INT. J. LANG. COMM. DIS., SEPTEMBER–OCTOBER 2007,
VOL. 42, NO. 5, 521–532

Analysis of formant frequencies in patients with oral or oropharyngeal cancers treated by glossectomy

Rehan Kazi[†‡], Vyas M. N. Prasad[†], Jeeve Kanagalingam[†], Christos Georgalas[†], Ramachandran Venkitaraman[†], Christopher M. Nutting[†], Peter Clarke[†], Peter Rhys-Evans[†] and Kevin J. Harrington[†‡]

†Head and Neck Unit, Royal Marsden Hospital, London, UK
‡The Institute of Cancer Research, London, UK

(*Received 26 January 2006; accepted 9 October 2006*)

Abstract

Aims: To compare voice quality as defined by formant analysis using a sustained vowel in patients who have undergone a partial glossectomy with a group of normal subjects.
Methods & Procedures: The design consisted of a single centre, cross-sectional cohort study. The setting was an Adult Tertiary Referral Unit. A total of 26 patients (19 males) who underwent partial glossectomy and 31 normal volunteers (18 males) participated in the study. Group comparisons using the first three formant frequencies (F1, F2 and F3) using linear predictive coding (Laryngograph Ltd, London, UK) were performed. The existence of any significant difference of F1, F2 and F3 between the two groups using the sustained vowel /i/ and the effects of other factors, namely age, first presentation versus recurrence, site (oral cavity, oropharynx), subsite (anterior two-thirds of the tongue, tongue base), stage, radiation, complication, and neck dissection, were analysed.
Outcomes & Results: Formant frequencies F1, F2 and F3 were normally distributed. F1 and F2 were significantly different in normal males versus females. F1, F2 and F3 were not different statistically between male and female glossectomees. Comparison of only women showed significant differences between normal subjects and patients in F2 and F3, but none in F1. This was the opposite in men where F1 was significantly different. Age, tumour presentation, site, subsite, radiation and neck dissection showed no significant difference. Postoperative complications significantly affected the F1 formant frequency.

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

International Journal of Language & Communication Disorders
ISSN 1368-2822 print/ISSN 1460-6984 online © 2007 Royal College of Speech & Language Therapists
http://www.informahealthcare.com
DOI: 10.1080/13682820601056566

Conclusions: The study found that the formant values in patients following a partial glossectomy were altered significantly as compared with the normal control subjects. Only gender and complications and not the age, site, subsite, radiation and neck dissection were seen to influence the formant scores.

Keywords: formant frequencies, partial glossectomy, LPC.

What this paper adds

What is already known

Oral and oropharyngeal tumours and their treatment can profoundly affect speech. Meaningful analysis of the voice is difficult as it is a complex multidimensional function. Formant frequencies represent an objective measure that may be useful in studying the effects of treatment on vocal function. Linear predictive coding (LPC) is a widely used, easily performed and reliable method incorporated in most acoustic software programs for the determination of formant frequencies.

What this study adds

The largest cross-sectional study of English-speaking partial glossectomy patients. A comparison with a large set of normal subjects to reveal differences. Formant frequencies are different in glossectomees with gender and complications having greatest influence.

Introduction

Tumours of the oral cavity and oropharynx have a relatively high incidence worldwide. The incidence and mortality rates for oral cancer, especially in younger men, have shown an increase in the UK and most other European countries over the last few decades (Patel *et al.* 2003). The tumour and the consequences of its treatment can profoundly affect one or more of the several important functions that these regions normally serve. Alteration of functions such as speech, taste, mastication, swallowing, oral sensation/continence and of body image can have a devastating impact on quality of life. Apart from the obvious goal of curing the patient, these factors must be considered in treatment planning.

Treatment for oral cavity or oropharyngeal cancers involving the tongue is increasingly aimed at preserving vocal and swallowing function and this has resulted in a greater role for chemoradiotherapy and organ-sparing surgery. Nonetheless, radical surgical excision, in the form of glossectomy (partial or total), reconstruction, and surgery to the neck, pharynx and/or larynx, continues to play a central role in the management of these tumours.

Patients who have undergone a glossectomy present a range of challenges to the clinical team. The tongue is an organ of central importance that integrates the functions of articulation, mastication, deglutition, taste, airway protection, maintenance of oral hygiene and enjoyment of food. From the point of view of speech rehabilitation, with the help of intensive speech and language therapy many patients achieve relatively good intelligibility through unusual labial, mandibular and pharyngeal speech compensations, modifications of normal articulatory patterns and may be heard as normal speakers (Rentschler and Mann 1980, Hamlet *et al.* 1992). However, many patients have distorted speech or may even be largely unintelligible. Even subtle changes in speech compared with the pre-morbid condition can have profound effects on quality of life.

