ACOUSTICS OF WHISPERED STOPS

Registration No. L0480002

A Dissertation Submitted in part fulfillment of
Master's Degree (Speech Language Pathology),
University of Mysore,
Mysore.

ALL INDIA INSTITUTE OF SPEECH AND HEARING
MANASAGANGOTHRI
MYSORE – 570 006

MAY – 2006

As I walk, I do not walk alone

You are always by my side.

When I pray, I do not pray alone

You are on your knees beside me.

When I sleep, I do not sleep alone

You stand over me and watch me sleep.

And when I am down and out

You are right there to give me joy.

When I have pain that runs deep in to my soul

You always take it away.

And when I'm in my darkest hour of the night

You show me the way.

And if by now you don't know what Iam trying to say,

Is that God is here beside us every day & every night

Dedicated to

My Dear

Achachan & Mummy!!!

Certificate

This is to certify that this Dissertation entitled **"Acoustics of whispered stops"** is a bonafide work in part fulfillment for the degree of master of (Speech Language Pathology) of the student (Registration No. L0480002). This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysore

May, 2006

Prof. M. Jayaram
Director
All India Institute of Speech and Hearing
Naimisham Campus
Manasagangothri
Mysore-570 006.

Certificate

This is to certify that this Dissertation entitled **"Acoustics of whispered stops"** is a bonafide work in part fulfillment for the degree of master of (Speech Language Pathology) of the student (Registration No. L0480002). This has been carried out under my guidance has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Guide

Mysore

May, 2006

Prof. S. R. Savithri
Professor and Head,
Department of Speech-Language Sciences
All India Institute of Speech and Hearing
Naimisham Campus
Manasagangothri
Mysore-570 006.

Declaration

This Dissertation entitled **"Acoustics of whispered stops"** is the result of my own study under the guidance of Prof. S. R. Savithri, Professor and Head, Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier in any other University for the award of any Diploma or Degree.

Mysore

May, 2006 Register No. L0480002

Acknowledgements

*I would like to express my sincere gratitude to my teacher and guide, **Prof, S.R . Savithri**, for her constant guidance and support. You have inspired me a lot. Thank you ma'm for the kindness and patience that you have shown right from the beginning. Without your help I wouldn't have been able to finish this study. Thank you for everything.*

*I thank **Prof. M. Jayaram**, Director, AIISH, Mysore, for allowing me to conduct this study.*

*My heart felt thanks to **Yeshodha ma'm, Rohini ma'm, Ananthi ma'm**, **Seema ma'm Kedarnath**, **Venugopal** and to all the staffs of Speech Language Science department for their timely help.*

Thanks to the subjects participated in this study.

Thanks to all my teachers from school days who enlightened the knowledge in me.

*Thank you, **Achachan** and **Mummy** for, giving me all strength , support, encouragement and for keeping faith in me.*

*Dear **Deepthi** & **Dipu,** Wonderful sister & brother. I thank God for giving you in my life.*

*I thank all my Friends, seniors and juniors especially **Sowmya**, **Ashly**, **Divya, Sr. Elizebeth, Sr.Cicilia, Maria, Karuna, Gayathri, Sneha, Shijitha**, **Deepthi**, **Rexy, Sangeetha**, **Feby, Anjana** and those who have helped me directly or indirectly for this dissertation.*

My huge thanks to all my dear batch mates in B.sc and M.sc. I am glad that I got you all as my class mates, who are there all the time, with stretching hands to help, to correct, to advise, amongst all to have gr8 fun. I cherish all the time we spent together. I 'll miss you all!!!!!!! Wish you all success in your life.

Above all, I thank Almighty for giving me the courage and strength to face all my joys and sorrows.

List of contents

List of tables

List of figures

CHAPTER – 1

Introduction

Speech is an integral part of human behavior. It is largely through speech that we are able to communicate with others and to make known our wants and needs. The source of speech is the air exhaled/ inhaled from/ through the lungs. The alternate opening and closing of the vocal cords result in the emission of successive puffs of air into the space above the larynx. When this alternate opening and closing or vibration of the vocal cords take place then the speech sound produced is termed 'voiced'. When there is no vibration of the vocal cords then the sound produced is termed as 'unvoiced'. Hence the primary source of speech is air.

Turbulence in the glottal air stream is an additional source of sound. It has a 'hissing' quality, which is labeled 'aspiration' when it is combined with vocal fold vibration and 'whisper' in the absence of vocal fold vibration (Titze, 1994).

Many of the advances in our understanding of normal speech perception have come from the study of speech perception in degraded listening conditions caused by masking noise or filtering or minimal synthesis. Through the process of elimination, such studies inform us to which properties of the speech signal are important for phonetic perception, and also which aspects of the speech signal are resistant to distortion.

Interestingly there have been few studies to examine the degradation produced by the normal vocal tract- whisper. Whisper is a mode of speech in which there is no vibration of the vocal folds. Unlike in normal speech, the ‘arytenoids’ are slightly abducted and “toed in” in whispered speech (Zemlin, 1988). This creates small chink, which can be of different shape, through which air gushes with turbulent force. Whispered speech is produced with more open glottis than in normal voicing and with longer syllable durations and stop closure intervals. Thus voicing is absent in whispered speech.

Whispering is also a vocal behavior of phonation. It has a special role in communication. It is used in situations, unlike the use of voice, to restrict the communication to a single listener or to a small group. Whisper may also be an indicator of laryngeal pathology.

Whisper is of interest for the same reasons as is the perception of mechanically produced distorted speech, and also because it is important to understand this naturally occurring communication mode.

An acoustic characteristic of whisper is a less studied area. Schwartz (1971) reported that the closure duration for /p/, /b/ under whispered speaking condition in young adult females was longer than that in whispered /m/. Dannenbring (1980) assessed perception of voicing in whispered consonants followed by vowels /a/, /i/, and /u/. Twelve subjects were able to discriminate between whispered consonants. Kallail & Emmanuel (1984) reported higher vowel formants, especially $F_{1,}$ in whispered vowels compared to phonated vowels. In contrast, Divya &Tanusree (2005) reported shorter vowel duration and higher F_2 in whispered speech compared to phonated speech. They did not find any significant difference between closure and consonant duration in two conditions. Tarter (1989) investigated the acoustic features important in identifying whispered consonants. The results indicated preservation of high frequency cues to voicing. Six untrained subjects identified 18 different whispered initial consonants significantly better than chance in nonsense syllable. The phonetic features of place and manner of articulation and to a lesser extent voicing were correctly identified. Confusion matrix and acoustic analysis indicated preservation of resonance characteristics for place and manner of articulation and suggested the use of burst, aspiration, or frication duration and intensity and a first formant cut back for voicing decisions.

