# Assessment of the Formant Frequencies in Normal and Laryngectomized Individuals Using Linear Predictive Coding

*†Rehan A. Kazi, *Vyas M.N. Prasad, *Jeeve Kanagalingam, *Christopher M. Nutting, *Peter Clarke, *Peter Rhys-Evans, and *†Kevin J. Harrington

*London, United Kingdom*

**Summary:** The objective of this study was to assess the difference in voice quality as defined by acoustical analysis using sustained vowel in laryngectomized patients in comparison with normal volunteers. This was designed as a retrospective single center cohort study. An adult tertiary referral unit formed the setting of this study. Fifty patients (40 males) who underwent total laryngectomy and 31 normal volunteers (18 male) participated. Group comparisons with the first three formant frequencies ($F_1$, $F_2$, and $F_3$) using linear predictive coding (LPC) (Laryngograph Ltd, London, UK) was performed. The existence of any significant difference of $F_1$, $F_2$, and $F_3$ between the two groups using the sustained vowel /i/ and the effects of other factors namely, tumor stage (T), chemoradiotherapy, pharyngectomy, cricothyroid myotomy, closure of pharyngoesophageal segment, and postoperative complication were analyzed. Formant frequencies $F_1$, $F_2$, and $F_3$ were significantly different in male laryngectomees compared to controls: $F_1$ ($P < 0.001$, Mann-Whitney $U$ test), $F_2$ ($P < 0.001$, Student's $t$ test), and $F_3$ ($P = 0.008$, Student's $t$ test). There was no significant difference between females in both groups for all three formant frequencies. Chemoradiotherapy and postoperative complications (pharyngocutaneous fistula) caused a significantly lower formant $F_1$ in men, but showed little effect in $F_2$ and $F_3$. Laryngectomized males produced significantly higher formant frequencies, $F_1$, $F_2$, and $F_3$, compared to normal volunteers, and this is consistent with literature. Chemoradiotherapy and postoperative complications significantly influenced the formant scores in the laryngectomee population. This study shows that robust and reliable data could be obtained using electroglottography and LPC in normal volunteers and laryngectomees using a sustained vowel.

**Key Words:** Formant—Laryngectomy—Linear predictive coding.

Accepted for publication June 28, 2006.

From the *Head and Neck Unit, Royal Marsden Hospital, 203 Fulham Road, Chelsea, London SW3 6JJ, UK; and the †The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK.

Address correspondence and reprint requests to Rehan A. Kazi, Head and Neck Unit, Royal Marsden Hospital, 203 Fulham Road, Chelsea, London SW3 6JJ, UK. E-mail: rehan.kazi@rmh.nhs.uk

*Journal of Voice*, Vol. 21, No. 6, pp. 661–668
0892-1997/$32.00

doi:10.1016/j.jvoice.2006.07.001

## INTRODUCTION

Treatment for laryngeal cancer is increasingly aimed at preserving vocal function and maintaining quality of life at the same time as delivering increased survival rates. In recent years, this has resulted in a greater role for radical chemoradiotherapy and organ (larynx) sparing surgery. Naturally, despite increasing use of organ preservation therapy, laryngectomy with or without surgery to the tongue base and pharynx is still necessary when recurrence or complete clearance is unachievable using the above conservative techniques.

The laryngectomized patient presents the clinician with a range of challenges pre-, peri-, and postoperatively. Loss of the natural voice is an obvious handicap to patients.[1] Voice restoration techniques vary and include esophageal speech, the electrolarynx, and tracheoesophageal (TO) speech after surgical voice restoration (SVR). TO speech involves the diversion of expired pulmonary air via a one-way valve (inserted into an iatrogenic fistula between the posterior wall of the trachea and the anterior wall of the esophagus) into the pharyngoesophageal segment.[1] This results in the vibration of the pharyngoesophageal/jejunal segment producing sound.[2–4] Prosthetic TO speech is an established and reliable technique for voice restoration and is often described as the ‘‘gold standard.’’[5] Although TO speakers are capable of limited intonation, their voices are still regarded as variable and different from normal speech.[1]

