Voicing Contrast in Tracheoesophageal Speakers

[1]Santosh M., [2]Priyanka Parakh & [3]Rajashekhar B.

Abstract

For optimal rehabilitation, it is essential to evaluate different factors which affect the intelligibility of alaryngeal speakers. Previous studies on Tracheo Esophageal (TE) speech production have mainly focused on aspects of neo-glottal phonation, speaking rate and pausing, and have ignored the changes in other speech characteristics. One such aspect which has not been studied extensively is the voiced/unvoiced distinction. Hence, the present study investigated the voicing contrasts in TE speakers. Two groups of subjects participated in the present study. Group I consisted of six TE speakers in the age range of 45 to 70 years. Group II consisted of six age and gender matched normal laryngeal speakers. Eight meaningful bisyllabic words containing all the voiced and unvoiced plosives (velar, palatal, dental and bilabial) in Kannada, uttered by the subjects were recorded and analyzed perceptually and acoustically. The results of the perceptual analysis revealed that, voicing contrast in TE speakers were near normal except for /p/. Acoustic analyses showed significant differences between acoustic parameters of voiced and unvoiced plosives, endorsing the view that TE speakers make use of multiple acoustic parameters for voicing contrast. The results of the present study have clinical relevance as reduced voicing contrast in case of alaryngeal speakers may indirectly reflect on intelligibility of speech.

Key words: Voice onset time, TE speakers, acoustic analysis

Total laryngectomy is the treatment of choice for individuals with carcinoma of larynx. This procedure alters speech production as there is removal of the larynx and rerouting of respiration through a stoma at the base of the neck. There are three ways of voice restoration after total laryngectomy: the esophageal speech, speech with the assistance of speech aids and the speech produced by the voice prosthesis i.e., tracheoesophageal (TE) speech production. In TE speech production, pulmonary air is shunted from the trachea to the esophagus to set pharyngo-esophageal (PE) segment into vibration.

Previous research with respect to TE speech production has mainly focused on aspects of neoglottal phonation, speaking rate and pausing (Singer & Blom, 1980; Debruyne, Delaere, Wouters & Uwents, 1994). However, changes in other speech characteristics have been largely ignored or have received very little attention. One such aspect is the voiced/unvoiced distinction. The knowledge about the ability of the TE speakers to produce voicing distinction is relevant as there is a change in the anatomy and physiology of voice production in these individuals. In TE speakers the source of voice production is PE segment, located between the third and sixth cervical vertebrae. After total laryngectomy, constriction of the cricopharyngeal muscle narrows the PE segment and results in the formation of a vibrator for pseudo-voice production. PE Segment is, at best, a quasielastic sphincter mechanism, not supported by an abductor –adductor system of muscles like that of the vocal folds within larynx. Rather, it functions as an unpaired, comparatively thick and inelastic fibromuscular mechanism (Dworkin & Meleca, 1997). It is obvious that, this would preclude the precise control of voicing.

The primary acoustic cue for voicing distinction in word-initial position is voice onset time (VOT). Voice onset time is defined as the time interval between the release of the burst and the onset of glottal pulse. Physiologically, VOT reflects the timing coordination between the articulatory and the phonatory systems. The release of stop closure is related to the supralaryngeal articulators such as lips, tongue tip, and tongue dorsum, while the onset of the phonation is a laryngeal event. In Indian context for normal speakers, voiced plosives are characterized by lead VOT and

[1]*Assistant Professor, Department of Speech and Hearing, Manipal College of Allied Health Sciences, Manipal University, Manipal, Karnataka-576104, email:santosh.m@manipal.edu,* [2]*Intern, Department of Speech and Hearing, Manipal College of Allied Health Sciences, Manipal University, Manipal, Karnataka-576104, email:priyanka_aslp@yahoo.co.in,* [3]*Professor and Head, Department of Speech and Hearing Manipal College of Allied Health Sciences, Manipal University, Manipal, Karnataka-576104. email:b.raja@manipal.edu,*

unvoiced plosives are characterized by lag VOT (Lisker & Abramson, 1964; Savithri, Sridevi & Santosh, 2003).

