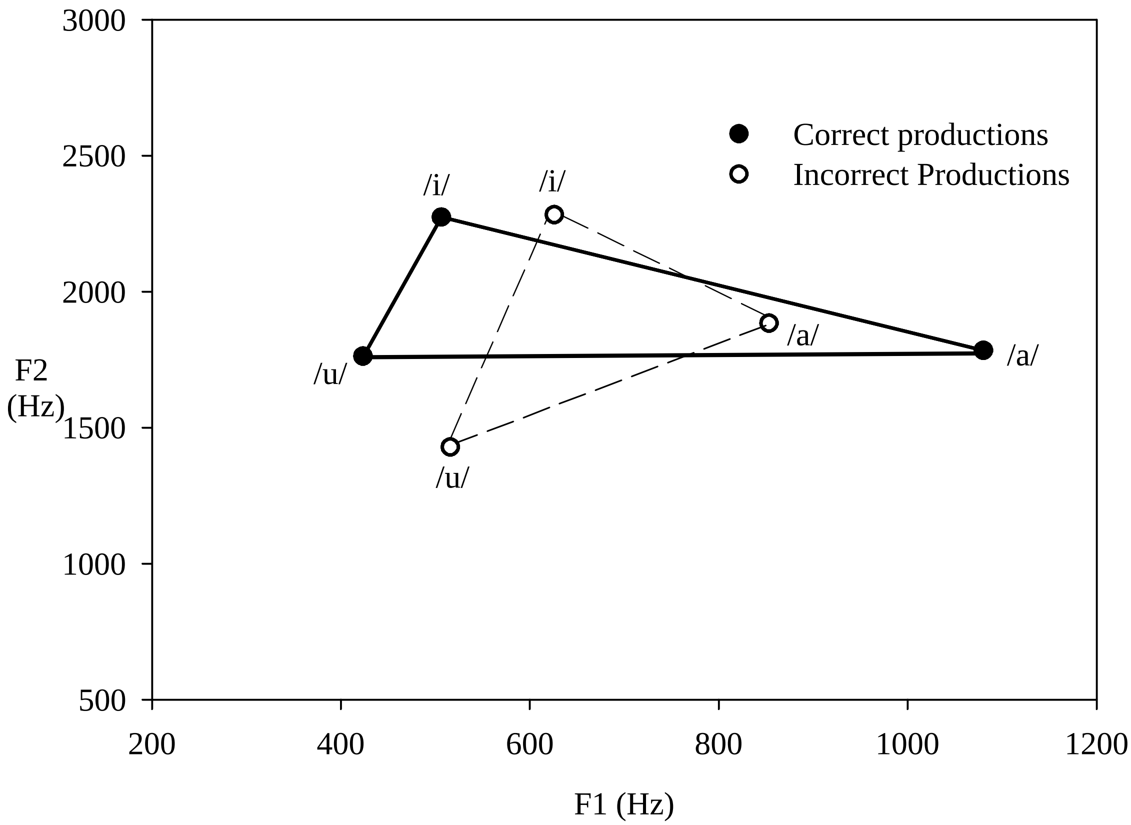

Figure 4. F1-F2 plot (vowel space) for correct and incorrect productions of /i/, /a/, and /u/.

As shown in Figure 4, the F1 frequencies in incorrect vowel productions were higher for /i/ and /u/ and lower for /a/ as compared with correct productions. Additionally, the values of F1 frequencies were closer together for incorrect productions compared to correct, indicating a lower frequency range and therefore less distinctiveness between vowels in incorrect vowel productions. The F2 values were similar between correct and incorrect vowel productions except that F2 for /u/ was unusually lower for incorrect productions. As shown in Table 4, the F1-F2 difference was smaller for incorrect vowel productions compared to correct vowel productions for high vowels, /i/ and /u/. Overall, calculation of the vowel working space area encompassing /i/, /a/, and /u/ showed a smaller working space area for incorrect productions (ABS=125738) than for correct productions (ABS=194237). This result agreed with Angelocci et al.'s (1964) comparison between hearing-impaired and normal-hearing speakers (see Figure 5) in that the vowel space derived from normal data was larger than that from the abnormal comparison groups.

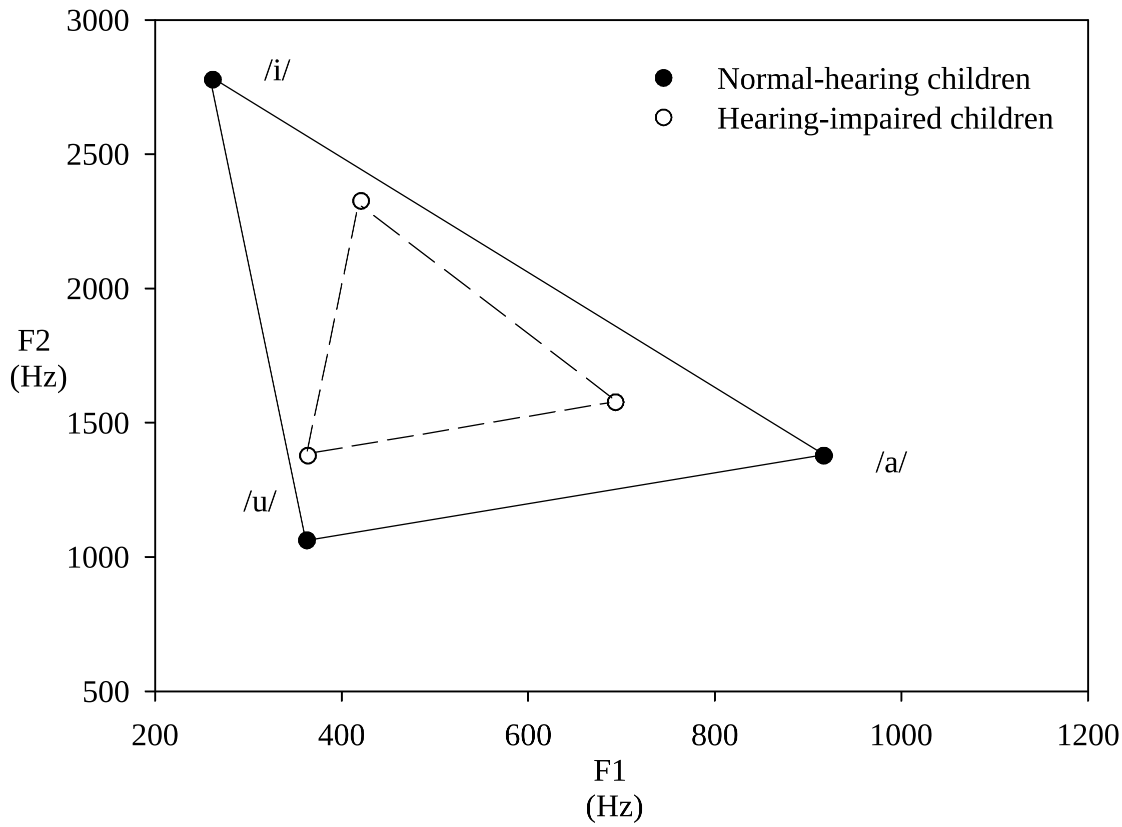

Figure 5. F1-F2 plot (vowel space) for /i/, /a/, and /u/ as produced by normal-hearing and hearing-impaired children (Angelocci et al., 1964, p.160).

2.2.4 Spectral Moment Analysis

<u>Moments 1 (Mean) and 2 (Standard Deviation)</u>. Figure 6 shows the M1-M2 plot for the average of correct and incorrect productions for each consonant.

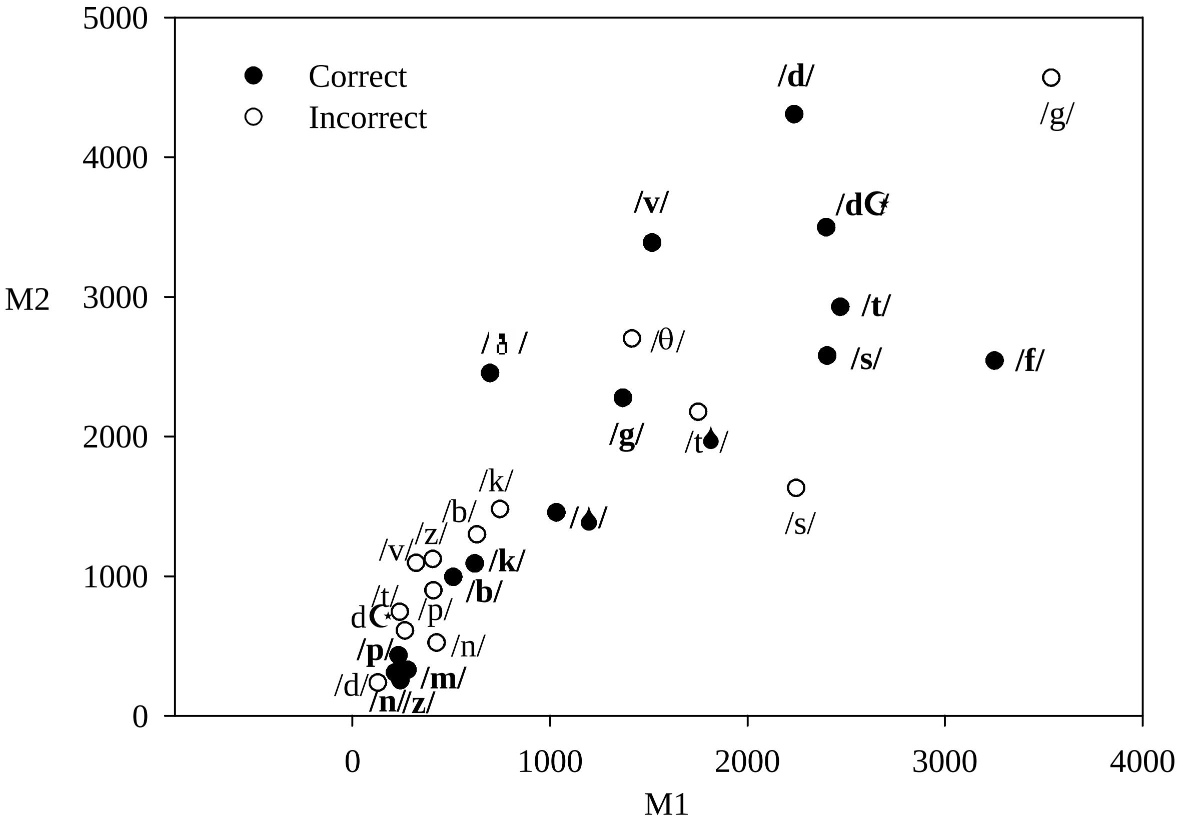

Figure 6. Average Moment 1 (Mean) and Moment 2 (Standard deviation) values for correct and incorrect consonant productions.

