Acoustic analysis of tracheo-oesophageal voice in male total laryngectomy patients

N Deore[1], S Datta[1], RC Dwivedi[2], R Palav[1], R Shah[1], SI Sayed[1], M Jagde[1], R Kazi[2]

[1]Grant Medical College, Mumbai, India
[2]Royal Marsden Hospital, London, UK

ABSTRACT

INTRODUCTION   The aim of this study was to test the null hypothesis that voice parameters of post-laryngectomy patients using tracheo-oesophageal (TO) prosthetic valves are similar to those of normal laryngeal subjects.
METHODS   Thirty total laryngectomy patients and thirty normal controls were subjected to acoustic analysis of single voice recordings using a sustained vowel. Acoustic parameters including fundamental frequency, jitter, shimmer, harmonics-to-noise ratio and maximum phonation time were analysed.
RESULTS   Poorer values were found as well as larger variability for all the voice parameters for the total laryngectomy patients using TO voice compared with those of normal subjects. There were statistically significant differences ($p<0.05$) for all studied parameters between the TO and normal speech.
CONCLUSIONS   Alaryngeal speech with TO voice prosthesis is not yet comparable to laryngeal speech.

CORRESPONDENCE TO

Nitin Deore, ENT Department, Grant Medical College, Dr Ambedkar Road, Byculla, Mumbai – 400 008, India
T: +91 (0)22 2373 5555; F: +91 (0)22 2379 1015; E: nitin399@gmail.com

With the advances in voice conservative surgeries and radiotherapy techniques, most of the patients with laryngeal cancer can be effectively cured.[1] However, in developing countries such as India, many patients unfortunately present late with advanced stage disease, making a total laryngectomy the only curative approach.[2] The prognosis of laryngectomised patients has remained relatively favourable over the years, with five-year survival rates of 65–75%.[3] Nevertheless, the procedure has its own set of functional, physiological and psychosocial consequences. In addition to the loss of voice, there can be loss of olfaction, a poor cough reflex, swallowing difficulties, pulmonary changes and complications associated with a permanent tracheostoma.[4]

The importance of speech is not appreciated until it is lost as speech forms an important part of day-to-day life in a world that relies heavily on verbal communication.[5] The functional rehabilitation of laryngectomised patients has been a major concern for head and neck cancer surgeons and speech and language therapists. Various developments in speech rehabilitation over the past three decades have led to improvements in the quality of life of these patients.[6,7]

Voice quality is a multidimensional component of fluent speech and has a varied application both culturally and physiologically. Voice quality plays an essential role in verbal communication. It can help or hinder intelligibility and

is a rich source of indexical information with linguistic, cultural and family determinants.[6,7] Rapid reestablishment of an acceptable voice and fluent, intelligible speech is critical to successful rehabilitation and psychological adjustment in laryngectomised patients. Ever since the introduction of surgical voice restoration using a prosthesis two decades ago, the tracheo-oesophageal (TO) voice rehabilitation has become the 'gold standard' method of voice restoration in post-laryngectomy patients.[8]

The technique of voice restoration involves creating a simple puncture between the posterior wall of the tracheostome and the anterior wall of the lower pharynx (TO puncture [TOP]), into which a one-way silicone valve is inserted. The prosthesis serves as a one-way valve to prevent soiling of the airway and opens to divert pulmonary air across the neoglottis on closure of the stoma during expiration. The basis of surgical voice restoration is that the shunted air vibrates the mucosa of the neoglottis to produce a sound that is finally articulated by the rest of the vocal tract to produce speech.[9]

In TO speech, as in a normal speech, the pulmonary air is used for voice production. Many acoustic measures of speech including fundamental frequency (F_0), jitter, shimmer and harmonics-to-noise ratio (HNR) were used by many researchers to study TO, oesophageal and nor-

Table 1   Comparison of acoustic parameters of normal subjects and tracheo-oesophageal (TO) speakers