The components of TO speech that dictate its quality are, as yet, incompletely understood. Formants are the resonant or dominant harmonics in the speech spectrum and are the characteristic partials that identify the vowel to the listener.[6] As such, they play a key role in determining the intelligibility of natural speech and are likely to influence the perceived quality of valved TO speech. Much of the limited previous work on formant analysis in laryngectomized patients has concentrated on esophageal speech in English speakers.[7] To the best of our knowledge, there have been only a couple of studies of formants in TO valved speech in Dutch and Spanish speakers.[8,9]

Electroglottography (EGG) is a robust and reliable method for voice analysis.[10,11] It allows for the accurate and objective determination of important voice parameters such as the fundamental frequency, jitter, shimmer, and normalized noise energy using the glottal waveform with both sustained vowels and connected speech.[11] EGG can also be used to analyze formants through the method of linear predictive coding (LPC). This technique is a widely used and reliable autoregressive method that attempts to predict future values of the input signal based on past signal values using lossy algorithms.[6,12] However, to the best of our knowledge, this method has not yet been used in the laryngectomee population for formant frequency estimation.

Therefore, in this study we attempted to analyze the effect of total laryngectomy on the production of formants in patients using TO valved speech by comparing these results with those pertaining to normal volunteers.

## PATIENTS AND METHODS

Fifty total laryngectomy patients using valved TO speech were identified from the Speech and Language Therapy database at the Royal Marsden Hospital. Of these, 40 were males and 10 were females with a mean age ± standard deviation (SD) of 63.9 years ± 10.5. The median time period post-treatment was 105 (range 34–225) months. All the participating patients in our study were using the Blom-Singer voice prosthesis. A neck dissection had been performed in 48% ($n = 24$), a myotomy in 70% ($n = 35$), partial pharyngectomy in 16% ($n = 8$), total pharyngectomy in 6.1% ($n = 3$) with flap reconstruction in 14.2% ($n = 7$) of patients. Eighty percent ($n = 40$) of patients had radiotherapy and 16% ($n = 8$) chemotherapy. In 78% ($n = 39$) of patients, primary puncture was performed.

Single voice recordings (sustained vowel of /i/ produced at a comfortable pitch and loudness for a duration of at least 5 seconds or as long as they could stably manage) of 31 normal control subjects (18 males) with a mean age of 40.9 years ± SD of 13.5 were also performed. The subjects were deemed to have a normal structure and function of the vocal cords with no history of laryngeal, neurological, or speech disorders.

The study was approved by the local research and ethics committee.

**EGG equipment**

Speech Studio (Laryngograph Ltd, London, UK) hardware and software were used for acoustic analysis. Data recordings were made in a quiet room with the subject comfortably seated using gold-plated electrodes attached on either side of the thyroid alae. Sound was picked up by a capsule electret microphone (Sony, Weybridge, UK), placed in front on the chest at a constant mouth to microphone distance of 15 cm and at an angle of 45°. The signal was transmitted to the Laryngograph Processor (Laryngograph Ltd, London, UK) consisting of a microphone preamplifier and a laryngograph-based fundamental frequency extractor. This was linked to a recorder (DAT) for the recording or playback of the Speech (Sp) and Laryngograph (Lx) waveforms. The Sp and Lx signals were acquired at 16 kHz, 16 bits signal resolution. Pitch was extracted by a dedicated hardware circuit and the period counted by a 12-MHz clock and rounded down to a single microsecond.