In general, studies on VOT values in TE speakers are few, with contradictory reports. Saito, Kinishi and Amatsu (2000); Searl and Carpenter (2002) reported that unvoiced stops produced by TE speakers were longer than produced by normal speakers. Further, longer VOT values were noticed for unvoiced stops as compared to voiced stops which is comparable to normal speakers. However, Robbins, Christensen, and Kempster (1986) reported that overall VOT values produced by TE speakers were shorter than those by normal speakers, while, Most, Tobin and Mimran (2000) investigating VOT values in Hebrew speakers, reported of no significant difference in VOT values between normal and TE speakers.

Acoustic and physiologic parameters other than VOT also play an important role in signaling a phoneme's voicing feature (Lisker & Abramson, 1967). In word - initial position, lead voice onset time (VOT) and shorter transition duration correlated with voicing. In the word-medial position, shorter closure duration and shorter transition duration correlate with voicing. In word-final position, larger preceding vowel duration and shorter transition duration correlate with voicing (Savithri, Sridevi & Santosh, 2003). There is a paucity of studies in literature on the other parameters and far less is known about these. The evidence of multiple parameters to the voicing feature may prove advantageous for TE speakers who are at risk for difficulty in making the voicing distinction via primary feature; VOT. The TE speakers conceivably could be trained to use secondary cues to enhance the voiced/unvoiced contrast. However, a primary need exists to describe what parameters are being employed by TE speakers to produce unvoiced and voiced phonemes. Also, most of the studies related to voicing distinction in alaryngeal speakers are in English, Mandarian and Hebrew languages. There are no reports in the Indian context. As Gandour, Weinberg, Petty, and Dardarananda (1987) point out, studies of alaryngeal speech in different languages are necessary, as they are expected to distinguish those features that are common across languages from those specific to particular languages. In addition, they contribute essential information to developing model of alaryngeal speech production applicable to all languages. In this regard, the present study was initiated. The present study investigated the voicing contrast in TE speakers.

Method

Subjects: Two groups of subjects participated in the present study. Group I consisted of six individuals who underwent total laryngectomy. All the subjects had undergone secondary tracheo-esophageal puncture (TEP) and used tracheo-esophageal voice as their primary mode of communication for minimum duration of two years. All the TE speakers were native speakers of Kannada and literates in Kannada and English. All of them were males in the age range of 45 to 70 years (mean – 59 years). The post-operative time ranged from three months to seven years (mean – 3.4 years). All the TE speakers used Blom Singer Low pressure (1.8cm) voice prosthesis (choice of the prosthesis was made by the Speech-Language Pathologists) and digital occlusion of the tracheostoma to produce voice. In the present study, proficiency criteria on the use of alaryngeal speaking method were not taken to facilitate generalization. Group II consisted of five age, gender and language matched laryngeal speakers.

Speech material: Eight meaningful Bisyllabic words in Kannada served as the material for the study. These eight words contained all the voiced and unvoiced plosives (velar, palatal, dental and bilabial) in Kannada. All the plosives in word-initial position were followed by vowel /a:/. The words were in a carrier phrase /Idhu ------ a:gidhe/. Table 1 shows the word list.

Place of articulation	Unvoiced	Voiced
Velar	/ka:ru/	/ga:re/
Dental	/tha:ru/	/dha:ri/
Retroflex	/ta:ru/	/dabbi/
Bilabial	/pa:ru/	/ba:ri/

Table 1: Bisyllabic word list in Kannada

Recording: Sentences were written on a flash card and visually presented to subjects in a sound treated room. They were instructed to read the sentences six times at their comfortable pitch and loudness into the microphone kept at a distance of 5 cm away from their mouth. Readings were directly recorded on to the computer memory using external module of Computerized Speech Lab. The recorded words were subjected to two experiments, Perceptual and acoustical analyses.