To summarize, there is hardly any information about the acoustic cues of voicing in whispered consonants. In this context, the present study was planned. The objective of this study was to investigate the voicing cues of stop consonants in whispered speech. Specifically temporal and spectral measures of stop consonants as occurring in CV and VCV syllables in whispered and normal conditions were extracted and compared.

Stop consonants are produced by occluding the oral cavity by an articulator and releasing the air stream after air pressure is built. The occlusion is usually formed at five sites: bilabial, dental, alveolar, retroflex, and velar. The occlusions are also formed at a variety of other places including palatal, uvular, pharyngeal and glottal. Stops are abundantly represented in the world’s languages and are often among the most frequently occurring consonants in a given language.

Stop consonants are also classified on the basis of voicing and aspiration. They are classified as voiced and unvoiced and as aspirated and unaspirated. When languages have three of the four kinds of stops- voiced unaspirated, voiced murmured, unvoiced unaspirated, and unvoiced unaspirated- they are termed 3-way languages. However, when languages have all the four categories of stops, they are termed 4-way languages. Table 1. Shows the phonetic classification of stop consonants.

	Unvoiced		Voiced	
Place of articulation	U	A	U	A
Velar	k	kh	g	gh
Retroflex	ṭ	ṭh	ḍ	ḍh
Alveolar	t	th	d	dh
Dental	t	d	d	dh

Table 1: Phonetic classification of stop consonants (U = unaspirated, A = aspirated, M = murmured).

Five qualitatively distinct acoustic segments can be identified for stop consonants as follows:

1. A period of occlusion (silent/voiced)
2 A transient explosion (usually less than 20 ms) produced by shock excitation of the vocal tract upon release of occlusion.
3 A very brief (0-10 ms) period of frication as articulators separate and air is blown through a narrow constriction as in the harmonic fricative.
4 A very brief period of aspiration (2-20 ms) in which may be detected noise excite formant transitions, reflecting shift in vocal tract responses as the main body of tongue moves towards a position appropriate for the following vowel.
5 Voiced formant transitions, reflecting the final stages of the articulatory movement into the vowel during the first few cycles of laryngeal vibration.

There are at least 10 acoustic cues of voicing in stop consonants Table 2 summarizes voicing cues of stop consonants

Sl. No.	Cue	Voiced stop	Unvoiced stop
1	VOT	<+20ms	>+20ms
2	Closure duration	Shorter	Longer
3	Total duration	Shorter	Longer
4	Burst duration	Shorter	Longer
5	Burst intensity	Low	High
6	Overall intensity	High	Low
7	Preceding vowel duration	Longer	Shorter
8	F_1 cut back	Absent	Present
9	F0	Rising	Falling
10	Voicing	Present	Absent

Table 2: Acoustic cues of voicing in stop consonants.

While this is the condition in normal speech, it may not so in whispered speech. Hence, it was hypothesized that there will be significant difference between conditions (whispered and voiced) on voicing cues

CHAPTER – II

Review of literature

Whispering is an aphonic laryngeal action. Whispering is a complex process, or more like phonation.” Under condition of partial laryngeal valving and with vocal folds stabilized sufficient air flow through the vocal tract on exhalation will produce laryngeal frication which is commonly labeled whispering (Bracket, 1957).

Travis (1957) reports that whispering differs from voiced phonation in the following manners: The glottis shows the shape of an inverted ‘y’ and the vocal cords are incompletely closed. Vocal cord tension is much lower than in phonation and the cord margins do not visibly vibrate. As a result, the escaping air is set in to non-periodic frictional turbulence so that a noise is produced instead of tones with periodic vibrations. The consumed air volume is greatly increased. Whispering is therefore much more strenuous than in speaking in normal voice. In consequence the subglottic air pressure is much higher than it is during phonation.

According to Zemlin (1981) a typical subglottic laryngeal constriction with the glottis open is observed in whispered phonation. Here an adduction of the false vocal folds takes place with the decrease in the size of the laryngeal cavity in anterior and posterior dimensions. The type of laryngeal constriction becomes more exaggerated for the strong so called ‘stage whispering’. This particular gesture for whispering is considered to contribute to the prevention of the vocal fold vibration by the transglottal airflow, as well as to facilitate the generation of turbulent noise in the laryngeal cavity. The physiological mechanism behind supra glottic laryngeal constriction is not clear. The EMG reports say that posterior cricothyroid and lateral cricothyroid muscles are responsible and related to supraglotic laryngeal constriction.

Whispered speech is different from normal speech in physiological, aerodynamical and acoustic aspects. A study done by Monson & Zemlin, (1984) concerning physiological aspects of whispered speech has revealed that, unlike in normal speech, the arytenoids are slightly abducted and "toed in". This creates a small chink, which can be of different shapes, through which air gushes with turbulent force. Airflow during low energy whisper averages to 205 ml/sec., which is approximately twice as high as that of phonated speech.

Solomon, McCall, Trosset & Gray (1989) did a study on laryngeal configuration and constriction during two types of whispering i.e. low effort whisper and high effort whisper. They examined the vocal fold configuration, glottal size and airway constriction by supraglottal structures during whispering. Two basic configurations of the vocal ford edges were exhibited during whispering: relatively straight and toed in at the vocal processes. Variations in vocal fold configuration reflect the contributions of intrinsic laryngeal muscles during whispering. The glottis was described as medium sized for most subjects during the whispered productions. There was no consistent pattern for different vowels as reported by Bracket (1971) where the glottis is larger for high vowels and smaller for low vowels. The supraglottal constriction by the lateral structures contributes to the production of high effort whisper than low effort whisper.