**Voice recording protocol**

The protocol was explained to all the patients in advance to allow familiarization with the process. Patients were allowed a few attempts of practice prior to recording. The EEG (Lx) signal on the machine was adjusted and calibrated to the optimal gain position for each subject individually. The protocol involved analyzing the sustained vowel /i/ produced at a comfortable pitch and loudness for a duration of at least 5 seconds (or as long as they could stably manage). All subjects provided synchronous acoustic and electrolaryngographic recordings of the sustained vowel /i/ in a single session.

**Formant measurement**

First, second, and third formant frequencies ($F_1$, $F_2$, and $F_3$) were measured using ‘‘Speech Studio.’’ The Laryngograph uses a pitch synchronized LPC method for formant extraction. A Fast Fourier Transformation spectrum is provided for visual inspection and cross checking of the LPC formants' values/position prior to finally recording the results to ensure accuracy. LPC spectra were computed for several pitch-synchronous windows for the mid portion of each sustained /i/ vowel, and the formant frequencies derived from the LPC coefficients were averaged to obtain an estimate of $F_1$, $F_2$, and $F_3$ for each speaker (see Figures 1 and 2, respectively).

**Main outcome measure**

The existence of any significant difference of $F_1$, $F_2$, and $F_3$ between the two groups using the sustained vowel /i/ and the effects of other factors namely, tumor stage (T), chemoradiotherapy, pharyngectomy, cricothyroid myotomy, closure of pharyngoesophageal segment, and postoperative complication were analyzed.

**Statistical analysis**

Statistical analysis was performed with the *Statistical Package for Social Sciences* (SPSS Inc, Chicago III, Illinois) using the Student's $t$ test and Mann-Whitney $U$ test. Student's $t$ test and Mann-Whitney $U$ tests were used for comparing the formants between the normal subject's speech and that of the TO speech. Formant frequencies in both normal and laryngectomized subjects for $F_1$ were not distributed normally, so the Mann-Whitney $U$ tests were used. On the other hand, $F_2$ and $F_3$ were normally distributed in both groups, so we used the Student's $t$ test. A $P$ value of less then 0.05 was taken as significant.

## RESULTS

All three formant frequencies ($F_1$, $F_2$, and $F_3$) for the vowel /i/ could be measured using LPC in all of the normal subjects. However, in the total laryngectomy group, the formant frequencies were variably assessable: $F_1$ in 100% ($n = 40$) of males and 90% ($n = 9$) of females, $F_2$ in 97.5% ($n = 39$) of males and 90% ($n = 9$) of females, and $F_3$ in 67.5% ($n = 27$) of males and 50% ($n = 5$) of females. The LPC spectra of a representative male normal volunteer and a representative male laryngectomy patient are shown in Figures 3 and 4, respectively.

Formant frequencies in both normal and laryngectomized subjects for $F_1$ were not distributed normally. On the other hand, $F_2$ and $F_3$ were normally distributed in both groups. The formant frequencies for the two groups are presented in Tables 1 and 2.