Voice parameters	Normal subjects (n=30)	TO speakers (n=30)	p-value
Average fundamental frequency	144.04Hz	110.31Hz	0.0004
Jitter	0.18%	2.18%	0.0003
Shimmer	0.95%	6.77%	<0.0001
Harmonics-to-noise ratio	25.03	11.41	<0.0001
Maximum phonation time	23.87 secs	6.87 secs	<0.0001

mal speech. They found that TO speech is more similar to normal speech than oesophageal speech for frequency and duration variables and that the intensity of TO speech is greater than oesophageal speech.[8]

Acoustic analyses of pathological voices have provided one of the most attractive methods for assessing vocal function. Such analyses not only have the advantage of being non-invasive but also provide quantitative information for assessment of voice function. The main aim of this study was to test the null hypothesis that speech parameters of post-laryngectomy patients using TO prosthetic valves are similar to those of normal laryngeal subjects. A further objective of the study was to test the vocal parameters of normal subjects and compare these to post-laryngectomy subjects using TO speech.

Methods

Thirty consecutive laryngeal cancer patients who underwent a total laryngectomy and were using TO voice as mode of communication were recruited for this study. The voice samples of these patients were compared with the voice samples of 30 normal subjects.

TO speakers

Thirty patients who underwent a total laryngectomy (with or without a partial pharyngectomy), primary closure and primary TOP between June 2007 and December 2009 at a multidisciplinary teaching hospital in Mumbai were included in this study. Prior approval of the local research ethics committee was obtained for the study.

All 30 patients had stage III/IV disease, of which 22 had laryngeal and 8 pyriform fossa cancers. None of the patients had distant metastasis. Salvage surgery for radiation failure was performed in five patients. Out of 30 TO speakers, 22 had a primary total laryngectomy and the remaining 8 had undergone a partial pharyngectomy with reconstruction using a pectoralis major myocutaneous flap. Postoperative radiation therapy (65Gy in 30 fractions) was given to 24 patients.

All patients underwent a primary TOP and had a 14Fr Foley catheter inserted in the surgically created fistula, which, after a gap of 10–14 days, was replaced by an appropriately sized indwelling Blom–Singer® voice prosthesis (InHealth Technologies, Carpinteria, CA, US).[2] A short cricopharyngeal myotomy was carried out in all the cases on the operating table. All TO speakers were disease free and were using TO valves of the Blom–Singer® type. The median voice prosthesis length was 8mm (range: 4–12mm). In total, 45 Blom–Singer® prostheses were required for the 30 patients over the period of study (2.5 years).

Normal subjects

The normal subjects were recruited from the social and work circle of the first author (ND). Subjects intended for the control group with a history of voice disorders, orofacial abnormalities, severe respiratory and allergic problems, inadequate hearing abilities and audible deviant voice qualities (as judged from their conversations with the author) were excluded from the study.

Equipment for voice recording and analysis

All the subjects underwent voice analysis in the voice laboratory of the hospital. The hardware and software of Dr Speech (Tiger DRS Inc, Seattle, WA, US) were used for acoustic analysis. The sound was picked up by a microphone (unidirectional condensed) placed in front at a constant mouth-to-microphone distance of around 5cm and at an angle of 45º. The signal was transmitted to the microphone pre-amplifier and converted from the analogue waveforms into digital form by Dr Speech software. The digital file was saved on a Pentium IV computer (sound card with 256MB RAM, 80GB free hard drive space) running Windows® XP.

Voice recording protocol

In this study, acoustic analysis of the sustained vowel /i/ was utilised as a tool to assess, document and investigate TO voice. The vowel /i/ was the preferred vowel for analysis as it puts the vocal folds in their most stressed configuration, giving greater contrast with normal production. Furthermore, the vowel /i/ specifically has been studied robustly in the literature.[10–12] The protocol was explained to all the patients in advance in order to familiarise them with the process and they were allowed a few attempts prior to commencing the recording.

The protocol was as follows:
1. Sustained vowel /i/ produced at a comfortable pitch and loudness for at least five seconds (or as long as the subject could manage stably);
2. Maximum phonation time (in seconds): sustained vowel /i/ produced at a comfortable pitch and loudness after a maximal deep breath.