Stathopoulos, Hoit, Hixon, Watson & Solomon (1991) did a study on respiratory and laryngeal mechanism during whisper. The study revealed a difference in respiratory function between whispering and speaking. The general configuration of the chest wall is the same as that of previous studies. The configuration characterized by a lower abdominal volume and a higher rib cage volume during utterance production compared to relaxation, suggests that the muscular pressures required for utterance production were provided by both the rib cage and abdomen, with the abdominal muscular pressures predominating (Hixon, Goldman, & Mead, 1976). The fact that the configuration of the chest wall was similar for whispering and speaking indicates that the muscular mechanisms operating were essentially the same. This observation offers sight in to the nature of respiratory- laryngeal interactions because it indicates that the general posturing of the chest wall is relatively independent of the state of the laryngeal airway (Hixon, 1988).

Higashikawa, Green, Moore, & Minifie (2003) examined the articulatory kinematic differences between whispered /p/ and /b/. Computerized video- tracking methods were used to evaluate kinematic differences between voiced and unvoiced stops. The results revealed that mean peak opening and closing velocities for /b/ were significantly greater than those for /p/ during whispered speech. This study supported the suggestion that whispered speech and voiced speech rely on distinct motor control processes.

There are few studies in literature related to the aerodynamic measurement in whispered speech. Arkebauer, Hixon & Hardy (1967) did study regarding air pressure variation during whisper and phonation. He says whispered speech differs from vocal speech in terms of the air pressure build up in the vocal tract during the production of stop consonant. In vocal speech the peak intra oral air pressure (P_0) for unvoiced stop /k/ is greater than it is for their voiced cognate /g/. Murray (1976) reported that, in whispered speech oral pressure is essentially the same for both voiced and unvoiced stops.

Weismer & Longstreth (1980) measured peak intra oral air pressure and flow for the syllabus /p/ and /b/ in two speaking conditions (syllable repetitions, and in carried phrases) and in two phonation modes (normal phonation and whisper). They found that the difference between the intra oral air pressure for /p/ and /b/ was statistically significant in normal phonation, but not in whisper. The difference in peak flow for /p/ and /b/ was statistically significant in both normal phonation and whisper.

Klich (1982) reported that, whispered speech differs from vocal speech in terms of the air pressure that build up in the vocal tract during the production of stop consonants. In vocal speech the peak intra oral air pressure for unvoiced stops is greater than for the voiced cognates. He also reports that mean airflow rates in breathy and whispered speech are significantly greater than those found in normal vocal speech and are more typical of those found during production of unvoiced fricative consonants. Hence inferences regarding speech with a breathy quality can be made by comparing the airflows found in vocal speech with those typical of whispered speech.

Hirano (1981) has pointed out that acoustic analysis of the voiced signal may be one of the most attractive methods for assessing phonatory function or laryngeal pathology because, it is non invasive and provides objective and qualitative data. The technique of acoustic analysis has promising future as a diagnostic tool in the management of voice disorders. Many acoustic parameters, derived by various methods have been reported to be useful in differentiating between pathological and normal voice.

Till date 6 studies have been conducted on acoustics of whisper. Schwartz (1971) studied the bilabial closure duration for /p/, /b/ and /m/ under voiced and whispered speaking conditions. The results obtained from 7 young adult females with normal speech indicated significantly greater whisper durations for /p/ and /b/, not /m/. An explanation for this finding is offered in terms of the hypothesis that individuals conserve air when whispering by prolonging air-arresting articulatory gestures.

Mc Glone & Manning (1979) investigated the perception of pitch for whispered and voiced vowels. This was accomplished by two procedures. First both whispered and voiced vowels were rank-ordered by judges and comparisons were made of these rankings to the spectral composition of each vowel (/i/, /I/, /ae/, /o/, and /u/). Secondly, all acoustic energy above formant 1 was electronically filtered out of the acoustic signal of whispered vowels and they were again rounded from highest to lowest pitch. Results revealed no relationship between the rank ordered pitch of the vowels by judges and their mean fundamental frequency. There was very little difference between the measured mean fundamental frequencies of these vowels. The results of this study suggest that the perceived pitch of vowel sounds, whether they be whispered or voiced, is more directly closed to F_2 than to the fundamental frequency or F_1. They also suggest that the major pitch detection clues are not found below F_2.

Dannenbring (1980) assessed perception of voicing in whispered consonants in front of (/a/, /i/, /u/) pronounced by a single speaker. Two experiments were conducted to investigate the ability of 12 subjects to discriminate between whispered consonants that are differentiated in normal speech on the basis of voicing. Subjects were presented with a single consonant on a trial, and asked to choose between it and its counter part on the voicing dimension (e.g., /d/ or /t/). In general, the subjects were able to make these dimensions.

Kallail & Emmanuel (1984) investigated the formant frequencies and amplitudes of phonated and whispered productions of five test vowels (/a/, /i/, /u/, /ae/ & /^/). Fifteen adult male subjects sustained each test three times in phonation. The phoneme represented by each recorded production was identified independently by 11 listeners. An acoustic spectrum of each sample was obtained through the use of real- time narrow band spectrum analyzer to permit formant measurements. They measured the formant frequencies and amplitudes (F_1 to F_3) from the acoustic vowel spectra and also the formant frequency and amplitude differences between the phonated and whispered productions. The results showed a trent for whispered vowel formants to be higher in frequency than phonated vowel formants, but the trent was only strongly evident for F_1.

A study conducted by Tarter (1989) investigated whether and how whispering affects consonant identification and he tried to find the acoustic features important for it. In this experiment, six untrained 20-40 year old listeners were given 18 different CV syllables produced by an adult male and adult female native speaker of English. [The vowel was /a/ and the consonants were (b, d, g, p, t, k, m, n, r, l, w, y, v, f, z, s, š, ž)]. The subjects identified 18 different whispered initial consonants significantly better than chance in nonsense syllable. The phonetic features of place and manner of articulation and to a lesser extent voicing were correctly identified. Confusion matrix and acoustic analysis indicated preservation of resonance characteristics for place and manner of articulation and suggested the use of burst, aspiration, or frication duration and intensity, and/or first formant cut back for voicing decisions. The results indicated that whispering preserves high frequency cues to voicing. Much acoustic information relevant to speech is transmitted in a whisper, in the absence of pitch and harmonic relationships.