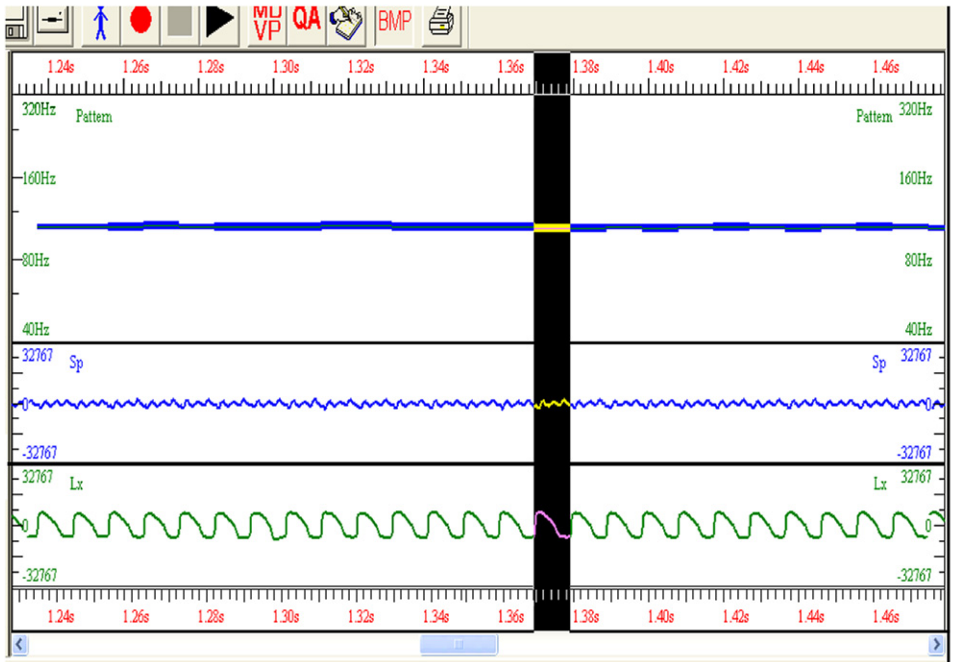


**FIGURE 1.** Representative male normal volunteer–Lx waveform and highlighted portion for formant analysis. Sp (blue), speech signal; Lx (green), laryngograph waveform. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

Comparison between alaryngeal and normal speakers showed that there were significant differences between male laryngectomees and male normal volunteers for $F_1$ ($P < 0.001$, Mann-Whitney $U$ test), $F_2$ ($P < 0.001$, Student's $t$ test), and $F_3$ ($P = 0.008$, Student's $t$ test). Similar comparisons for the females in both groups demonstrated no significant differences for $F_1$, $F_2$, or $F_3$.

Correlation of formants with other demographic and treatment-related variables was performed for the males in the laryngectomy group, but was not possible in females due to the small number of subjects. The $F_1$ formant frequency was found to be significantly lower in patients who had had a postoperative complication (305 Hz vs 393 Hz, $P = 0.02$, Mann-Whitney $U$ test) or had received chemoradiotherapy ($F_1$ 314 Hz vs 395 Hz, $P = 0.01$, Mann-Whitney $U$ test) as compared to those who did not. There were no other significant differences of $F_1$ based on age, T, closure of the pharyngoesophageal segment, cricopharyngeal myotomy, pharyngectomy, neck dissection, reconstruction, or radiotherapy. No significant differences were found in the $F_2$ and $F_3$ formant frequencies for any of the above variables.

## DISCUSSION

Contemporary voice rehabilitation following total laryngectomy has reached an advanced stage with many patients achieving intelligible and functional speech after appropriate rehabilitation with a speech and language pathologist.[1–4] However, patients who have undergone a laryngectomy remain a unique population amongst head and neck cancer sufferers because of their prosthetic speech.[5] It is well recognized that TO speech is superior to other alaryngeal phonation techniques such as esophageal speech or the use of the electrolarynx. However, TO speech is still regarded as perceptually worse than normal speech.[13] A key step in identifying factors that dictate the quality of the reconstructed voice will be an accurate characterization of the components of the voice in TO speakers in comparison to normal subjects. As such, an analysis of formants is important because they are