Inspection of Figure 6 revealed that most incorrect consonant productions consistently exhibited lower M1 values than correct consonant productions, which covered a greater frequency range. In addition, all fricatives had M1 values lower than those found in normal hearing speakers, as reported in Fry (1979) (see Table 6) for both correct and incorrect productions. However, those consonants that had a higher percentage of correct production had M1 values closer to Fry's (1979) values compared to those with low accuracy. For example, /f/ which was consistently perceived as correct had a mean of 3251 Hz, while /! / which was perceived correctly only 20% of the time had a mean value of only 17 Hz. A similar pattern was seen for plosives, although

M1 values were more variable. For example, /b/, which had a high percentage of correctly perceived productions, had a mean within the expected normal limits, while /t/, which was only perceived correctly 40% of the time, had a mean well below that expected for the normal range.

Table 6. Acoustic Characteristics of English Consonants (Values of the frequency range were taken from Fry, 1979)

Consonant	Place of Articulation	Manner of Articulation	Frequency range
/p/ and /b/	Bilabial	Plosive	600-800 Hz
/t/ and /d/	Alveolar	Plosive	4000 Hz
/k/ and /g/	Velar	Plosive	1800 – 2000 Hz
/f/ and /v/	Labio-dental	Fricative	6000-8000 Hz
/! / and /! /	Inter-dental	Fricative	6000-8000 Hz
/s/ and /z/	Alveolar	Fricative	4000-8000 Hz
/! /	Palato-alveolar	Fricative	1800-6000 Hz
/t! / and /d! /	Palato-alveolar	Affricate	1800-2000 Hz
/m/	Bilabial	Nasal	800-2000 Hz
/n/	Alveolar	Nasal	800-2000 Hz

Regarding the spectral differences between /s/ and /! / often reported in the literature (Forrest et al., 1988; Jongman et al., 2000; Matthies, Svirsky, Lane & Perkell et al., 1994; Uchanski & Geers, 2003), the /s/ was found in this study to have a higher M1 than /! / in the correct consonant productions as would be expected in normal speech, while both M1 and M2 values for the incorrectly produced /s/ were lowered, making it more similar to /! /. Inspection of Figure 6 also revealed that incorrectly produced fricatives exhibited lower M2 values and incorrectly produced plosives higher M2 values than those of their correct counterparts, indicating that incorrectly produced fricatives and plosives tended to deviate from a normal pattern.

Moment 3 (Skewness). The average M3 values for individual consonants correctly and incorrectly produced are shown in Table 7.

Table 7. Moment 3 (M3) and Moment 4 (M4) for correct and incorrect production of each consonant in word-initial position.

	Correct					Incorrect				
		M3		M4			M3		M4	
	n	Mean	SD	Mean	SD	n	Mean	SD	Mean	SD
/p/	4	28	10	1165	636	5	14	10	678	731
/t/	2	10	--	198	--	3	16	8	512	479
/k/	5	7	15	782	1038	5	14	10	408	425
/b/	5	15	11	1533	1804	1	5	--	35	--
/d/	3	2	1	3	3	2	28	--	1745	--
/g/	3	6	5	58	57	2	5	--	3	--
/s/	3	4	4	29	26	5	14	13	859	841
/B/	5	12	11	582	76	0	--	--	--	--
/z/	1	28	--	1317	--	4	14	10	326	345
/f/	5	26	48	3010	6175	0	--	--	--	--
/v/	3	4	3	25	28	2	17	--	477	--
/! /	0	--	--	--	--	5	7	7	101	186
/ ð /	5	5	1	25	7	0	--	--	--	--
/t! /	0	--	--	--	--	5	4	2	22	26
/d! /	1	2398	--	3	--	4	13	5	344.1	376
/m/	5	278	48	5982	8489	0	--	--	--	--
/n/	2	42	--	2934	2377	3	32	30	1854	2571

Many abnormally high positive values for M3 were present in both correct and incorrect productions, indicating a concentration of noise energy in the lower frequencies. Previous studies have reported M3 for /s/ to be more negatively skewed than /! / (Uchanski & Geers, 2003; Matthies et al., 1994). In the current study, correctly produced /s/ were indeed found to be more negatively skewed than /! /, however, for incorrect productions, both /s/ and /! / had similar M3 values.

Moment 4 (Kurtosis). Inspection of Table 7 did not reveal any consistent difference between correct and incorrect consonant productions on the measure of M3. While previous studies (Uchanski & Geers, 2003; Matthies et al., 1994) have documented a higher M4 for /s/ compared to /! / in hearing-impaired children, the correct /s/ productions were found in this study to have a lower M4 measure than the correct /! /.

2.3 Discussion

The hypothesis of Study One was that the use of objective acoustic measures would help identify acoustic characteristics which differentiate whether a speech production is perceived as correct or incorrect. A number of acoustic properties found to differentiate between a hearing-impaired child's correct and incorrect speech productions showed an error pattern supporting the acoustic theory of speech production. These acoustic patterns will be discussed, including the temporal properties and spectral characteristics of consonant and vowel productions.

2.3.1 Temporal Properties

Segmental lengths of both vowels and consonants were found in this study to differ for correct and incorrect phoneme productions.

Vowel Duration. Researchers (Monsen, 1974; Gulian, Hinds, Fallside & Keiller, 1983) have identified vowel prolongation as one of the speech characteristics of the hearing-impaired. Participant A in this study exhibited longer vowel durations for all incorrect productions of the three corner vowels, /i/, /a/, and /u/, while correct vowel productions had similar vowel durations to those documented from normal-hearing speakers. Since time duration, as Monsen (1974) suggested, is one of the more accessible speech parameters for hearing-impaired speakers, hearing-impaired speakers' tendency to prolong vowels might mostly be related to their greater reliance on vowel duration to compensate for deficiencies in achieving other phonemic contrasts.

Consonant Duration. Consonant duration was not found to be consistently prolonged in incorrect consonant productions as compared with either correct consonant productions or normal speaker's data. This greater variability in timing for consonants compared to vowels may be due to the more complex and diverse combinations of place and manner of articulations required for forming a phonemic contrast. For example, in terms of noise duration, an important cue in the differentiation between fricatives and affricates, correct productions by Participant A were found to be longer for fricatives compared to affricates.

In terms of the voiced-voiceless contrast, a commonly reported error in the speech of the hearing-impaired, (Monsen, 1974; Gulian et al., 1983), Participant A showed no consistent difference in VOT length between the three voiced-voiceless pairs (/p/-/b/, /t/-/d/, and /k/-/g/). The VOT length provides an important cue for the phonemic contrast between voiced plosives and their voiceless counterparts. The distinction requires fast movements of the articulators and good coordination of motor control between the larynx and upper articulators. As already reported, VOT of normal-hearing speakers has been reported to be 10 to 35ms for voiced plosives (/b/, /d/, /g/) and 35 to 100ms for voiceless plosives (/p/, /t/, /k/) (Gulian et al., 1983). In this study, VOT for correct productions of voiced and voiceless plosives were within these normal ranges. However, incorrect productions were more variable and the alveolar and velar plosives (i.e., /d/, /g/, /t/, and /k/) were prolonged outside these ranges. Follow-up physiological studies are needed to investigate the timing relationship required for correct consonant productions between the release of the oral constriction, the rate of change in laryngeal behaviours, and the regulation of airflow to allow for provision of visual instrumented feedback to enhance the hearing-impaired individual's control of these contrasting mechanisms.

2.3.2 Vowel Production

The F1 and F2 analysis of the corner vowels /i/, /a/ and /u/, revealed a relationship between formant values and perception of accuracy of the vowels. Values for correct and incorrect productions were similar to previously documented results for normal-hearing and hearing-impaired speakers, respectively (Angelocci et al., 1964). For the F1 measure, correct productions showed a greater range in F1 values across vowels compared to incorrect productions. Variation of F1 is associated with tongue height and thus the narrower range of F1 values for the incorrect productions suggests more limited jaw and tongue movement compared to correct productions (Monsen, 1976). Conversely, F2 measures were found to occupy a more limited range for correct productions compared to incorrect productions. However, this was likely due to the diphthongization of /u/, which led to a relatively low F2. The F2-F1 difference might provide a more straightforward explanation than F2 measure regarding factors contributing to incorrect vowel productions in this study. The F2-F1 difference varies mostly with the forward and backward movement of the tongue and thus a higher F2-F1 difference indicates a more anterior tongue position. As expected, the F2-F1 difference was found to be smaller for incorrect productions, suggesting restricted tongue movement and more posterior tongue positioning than for correct productions. This is not surprising because the second formant is difficult for hearing-impaired children to correctly perceive auditorily or visually since tongue movement for vowels is not visible (Monsen, 1976).

Calculations of the vowel space area revealed a reduced vowel space associated with incorrect productions. A reduced vowel space area represents a restriction of tongue elevation and front-back tongue movement. Previous research (Liu et al., 2005) has documented a reduced vowel space in the speech of cerebral palsied young adults and substantiated the relationship between speech intelligibility and range of vowel space with a perceptual study. It can be inferred from these results that misarticulation is generally associated with a smaller vowel working space, reflecting a more reduced articulation range or a more restricted range of tongue movements. The clinical implication of our finding was that expansion of vowel space could be potentially beneficial to improvement of speech intelligibility of the hearing-impaired, especially for those speakers whose limited range of articulatory movement mainly resulted from inadequate perceptual mapping rather than speech motor constraints.

2.3.3 Consonant Production

Consonants with the highest accuracy of production were found in Participant A to be those with a labial place of production as well as those with a more visible place of articulation. This is not surprising because children with severe hearing impairment are most likely to rely heavily on visible elements of speech production from early in life (Hudgins & Numbers, 1942). The present finding supports the theory that place of articulation is likely to be one of the earliest learnt and most developed speech behaviours (Smith, 1975). As for consonants with less visible place of articulation, those with concentration of noise energy mainly in the higher frequency range, mostly affricates, appeared to be most susceptible to misarticulation. According to Jongman et al. (2000), all four spectral moments could be related to place of articulation. In this study, the measures of all four moments were found useful in reflecting some patterns of misarticulation, with M1 providing the most interpretable information.

<u>Spectral Moment 1</u>. The M1 provides information about placement of articulation and is thus important for distinguishing between consonants with the same manner of articulation. For example, M1 values must be wide apart to differentiate between /p/, /t/ and /k/. In this study, plosives identified as “correct” were indeed found to exhibit M1 values spreading across a greater

frequency range and thus greater distinction between phonemes than those identified as "incorrect".