Divya & Tanusree (2005) investigated the difference between normal and whispered speech on specific acoustical parameters [vowel duration, first three formant frequencies of vowel /a/ in CVCV words and closure duration and first three formant frequencies of /a/ (preceding and following the consonant) in the VCV word]. Two, 23-year-old male native Kannada speakers participated in the study. The material consisted of 12 bisyllabic Kannada CVCV words and 10 bisyllabic Kannada VCV words. The words were read by the subjects three times in whispered and normal conditions. The parameters extracted were duration of vowel /a/, formant frequencies (F_1, F_2, F_3) and closure and consonant duration. The result showed that the duration of vowel was significantly shorter in whispered speech than in normal speech. Amongst the formants, F_2 was significantly higher in whispered speech. There was no significant difference between the closure and consonant duration in the two conditions.

Ballenger (1971) said that, aphonia is nothing but a kind of whisper. In cases of aphonia and dysphonia, whispering is present. In cases of intermittent aphonia also some whisper is present.

Of these, studies by Schwartz (1971), Kallail & Emmanuel (1984), and Divya & Tanusree (2005) are on acoustics of vowels/ consonants in whispered speech. Mc Glone & Manning (1979) have investigated pitch perception on whispered speech. Dannenbring (1980) has used a discrimination task for whispered consonants and Tarter (1989) has used identification of whispered stops. There is hardly any information on acoustic cues of voicing in whisper. Therefore the present study investigated the acoustic cues of voicing in whisper.

Stop consonants are produced by occluding the oral cavity by an articulator and releasing the air stream after air pressure is built. The occlusion is usually formed at five sites: bilabial, dental, alveolar, retroflex, and velar. The occlusions are also formed at a variety of other places including palatal, uvular, pharyngeal and glottal. Stops are abundantly represented in the world's languages and are often among the most frequently occurring consonants in a given language.

Stop consonants are also classified on the basis of voicing and aspiration. They are classified as voiced and unvoiced and as aspirated and unaspirated. When languages have three of the four kinds of stops- voiced unaspirated, voiced murmured, unvoiced unaspirated, and unvoiced unaspirated- they are termed 3-way languages. However, when languages have all the four categories of stops, they are termed 4-way languages. Table 3 shows the phonetic classification of stop consonants of Kannada.

Place of articulation	Unvoiced		Voiced	
	U	A	U	A
Velar	k	kh	g	gh
Retroflex	ṭ	ṭh	ḍ	ḍh
Alveolar	t	th	d	dh
Dental	t	d	d	dh

Table 3: Phonetic classification of stop consonants in Kannada (U = unaspirated, A = aspirated, M = murmured).

Five qualitatively distinct acoustic segments can be identified for stop consonants as follows:

1 A period of occlusion (silent/voiced)

2 A transient explosion (usually less than 20 ms) produced by shock excitation of the vocal tract upon release of occlusion.

3 A very brief (0-10 ms) period of frication as articulators separate and air is blown through a narrow constriction as in the harmonic fricative.

4 A very brief period of aspiration (2-20 ms) in which may be detected noise excite formant transitions, reflecting shift in vocal tract responses as the main body of tongue moves towards a position appropriate for the following vowel.

5 Voiced formant transitions, reflecting the final stages of the articulatory movement into the vowel during the first few cycles of laryngeal vibration.

There are at least 9 acoustic cues of voicing in stop consonants Table 4 summarizes voicing cues of stop consonants in Kannada.

Sl. No.	Cue	Voiced stop	Unvoiced stop
1	VOT	<+20ms	>+20ms
2	Closure duration	Shorter	Longer
3	Total duration	Shorter	Longer
4	Burst duration	Shorter	Longer
5	Burst intensity	Low	High
6	Overall intensity	High	Low
7	Preceding vowel duration	Longer	Shorter
8	F0	Rising	Falling
9	Voicing	Present	Absent

Table 4: Acoustic cues of voicing in stop consonants of Kannada.

Many of the advances in our understanding of normal speech perception have come from the study of speech perception in degraded listening conditions caused by masking noise or filtering or minimal synthesis. Through the process of elimination, such studies inform us to which properties of the speech signal are important for phonetic perception, and also which aspects of the speech signal are resistant to distortion.

Interestingly there have been few studies to examine the degradation produced by the normal vocal tract- whisper. Whisper is a mode of speech in which there is no vibration of the vocal folds. Unlike in normal speech, the ‘arytenoids’ are slightly abducted and “toed in” in whispered speech (Zemlin, 1988). This creates small chink, which can be of different shape, through which air gushes with turbulent force. Whispered speech is produced with more open glottis than in normal voicing and with longer syllable durations and stop closure intervals. Thus voicing is absent in whispered speech.

Whispering is also a vocal behavior of phonation. It has a special role in communication. It is used in situations, unlike the use of voice, to restrict the communication to a single listener or to a small group. Whisper may also be an indicator of laryngeal pathology.

Whisper is of interest for the same reasons as is the perception of mechanically produced distorted speech, and also because it is important to understand this naturally occurring communication mode.

CHAPTER - III

Method

Subjects: Ten male and ten female normal Kannada speakers in the age range of 18-25 participated in this experiment. The mean age of male subjects was 19 years and that of female subjects was 18 years

Material: Thirty-six CV syllables with stop consonants of Kannada in /a/, /i/, /u/ vowel combination, and eight VCV syllables with vowel /a/ formed the material.

Kannada has voiced and unvoiced stop consonants that occur in word initial and word -medial positions. Kannada is a Dravidian language spoken by 5 crore people in the state of Karnataka(sensus 2001). Table 5 shows the stop consonants of Kannada and table 6 shows material of the study

Place of articulation	Velar	Retroflex	Dental	Bilabial
Unvoiced	k	ṭ	t	p
Voiced	g	ḍ	d	b

Table 5: Stop consonants of Kannada.

Manner	Stops	CV	CV	CV	VCV
Velar	k	ka	ki	ku	aka
	g	ga	gi	gu	aga
Retroflex	ṭ	ṭa	ṭi	ṭu	aṭa
	ḍ	ḍa	ḍi	ḍu	aḍa
Dental	t	ta	ti	tu	ata
	d	da	di	du	ada
Bilabial	p	pa	pi	pu	apa
	b	ba	bi	bu	aba

Table 6: Material used in the study.