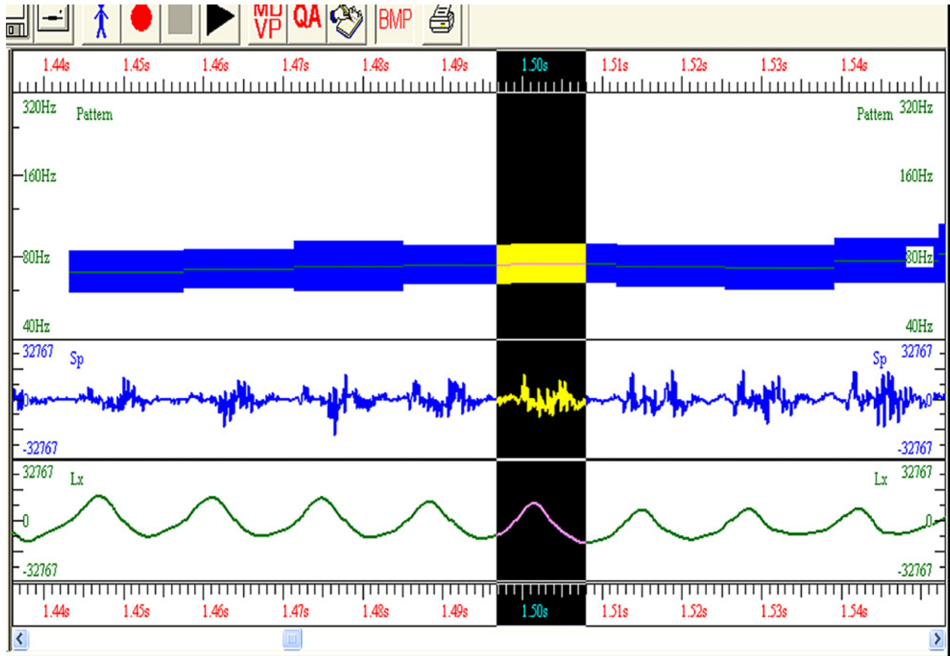


**FIGURE 2.** Representative male total laryngectomy patient–Lx waveform and highlighted portion for formant analysis. Sp (blue), speech signal; Lx (green), laryngograph waveform. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

believed to be responsible for the phonetic characterization of vowel quality and are seen as essential components in the intelligibility of speech.[5] Therefore, this study aimed to compare the formant frequencies in laryngectomees and normal subjects using the sustained vowel /i/ since this has been robustly studied in the literature. Future studies will look at correlating the measured formant frequencies with perturbation measures and perceptual voice analysis in patients who have undergone laryngectomy.

EEG analysis of voice has been shown to provide a reliable method of procuring the various voice parameters.[10,11] This method was used in our study to analyze valved speech in laryngectomy patients and is the focus of a number of ongoing studies at our institution. For formant analysis, we used the LPC method of spectral analysis because this has been shown to be reliable and robust.[9] To be accurate, we examined pitch-synchronous windows for the mid portion of each sustained /i/ vowel, and the formant frequencies derived from the LPC coefficients were averaged to obtain an estimate of $F_1$, $F_2$, and $F_3$ for each speaker. We examined the mid portion of the vowel because this was the most stable part. One of the sources of error in LPC analysis can be the fact that the voice signal can be markedly aperiodic.[9] We were able to measure most of the formant frequencies in all but one of our subjects. This particular patient had a markedly aperiodic signal with significant noise that precluded the use of this technique. The $F_1$ and $F_2$ formant frequencies obtained in our control group of normal speakers were in general agreement with those obtained in other studies in the normal population.[12,13] Minor differences may be attributed to the fact that formant structure can vary slightly due to consonantal context. Thus, minimal differences found among different studies can be attributed to coarticulation effects and the analytical method used.[14] Our study showed significantly higher $F_1$, $F_2$, and $F_3$ formant frequency values in laryngectomized males as compared to normal male subjects. In this regard, our results are consistent with the limited previous work carried out.[10–12] However, to the best of our knowledge, this is the first large study in the English-speaking population to use EGG and LPC method for estimation of the formant frequencies in laryngectomees.