In addition, the M1 values of consonants, especially for fricatives, were found in this study to be lower than the published norms (Fry, 2001). Furthermore, a large proportion of the participant's incorrect consonant productions yielded M1 values much lower than those in her correct consonant productions. In particular, the average M1 value for Participant A's correct /s/ productions was found to be a higher frequency (2403 Hz) than that for her incorrect /s/ productions (2247 Hz), which was closer to the average M1 value for /! /. This finding supports the theory that the two fricative sounds /s/ and /! / are easily confused due to the limited auditory access to high frequency consonant information, leading to difficulties acquiring the necessary production scheme. Since a higher M1 value typically reflects a more anterior tongue positioning in forming a constriction in the oral cavity, the low M1 value exhibited by Participant A's incorrect /s/ productions and the comparatively low M1 values across consonants suggested that Participant A's tongue placement tended to be more posterior in incorrect consonant productions. Since the M1 measure appeared to be sensitive in differentiating correct and incorrect consonant productions, it could be used in clinical application to provide feedback in speech training and monitor progress.

<u>Spectral Moment 2</u>. The rest of the spectral moments were also useful, although less directly, for detecting a speech error and possibly for relating to a specific error process. For example, the finding that M2 values for all consonants were relatively high for both correct and incorrect productions indicates that the noise energy was generally spread over a large range. This results might be due to the imprecise articulation of the speaker as well as the fact that these recordings were taken over treatment period when the participant's articulation may have been undergoing changes. In addition, although M2 values were variable, some patterns between

correct and incorrect productions were worth noting. In particular, the M2 values for fricatives, which are expected to be high for normal-hearing speakers, were found to be lower for incorrect productions compared to correct productions. For example, the correct production of the affricate /d! / had a higher M2 value than the incorrect production of the affricate /t! /. Similarly, M2 values for plosives /p/, /b/, /k/ and /g/, which are expected to be low for normal-hearing speakers (Forrest, Weismer, Elbert & Dinnsen, 1994), were also found to be abnormally high for incorrect productions. These findings suggest that perception of correct consonant production was related to the spread of noise energy.

Spectral Moment 3. As for M3, relatively high values were found for both correct and incorrect productions, indicating a concentration of energy in the lower frequencies. This may be explained again by the acoustic theory, whereby Participant A's hearing loss has restricted her accessibility to higher frequencies of the speech spectrum and consequently her speech production was concentrated to the more audible lower frequency ranges. Correct productions of /s/ were more negatively skewed, similar to previously documented results of normal-hearing speakers. Conversely, incorrect productions of /s/ were more positively skewed and similar to M3 values for /! /.

Spectral Moment 4. Measures of kurtosis (M4) provided limited information to discriminate between correct and incorrect productions. The conflicting results in comparing findings of M4 values between studies may be related to methodological differences that need further investigation.

2.3.4 Summary

Overall, M1 and M3 appeared to provide the most information to distinguish placement of articulation and between correct and incorrect productions. These findings supported Tjaden and

Turner's (1997) suggestion that M1 and M3 are sensitive to similar articulatory behaviours. Some of the varied results obtained for spectral moment analysis may be due to the stage of development of many consonant sounds and the fact that recordings were taken over a treatment period. Previous studies (Kent, 1976; Forrest et al., 1994) have shown that as speech motor control is developing during early childhood, the spectral characteristics have high variability. Although Participant A in this study is at a developmental age in which she should have acquired all the consonants of speech, her severe-profound hearing loss has evidently resulted in delayed speech acquisition and thus the highly unstable productions as shown in younger speakers.

3 Study Two

The second study was conducted to evaluate, through both subjective and objective assessment, the effectiveness of using acoustic signals as visual biofeedback in the speech training of hearing-impaired children.

3.1 Method

A single-subject multiple baseline design was implemented across behaviours with replication across three participants. The design involved traditional non-visual treatment followed by visual treatment working on misarticulated sounds or error processes in a time-staggered fashion.

3.1.1 Participants

Three primary-school-aged children with hearing loss participated in this study. One of the participants was Participant A from Study One. For additional participant recruitment, a letter was sent to the Christchurch office of the Advisors of Deaf Children (AODC) detailing the study and providing information for parents of potential candidates. The AODC was asked to identify children suitable for the study and send the letter to their parents inviting them to contact the researcher if they were interested in having their child participate. The participants recruited from this process were two brothers aged seven and nine years, with hearing losses of unknown etiology. These two participants will be referred to as Participant B and Participant C respectively.

Participant B, aged 9, was reportedly diagnosed at five years of age with a profound sensorineural hearing loss in the right ear and a moderate sensorineural hearing loss in the left ear. Based on the hearing test conducted at the time of the study, Participant B's PTA was 100dB HL in the right ear and 60 dB HL in the left ear. Participant B, aged 7, wore a hearing aid in the right ear only. Participant C was reportedly diagnosed at three years of age with a bilateral severe sloping to profound sensorineural hearing loss. At the time of the study, Participant C's PTA was 80 dB HL in the right ear and 100 dB HL in the left ear. Participant C wore bilateral hearing aids.

Both Participant B and C had previously received speech-language therapy for two years, which had ceased approximately two years prior to the study. They both attended a mainstream school and used verbal communication. Based on the experimenter's perceptual assessment, their speech was intelligible most of the time, despite exhibiting neutralized vowels and misarticulated consonants such as frequent substitutions for the high-frequency fricative /s/. Like Participant A, both Participant B and C showed no physical or learning handicaps other than hearing impairment and wore their hearing aids during all experimental sessions.

3.1.2 Participant's Task

Assessment Sessions. All participants were asked to say the word list from the Goldman-Fristoe Test Articulation Test. Since Participants B and C showed high scores (% consonants correct Participant B = 87%; Participant C = 85%) on the Goldman-Fristoe Test, they were also asked to read in a different recording session the word list from the New Zealand Articulation Test (see Appendix 1). The NZ Articulation Test presents a higher level of challenge because it contains a greater number of multisyllabic words and consonant clusters. A probe word list (see Appendix 1), which contained 'tokens' (opportunities for production) of each target as well as the error processes used as controls, was also employed to elicit speech production from a participant immediately before and after a training session.

Training Sessions. During a training session, a participant was asked to say or repeat after the experimenter, who was a trained speech therapist, words on the training word list (see Appendix 2) with or without visual feedback. 'Traditional training' without visual feedback involved a traditional hierarchical training approach with the participant (Participant A only) receiving verbal instruction using simple visual and tactile cues on how to improve and monitor her speech. 'Visual training' involved the participant looking at a computer screen displaying real-time pitch and intensity changes using the VisiPitch. In 'visual training', initial orientation to the spectrographic display or intensity or pitch traces was given for the participant to practice identifying the visual structure of different speech segments as modelled by the experimenter. Models of target words were then displayed and the participants were required to match the models and judge the accuracy of their own production of the target speech segment embedded in words or sentences. The participant was reminded constantly to say the target sound in a way that its related computer display would achieve a close match with that modelled by the experimenter.

3.1.3 Instrumentation

Audio-recordings during all experimental phases were made using the same instrumentation as detailed in Study One. For speech training, a Visi-Pitch Model6087AT (Kay Elemetrics, USA) was used to provide a real-time display of the intensity and pitch trace of the speech signals. In addition, a laptop computer (NEC, Japan) installed with a time-frequency analysis software (TF32; copyright: Paul Milenkovic, 2000) was used to display the spectrograms of the recorded speech signals, with frequency (100 to 5000 Hz) on the vertical axis and time (with flexible length adjustment) on the horizontal axis. Intensity variations over a 30-dB range were reflected in the relative darkness of the trace.

3.1.4 Procedures

This section describes procedures used for sessions taking place during pre-treatment phase, treatment phase, and follow-up phase. All sessions were conducted in a quiet room by one of the researchers, a fully trained and experienced speech-language therapist.

Pre-treatment Phase. During the initial session, all participants were asked to perform the speech task using the word list from a standardized articulation test as mentioned above. Perceptual analyses were performed to identify error processes for training. Error processes were chosen as targets based on their high occurrence in the initial recordings as well as their spectrographic visibility for training.

Treatment Phase. Participant A participated in individual, half-hour training sessions, twice a week for three months with a break mid-way for a holiday period. Two error processes were targeted for treatment in subsequent sessions, scheduled one to two weeks apart. Participant B participated in individual, half-hour training sessions once a week for a six-week block. Participant C participated in individual, half-hour training sessions once a week for a two-week block. For each of the targets, Participant A was initially trained without visual feedback and then with visual feedback. Due to time constraints and the consideration that Participant B and C had previously received extensive traditional speech training, both Participant B and C received visual training only. Following the completion of each training session, the probe list was administered and recorded without any verbal or visual feedback.

Follow-up Phase. Following completion of the training phase, acoustic recordings for the word list from the Goldman-Fristoe Articulation Test (for Participant A) and the New Zealand

Articulation Test (for Participants B and C), as well as the probe lists (for all participants) were repeated without any feedback.

3.1.5 Measurement

The dependent variables measured from the recorded speech samples included the segmental length of each target consonant and consonant cluster, the frequency of the first two formants for each of the vowels /i, a, u/, calculation of the vowel space, and four measures from spectral moment analysis for each of the plosives, fricatives, affricates, and nasals.

3.1.6 Data Analysis

Perceptual Analyses. All baseline and training probes were assessed in a blinded pseudo-random order to avoid rater bias. The analysis method used to derive percentage correct scores for each English phoneme and identify error processes has been described in Study One. To obtain percentage correct scores for each target sound from the probe list recordings, the experimenter used the TF32 software to play back the digitized speech signals, listen to the signals, and identify the misarticulated segments. The scores were then plotted for multiple-baseline analysis. The effectiveness of each training method was determined by examining changes in trend, slope, level, and variability from baseline to treatment phases. A significant trend of speech improvement was defined as an increase in the slope of the line for the training accuracy data compared to that established by the baseline accuracy data. A level shift was demonstrated when there was an abrupt upward shift in accuracy following the introduction of training. Maintenance of improvement was evaluated by comparing scores from follow-up sessions to those from baseline sessions. Maintenance was demonstrated when all scores from the follow-up session exceeded early baseline scores. Generalisation was assessed by comparing scores from the probe list to determine whether the improvements achieved for trained words were transferred to untrained words without direct instruction. For Participant A, the split-middle median technique was employed to establish a trend line and identify the rate of change in target production accuracy during the visual training phase. For Participants B and C, due to the limited number of repeated recordings, only visual inspection of the results was used to detect trends.

Acoustic Analyses. Objective acoustic analyses were undertaken to identify the more subtle changes in acoustic characteristics of each participants' consonant and vowel productions. Comparisons were made between early baseline and follow-up sessions for all participants, as well as between traditional training sessions and visual training sessions for Participant A. Temporal analyses, including voice onset time and consonant and vowel durations, were performed in the same manner as described in Study One. Spectral analysis of vowels, including Formants 1 and 2, and vowel space, were also performed in the same manner as described in Study One.