Procedure: Subjects were individually tested. Each syllable was written on a 3” x 3” flash card. Subjects read the syllables presented visually under two conditions- normal and whisper- three times. All the productions were audio recorded with the microphone positioned 10 cm away from the mouth of the speakers on to an audiocassette using Sony tape recorder. The syllables were then further recorded on to the memory of the computer at 16000 k Hz sampling frequency.

Perceptual and acoustic analysis was done. The tokens were given for perceptual analysis to determine the perception of intended syllables. Three female Kannada speaking Speech Language Pathologists perceptually analyzed the samples. Samples were presented to them trough headphones. They recorded the syllables on response sheet. Percent correct response for each stop consonant was calculated using the following formula:

$$\text{Percent correct response} = \frac{\text{No. of correct response}}{\text{Total no. of tokens}} * 100$$

Stops that were identified more than 66.6% of times were considered for acoustical analysis. There were 1920 tokens in normal condition and 1101 in whispered condition .All the tokens in normal condition were identified correctly but among 960 whispered unvoiced stops 920 were identified correctly and among 960 whispered voiced stops 181 were identified correctly. The speech filing system (SFS) Software and SSL Pro_2V_2 (Voice and Speech Systems, Bangalore) were used for the extraction of the acoustic parameters. Voice onset time, burst duration, burst amplitude, closure duration and total duration were extracted through SFS and transition duration was calculated using SSL Pro_2V_2. Following were the acoustic parameters extracted in word initial and word-medial position of the syllables.

Word -initial position

Voice onset time (VOT): It was measured as the time duration between the articulatory release and the onset of voicing. Measurements of VOT have been widely used for contrasting voiced with unvoiced stops. Usually voiced stops have voicing before the articulatory release and hence the VOT is leading. This is termed lead VOT or - VOT. In the production of unvoiced stops, voicing occurs after the release. That is, voicing is laging. Therefore, VOT is termed lag VOT or + VOT. In figure 1 vertical bars shows the measurement of VOT for voiced and unvoiced stops.

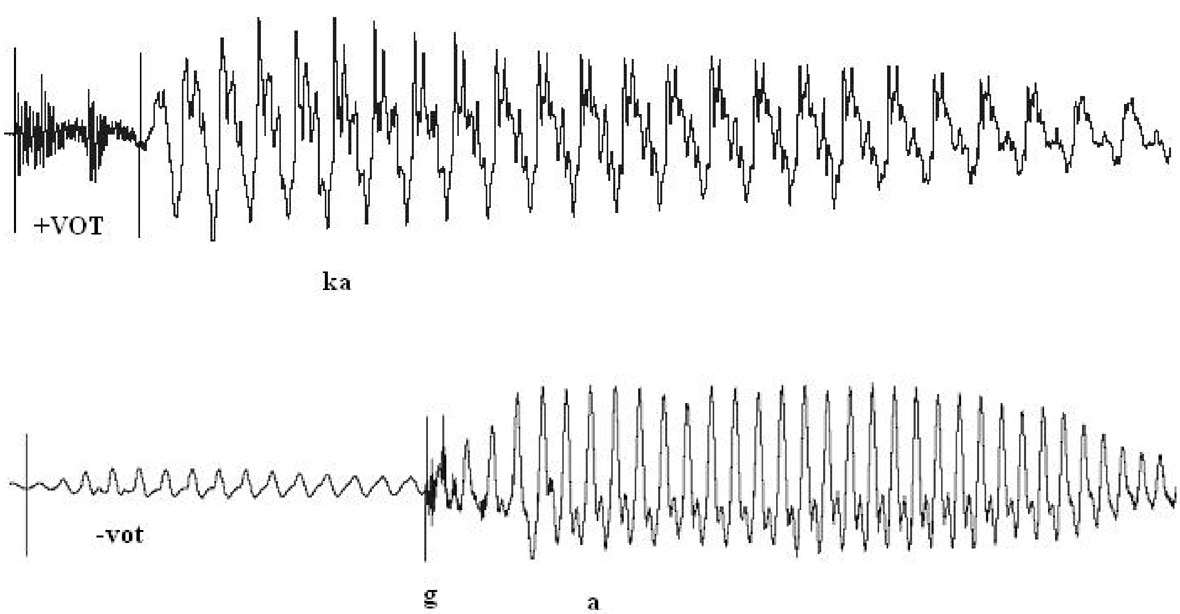

Figure 1: Illustration of measurement of VOT for /k/ and /g/.

Burst duration (BD): It is the time duration for which the burst or the transition excitation occurs. The burst duration is longer in unvoiced plosives compared to that of voiced plosives. Figure 2 depicts the burst duration

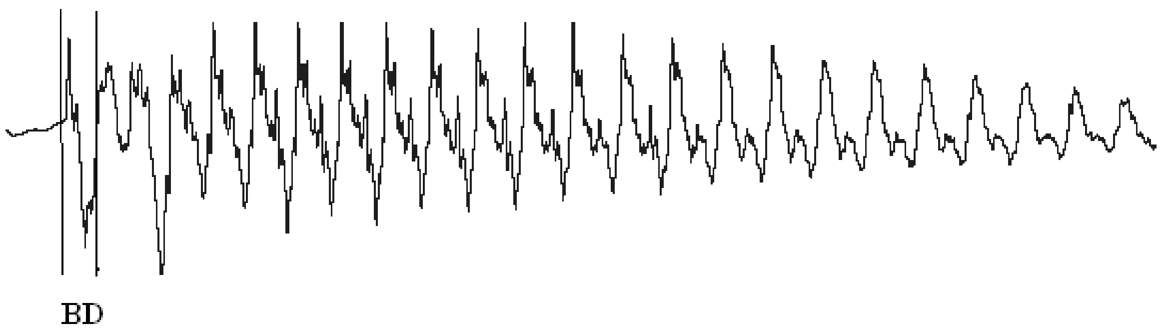

Figure 2: Illustration of measurement of burst duration (BD) for /p/.

F_2 Transition duration (TD): It is the time duration between the onset of F_2 to the steady state of F_2. In figure 3 vertical bar shows the measurement of transition duration of vowel /a/ following /g/.