Table 2 Mean acoustic parameters of normal subjects and tracheo-oesophageal (TO) speakers in the literature and this study

Studies	Parameter									
	Fundamental frequency (Hz)		Jitter		Shimmer		Words per minute		Maximum phonation time (secs)	
	Normal	TO speakers	Normal	TO speakers	Normal	TO speakers	Normal	TO speakers	Normal	TO speakers
van As, 2001[6]	110	103	NR	6.8	NR	NR	NR	NR	26.0	12.8
Pindzola, 1988[7]	128.4	107.7	2.0	4.6	NR	NR	158.8	152.2	24.9	16.4
Blood, 1984[15]	120.8	88.3	NR	NR	NR	NR	NR	NR	NR	NR
Robbins, 1984[16]	102.8	101.7	0.8	0.8	0.3	0.8	172.8	127.5	NR	NR
Qi, 1995[17]	131.9	86.0	NR	NR	NR	NR	NR	NR	NR	NR
Kazi, 2006[18]	171.3	103.8	0.4	5.9	0.9	2.1	165.8	134.5	23.9	11.8
Baggs, 1983[19]	NR	NR	NR	NR	NR	NR	182.5	132.4	19.9	10.9
Robbins, 1984[20]	NR	NR	NR	NR	0.4 (ratio)	10.6 (ratio)	NR	NR	21.8	12.2
Debruyne, 1994[21]	NR	NR	NR	NR	NR	NR	NR	NR	NR	6.0
Present study	144.04	110.31	0.18%	2.18%	0.95%	6.77%	NR	NR	23.87	6.87

NR = not reported

Acoustic parameters evaluated during the study

All subjects provided acoustic recordings of the sustained vowel /i/ at a comfortable pitch and loudness. The stable mid-portion of the recording for analysis in a single session was used for analysis. Voice parameters included F_0, jitter, shimmer, HNR and maximum phonation time (MPT). F_0 is the lowest frequency (first harmonic) of a periodic signal. Jitter is the cycle-to-cycle variability of the pitch period or F_0. It is a measurement of how much a given pitch period differs from the one or several pitch periods that immediately precede or follow it. Shimmer or amplitude perturbation is a measure of cycle-to-cycle fluctuation in waveform amplitude. This measure is sometimes made with peak-to-peak amplitude or sometimes peak amplitude. In vocal assessment we use peak amplitude. HNR is the ratio of harmonic energy to noise energy.

Statistical analysis

All data were analysed using SPSS® (SPSS Inc, Chicago, IL, US) v14. The data from normal subject and TO speaker groups were compared using a t-test. A p-value of <0.05 was taken as significant.

Results

The mean age of the TO patient group at the time of voice assessment was 61 years (standard deviation [SD]: 8.0 years, range: 29–74 years). The median time from completion of treatment to voice assessment was 7 months (range: 3–30 months). The mean age of the control cohort was 61 years (SD: 7.3 years, range: 34–72 years).

Acoustic analysis

Data for acoustic analysis are presented in Table 1. The average fundamental frequency for the normal subjects was 144.04Hz. In comparison, the TO speakers had an average fundamental frequency of 110.31Hz, which is significantly lower. The mean jitter value for TO speakers was 2.18%, which was significantly higher than that of the control group (0.18%). Similarly, the mean shimmer value for patients with TO speech was significantly higher than that of normal subjects (6.77% vs 0.95%). The mean HNR for TO speakers was 11.41 whereas for normal speakers it was 25.03. The mean MPT for TO speakers was 6.87 seconds, which was significantly lower than that of the control group (23.87 seconds).

Discussion

Ever since the first laryngectomy was performed, postoperative voice rehabilitation has been of major concern to head and neck surgeons. Gussenbauer, Billroth's assistant, used a reed valve placed through a temporary pharyngostome so that expired air could be set into vibration by the valve and redirected through the pharynx and buccal cavity, allowing speech production.[15] The main purpose of the pharyngostome was to prevent pulmonary complications secondary to wound breakdown, a common occurrence at that time. With the development of improved operative techniques avoiding the necessity of a temporary pharyngostome, this method of voice production was lost. Since then various alternative techniques have evolved.