Experiment I: Perceptual analyses

Procedure: The first syllable of all the bisyllabic words (example /ka:/ of /ka:ru/) was selected and copied using COOLEDIT software and was made as a separate token for perceptual evaluation. Each token was presented thrice in random order. Further, 10% of the samples were re-recorded to check for the reliability. These recorded samples were presented to three trained listeners (SLP with minimum experience of 10 years). The listeners

transcribed the consonants using open response paradigm. Listener's pooled responses were converted to confusion matrices and analyzed for voicing, manner and place of articulation.

Experiment II: Acoustic analyses

Procedure: Waveform display and spectrogram of Computerized Speech Lab (CSL 4500, Kay Elemetrics), which permitted digitization and storage was used for analysis. Each word was displayed as a broadband spectrogram with a pre emphasis factor of 0.80. The analysis size and bandwidth were set to 50 points and 'Hamming' window was used. Spectrograms were displayed as monochrome (black on white) with a grid size of 8x8 pixels (x grid -8 pixels and y grid -8 pixels) with a linear vertical axis. Words were displayed on broadband spectrogram and the target syllables (stop consonant vowel) were 'zoomed in'. The segment was visually and auditorily verified to make sure of the target syllable. Acoustic measures were made using the cursors as follows.

Acoustic measures

(1) Voice onset time (VOT): It is the time difference between the onset of the burst and the onset of the voicing depicted as voice bars on the baseline. During VOT measurement for voiced stops in TE speakers, it was noted that for two speakers pre-voicing was present. In them VOT was measured as showed in figure 1 (lead VOT). However, in rest of subjects there was no pre-voicing. In them VOT was measured as shown in figure 2 (Lag VOT).

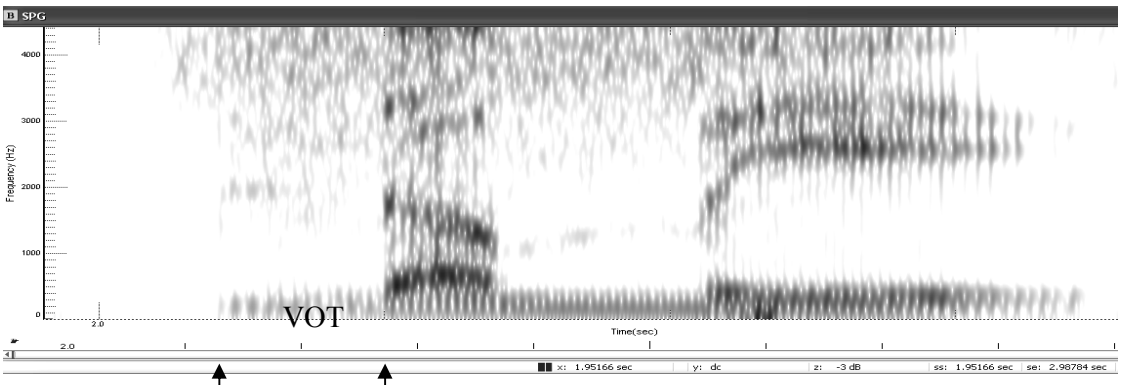

Figure 1: Measurement of VOT for voiced stop /da/ in the word /dabbi/.

(2) Vowel duration (VD): It is the time difference between the onset and offset of voicing of the vowel.

(3) Burst duration (BD): It is the time difference between the onset and offset of the articulatory release.

(4) F_2 transition duration (F_2 TD): It is the time difference between the onset and steady state of the second formant of the following vowel.

(5) Errors on the spectrograms: For TE speakers, different types of errors were identified and classified as visible on wide-band spectrograms like absent voicing, absent burst and weak burst.