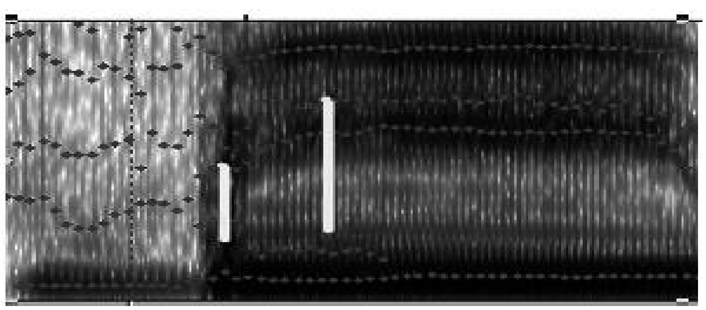

Figure 3: Illustration of measurement of F_2 transition duration in /a/ following /g/.

Burst Amplitude: Burst amplitude was measured as the peak amplitude of burst on amplitude curve.

Word-medial position

Closure duration: It is the interval of stop closure indicating the time for which the articulators are held in position for a stop consonant. Figure 4 illustrates measurement of closure duration.

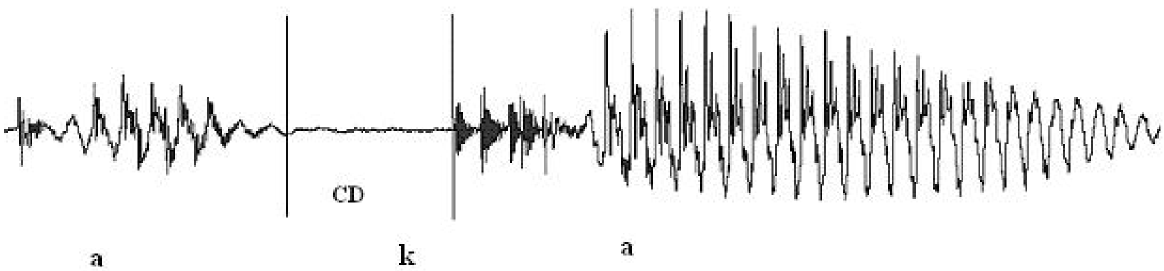

Figure 4: Illustration of measurement of closure duration in CVC /aka/.

Total duration: It is the total time taken for the production of bisyllabic word (e.g. aka). Vertical lines in figure 5 illustrates the measurement of total duration.

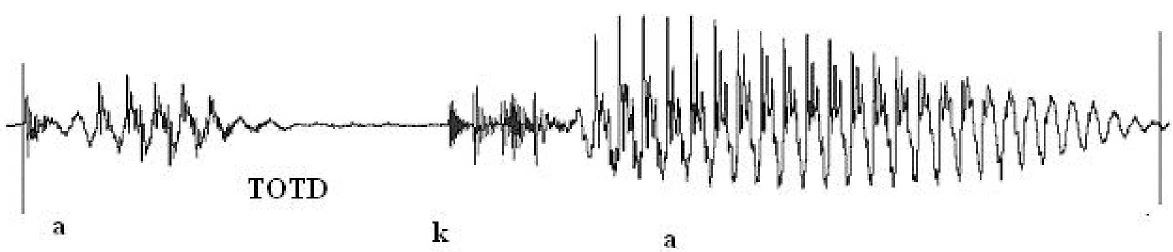

Figure 5: Illustration of measurement of total duration of /aka/

Statistical Analysis: The mean and standard deviation of all the parameters of 20 subjects under 2 conditions were calculated. The S-ratio of each parameter was measured as the difference in the measure of the parameter for voiced and unvoiced stop consonant. For example, if the closure duration of the voiced stop consonant is 60 ms and that of the unvoiced stop consonant is 90 ms, S-ratio for closures duration is equal to 30 ms (90-60 ms) S ratio will indicate whether a given parameter can contrast voicing. For example if the S-ratio for transition duration is 3 ms and that for closure duration is 30 ms in word-medial position, then closure duration would be considered as a stronger parameter contrasting voicing. Parameters that had better S-ratio in whispered condition than in normal condition will cu voicing in whisper. Table 7 gives an example of this.

Parameters	Normal	Whisper
VOT	130	0
Burst duration	5	20

Table 7: Illustration of an example of S- ratio of two parameters in normal and whisper conditions.

CHAPTER - IV

Results and Discussion

S-ratio of transition duration, burst duration, burst amplitude, closure duration and total duration were higher in whispered condition compared to normal condition. Table 8 shows the mean and standard deviation of S- ratio in normal and whispered conditions.

Parameters	Normal	Whisper
VOT	158	77
Transition duration	0	0.25
Burst duration	0.50	1
Burst amplitude	0.5	3.5
Closure duration	32	44
Total duration	2.5	27.78

Table 8: S-ratio in normal and whispered condition.

Voice onset time: S-ratio in normal condition was higher than that in whispered condition on all place of articulation. Bilabial stops had the higher S-ratio and retroflex stops had the lowest S-ratio in normal condition. But velar stops had the highest S-ratio and retroflex stops had the lowest S-ratio in whispered condition Tables 9 and 10 show the mean and S-ratios for voiced and unvoiced stops in normal and whispered voice conditions.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	43	21	28	29	29
Voiced	-111	-121	-122	-155	-117
S-ratio	154	142	150	184	158

Table 9: Mean and S-ratio of VOT in normal condition.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	39	19	23	28	27
Voiced	-79	-40	-45	-36	-50
S-ratio	118	59	68	64	77

Table 10: Mean and S-ratio of VOT in whispered condition.

Transition duration:_S-ratio for transition duration in whispered condition was greater than the normal condition. Tables 11 and 12 show mean transition duration and S-ratio in 2 conditions.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	65	67	68	62	66
Voiced	69	66	67	62	66
S-ratio	-4	1	1	2	0

Table 11: Mean and S-ratio of transition duration in normal condition.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	61	63	64	61	62
Voiced	62	67	68	53	62
S-ratio	-1	-4	-4	8	-0.25

Table12: Mean and S-ratio of transition duration in whispered condition.

Burst duration: S-ratio was higher in whispered condition compared to normal condition. Tables 13 and 14 show mean burst duration and S-ratio in 2 conditions.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	9	7	6	7	7
Voiced	11	6	7	7	8
S-ratio	2	-1	1	0	.50

Table 13: Mean and S- ratio of burst duration normal condition.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	10	6	6	7	7
Voiced	6	6	6	7	6
S-ratio	-4	0	0	0	-1

Table 14: Mean and S-ratio of burst duration in whispered condition.