3.1.7 Statistical Analysis

The VOT measures from the probe list recordings for each of the three participants were submitted separately to a one-way Analysis of Variance (ANOVA) test to compare accuracy scores for pre-and post-treatment phases as well as traditional and visual training approaches. This method was used because no systematic temporal relationship between the sampling points collected over time was found based on visual inspection of the time plots and thus independence between these sets of time-series data was assumed. Significance level was set at 0.05.

3.2 Results

This section describes results of the initial analyses used to identify error processes for speech training and results related to the speech changes of Participant A, B, and C respectively over the course of the experiment as assessed through perceptual and acoustic analyses.

3.2.1 Misarticulation Identified for Speech Training

For Participant A, seven error processes were identified based on the perceptual analysis of the standardized test administered in the initial session. These included deletion of final consonant (DFC), consonant cluster reduction (CCR), /! / substitution (i.e., substituting /s/ with /! /), weak syllable deletion, stopping (i.e., using plosives in place of fricatives), velar fronting, and diphthong errors. Two of these, DFC and CCR, were targeted during treatment whilst the remaining processes remained untreated and acted as controls. The error processes were chosen as targets because of their high occurrence rate as well as their spectrographic visibility for training purposes.

Based on the perceptual analysis of the two standardized articulation tests administered in the initial session, five error processes were identified for Participant B, including DFC, CCR, de-affrication (producing only the fricative component of an affricate), interdental /s/ and /z/, and nasalization. Four error processes were identified for Participant C, including stopping, DFC, dental /s/ and CCR. For both Participant B and C, DFC was targeted for treatment. This error process had a significantly higher occurrence in sentences than in single words for both Participant B and C, therefore the target process was trained in sentences and all probe and training lists consisted of sentences.

3.2.2 Perceptual Analyses

Scores for each participant were plotted individually for visual inspection of the trend and variability within each phase and the difference in trends between phases.

Participant A. Accuracy scores for both the Deletion of Final Consonant (DFC) and Consonant Cluster Reduction (CCR) targets in the Goldman-Fristoe and probe list recordings for Participant A can be seen in Figure 7. As shown in Figure 7, scores from the Goldman-Fristoe list failed to reveal any treatment effect. However, data from the probe list, which contained several more tokens, showed a clear trend of improvement in the production of both DFC and CCR targets during the visual training phase. In particular, the positive slope of the split-median line within the visual training period for the DFC data indicated a small improvement on the percent correct score over the training period. Figure 7 also showed that the visual training phase generally yielded a higher level of percentage correct scores than the one probe score obtained from the traditional training phase, suggesting that instruction with visual feedback might be more effective in improving target production compared to traditional training.

The percentage correct scores obtained from the untrained generalisation probes for the CCR target are also shown in Figure 7. No change in scores can be seen with traditional training. In contrast, the introduction of visual training resulted in immediate and sustained improvement on the percent correct scores. This finding suggests that instruction with visual feedback resulted in better transfer of learning to untrained consonant clusters compared to traditional training.

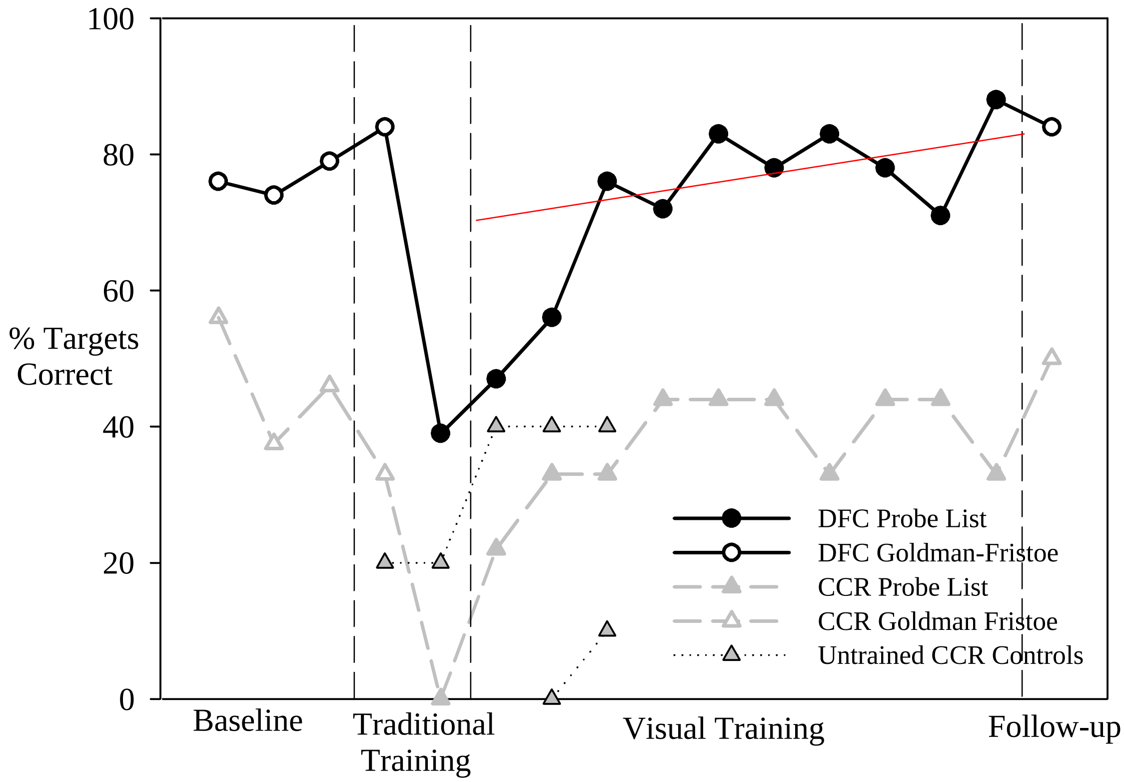

Figure 7. Percentage of targets correct for Participant A's deletion of final consonant (DFC) and consonant cluster deletion (CCR) targets over the training period. The red line is the split median line for the visual training period.

Participant B and C. Findings for Participants B and C, which can be seen in Figure 8, are not as clear as for Participant A. Results for both participants showed minimal change between pre and post-treatment recordings. This is likely due to the small number of tokens and the high percent correct scores that they both achieved in the baseline phase, which restricted the opportunity for improvement in subsequent recordings.

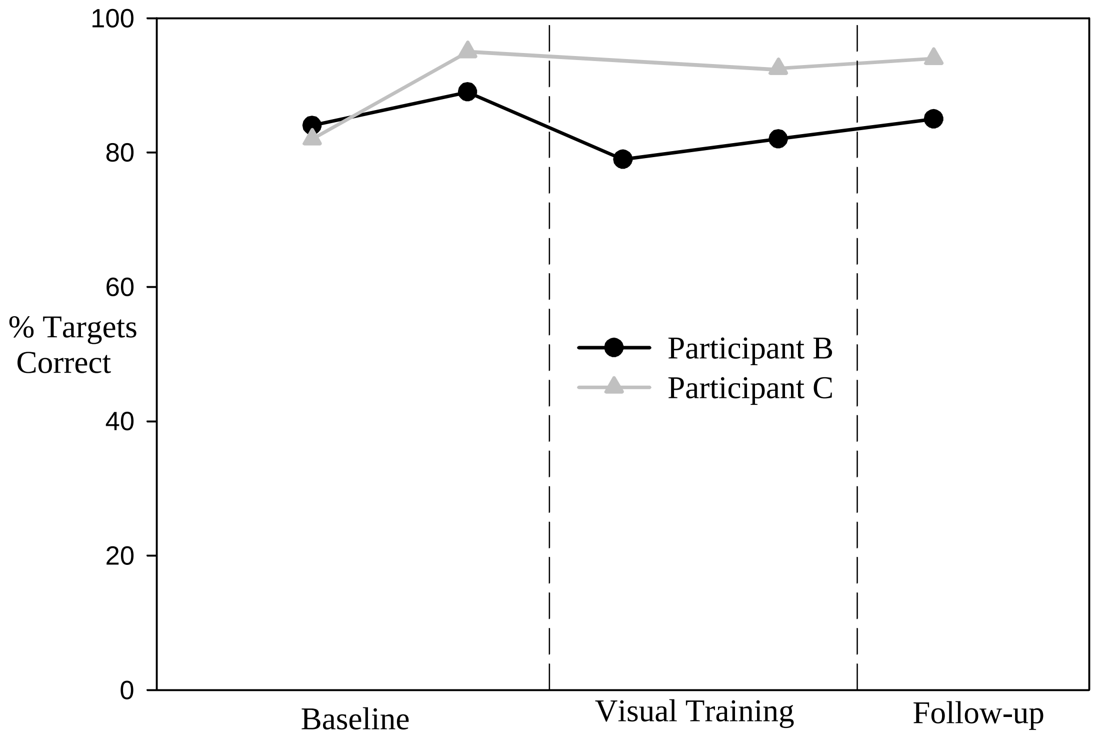

Figure 8. Percentage of targets correct for Deletion of Final Consonant (DFC) for Participants B and C over the training period

3.2.3 Voice Onset Time

Results from analysis of the VOT measures for all three participants suggested some reduction in VOT with the use of visual training.

Participant A. Changes in the VOT measure of voiceless plosives for Participant A can be seen in Figure 9. Although an ANOVA test revealed that changes in VOT across sessions were not statistically significant [$F(4, 10) = 0.432$, $p = 0.782$], visual inspection of Figure 9 revealed some trends. As shown in Figure 9, VOT measures for /p/ reduced with both traditional and visual training. However, the trend was variable within the training phases and the limited number of tokens restricted the ability to compare the two training approaches. The VOT measures for /t/ were highly variable, with no reliable trends, mostly due to an outlier present in the traditional training phase. The VOT measures for /k/ showed a reduction in VOT with the traditional training which continued with visual training.

In summary, although further studies with a longer observation period are required to make a comparison between training approaches, results obtained for Participant A indicate an overall reduction in VOT for voiceless plosive consonants, reflecting a positive shift toward normal plosive productions in terms of better voiced-voiceless distinction, with both training approaches in comparison with baseline recordings.