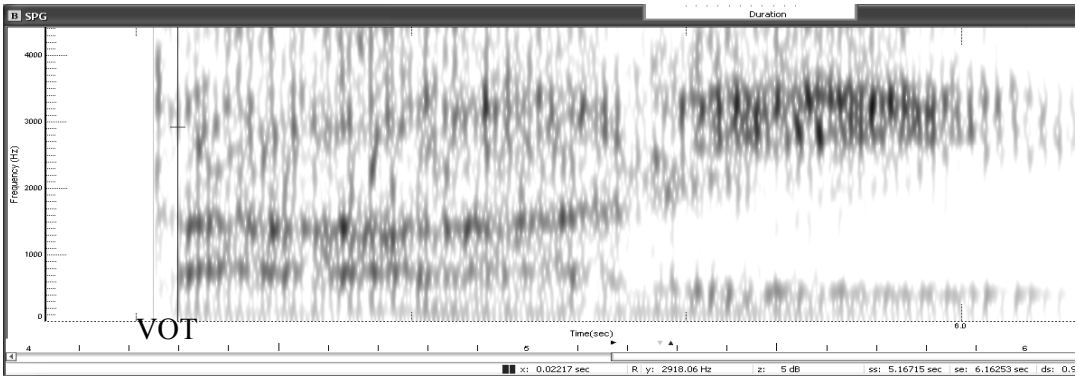

Figure 2: Measurement of VOT for voiced stop /dh/ in word /dha:ri/.

Intra and Inter Judge reliability: For perceptual analysis, both intra- and inter-judge agreement was greater than 80%. For acoustical analysis, the samples were re-analyzed after a time gap of 1 week to check for intra judge reliability. The differences between the two values were less than 5 ms.. Another experienced speech language pathologist who was unaware of the purpose of the study analyzed 10% of the samples. The differences between the two examiners were less than 5 ms.

Statistical analysis: The analyzed data was tabulated for each subject and subjected to statistical analysis. SPSS (Version 11) was used for the statistical analysis. Means and standard deviations were calculated. Paired sampled t-test was done to find the significant difference between VOT of the voiced and unvoiced plosives. Independent sampled t-test was used to find the significant difference between the groups.

Results

Experiment I: Perceptual analyses

The results of the perceptual analysis indicated that in normals all the consonants were identified with 100% accuracy. Whereas in alaryngeal speakers, all the phonemes were identified with high level of accuracy (≥ 70%) except for /p/. Phoneme /p/ was confused with its voiced counterpart /b/, 20% of the time, and also with other phonemes like /m/, /v/, /a/, and /n/. Table 2 shows percent correct identification of plosives in word initial position.

	Stimulus									
		/k/	/g/	/t/	/d/	/th/	/dh/	/p/	/b/	Others
Response	/k/	**74.6**	15.49	0.0	0.0	0.0	1.4	0.0	4.2	/n/= 1.4 /n/= 1.4 /ch/ = 1.4
	/g/	5.4	**80.43**	0.0	0.0	1.08	3.26	0.0	1.08	/r/ = 7.6, /n/ = 1.08, /h/ = 2.15, /a/ = 2.15
	/t/	3.22	0.0	**80.06**	3.22	2.15	6.45	1.07	1.07	/m/ = 1.07
	/d/		0.0	0.0	**100**	0.0	0.0	0.0	0.0	-
	/th/	7.2	0.0	13.54	0.0	**73.95**	4.16	1.04	0.0	-
	/dh/	1.04	0.0	2.08	0.0	9.37	**82.2**	0.0	2.08	/m/ = 1.04, /v/ = 1.04, /a/ = 1.04
	/p/	4.2	0.0	1.05	0.0	0.0	0.0	**66.33**	20	/n/ = 2.1, /v/ = 2.1, /a/ = 1.05, /m/ = 3.10
	/b/		1.04	0.0	0.0	2.08	6.25	0.0	**86.45**	/m/ = 1.04, /n/ = 1.04, /v/ = 2.08

Table 2: Percent correct identification of plosives in word initial position in TE speakers.

Experiment II: Acoustical analyses

1. **Voice onset time:** The mean VOT values were significantly longer for voiced plosives (lead) compared to unvoiced plosives in both the groups. However, in those individuals in group I who had short lag VOT values for voiced plosives, VOT values were longer in unvoiced plosives compared to voiced counterparts. Between groups comparisons showed that the mean VOT values were longer in group I compared to group II. Table 3 shows the mean and SD values for VOT in group I and group II.