TO voice differs significantly from pathological voice in that it is more aperiodic and of a lower fundamental frequency.[8] Many groups have attempted to analyse the voice using a range of methods. One commonly followed method is the use of acoustic signal typing, as described by Titze,[14] which has been widely applied in TO speakers.[6,10,14] In our study, acoustic analysis of the sustained vowel /i/ was used as a tool to assess, document and investigate TO voice. The cohort of TO speakers in this study yielded results similar to other studies using speech signal analysis (Table 2).

The TO speakers had an average fundamental frequency of 110.31Hz, which was significantly lower than that of the normal controls (144.04Hz). This is comparable with findings of other studies.[6,7,15,17,18,21] In the present study, the mean jitter value of TO speakers was 2.18%, which was significantly higher than that of the control group (0.18%). Shimmer was similarly higher for TO speakers (6.77%) than for normal subjects (0.95%). Both of these differences were statistically significant ($p<0.0001$). These findings are consistent with other studies available in the literature.[6,7,16,17,19,21] The results suggest that although a fairly regular and stable vibratory mode is obtained in TO speakers, it is still significantly poorer than that of laryngeal speakers.

In our study we observed that HNR was lower for TO speakers (11.41) than for normal speakers (25.03), meaning there was more noise or aperiodicity on sounds than the harmonics produced by the neoglottis. This is understandable as sound is also produced by leakage at the time of stoma occlusion.

The temporal measure of MPT for TO speakers was 6.87 seconds, which was significantly shorter than that of normal subjects (23.87 seconds). These results are consistent with the findings in other studies[6,7,17–21] and can be explained by the fact that TO speakers have reduced breath support due to varying amounts of air leakage at the stoma occlusion. Furthermore, they have to alternate constantly between conspicuously drawing air into the lungs through the stoma and stoma occlusion with a finger to produce voice naturally, resulting in slower speaking rates.[21] The association of an increased MPT with a surgical myotomy during a laryngectomy is very interesting and is possibly due to its influence on tonicity of the neoglottis. A tonic neoglottis would be capable of increased MPT and is something very desirable.[21]

Normal speech is characterised by smooth onset, offset and absence of pitch breaks while the same cannot be said for alaryngeal speech, which lacks fine motor control.[8,12] This variability is a characteristic feature of TO speech and could be a result of the larger anatomical and morphological variation of the neoglottis compared with the vocal folds.[6,12] Another possible reason could be the inclusion of the patients who had undergone partial or total pharyngeal reconstruction.

There appears to be an association between F_0 and treatment variables such as tumour stage, neoglottis closure, complications and reconstruction. A higher fundamental frequency is seen with patients with a horizontal neoglottis closure, with no complications and no reconstruction. F_0 is determined by the equation

$$Fx = \frac{1}{2Lv(T/P)}$$

where Lv is the length of the vocal fold, T is the mean longitudinal stress and P is the tissue density.[10] In the case of the neoglottis, the myoelastic properties are clearly different. The mean longitudinal stress (T) is small and the tissue density (P) is large, producing a lower fundamental frequency. The data also suggest that where there has been pharyngeal reconstruction, F_0 is substantially lower.

Conclusions

This study shows that robust, reliable and consistent data could be obtained using acoustic measures of voice in normal volunteers and laryngectomees with a sustained vowel. These are of particular value in demonstrating both change to clinician and patient by means of an interactive visual feedback and the efficacy of treatment during rehabilitation of the patient. Acoustic measures of voice such as F_0, jitter, shimmer and HNR as well as a temporal measure (MPT) provide objective and quantifiable measures that can be useful in substantiating subjective perception of voice quality.