Burst amplitude: S-ratio was higher in whispered condition compared to normal condition. Also, in normal condition no difference in burst amplitude between voiced and unvoiced stops was observed. But in whispered condition unvoiced stops had higher burst amplitude compared to voiced stops. Tables 15 and 16 show mean burst amplitude and S-ratio in 2 conditions.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	63	64	60	65	63
Voiced	62	65	61	66	64
S-ratio	-1	1	1	1	.50

Table 15: Mean and S-ratio of burst amplitude normal condition.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	62	62	58	64	61
Voiced	57	59	55	61	58
S-ratio	-5	-3	-3	-3	-3.5

Table 16: Mean and S-ratio of burst amplitude in whispered condition.

Closure duration: S-ratio was higher in whispered condition compared to normal condition. Unvoiced stops had larger closure duration compared to voiced stops in both condition. Bilabials had the longest closure duration in normal condition and voiced bilabial stops had the longest closure duration (among voiced) in whispered condition. Also, voiced stops in whispered condition had shorter closure duration compared to voiced stops in normal condition. Tables 17 and 18 show the mean closure duration and S-ratio in 2 conditions.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	102	108	111	113	108
Voiced	75	63	80	89	77
S-ratio	27	45	31	24	32

Table 17: Mean and S-ratio of closure duration in normal condition.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	107	111	111	111	110
Voiced	70	48	69	76	66
S-ratio	37	63	42	35	44

Table 18: Mean and S-ratio of closure duration in whispered condition.

Total duration: S-ratio in whispered condition was higher compared to that in normal condition. Voiced stops had shorter closure duration in whispered condition compared to that in normal condition. Tables 19 and 20 show the mean total duration and S-ratio in 2 conditions.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	440	444	440	438	440
Voiced	446	425	447	434	438
S-ratio	-6	19	-7	4	2.5

Table 19: Mean and S-ratio of total duration in normal condition.

Place of articulation	Velar	Retroflex	Dental	Bilabial	Average
Unvoiced	428	415	422	391	414
Voiced	442	401	394	308	287
S-ratio	-14	14	28	83	27.78

Table 20: Mean and S-ratio of total duration in whispered condition.

Discussion

The results of the study indicate that all parameters investigated, except VOT, had higher s-ratio in whispered condition compared to normal condition. This is in agreement with the results of Murray (1990) in that VOT did not cue voicing in whispered speech.

Among the 5 parameters - transition duration, burst duration, burst amplitude, closure duration and total duration - closure duration appears to contrast voicing in whisper followed by total duration. The reason for this higher S-ratio is attributable to the shorter closure duration in 'voiced' stops and longer closure duration in unvoiced stops in whispered condition. A question arises as to why the same contrast was not evident in total duration as total duration is the sum of consonant and vowel duration (in this study). One probability is that all stops, irrespective of voicing, were aspirated in whispered condition. Aspiration will have lengthened the total duration of voiced stops. Hence, the contrast between 'voiced' and 'unvoiced' stops might be reduced in whispered condition. The other three parameters did not strongly contrast voicing in whisper.

In the present study only the production aspects were investigated. It will be interesting to to see whether closure duration cues voicing in perception also. Future experiment may lengthen closure duration in the 'voiced' stops in whispered condition to see if the perception changes from 'voiced' to 'unvoiced'. Future studies may also investigate whether aphonics use longer phoneme durations to cue voicing

CHAPTER- V

Summary and conclusions

Whisper is a mode of communication, which is more like phonation It has a special role in communication. It is used in situations, unlike the use of voice, to restrict the communication to a single listener or to a small group. Whisper may also be an indicator of laryngeal pathology.

Air is the primary source for both the condition. Even though they share some similar feature there are many differences between whisper and normal phonation. Many authors have tried to study this naturally occurring communication mode mainly to understand the disordered condition like aphonia and dysphonia. It was interesting to know the parameters that were helping to cue voicing in whisper.

We are able to perceive whispered speech in the absence of vibration of the vocal folds. But, there is hardly any information about the acoustic cues of voicing in whispered consonants. Six studies on acoustics of whisper have been conducted sofar. Schwartz (1971) reported that the closure duration for /p/, /b/ under whispered speaking condition in young adult females was longer than that in whispered /m/. Dannenbring (1980) assessed perception of voicing in whispered consonants followed by vowels /a/, /i/, and /u/. Twelve subjects were able to discriminate between whispered consonants. Kallail & Immanuel (1984) reported higher vowel formants, especially $F_{1,}$ in whispered vowels compared to phonated vowels. In contrast, Divya & Tanusree (2005) reported shorter vowel duration and higher F_2 , in whispered speech compared to phonated speech. They did not find any significant difference between closure and consonant duration in two conditions. Tarter (1989) investigated the acoustic features important in identifying whispered consonants. The results indicated preservation of high frequency cues to voicing. Six untrained subjects identified 18 different whispered initial consonants significantly better than chance in nonsense syllable. The phonetic features of place and manner of articulation and to a lesser extent voicing were correctly identified. Confusion matrix and acoustic analysis indicated preservation of resonance characteristics for place and manner of articulation and suggested the use of burst, aspiration, or frication duration and intensity and a first formant cut back for voicing decisions.

Of these, studies by Schwartz (1971), Kallail & Emmanuel (1984), and Divya & Tanusree (2005) are on acoustics of vowels/ consonants in whispered speech. Mc Glone & Manning (1979) have investigated pitch perception on whispered speech. Dannenbring (1980) has used a discrimination task for whispered consonants and Tarter (1989) has used identification of whispered stops. There is hardly any information on acoustic cues of voicing in whisper. Therefore the present study investigated the acoustic cues of voicing in whisper. Specifically temporal and spectral measures of stop consonants as occurring in CV and VCV syllables in whispered and normal conditions were extracted and compared.