Groups	Unvoiced		Voiced (short Lag)		Voiced (Lead)	
	Mean	SD	Mean	SD	Mean	SD
Group I	28.00	17.40	22.35	12.35	108.97	70.31
Group II	25.93	14.24	-	-	90.60	33.17

Table 3: Mean and SD of VOT for unvoiced and voiced plosives in two groups.

2. **Vowel duration:** The results indicated significant difference (group I- (t=3.77, df= 102, p<0.05), group II- (t= 6.46, df=90, p<0.05)) between unvoiced and voiced plosives in both the groups. The mean vowel duration was significantly longer following unvoicedbv plosives compared to voiced plosives in both the groups. The mean vowel duration was significantly longer in group I compared to group II. The results of independent samples't' test showed significant difference (unvoiced- (t= 5.3, df= 192, p<0.05), voiced- (t= 4.25, df= 199, p<0.05)) between groups for both voiced and unvoiced plosives. Table 4 shows the mean and SD values of vowel duration in group I and group II.
3. **Burst duration:** The results showed no significant difference [group I- (t= 1.27, df= 79, p<0.05), group II-(t= 0.654, df= 90, p<0.05)] between voiced and unvoiced

Groups	Unvoiced		Voiced	
	Mean	SD	Mean	SD
Group I	309.19	82.21	264.77	15.76
Group II	260.88	30.03	206.07	76.77

Table 4: Mean and SD values of vowel duration (ms) for preceding unvoiced and voiced plosives in two groups.

plosives. The mean burst duration was significantly longer in unvoiced plosives compared to voiced plosives in both the groups.The between groups comparison showed that the mean burst duration values were significantly longer in group I compared to group II. The results of the paired sampled t test showed no significant difference (unvoiced- (t= 4.28, df= 182, p>0.05), voiced- (t= 3.4, df= 173, p<0.05)) between unvoiced and voiced plosives in both group I and II. Table 5 shows the mean and SD values for burst duration in both the groups.

Groups	Unvoiced		Voiced	
	Mean	SD	Mean	SD
Group I	18.35	8.75	16.31	10.99
GroupII	12.05	7.25	11.46	8.42

Table 5: Mean and SD values of burst duration (ms) for unvoiced and voiced plosives in both groups.

4. **Transition duration:** The results indicated no significant difference [group I- (t= 0.835, df= 102, p>0.05), group II- (t= 0.031, df= 89, p>0.05)] between unvoiced and voiced plosives in both groups. The mean transition duration was longer in unvoiced plosives compared to voiced plosives. Also, the mean transition duration was significantly longer in group I compared to group II. The results of independent samples t test showed significant difference [unvoiced- (t= 2.9, df= 191, p<0.05), voiced- (t= 1.66, df= 196, p> 0.05)] between groups only for unvoiced plosives. Table 6 shows the mean and SD values for transition duration in group I and group II.

Groups	Unvoiced		Voiced	
	Mean	SD	Mean	SD
Group I	48.09	17.89	45.96	20.83
Group II	40.40	20.73	40.50	21.89

Table 6: Mean and SD values of transition duration (ms) for unvoiced and voiced plosives in two groups.

5. **Errors on spectrograms:** Observation of the spectrogram revealed the following aberrant errors.

A. **Absent voicing**: Normal voiced phoneme is characterized by voice bars preceding the burst on the baseline of the spectrograms. The absence of voice bars was noted for voiced phonemes in three of the five subjects. Figure 3 and 4 show the spectrograms of the word /ga:re/ and /dabbi/ in word- initial position indicating absent voicing.

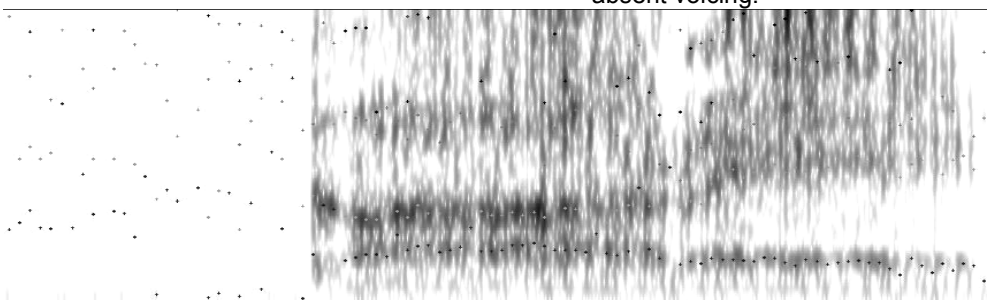

Figure 3: Spectrogram of the word /ga:re/ indicating absence of voice bars before the burst.