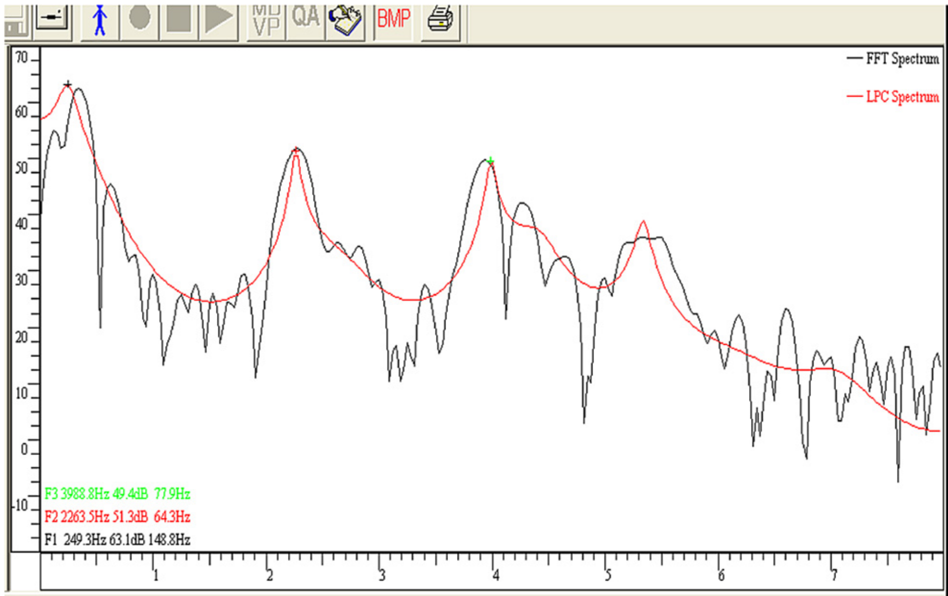


**FIGURE 3.** LPC spectrum of a representative male normal volunteer showing formant peaks. *X*-axis, frequency in kHz; *Y*-axis, amplitude in dB. LPC spectrum is shown in red. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

Vocal tract length appears to be the most important factor determining the positions of the formant frequencies. With shorter vocal tracts, higher formants are found.[15] Another hypothesis put forward to explain these phenomena is that the longer vowel duration found in the speech of alaryngeal patients is attributable to the slower decay in pharyngoesophageal segment vibrations as compared with

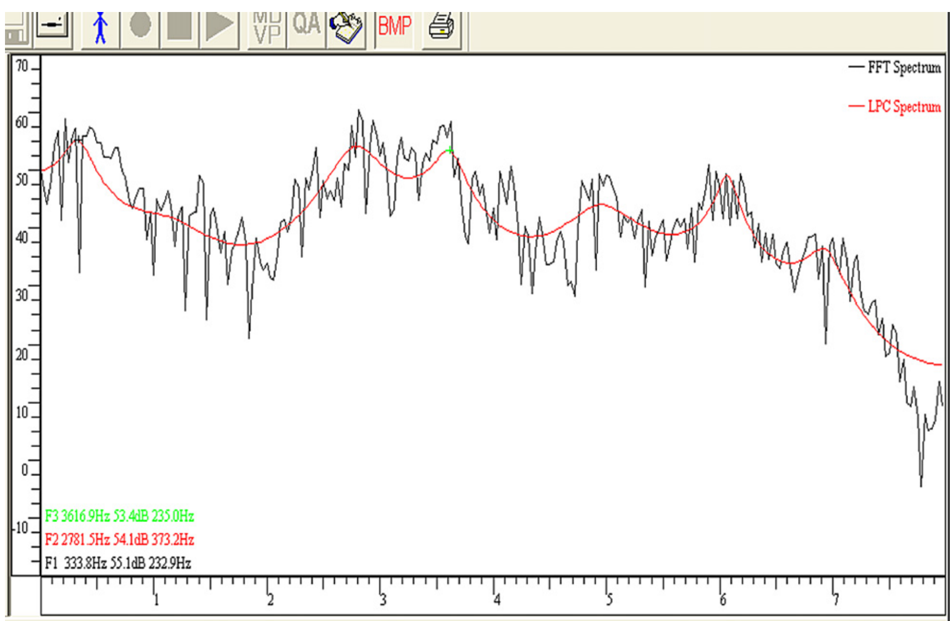


**FIGURE 4.** LPC spectrum of a representative male total laryngectomy patient showing formant peaks. *X*-axis, frequency in kHz; *Y*-axis, amplitude in dB. LPC spectrum is shown in red. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