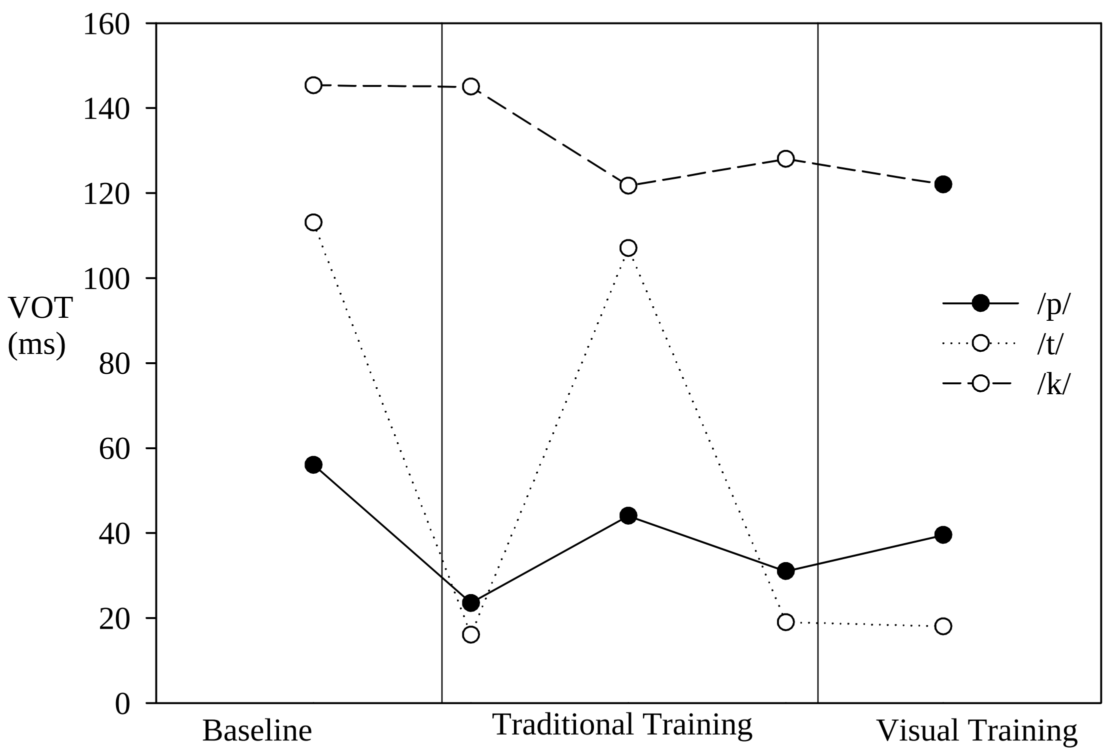

Figure 9. Change in Voice Onset Time (VOT) for Voiceless Plosives for Participant A over the Training Period

Participant B and C. The VOT measures observed over time for Participants B and C are shown in Figure 10. Results for Participant B were variable, and ANOVA scores failed to detect a statistically significant change [$F\ (4, 10) = 0.0366$, $p = 0.997$). However, based on visual inspection of Figure 10, a downward trend was evident following the introduction of visual training of /p/ and /t/. As for Participant C, although only a limited number of measures were obtained, a downward trend was also seen following visual training as compared to scores for baseline recordings.

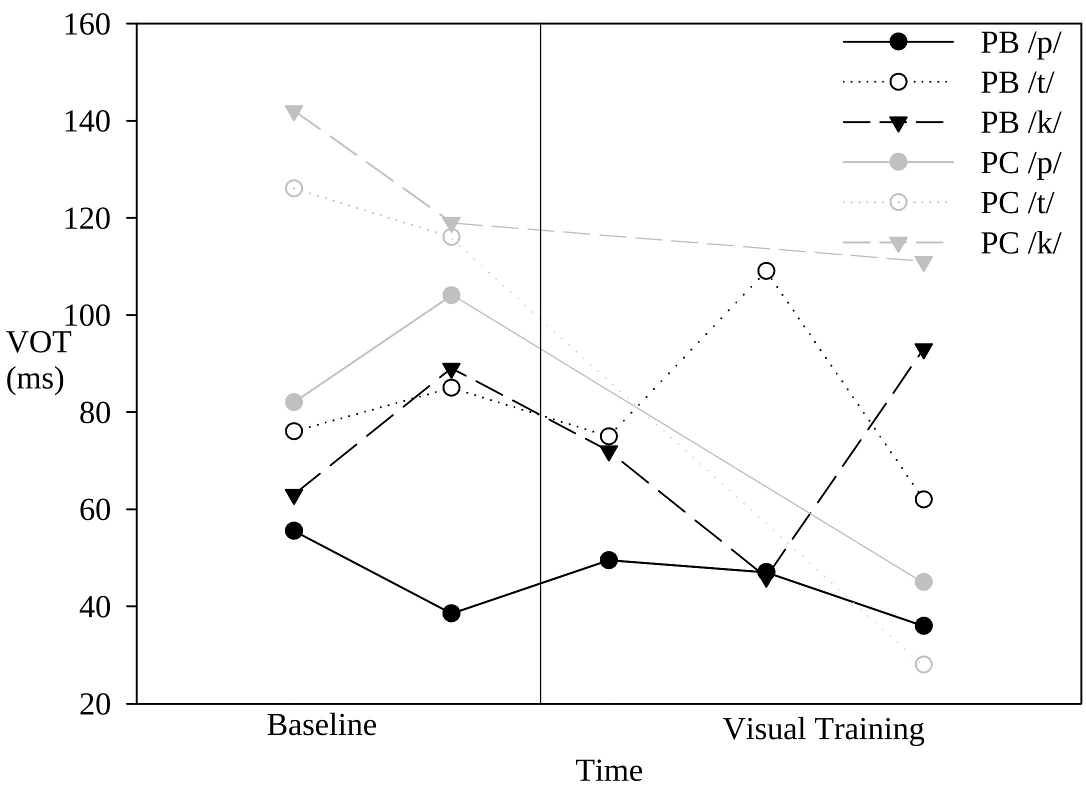

Figure 10. Change in Voice Onset Time (VOT) for Voiceless Plosives over the Training Period for Participants B and C (PB and PC)

3.2.4 Segmental Length

Length of the targeted consonant cluster /fl/ was measured for Participant A. An increase in consonant cluster length was presumed to be an acoustic indication of an improved awareness and production of the two consonant components of the cluster. The initial baseline measure was abnormally short at approximately 70 ms. Following the introduction of traditional training, an immediate increase in consonant cluster length could be seen for the trained /fl/ target, to within a more normal range of 158ms (see Table 8 for normal range). With the introduction of visual training, the length continued to be maintained within the high value range. A one-way ANOVA test revealed a statistically significant change in mean values over the training period [$F(8, 16) = 2.637$, $p = 0.047$]. These results suggest that both training approaches resulted in an increase in consonant cluster length, suggesting improved awareness and production of consonant clusters. Measures for the untrained control /pr/ were variable over the training period and abnormally low, suggesting no generalisation effect. This suggests that further training is required to facilitate generalisation to other consonant clusters. Temporal measures of Participant A's second target, final consonant deletion, were highly variable, however a trend of increasing length was suggested. ANOVA scores for final consonant length over the treatment period were not statistically significant [$F(4, 30) = 0.311$, $p = 0.868$].

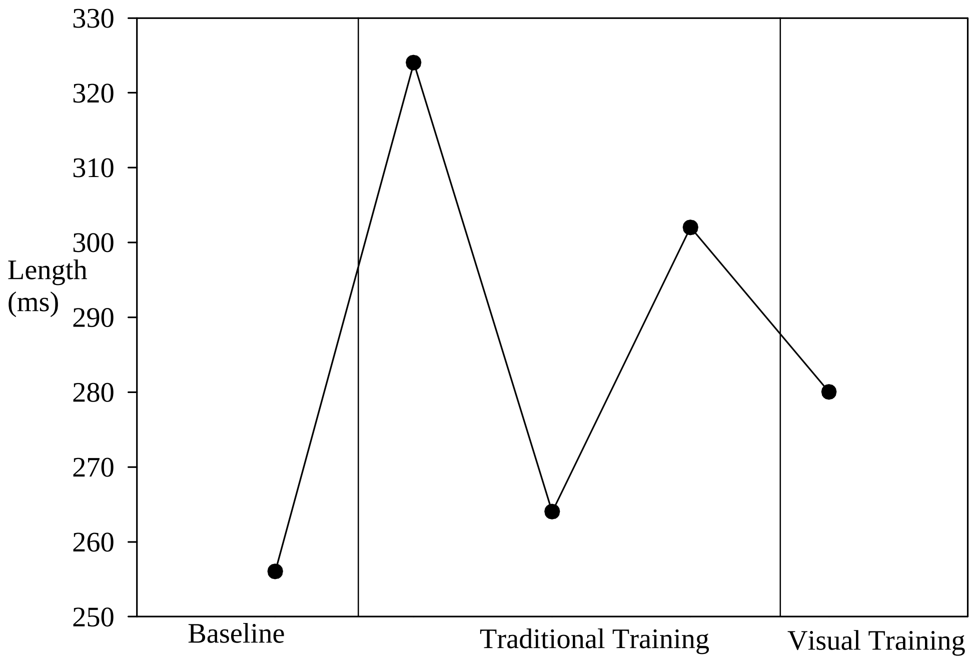

Figure 11. Average Length of Final Consonants for Participant A

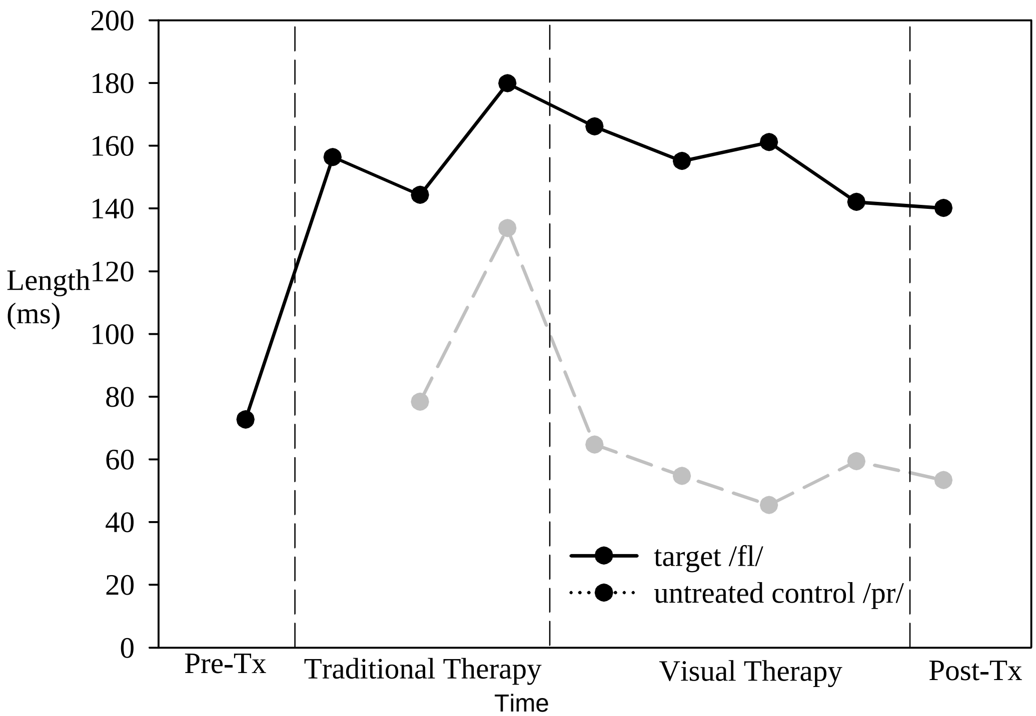

Figure 12. Average Length of Target Consonant Cluster and Controls over the Training Period for Participant A

Table 8. Measured duration of consonants in natural clusters in word-initial, stressed situations, for a normal hearing speaker (Umeda, 1977).