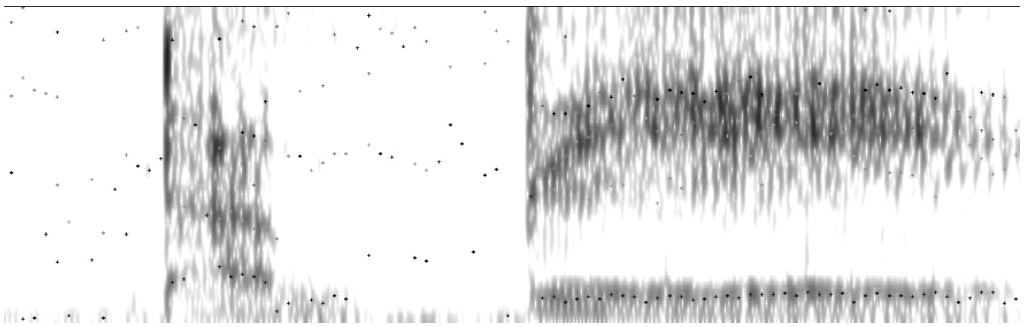

Figure 4: Spectrogram of the word /dabbi/ indicating absence of voice bars before the burst.

B. **Absent burst:** Burst is indicated by noise energy spread as irregular vertical striations across the frequency spectrum on the broadband spectrogram. Figure 5 shows the spectrogram of the word /pa:ru/ indicating absent burst.

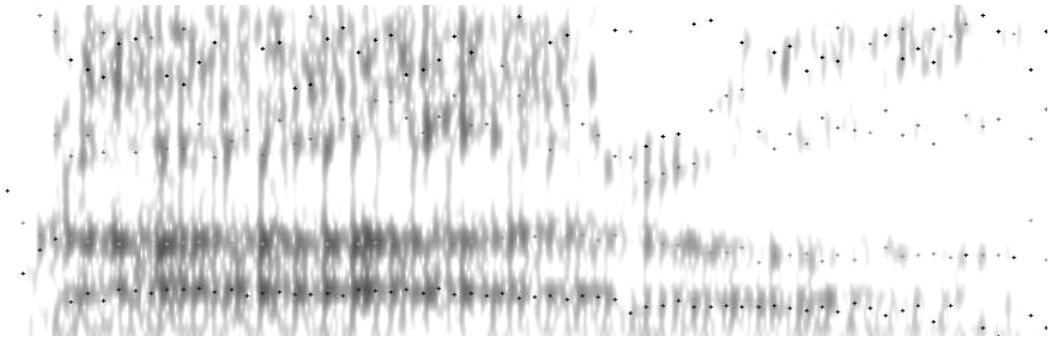

Figure 5: Spectrogram of the word /pa:ru/ indicating absence of burst before voicing bars of the vowel /a/ in the word-initial position.

C. **Weak burst:** Weak burst is characterized by less energy (light vertical striations) on the spectrogram. Weak burst is typically seen in normal speakers for voiced plosives. Weak burst was noted in four of the five TE speakers for both unvoiced and voiced phonemes. Figure 6 shows the spectrogram of the word /ta/ indicating weak burst.

