Acoustics of Whispered Stops

Deepa D & S R Savithri*

Abstract

This study investigated voicing cues of stop consonants in whispered speech. Whispered speech is perceived in the absence of fundamental frequency. But there is hardly any information on the acoustic cues of voicing in whispered consonants. Twenty normal Kannada speaking subjects read visually presented syllables in normal and whispered voice conditions. Acoustic parameters were extracted and S ratio was calculated. The results indicated that transition duration, burst duration, burst amplitude, closure duration and total duration were longer in whispered condition compared to normal condition.

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 as to which properties of the speech signal are important for phonetic perception and 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 communication role, although 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

* Professor of Speech Science, All India Institute of Speech and Hearing, Mysore, India
e-mail: savithri_2k@yahoo.com

of mechanically produced distorted speech and also because it is important to understand this naturally occurring communication mode.

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 whispered consonants. Kallail and Emmanuel (1984) reported higher vowel formants, especially F_1, in whispered vowels compared to phonated vowels. In contrast Divya and 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 investigated the voicing cues of stop consonants in whispered speech. It was hypothesized that there will be no significant difference between the conditions (whispered and voiced) on voicing cues.

Method

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

Materials: 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 (census 2001). Table 1 shows the material of the study

Table 1: Material used in the study.

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

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 readings were audio recorded with the microphone positioned 10 cm away from the mouth of the speaker on to an audio cassette using Sony tape recorder. The syllables were then further recorded on to the memory of the computer at 16000 KHz sampling frequency. Perceptual and acoustic analysis was done. Three female Kannada speaking speech-language pathologists identified the syllables. Samples were presented to them through headphones and subjects recorded their responses on a 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 920 whispered unvoiced stops and 181 whispered voiced stops were identified correctly in whispered condition. 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.

1. Word initial position

i. 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 show the measurement of VOT for voiced and unvoiced stops.

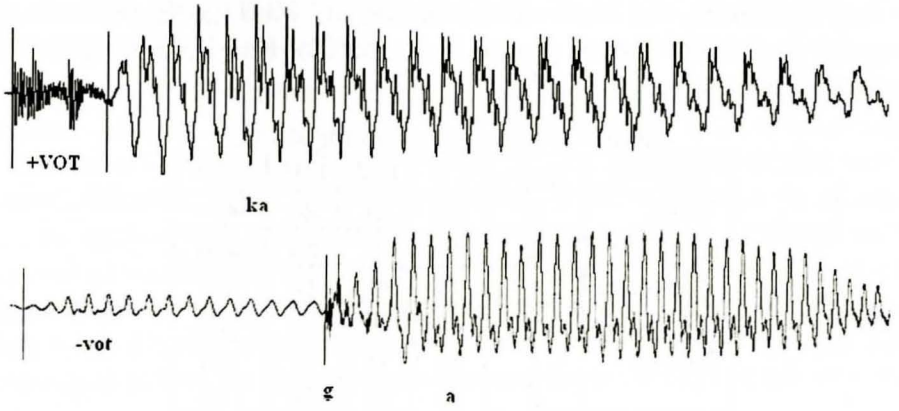

Figure 1: Illustration of measurement of VOT for /k/ and /g/

ii. 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

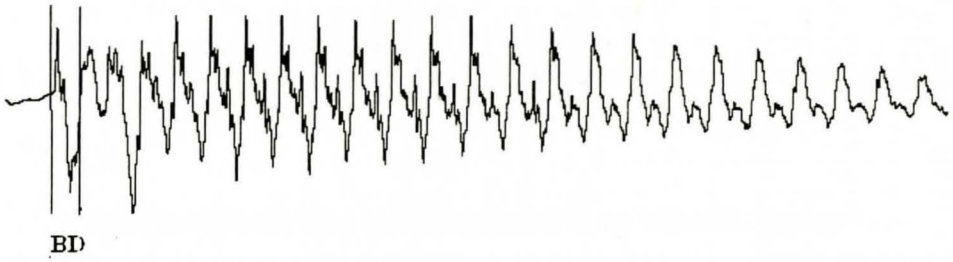

Figure 2: Illustration of measurement of burst duration (BD) for /p/

iii. 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 bars show the measurement of transition duration of vowel /a/ following /g/.

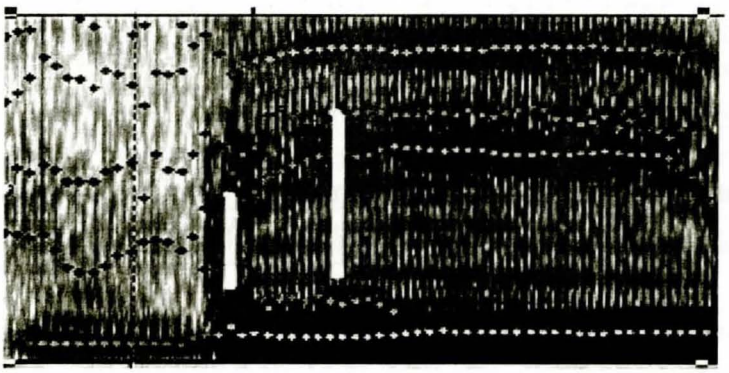

Figure 3: Illustration of measurement of F_2 transition duration in /a/ following /g/

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

2. Word-medial position

i. 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.