Consonant	Length	Followed by						Preceded by		
	(msec)	r	l	p	t	k	s	s, f	b, g	p, k
f	122	+2	+15							
s	129			-43	-42	-39				
p	89	-19	-14				-13			
t	77	-30					-36			
k	69		-4				-21			
b	90									
d	83	-18								
g	67		+6							
l	66							-19	-8	+3

For Participants B and C, final consonant length was measured as an acoustic indication of the participants' awareness and production of final consonants. An increase in final consonant length was presumed to be an acoustic indication of improved awareness and production of final consonants. Participant B showed an increase in final consonant length with the introduction of visual training, however the increase was not maintained in the follow-up recording. An ANOVA test failed to detect statistically significant changes in length over the training period [$F(4, 70) = 0.453$), $p = 0.770$]. Participant C showed an unexpected reduction in final consonant length in both the training and follow-up phases. However, when comparing Participant C's final consonant lengths to Umeda's 1977 norms (Table 8), the lengths from the initial recording were abnormally high. Following training the lengths reduced to more normal levels, indicating a positive effect of visual training.

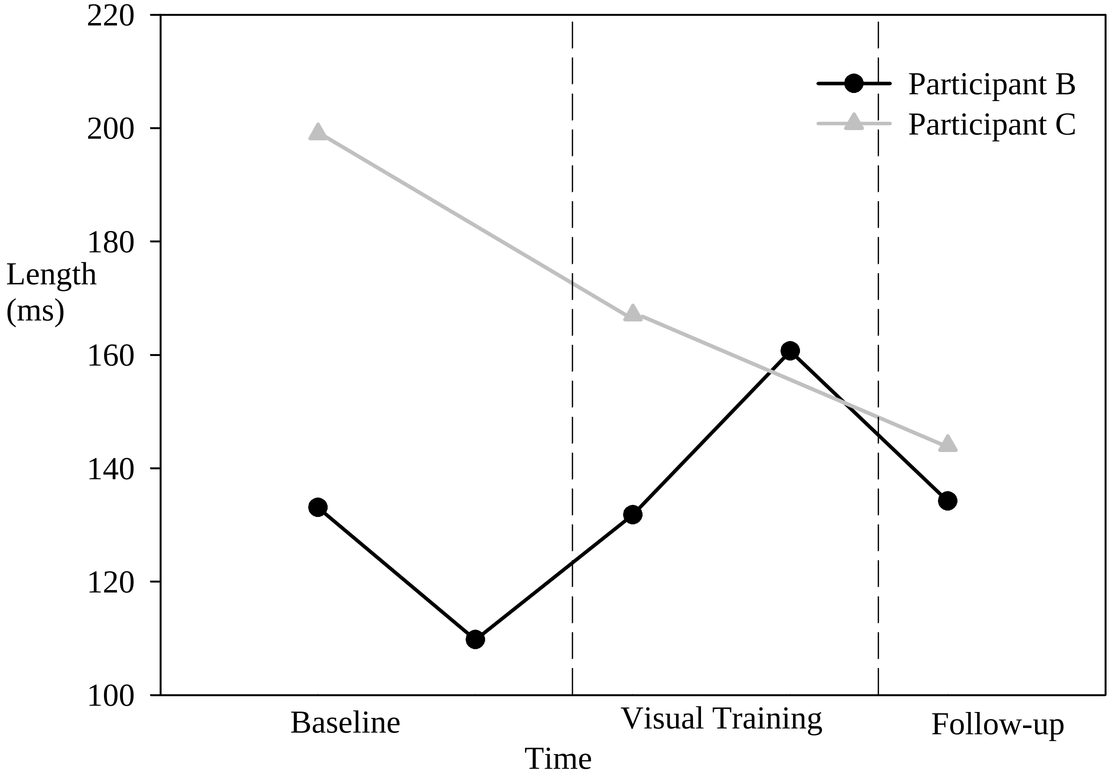

Figure 13. Average Length of Final Consonants over Training Period for Participants B and C

Table 9. Final consonant Length for a Normal-Hearing Speaker (Umeda, 1977)

Consonant	Length (ms)
t	46
k	78
d	76
dg	58
f	152
v	70
z	91
!	85
m	90
n	81
!	83
l	66

3.2.5 Formant Frequencies and Vowel Space

F1-F2 plots were used to examine changes in formant frequencies and formant variability of vowels over the training period. Frequency values for the first and second formants are plotted in Figure 14. The triangle shapes represent the vowel space area, which are also documented in Table 10. Visual inspection of the plots showed similar vowel production between baseline and follow-up recordings for Participant A, however calculation of the vowel triangle area revealed an increase in vowel space following training.

Results for Participant B showed an overlap of vowel formant regions in the baseline recording, but the post-visual training recording showed a significantly larger vowel space with more distinction between vowels.

Results for Participant C indicated adequate distinction between vowel formants in the baseline recording, however post-training there was an unexpected reduction in vowel space, due to a reduced F2 range between /i/ and /u/.

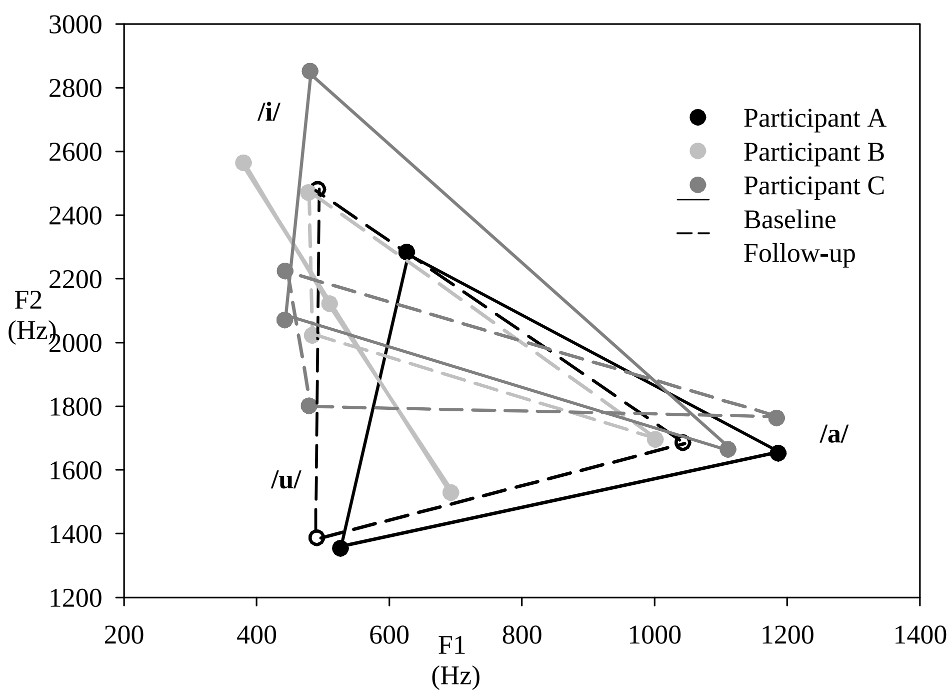

Figure 14: Vowel space pre and post-training for each participant

Table 10. Vowel Triangle Area (Hz^2) for each Participant Pre and Post-Training

	Pre-Training	**Post-Training**
Participant A	292025	301551
Participant B	2921	114829
Participant C	268694	148406

3.3 Discussion

The hypothesis for Study Two was that visual training using spectrograms would be more effective than traditional training because visual feedback provides objective and immediate feedback, compared to traditional training, which provides more subjective, delayed feedback via the clinician. Individually, the three participants showed positive but different effects of training with visual training using spectrograms. Participant A, who received the longest treatment period displayed greater improvements compared to Participants B and C. However considering the amount of prior speech therapy that Participants B and C had received, their gains made in a brief time period were also positive. All three participants in the study appeared to be more motivated during the visual training phases compared to traditional training.

Results confirmed that, in most cases, the acoustic measures were more sensitive than perceptual measures in identifying changes & highlighting differences between training approaches. A number of studies have also reported greater measurable differences between conditions using acoustic measures, compared to perceptual measures which listeners often hear as neutralized (Uchanski & Geers, 2003). This is due to the ability of acoustic analyses to provide a more detailed acoustic description of speech characteristics.

3.3.1 Participant A

Participant A, who received the longest training period out of all three participants displayed greater improvement of targeted error processes on both perceptual and acoustic measures. Perceptual measures revealed minimal change in perceived accuracy of targets in the Goldman-Fristoe baseline recordings but the probe list was more sensitive in revealing improvement in accuracy of targets. The lack of change in the Goldman-Fristoe measures was likely due to two factors: firstly, there were only a small number of recordings of the Goldman-Fristoe taken; and secondly, the Goldman-Fristoe contained only a minimal number of target tokens, which affected the ability to create a representative trend. In contrast, the probe list recordings, which were taken throughout the training period, were considerably more sensitive to changes in accuracy of target production and revealed greater improvement in both targets with visual training compared to traditional training, as well as greater generalisation. These results suggested that visual training resulted in improved perceived accuracy of production for both final consonant deletion and consonant cluster reduction, as well as the transfer of newly learned target processes to other consonants.

The VOT measure provides an important cue for the phonemic contrast between voiced & voiceless plosives. The distinction requires fast movements of the articulators and good coordination of motor control between the larynx and the articulators (Monsen, 1976). The VOT measure for normal-hearing speakers has been reported to be 35 to 100ms for voiceless plosives (/p/, /t/, /k/) (Gulian et al., 1983). Prior to training, VOT for Participant A was outside these norms; however, following training Participant A demonstrated a reduction closer to the normal range, with the exception of /k/. The result for /k/ may be due to its limited visibility of production within the vocal tract, making the sound difficult for a hearing-impaired child to perceive. In general, Participant A showed a reduction in VOT following both traditional and visual training. However, a restricted number of tokens meant that the two training approaches could not be reliability compared. Future studies with more assessment sessions within the treatment phases would be useful to determine which training approach would result in more stable speech improvement.

Temporal measures revealed an increase in the length of both target error processes following training. Measures of final consonant length were highly variable with both training approaches, which limited interpretation, although a trend of increasing length was suggested. This suggests that there may have been an improvement in the production of final consonants following treatment. For consonant cluster length, the trained target /fl/ increased immediately with traditional training, and was maintained with visual training, suggesting improved awareness and production of the two components of the consonant cluster. However, the untrained target /pr/ remained unchanged, indicating that further treatment was necessary to facilitate generalisation to other consonant clusters.