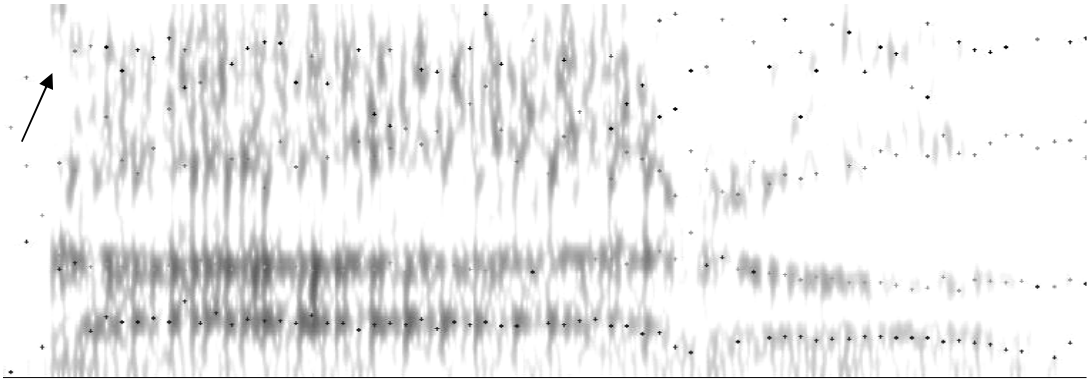

Figure 6: Spectrogram of the word /taru/ indicating weak burst in the word-initial position.

Discussion

The present study investigated the voicing contrasts in TE speakers through perceptual and acoustical analyses. Several points of interest evolved from the study. First, results of the perceptual analysis indicated that voicing contrast in TE speakers were perceived correctly except /p/, which was confused with its voiced counterpart. The results of the present study highlight that TE speakers with good intelligibility and adequate loudness, have the ability to contrast voicing, unlike their esophageal counterparts (Hyman, 1955; Shames, Font, Matthew, 1963; Sacco, Mann, Schultz, 1968; Nicholas, 1976; Hirose, 1996). This difference between TE and esophageal speakers may be due to the difference in air reservoir in TE speakers. In esophageal speakers, due to the limitation of air volume in the oral-pharyngeal air reservoir, it is difficult for them to produce high intraoral pressure, which results in voicing confusion. Whereas in TE speakers, as they make use of lung air, they are able to build sufficient intraoral air pressure which results in voicing distinction. However, it will be interesting to investigate TE speakers' production for aspirated and unaspirated plosives.

Second, significant differences between unvoiced and voiced plosives were observed in both the groups for voice onset time, vowel duration and burst duration in the word-initial position. The mean VOT values were significantly longer in voiced plosives (lead VOT) compared to unvoiced plosives, and mean vowel duration and burst duration values were significantly longer in unvoiced stops compared to voiced stops. This supports the findings of Lisker & Abramson, (1967); Savithri, Sridevi & Santosh, (2003), that there are multiple acoustic cues for distinguishing voicing contrast in the word-initial position. Another important point is about functioning of PE segment. The results of the present study highlight that TE speakers also make use of multiple

acoustic parameters to contrast voicing in word-initial position.

Third, the results also showed significantly longer vowel duration and burst duration (both voiced and unvoiced plosives) and longer transition duration (unvoiced plosives) in TE speakers compared to normal speakers. The reason for longer duration of parameters may be attributed to attempts of alaryngeal speakers to increase articulatory precision (Searl & Carpenter, 2001). The longer VOT in TE speakers when compared to normals can be due to functioning of PE segment. Unlike vocal folds which have to be adducted to begin vibration, the PE segment will be relaxed to assume vibration. In the production of unvoiced plosives in which the intraoral pressure is lower, the time for the release of intraoral pressure is short. Therefore, the onset of vibration depends largely on the timing of PE segment relaxation. Since motor control of neoglottis is limited, it is speculated that the time needed to relax the neoglottis in TE speakers is longer when compared to adduction of vocal folds in normal speakers, yielding a longer VOT in TE speech (Ng & Wong, 2009).

The presence of weak burst and absence of burst reflects on the insufficient oral release of the plosives, which is attributable to insufficient respiratory support. Also, the aberrant spectrograms indicate that three out of the six TE speakers had absent voicing bars for voiced plosives in word-initial positions. This further reveals that PE segment does not appear to have motor control like the vocal folds, for quick abduction and adduction and appropriate coordination with other speech structures.