**TABLE 1.** *Formant Frequencies ($F_1$, $F_2$, and $F_3$) in Normal Subjects*

| | Males | | | Females | | |
|---|---|---|---|---|---|---|
| Formants | No. | Mean/Median | SD/Range | No. | Mean/Median | SD/Range |
| $F_1$ | 18 | 260.5 | 237.9–343.8 | 13 | 368.2 | 250.9–460.2 |
| $F_2$ | 18 | 2223.9 | 158.1 | 13 | 2678.7 | 163.4 |
| $F_3$ | 18 | 3226.9 | 414.6 | 13 | 3301.2 | 244.5 |

Data for $F_1$ were not normally distributed and are, therefore, presented as median and range. Data for $F_2$ and $F_3$ were normally distributed and are presented as mean and standard deviation.

laryngeal vibrations in normal speakers.[16,17] In alaryngeal speech, the motor control of the pharyngoesophageal segment is not comparable with the motor control capabilities of the normal speaker.[18,19] Our findings correspond with those of others in that a possible explanation for the increased formants in laryngectomees could be the hypothesis that the vocal tract length appears to be an important factor in determining the average formant positions.[10–12] TO and esophageal speakers articulate vowels with fronted and higher tongue positions relative to the tongue position in normal speakers.[7] Another explanation provided in the literature is that the back of tongue might be a little lowered, due to the removal of the larynx in laryngectomees.[8]

The difference in formant frequency between individual TO speakers may be larger than those between individual normal speakers, since the anatomy of the voice source and the vocal tract depends on the type and extent of the surgical intervention. These differences in vocal tract most probably also explain the interspeaker differences that were found between the TO speakers. An uneven spread of $F_1$ was seen in all subjects. However, we did not show larger standard deviations of $F_1$ and $F_2$ formant values in the alaryngeal group compared to controls. This is most likely due to not measuring the vowel space and speaking rate in our subjects.[7] Because of the small number of female laryngectomees it was difficult to show any difference between this group and controls. Interestingly, male laryngectomees who underwent pharyngectomy and reconstruction using jejunal flaps, gastric pull up, or musculocutaneous flaps had no significant difference between their formant frequencies compared with controls. The effect of chemoradiotherapy and postoperative complications (pharyngocutaneous fistula) on the formants is difficult to explain conclusively in the light of any previous experience.

Our experience with acoustical analysis measuring formant frequency with EGG in English speakers is consistent with the results obtained by other groups using different methods, and to this end is confirmatory to the further use of EGG in other selected groups of patients with voice handicap.[7–9]

We recommend performing videofluoroscopy recordings during speech as this may give more information about the length of the vocal tract and the position of the back of the tongue in laryngectomees as compared to normals. This may further help in clarifying the effect of laryngectomy on formant frequencies. Also the influence of the

**TABLE 2.** *Formant Frequencies ($F_1$, $F_2$, and $F_3$) in Laryngectomees*

| | Males | | | Females | | |
|---|---|---|---|---|---|---|
| Formants | No. | Mean/Median | SD/Range | No. | Mean/Median | SD/Range |
| $F_1$ | 40 | 391.5 | 290.6–1448.6 | 9 | 413.0 | 288.1–814.8 |
| $F_2$ | 39 | 2542.3 | 200.9 | 9 | 2706.9 | 220.0 |
| $F_3$ | 27 | 3544.2 | 257.9 | 5 | 3333.6 | 231.0 |

Data for $F_1$, $F_2$ and $F_3$ are the same as provided in the legend to Table 1.

different vowel formant frequencies and ratios on the intelligibility of the vowels should also be studied. It would indeed be useful to study the vowel space of the individual speakers with /i/, /a/, and /u/ as useful reference points. Perhaps a shortcoming of this study is its retrospective nature. A longitudinal analysis is needed to evaluate specific alterations in the vowel formant structure within individual patients prospectively over time. There is also an element of selection and survival bias that is inherently seen in any retrospective design study.