Although vowels were not targeted in the study, formant measures for Participant A revealed a slight increase in vowel space following training, indicating greater differentiation in the production of different vowels. This improvement in vowel production may have been a result of Participant A's increased awareness of speech production and greater attempts to improve her speech clarity.

3.3.2 Participant B

Participant B, who received visual training only, also displayed some improvements in his speech production with visual training. Perceptual analyses using the Goldman-Fristoe and New Zealand Test of Articulation did not reveal any perceived improvement in production accuracy of the target final consonant deletion. This was likely due to the high scores obtained in the baseline, which restricted his ability to demonstrate improvement. However acoustic measures of the probe sentence list were more sensitive in identifying subtle changes in production over the training period.

Voice Onset Time (VOT) measures for Participant B's voiceless plosives were within Gulian et al.'s (1983) normal range prior to training, however reduced further for /p/ and /t/ following training. This suggested improved coordination of phonation and articulation of these two sounds. As with Participant A, the lack of improvement for /k/ is not surprising due to its lack of visibility.

Temporal measures of final consonant length for Participant B increased over the training period suggesting an improved awareness and production of final consonants, however the improvement was not maintained in the follow-up recording. This result suggested that further training was required for Participant B to maintain improvements, which was not surprising given the short training period that he received.

Formant measures of vowels revealed an improved vowel space following training compared to baseline recordings. This suggests that Participant B produced a greater range of formant frequencies following visual training, resulting in greater distinction between vowels. As already discussed, vowels were not targeted in the study therefore the improved vowel production may be due to an increased attempt at speech production clarity as a result of the training.

3.3.3 Participant C

Participant C received only two training sessions and as a result, both his perceptual and acoustic measures were variable and not as clear as results for Participants A and B. Perceptual analyses did not reveal any change in perceived accuracy of Participant C's target production. As for Participant B, this may be due to the high scores obtained in the baseline recording, as well as the limited number of recordings taken. However objective acoustic measures were sensitive in identifying some changes in production.

Voice Onset Time (VOT) for voiceless plosives reduced following visual training. Post-visual training measures of VOT reduced to a range within Gulian et al.'s (1983) norms. This indicated improved coordination of phonation and articulation, which is important for production of the voiced-voiceless distinction as already discussed.

Temporal measures of final consonant length in the baseline recording revealed abnormally long measures compared to Umeda's (1977) published norms. However following visual training, final consonant length reduced to within Umeda's normal range. This may be due to Participant C's improved coordination, as demonstrated in improved VOT, as well as improved awareness of speech clarity.

Participant C showed an unexpected reduction in vowel space following the training period. A reduced vowel space area represents a restriction of tongue elevation and front-back tongue movement (Liu et al., 2005). This result may be due to the short training period he experienced, which caused variability in speech production.

3.3.4 Summary

In summary, all three participants showed improvement following the visual training. Participant A showed improvement in both her perceptual and acoustic measures for her two target processes, final consonant deletion and consonant cluster reduction. Both training approaches appeared to be effective in improving her accuracy, however she appeared to have more interest in the visual training sessions. Perceptual measures suggested greater improvement with visual training, however acoustic measures failed to detect a difference between the two approaches. Further measures of each training approach are required before the effectiveness of each training approach can be reliably compared. Perceptual measures for Participant B failed to identify improvements in the perceived accuracy of final consonant productions. However the acoustic measures were more sensitive to changes and highlighted improvements in both consonant and vowel production following the visual training. Although Participant C received minimal training, his acoustic measures also suggested some improvements in production of final consonants. The limited training period for Participant C meant that results were variable which limited interpretation, however trends of improvement were evident in voice onset time and temporal measures. Overall, all three participants showed improvements with the visual training and appeared to be interested and motivated to learn with the use of the spectrograms.

4 Summary of Main Findings and Discussion

The objectives of this thesis were to use both perceptual and objective acoustic measures to identify acoustic characteristics related to the perception of hearing impaired children's speech and to determine the effectiveness of utilizing visual feedback in their speech training. Study One investigated which temporal and spectral characteristics in the speech of hearing-impaired children can be associated with perceptual accuracy. It was found that the perception of vowel accuracy appeared to be related closely to formant frequencies and vowel duration while consonant accuracy to voice onset time for plosives, and moment 1 (mean) and moment 3 (skewness) spectral measures for plosives, fricatives, and affricates. Study Two aimed to determine whether visual training using spectrograms was more effective than traditional training in improving the speech production of hearing impaired children. The use of spectrograms as visual biofeedback in speech training of hearing impaired children appeared to facilitate learning, and increased the participants' motivation to learn as compared to traditional training approaches. Although subjective perceptual assessment did not reveal a significant treatment effect, several aspects of speech improvement made by Participants B and C could be observed through objective acoustic measures. Results from both studies underscore the importance of objective measures for documenting speech errors and the training effect.

4.1 Limitations of Study

Study One and Study Two together have provided insight into the acoustic characteristics of hearing impaired children's speech that may be useful in their aural rehabilitation, and have also provided support for the use of spectrograms in their speech training. However there are certain limitations. Firstly, the findings are based on single case study designs with a small number of participants, therefore we are unable to generalise findings to other hearing impaired children. Studies employing a larger number of participants are necessary to further investigate the efficacy of visual training compared to traditional training approaches. Secondly, each child received a different length of training. This was due to time constraints of the study as well as commitments of the participants. The shortened training periods for Participants B and C meant that training sessions were limited to visual training only, limiting the ability to make comparisons between training approaches. The ability to make comparisons between training approaches was also restricted in some cases for Participant A, due to the limited number of acoustic measures following visual training. Thirdly, no long-term follow-up recordings were taken, therefore maintenance effects of training could not be measured. Additionally, no spontaneous recordings were taken, therefore the effect of training on intelligibility of spontaneous speech was unable to be measured. Fourthly, in many measures there were an insufficient number of tokens to perform statistical tests or reliably draw interpretations. Finally, values obtained from the spectral moment analyses for M3 and M4 were much larger than those reported by others. The Praat programme states that typical values for M3 and M4 are $\pm$ 1 to $\pm$3 for talkers with normal hearing. Uchanski and Geers (2003) reported skewness values that range from about -4.5 to +1 for /s/ and /! / for children with normal hearing. The M3 values for Participant A are unusually higher, ranging from 4 to 14 for her /s/ and /! / sounds respectively. Additionally, the M4 values are very large and have large standard deviations (for /f/, M4 mean is 3010, M4 SD is 6175).

4.2 Clinical Implication

Although the thesis findings are not yet viable at a clinical level, possible clinical implications for the future can be predicted. Based on the theoretical understanding of the acoustic-articulatory relationship in vowel and consonant productions, acoustic analysis was found to be useful for revealing the underlying constraints responsible for the speech deficits of the hearing-impaired. The acoustic characteristics identified may be useful in helping clinicians to guide aural habilitation and speech training. For example, improved perception of vowels was found to be related to an increased vowel space area as well as reduced vowel duration. These results suggest that training aimed at increasing the vowel space may improve the perceived accuracy for listeners. Similarly, consonant accuracy appeared to be most closely related to moment 1 (mean of the energy spread). Since M1 measures are affected by tongue position, this acoustic measure could be used in clinical application to provide feedback in speech training and monitor progress.

Although speech training with spectrograms as visual feedback were not found to be more effective than traditional training techniques, the learning outcomes observed, as well as the participants' improved motivation, suggested that spectrograms may be a useful supplementary tool for training particular speech targets. The spectrographic displays are sensitive to changes in manner of articulation. Since traditional training tends to focus more on placement of articulation, spectrographic displays can provide an additional cue for speech training.

4.3 Future Studies

Since these studies were both single-case designs, the findings are not viable at a clinical level, and future research is required on their role as clinical tools. Study One has provided new information on the spectral characteristics of vowels and consonants produced by one hearing-impaired child. The spectral characteristics need to be examined in more participants to determine whether these characteristics are typical of the population. Findings from Study Two suggested that spectrograms may be a helpful supplementary clinical tool for aural habilitation and speech training. Further research is required with more participants to provide evidence for the effectiveness. Additionally, research is required to determine which error processes and/or speech sound errors benefit most from training using spectrograms. A larger sample size and longer experimental period are required to better understand the general and long-term effectiveness of visual treatment using spectrograms before they can be accepted as viable clinical tools.

4.4 Conclusion

The results of this thesis have provided further information on the underlying acoustic characteristics of hearing impaired children's speech. These characteristics have also provided useful clues to the type of feedback required by hearing impaired children in their speech training. The results have also suggested that the use of spectrograms as visual feedback is an effective supplementary clinical tool for speech training of this population. Since these studies were both single-case designs, the findings are not viable at a clinical level, and future research is required on their role as clinical tools.

References

Angelocci, A.A., Kopp, G.A., & Holbrook, A. (1964). The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys. *Journal of Speech and Hearing Disorders, 29(2),* 56-170.

Boothroyd, A. (1984). Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *Journal of Speech and Hearing Research, 27,* 134-144.

Carney, A.E. & Moeller, M.P. (1998). Treatment efficacy: Hearing loss in children. *Journal of Speech-Language and Hearing Research, 41,* S61-84.

Crawford, R. (1995). Teaching voiced velar stops to profoundly deaf children using EPG – two cases. *Clinical Linguistics and Phonetics, 9(3),* 255-269.

Culbertson, D.N. & Kricos, P.B. (2002). Language and speech of the deaf and hard of hearing. In R.L. Schow & M.A. Ling (Eds.), *Introduction to Audiologic Rehabilitation* (4th ed.). (pp.183-224). Boston: Allyn and Bacon.

Dagenais, P.A., Citz-Crosby, P., Fletcher, S.G., & McCutcheon, M.J. (1994). Comparing abilities of children with profound hearing impairments to learn consonants using electropalatography or traditional aural-oral techniques. *Journal of speech and hearing research, 37,* 687-699

Dodd, B. (1976). The phonological systems of deaf children. *Journal of Speech and Hearing Disorders, 41(2),* 185-197.

Downs, M.P. & Yoshinaga-Itano, C. (1999). The efficacy of early identification and intervention for children with hearing impairment. *Pediatric Clinics of North America, 46(1),* 79-87.

Dunn, C & Newton, L. (1986). A comprehensive model for speech development in hearing-impaired children. *Topics in Language Disorders, 6(3),* 25-46.

Elfenbein, J.L., Hardin-Jones, M.A., & Davis, L.M. (1994). Oral communication skills of children who are hard of hearing. *Journal of Speech and Hearing Research*, 37, 216-226.

Ertmer, D.J. & Maki, J.E. (2000). A comparison of speech training methods with deaf adolescents: Spectrographic versus noninstrumental instruction. *Journal of Speech, Language, and Hearing Research, 43(6)*,1509-1523.