Ten male and ten female adult subjects with mean age of 19.5 years participated in the study. They read 36 CV syllables with stop consonants of Kannada in /a /, / i / and /u /vowel environments and 8 VCV syllables with vowel /a/ written on 3”x3” flash cards three times in normal and whispered conditions. This was audio-recorded using a Sony high fidelity tape recorder and then transferred on to the computer memory. The sample was then subjected to perceptual analysis by 3 Kannada speaking Speech Language Pathologists. Subjects identified the syllables. Those stops identified in the syllables by more than 66.6% of subjects were considered for acoustic analysis. Voice onset time, transition duration, burst duration, burst amplitude; closure duration and total duration were measured using Speech Filing System and SSL Pro_2V_2 software.

Mean and S-ratio of each parameter was calculated and compared. The results indicated that S-ratio in whispered stops were higher compared to normal stops on all parameters except voice onset time. This is in agreement with the results of Murray (1990) in that VOT did not cue voicing in whispered speech. Table 21 shows the mean S-ratio in normal and whispered conditions.

Parameters	Normal	Whisper
VOT	158	77
Transition duration	0	0.25
Burst duration	0.50	1
Burst amplitude	0.50	3.5
Closure duration	32	44
Total duration	2.5	27.78

Table 21: S-ratio in normal and whispered condition.

Among the 5 parameters - transition duration, burst duration, burst amplitude, closure duration and total duration - closure duration appears to contrast voicing in whisper followed by total duration. The reason for this higher S-ratio is attributable to the shorter closure duration in ‘voiced’ stops and longer closure duration in unvoiced stops in whispered condition. A question arises as to why the same contrast was not evident in total duration as total duration is the sum of consonant and vowel durations. One probability is that all stops, irrespective of voicing, were aspirated in whispered condition. Aspiration will have lengthened the total duration of voiced stops. Hence, the contrast between ‘voiced’ and ‘unvoiced’ stops might be reduced in whispered condition. The other three parameters did not strongly contrast voicing in whisper.

In the present study only the production aspects were investigated. It will be interesting to to see whether closure duration cues voicing in perception also. Future experiment may lengthen closure duration in the ‘voiced’ stops in whispered condition to see if the perception changes from ‘voiced’ to ‘unvoiced’.

Future studies may also investigate whether aphonics use longer phoneme durations to cue voicing. When working with aphonic clients, who almost use a whispered speech, these data could be kept in mind. Further investigations are needed to find out the acoustic characteristics of whisper in children and also a perceptual study to find out what could be the primary cue for identifying voicing in whisper. Studies could be focused on different phonemes other than stops.

REFERENCES

Arkebauer, H. J., Hixon, T. J., & Hardy, J. C. (967). Peak intra oral air pressures during speech. *Journal of Speech and Hearing Research*, 10, 196-208.

Ballenger, (1971*). Disease of the nose, throat, and ear*. Great Britain: London.

Bracket, (1957). Parameters of voice quality, In L. E. Travis (Ed.). *Handbook of speech pathology and Audiology*, 445, New Jersey: Prentice-Hall Inc.

Dannenbring, G. L. (1980). Perceptual discrimination of whispered phoneme pairs. *Perceptual Motor Skills*, 51, 979-985.

Divya, C., & Tanusree, S. (2005). *Acoustics of whisper*. *A paper presented at the 37th Conference of Indian Speech and Hearing Association.*

Higashikawa, M., Green, J. R., Moore, C. A., & Minifie, F. D. (2003). Lip kinematics for /p/ and /b/ production during whispered and voiced speech. *Folia Phoniatrica,* 55, 17-27.

Hirano, M. (1981). *Clinical examination of voice*. Springer-Verlagwein, New York.

Hixon, T., Mead, J., & Goldman, M. (1976). Dynamics of the chest wall during speech production: Function of the thorax, ribcage, diaphragm, and abdomen. *Journal of Speech and Hearing Research*. 19, 297-356.

Hixon,T. (1988). Comment on Cavallo and Baken (1985). *Journal of Speech and Hearing Research,* 31, 726-728.

Kallail, K. J., & Emmanuel, F. W. (1984). An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics*, 12, 175-186.

Klich, R. J. (1982). Effects of speech level and vowel context on intraoral air pressure in vocal and whispered speech. *Folia phoniatrica,* 34, 33-40.

Luchsinger, R., & Arnold, G. (1965). *Voice- Speech- Language- Clinical communicology: Its physiology & pathology*. Belmont, California: Wads Worth Inc.

McGlone, R. E., & Manning, W. H. (1979). Role of second formant in pitch perception of whispered and voiced vowels. *Journal of Phonetics*, 4, 183-187.

Monoson, P., & Zemlin, W. R. (1984). Quantitative study of whisper. *Folia Phoniatrica*, 36, 53-65.

Murray, J.M. (1990). Perception of voicing in whispered stops. *Phonetica*, 47, 173-181.

Murray, T., Brown, W. (1976). Peak intra oral air pressure in whispered stop consonants. *Journal of phonetics*, 4, 183-187.

Schwartz, M. F. (1971). Bilabial closure durations for /p/, /b/, and /m/ in voiced and whispered vowel environments. *Journal of the Acoustical Society of America*, 51, 2025-2029.

Solomon, N. P., McCall, G. N., Trosset, M. W., & Gray, W. C. (1989). Laryngeal configuration and constriction during two types of whispering. *Journal of Speech and Hearing Research,* 32, 161-174.

Stathopoulos, E. T., Hoit, J. D., Watson, P. J., & Solomon, N. P. (1991). Respiratory and Laryngeal Function During Whispering. *Journal of Speech and Hearing Research*, 34, 761-767.

Tarter, V. C. (1989). What's in a whisper? *Journal of the Acoustical Society of America,* 86(5), 1678-1683.

Titze, I. (1994). *Principles of voice production*. New Jersey: Prentice-Hall Inc.

Travis, L. E. (1957). *Handbook of Speech pathology and Audiology*. New Jersey: Prentice-Hall Inc.

Weismer, G., & Longstreth, D. (1980). Segmental gestures at the laryngeal level in whispered speech: Evidence from an aerodynamic study. *Journal of Speech and Hearing Research*, 23, 382-392.

Zemlin, W. R. (1981). *Speech and Hearing Services*. New Jersey: Prentice-Hall Inc.

Zemlin, W. R. (1988). *Speech and Hearing Science: Anatomy and Physiology*. New Jersey: Prentice-Hall Inc.