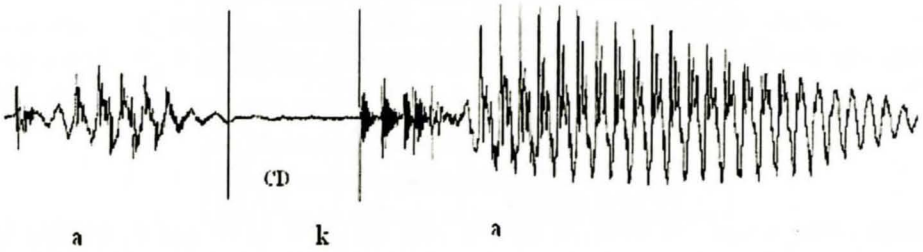

Figure 4: Illustration of measurement of closure duration in CVC /aka/

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

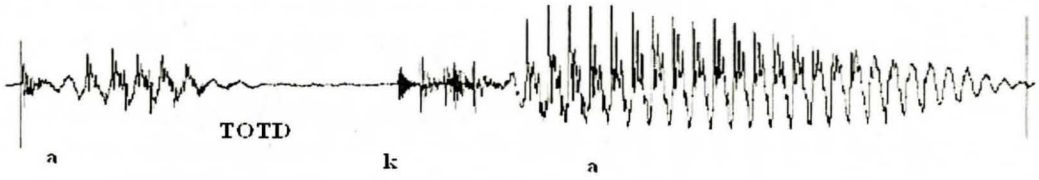

Figure 5: Illustration of measurement of total duration of /aka/

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 closure duration is equal to 30 ms (90-60 ms). S ratio will indicate whether a given parameter can contrast voicing or not. 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 cue voicing in whisper. Table 2 gives an example of this.

Table 2: Illustration of S- ratio of two parameters in normal and whisper conditions

Parameters	Normal	Whisper
VOT	130	0
Burst duration	5	20

Results

S-ratio of transition duration, burst duration, burst amplitude, closure duration and total duration were higher in whispered condition compared to normal condition. Table 3 shows the mean and standard deviation of S-ratio in normal and whispered conditions. Table 4 shows the Mean and S-ratio of all parameters in all places of articulation.

Table 3: S-ratio in normal and whispered condition

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 4: Mean and S-ratio of Acoustic parameters (V = Velar, R = Retroflex, D = Dental, B = Bilabial, N = Normal, W = Whisper, U = Un voiced, V = Voiced).

Parameters/		V		R		D		B		Average	
Voice onset time	U	N	W	N	W	N	W	N	W	N	W
		43	39	21	19	28	23	29	28	29	27
	V	-111	-79	-121	-40	-122	-45	-155	-36	-117	-50
	S-ratio	154	118	142	59	150	68	184	64	158	77
Transition duration	U	65	61	67	63	68	64	62	61	66	62
	V	69	62	66	67	67	68	62	53	66	62
	S-ratio	-4	-1	1	-4	1	-4	2	8	0	-0.25
Burst duration	U	9	10	7	6	6	6	7	7	7	7
	V	11	6	6	6	7	6	7	7	8	6
	S-ratio	2	-4	-1	0	1	0	0	0	.5	-1
Burst amplitude	U	63	62	64	62	60	58	65	64	63	61
	V	62	57	65	59	61	55	66	61	64	58
	S-ratio	-1	-5	1	-3	1	-3	1	-3	.5	-3.5
Closure duration	U	102	107	108	111	111	111	113	111	108	110
	V	75	70	63	48	80	69	89	76	77	66
	S-ratio	27	37	45	63	31	42	24	35	32	44
Total duration	U	440	428	444	415	440	422	438	391	440	414
	V	446	442	425	401	447	394	434	308	438	287
	S-ratio	-6	-14	19	14	-7	28	4	83	2.5	27.8

i. Voice onset time: Bilabial stops had the highest 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.

ii. Transition duration: S-ratio for transition duration in whispered condition was greater than the normal condition.

iii. Burst duration: S-ratio was higher in whispered condition compared to normal condition.

iv. 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.

v. 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.

vi. 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.

Discussion

The results of the study indicate that all the 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 appear 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.

Conclusions

Whisper is a mode of communication which is more like phonation. It has a special role in communication. Whisper may also be an indicator of laryngeal pathology. 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 is interesting to know the parameters that cue voicing in whisper. In the present study only the production aspects were investigated. It will be interesting 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

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.*

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

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.

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

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

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