Ertmer, D.J., Stark, R.E. & Karlan, G.R. (1996). Real-time spectrographic displays in vowel production training with children who have profound hearing loss. *American Journal of Speech-Language Pathology, 5(4),* 4-16.

Forrest, K & Elbert, M. (1990). Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clinical Linguistics and Phonetics, 4(4),* 327-340.

Forrest, K., Weismer, G., Elbert, M., & Dinnsen, A. (1994). Spectral analysis of target-appropriate /t/ and /k/ produced by phonologically disordered and normally articulating children. *Clinical Linguistics and Phonetics, 8(4),* 267-281.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R.N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84(1),* 115-123.

Fry, D.B. (1979). *The Physics of Speech*. Cambridge: Cambridge University Press.

Goldman, R. & Fristoe, M. (1969). *Goldman-Fristoe Test of Articulation*. Minnesota: American Guidance Service.

Goodman, A. (1965). Reference levels for pure-tone audiometer. *ASHA, 7*, 262-263.

Grant, K.W., Walden, B.E., & Seitz, P.F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America, 103 (5),* 2677-2690.

Gulian, E., Hinds, P., Fallside, F., & Keiller, C. (1983). Acquisition of the voicing distinction by profoundly hearing-impaired children, *British Journal of Audiology*, *17(4)* 233-244.

Harrell, R.W. (2002). Puretone evaluation. In J. Katz (Ed.), *Handbook of Clinical Audiometry* (5th ed.). (pp. 71-87). Baltimore: Lippincott, Williams & Wilkins.

Hudgins, C. & Numbers, F. (1942). An investigation of the intelligibility of the speech of deaf and normal subjects. *Genetic Psychology Monographs*, *25*, 286-392.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108(3),* 1252-1263.

Katz, J. (2002). *Handbook of Clinical Audiometry* (5th ed.). Baltimore: Lippincott, Williams & Wilkins.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research, 19*, 421–447.

Kidd, G. (2002). Psychoacoustics. In J. Katz (Ed.), *Handbook of Clinical Audiometry* (5th ed.). (pp. 33-49). Baltimore: Lippincott, Williams & Wilkins

Kuhl, P.K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70*, 340-349.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431-461.

Ling, D.L. (1976). *Speech and the Hearing-Impaired Child: Theory and Practice*. Washington: Alexander Graham Bell Association for the Deaf.

Ling, D. (1989). *Foundations of spoken language for hearing-impaired children*. Washington: Alexander Graham Bell Association for the Deaf.

Ling, D. (2001). Speech development for children who are hearing impaired. In R.H. Hull (Ed.), *Aural Rehabilitation: Serving adults and children* (4th ed.). (pp.). San Diego: Singular, Thomson Learning.

Ling, D. & Ling, A.H. (1978). *Aural Habilitation: The Foundations of Verbal Learning in Hearing-Impaired Children*. Washington: The Alexander Graham Bell Association for the Deaf.

Liu, H., Tsao, F., & Kuhl, P.K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Journal of the Acoustical Society of America, 117(6),* 3879-3889.

Maki, J.E., & Streff,M.S. (1978 November). *Clinical evaluation of the speech spectrographic display with hearing impaired adults.* Paper presented at the annual convention of the American-Speech-Language-Hearing-Association, San Francisco.

Martin, F.N. & Clark, J.G. (2003). *Introduction to Audiology* (8th ed.). Boston: Allyn & Bacon.

Matthies, M., Svirsky, M.A., Lane, H.L., & Perkell, J.S. (1994). A preliminary study of the effects of cochlear implants on the production of sibilants. *Journal of the Acoustical Society of America, 96(3),* 1367-1373.

Miller, J.D. (1977). Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T.H. Bullock (Ed.), *Recognition of Complex Acoustic Signals* (Life Sciences Research Report 5), (pp. 49). Berlin: Dahlem Konferenzen.

Monsen, R.B. (1974). Durational aspects of vowel production in the speech of deaf children. *Journal of Speech and Hearing Research, 17*, 386-398.

Monsen, R.B. (1976). Normal and reduced phonological space: The production of English vowels by deaf adolescents. *Journal of Phonetics*, *4*, 189-198.

Moore,J.K. Perazzo,L.M. & Braun,A. (1995). Time course of axonal myelination in the human brainstem auditory pathway. *Hearing Research, 87(1-2),* 21-31.

Moyle, J. (2004). *New Zealand Articulation Test*. Lower Hutt: Special Education.

Pantelemidou, V., Herman, R., & Thomas, J. (2003), Efficacy of speech intervention using electropalatography with a cochlear implant user. *Clinical Linguistics & Phonetics, 17(4-5),* 383–392

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28,* 233-272.

Pratt, S.R., Heintzelmann,A.T., & Deming, S.E. (1993). The efficacy of using the IBM SpeechViewer vowel accuracy module to treat young children with hearing impairment. *Journal of Speech and Hearing Research, 36(5),* 1063-1074.

Ruffin-Simon,C. (1983). Relationship between production and perception of speech: Some developmental issues. In I. Hochberg, H. Levitt & M.J. Osberger (Eds.). *Speech of the Hearing Impaired: Research, Training, and Personnel Preparation.* (pp.233-250). Baltimore: University Park Press.

Ryalls, J., Le Dorze, G., Boulanger, H., & Laroche, B. (1995). Speech therapy for lowering vocal fundamental frequency in two adolescents with hearing impairments: A comparison with and without SpeechViewer. *The Volta Review, 97(4),* 243-250.

Shuster, L.I., Ruscello, D.M., & Smith, K.D. (1992). Evoking [r] using visual feedback. *American Journal of Speech-Language Pathology, 1(3),* 29-34.

Smith, C.R. (1975). Residual hearing and speech production in deaf children. *The Journal of Speech and Hearing Research, 18,* 795-811.

Smith, B.L. (1982). Some observations concerning pre-meaningful vocalisation of hearing-impaired infants. *Journal of Speech and Hearing Disorders, 47*, 439-441.

Stark, R.E. (1971). The use of real-time visual displays of speech in the training of a profoundly deaf non-speaking child: A case report. *Journal of Speech and Hearing Disorders, 36(3),* 397-409.

Stevens, K.N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America, 111*, 1872-1891.

Stoel-Gammon,C. & Otomo,K. (1986). Babbling development of hearing-impaired and normally hearing subjects. *The Journal of Speech and Hearing Disorders, 51(1),* 33-41.

Tjaden, K. & Turner, G.S. (1997). Spectral properties of fricatives in amyotrophic lateral sclerosis. *Journal of Speech, Language and Hearing Research, 40(6),* 1358-1372.

Uchanski, R.M. & Geers, A.E. (2003). Acoustic characteristics of the speech of young cochlear implant users: A comparison with normal-hearing age-mates. *Ear and Hearing, 24(1S),* 90S-105S.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America, 61(3),* 846-858.

Verbrugge, R.R & Strange, W. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America, 60(1),* 198-211.

Weiss, C.E., Gordon, M.E., & Lillywhite, H.S. (1980).*Clinical Management of Articulatory and Phonologic Disorders (* 2nd ed.). Baltimore: Williams & Wilkins.

Yost, W.A. (2000). *Fundamentals of Hearing: An Introduction* (4th ed.). San Diego: Academic Press.

Appendix 1

Assessment Word Lists

Goldman-Fristoe Articulation Test	New Zealand Articulation Test	Probe List Participant A	Probe List Participants B and C
1. house	1. Pencil	1. drop	1. She has chicken pox
2. telephone	2. Apple	2. cleaner	2. Let’s have the medium size
3. cup	3. Cup	3. bicycle	3. The bank was closed on the weekend
4. gun	4. Ball	4. zebra	
5. knife	5. Table	5. kite	4. It’s on next month
6. window	6. Web	6. down	5. Pick up that paper
7. wagon	7. Tap	7. castle	6. Mum made a shopping list
8. chicken	8. Bottle	8. listen	7. Don’t stand on the edge of the cliff
9. zipper	9. Boat	9. playhouse	
10. scissors	10. Door	10. mouth	8. Close your mouth when you eat
11. duck	11. Ladder	11. sock	
12. vacuum	12. Bed	12. drive	9. I have a million balls
13. matches	13. Keys	13. cat	10. The work has one syllable
14. lamp	14. Circle	14. hotdog	
15. shovel	15. Bike	15. float	
16. car	16. Girl	16. angel	
17. rabbit	17. Burger	17. ice cream	
18. fishing	18. Pig	18. train	
19. church	19. Fish	19. music	
20. feather	20. Dolphin	20. dragon	
21. pencils	21. Leaf	21. fireman	
22. carrot	22. Van	22. toaster	
23. bathtub	23. TV	23. steak	
24. thumb	24. glove		
25. jumping	25. mouse		
26. pajamas	26. hammer		
27. plane	27. worm		
28. brush	28. knife		
29. drum	29. money		
30. flag	30. train		
31. Santa Claus	31. lamp		
32. Christmas tree	32. toilet		
	33. hat		
33. squirrel	34. sun		
34. sleeping	35. whistle		
35. stove	36. house		
36. wheel	37. zip		
37. yellow	38. puzzle		
38. this	39. cheese		
39. orange	40. shoe		
40. bath	41. fishing		
41. finger	42. brush		
42. ring	43. chair		
43. blue	44. watching		

	45. witch 46. jam 47. magic 48. fridge 49. thumb 50. nothing 51. bath 52. there 53. feather 54. rabbit 55. carrot 56. yoghurt 57. wheel 58. singing 59. ring 60. present 61. bread 62. frog 63. truck 64. drum 65. crab 66. green 67. blue 68. plane 69. clown 70. glasses 71. flower 72. slide 73. swing 74. smoke 75. snail 76. spider 77. school 78. star 79. spray 80. straw 81. scratching 82. square 83. sausages 84. gorilla 85. animals 86. pacific 87. motorbike 88. umbrella 89. washing machine 90. caterpillar 91. calculator 92. Australia 93. helicopter Tairanosaurus		

Appendix 2

Training Word Lists

Participant A

Target: Deletion of Final Consonant

1. book
2. cup
3. put
4. bag
5. tub
6. mad
7. late
8. break
9. need
10. sharp

Target: Consonant Cluster Reduction

1. flee
2. flop
3. flu
4. flip
5. flame
6. flat
7. flow
8. flesh
9. fly
10. flower

Participants B and C

Target: Deletion of Final Consonant

1. The parks were green.
2. The scissors are sharp.
3. He is second in line.
4. Someone is saying something.
5. Put the tablecloth on the table.
6. Black widows are spiders.
7. The plane landed on the ground.
8. The wind is blowing the trees.
9. Eggs are in the henhouse.
10. We planted five flowers last spring.
11. I like to drive to the West Coast.
12. We speak the English language.
13. You need to practice catching the ball.
14. We picked a bunch of carrots for dinner.
15. We are going to have chicken for lunch.