This has enormous potential for further investigations in laryngectomees and other patients with head and neck cancer. Various advances have tried to tackle the loss of voice associated with a laryngectomy, including voice conservation surgeries and restoration. They all aim to give a better quality of life to the patient. The TO voice prosthesis has become the gold standard in various centres for voice rehabilitation since its introduction in 1980. However, alaryngeal speech with TO voice prosthesis is not yet comparable with laryngeal speech.

Acknowledgements

References

1. Steiner W, Aurbach G, Ambrosch P. Minimally invasive therapy in otolaryngology and head and neck surgery. *Minim Invasive Ther Allied Technol* 1991; **1:** 57–70.

2. Bosch A, Kademian MT, Frias ZC, Caldwell WL. Failures after irradiation in early vocal cord cancer. *Laryngoscope* 1978; **88:** 2,017–2,021.

3. Manni JJ, Terhaard CH, de Boer MF *et al*. Prognostic factors for survival in patients with T3 laryngeal carcinoma. *Am J Surg* 1992; **164:** 682–687.

4. Pawar PV, Sayed SI, Kazi R, Jagade MV. Current status and future prospects in prosthetic voice rehabilitation following laryngectomy. *J Cancer Res Ther* 2008; **4:** 186–191.

5. Jassar P, England RJ, Stafford ND. Restoration of voice after laryngectomy. *J R Soc Med* 1999; **92:** 299–302.

6. van As CJ. *Tracheoesophageal speech. A multidimensional assessment of voice quality* [PhD thesis]. Amsterdam, Netherlands: University of Amsterdam; 2001.

7. Pindzola RH, Cain BH. Acceptability ratings of tracheoesophageal speech. *Laryngoscope* 1988; **98:** 394–397.

8. Rhys Evans PH, Blom ED. Functional Restoration of Speech. In: Rhys Evans PH, Montgomery PQ, Gullane PJ, eds. *Principles and Practice of Head and Neck Oncology*. London: Martin Dunitz; 2003. pp999–1,063.

9. Singer MI, Blom ED. An endoscopic technique for restoration of voice after laryngectomy. *Ann Otol Rhinol Laryngol* 1980; **89:** 529–533.

10. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. 2nd edn. New York, US: Singular; 2000.

11. Fourcin AJ, Abberton E. First applications of a new laryngograph. *Med Biol Illus* 1971; **21:** 172–182.

12. Moon JB, Weinberg B. Aerodynamic and myoelastic contributions to tracheosophageal voice production. *J Speech Hear Res* 1987; **30:** 387–395.

13. Gussenbauer C. Über die erste durch Th. Billroth am Menschen ausgeführte Kehlkopf-Extirpation und die Anwendung eines künstlichen Kehlkopfes. *Arch Klin Chir* 1974; **17:** 343–356.

14. Titze IR. *Workshop on Acoustic Voice Analysis: Summary Statement*. Iowa City, IA, US: National Centre for Voice and Speech; 1995.

15. Blood GW. Fundamental frequency and intensity measurements in laryngeal and alaryngeal speakers. *J Commun Disord* 1984; **17:** 319–324.

16. Robbins J. Acoustic differentiation of laryngeal, esophageal, and tracheoesophageal speech. *J Speech Hear Res* 1984; **27:** 577–585.

17. Qi Y, Weinberg B. Characteristics of voicing source waveforms produced by esophageal and tracheoesophageal speakers. *J Speech Hear Res* 1995; **38:** 536–548.

18. Kazi R, Kiverniti E, Prasad V *et al*. Multidimensional assessment of female tracheoesophageal prosthetic speech. *Clin Otolaryngol* 2006; **31:** 511–517.

19. Baggs TW, Pine SJ. Acoustic characteristics: tracheoesophageal speech. *J Commun Disord* 1983; **16:** 299–307.

20. Robbins J. Acoustic differentiation of laryngeal, esophageal, and tracheoesophageal speech. *J Speech Hear Res* 1984; **27:** 577–585.

21. Debruyne F, Delaere P, Wouters J, Uwents P. Acoustic analysis of tracheo-oesophageal speech versus oesophageal speech. *J Laryngol Otol* 1994; **108:** 325–328.