Conclusions

In the present study, TE speakers could contrast voicing similar to normals with high level of accuracy. Both TE and normal speakers used multiple acoustic parameters for voicing contrast. However, in TE speakers, the values of acoustic parameters were longer when compared to normals. This may be due to absence of the precise control of voicing, which may not be present due to limited motor control of PE segment. The results of the present study have clinical relevance as reduced voicing contrast in alaryngeal speakers may indirectly affect the intelligibility of speech. This study, in addition, gives an overview of the nature of voicing contrast in case of alaryngeal speakers, where the vibrating segment is the PE segment and not the vocal cords as in normals. However, results cannot be generalized owing to the limited number of subjects. Also, trained SLP's were used for the consonant identification which could have led to the higher percent consonant accuracy. Further research incorporating large group of subjects is warranted to substantiate the present study. Further, comparison of acoustical differences between accurately identified and inaccurately identified consonants, between proficient and non-proficient speakers, and comparison of values across place of articulation needs to be done.

References

Debruyne, F., Delaere, P., Wouters, J., & Uwents, P. (1994) Acoustic analysis of tracheoesophageal versus esophageal speech. *Journal of Laryngology and Otology*, *108*, 325-328.

Dworkin, J.P, & Meleca R.J. (1997). *Vocal Pathologies: Diagnoses, treatment and case studies*. San Deigo, CA: Singular.

Gandour, J., Weinberg, B., Petty, S.H., & Dardarananda R. (1987) Voice onset time in Thai alaryngeal speech. *Journal of Speech and Hearing Disorders, 52,* 288-294.

Hirose, H. (1996). Voicing distinction in esophageal speech. *Acta Otolaryngol Suppl*, *524*, 56-63.

Hyman, M. (1955). An experimental study on artificial larynx and esophageal speech. *Journal of Speech and Hearing Disorders, 20,* 291-299.

Lisker, L., & Abramson, A.S. (1964). A Cross language study of voicing in initial stops: Acoustical measurements. *Word*, *20*: 384-422.

Lisker, L., & Abramson, A.S.(1967). Some effect of context on voice onset time in English stops. *Language and Speech, 10*, 1-28.

Most, T., Tobin, Y., & Mimran, R.C. (2000). Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production. *Journal of Communication Disorders, 33,* 165-181.

Ng, M. L. & Wong, J. (2009). Voice onset time characteristics of Esophageal, Tracheoesophageal, and Laryngeal speech of Cantonese. *Journal of Speech and Hearing Research*, *52,* 780-789.

Nicholas, A. (1976) Confusions in recognizing phonemes spoken by esophageal speakers: initial consonants and clusters. *Journal of Communication Disorders, 9*, 27-41.

Robbins, J., Christensen, J., & Kempster, G. (1986). Characteristics of speech production after tracheoesophageal puncture: voice onset time and vowel duration. *Journal of Speech and Hearing Research 29*,499-504.

Sacco, P., Mann, M., & Schultz, M. (1968) Perceptual confusions in selected phonemes in esophageal speech. *Journal of Indiana Speech and Hearing Association*, *6,* 196-203.

Saito, M., Kinishi, M., & Amatsu, M. (2000) Acoustic analyses clarify voiced-voiceless distinction in tracheoesophageal speech. *Acta Otolaryngologica, 120,* 771-777.

Savithri, S.R., Sridevi, M., & Santosh, M. (2003). Voicing Contrasts in Indian Languages: Acoustic measurements. A paper presented at the National Symposium of Acoustics, Pune.

Searl, J, & Carpenter, M.A. (2002).Acoustic cues to the voicing feature in tracheoesophageal speech. *Journal of Speech Language and Hearing Research, 45, 282*-294

Shames, G.H., Font, J.M., & Matthews, J. (1963) Factors related to speech proficiency in the laryngectomized. *Journal of Speech and Hearing Disorders*, *28*, 273-278.

Singer, M.I., & Blom, E.D. (1980) An endoscopic technique for restoration of voice after laryngectomy. *Annals of Otology, Rhinology and Laryngology*, *89*, 529-532.