## CONCLUSION

This study found that the formant frequencies in adult male patients who have undergone total laryngectomy and SVR were increased compared to the normal subjects. Chemoradiotherapy and postoperative complications significantly influenced the formant scores in the laryngectomee population. This study shows that robust and reliable data could be obtained using EGG and LPC in normal volunteers and laryngectomees with a sustained vowel.

## REFERENCES

1. Blom ED, Singer MI, Harmaker RC. A prospective study of tracheoesophageal speech. *Arch Otolaryngol Head Neck Surg*. 1986;112:440–447.
2. Mendelsohn M, Morris M, Gallagher R. A comparative study of speech after total laryngectomy and total laryngopharyngectomy. *Arch Otolaryngol Head Neck Surg*. 1993; 119:508–510.
3. Perry A. Surgical voice restoration following laryngectomy: the tracheoesophageal fistula technique (Singer-Blom). *Br J Disord Commun*. 1988;23:23–30.
4. Bates GJ, McFeeter L, Coman W. Pharyngolaryngectomy and voice restoration. *Laryngoscope*. 1990;100:1025–1026.
5. Kazi R, Singh A, De-Cordova J, Clarke P, Harrington K, Rhys-Evans P. A new self-administered questionnaire to determine patient experience with the Blom-Singer valve. *J Postgrad Med*. 2005;51:253–258.
6. Baken RJ, Orlikoff RF. Sound spectrography. *Clinical Measurement of Speech and Voice*. 2nd ed. San Diego, CA: Singular; 2000. 258–266.
7. Sisty NL, Weinberg B. Formant frequency characteristics of esophageal speech. *J Speech Hear Res*. 1972;15:439–448.
8. Van As CJ, van Ravesteijn MA et al. Presented at: IFA Proceedings. Institute of Phonetic Sciences, University of Amsterdam; 1997;21:143–153.
9. Miralles JL, Cervera T. Acoustical analysis of Spanish vowels produced by laryngectomised subjects. *J Speech Lang Hear Res*. 2001;44:988–996.
10. Fabre P. Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence. *Bull Nat Med* 1957;66–69.
11. Fourcin AJ. Laryngographic assessment of phonatory function. *ASHA Rep*. 1981;11:116–127.
12. Atal BS, Hanauer SL. Speech analysis and synthesis by linear prediction of the speech wave. *J Acoust Soc Am*. 1971;50:637–655.
13. Van As CJ, Hilgers FJM, Verdonek-de Leeuw IM, Loopman-van Beinum FJ. Acoustical analysis and perceptual evaluation of tracheoesophageal prosthesis voice. *J Voice*. 1998;12:239–248.
14. Bradlow AR. A comparative acoustic study of English and Spanish vowels. *J Acoust Soc Am*. 1995;97:1916–1924.
15. Stevens KN, House AS. Development of a quantitative description of vowel articulation. *J Acoust Soc Am*. 1955;27: 484–493.
16. Peterson G, Barney H. Control methods used in a study of the vowels. *J Acoust Soc Am*. 1952;24:175–184.
17. Christensen JM, Weinberg B. Vowel duration characteristics of esophageal speech. *J Speech Hear Res*. 1976;19: 678–689.
18. Robbins J, Christensen J, Kempster G. Characteristics of speech production after traqueoesophageal puncture: voice onset time and vowel duration. *J Speech Hear Res*. 1986; 29:499–504.
19. Moon JB, Weinberg B. Aerodynamic and myoelastic contributions to traqueoesophageal voice production. *J Speech Hear Res*. 1987;30:387–395.