Meaningful analysis of the voice is difficult as it is a complex multidimensional function. In studying the outcome of cancer treatment and speech rehabilitation, it is often difficult to find robust, discriminating functional measures. Formant frequencies represent an objective measure that may be useful in studying the effects of treatment on vocal function. Formants are the resonant harmonics in the speech spectrum and are described as being the characteristic partials that help identify the vowel to the listener (Atal and Hanauer 1971, Baken and Orlikoff 2000). The movements of the articulators set the limits of variation, which in the spectral domain are typically seen as variation of the formant frequencies. As a result, a variety of methods have been developed to measure speech and voice outcomes. One such reliable and robust method is electroglottography (EGG) (Whitehill *et al.* 2006). It allows for the accurate and objective determination of important voice parameters such as the fundamental frequency, jitter, shimmer and normalized noise energy using glottal waveform with both sustained vowel and connected speech (Fourcin 1981, Childers and Krishnamurthy 1985). Linear predictive coding (LPC) is a widely used, easily performed and reliable autoregressive method incorporated in most acoustic software programs for the determination of formant frequencies. It attempts to predict future values of the input signal based on past signal values using lossy algorithms (Atal and Hanauer 1971, Baken and Orlikoff 2000).

Although a number of studies have assessed articulation and speech in patients who have undergone partial glossectomy, very few have looked at the effect on formant frequencies and none has used the LPC methodology (Heller *et al.* 1991, Korpijaakko-Huuhka *et al.* 1995, Wakumoto *et al.* 1996, Knuuttila *et al.* 1999). Moreover, the lack of sufficient case numbers and absence of correlation with treatment data has detracted from the overall value of the previous studies. The underlying hypothesis behind this study was that partial glossectomy alters the production of formant frequencies in oral and oropharyngeal cancer patients. Therefore, the aim was to analyse the effect of a partial glossectomy operation on the vowel /i/ formant frequencies. Specifically, it was to identify if there is an alteration and, if so, the extent of that alteration and its correlation with treatment variables.

This study reports the first large cross-sectional analysis of vowel formant frequencies measured by the LPC method in patients who have undergone a partial glossectomy for oral and oropharyngeal cancers.

Methods

A total of 26 squamous cell carcinoma patients who underwent partial glossectomy between May 1987 and June 2004 were identified from the Speech and Language Therapy database of the Royal Marsden Hospital. Of these, 20 were males and seven were females, with a mean age of 53.1 ($\pm$ standard deviation (SD)$=$8.7) years. The median period from treatment to voice analysis was 33 (range 4–210) months. Thirteen patients (48.1%) had cancer of the oral cavity (12 anterior two-thirds of the tongue, one floor of the mouth) and 14 (51.9%) of the oropharynx (three tonsil, 11 tongue base). A total of 21 patients (80.7%) were treated for primary disease while five (19.2%) had recurrent disease. All patients had underwent a partial glossectomy operation (surgical excision of half or less than half of the tongue as governed by the size/depth of the cancer). A reconstruction was performed in six patients (23%) (three patients underwent radial forearm free flap, two a pectoralis major flap and one a latissimus dorsi flap). Neck dissection was performed in 21 patients (80.7%, selective neck dissection in 14 and modified radical neck dissection in seven patients) and no patient had a mandibular resection. Nineteen patients (73%) received adjuvant radiotherapy and of these four patients (14.8%) had also received chemotherapy.

Single voice recordings of 31 normal control subjects (18 males and 13 females) with a mean age of 53.4 ($\pm$ SD=17.3) years were also performed. These data have previously been reported in part in a previous study of formant frequencies in laryngectomees. Kazi *et al.* (2006) The control subjects had 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.

Electroglottography 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) consisting of a microphone pre-amplifier 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; period was counted by a 12-MHz clock and rounded down to 1 ms.

Voice-recording protocol

Before any recordings were made, the protocol was explained to the patients to allow them to familiarize themselves with the process. Patients were allowed to practise the technique before commencing the recording. The electroglottograph (Lx) signal on the machine was adjusted and calibrated to the optimal gain position individually for each subject. The protocol involved analysing the sustained vowel /i/ produced at a comfortable pitch and loudness for at least 5 s (or for as long as stably possible). All subjects provided synchronous acoustic and electrolaryngographic recordings of the sustained vowel /i/ in a single session.

Formant measurement

Speech Studio was used to derive the first, second and third formant frequencies (F1, F2 and F3). The laryngograph uses a pitch synchronized LPC method for formant extraction. A fast Fourier transformation (FFT) spectrum is provided for visual inspection and cross-checking of the LPC formants values/position before finally recording the results to ensure accuracy. LPC spectrums were computed for several pitch-synchronous windows for the mid-portion of the sustained /i/ vowel and the formant frequencies derived from the LPC coefficients were averaged to obtain an estimate of F1, F2 and F3 for each speaker (figures 1 and 2, respectively).

Statistical analysis

Statistical analysis was performed using the Statistical Package for Social Sciences (SPSS, Inc., Chicago, IL, USA) using analysis of variance (ANOVA) and a Student's *t*-test. Since assumptions of normality could be met, ANOVA was performed to correlate formant frequencies and socio-demographic data. A Student's *t*-test for independent samples was used to compare the formant frequencies between males and females of both study groups and for correlation with treatment data. A *p*-value of less then 0.05 was taken as being statistically significant. In view of the fact that multiple comparisons were performed on the data, Bonferroni's correction was applied through the commercial SPSS statistical package so as to reduce the chance of a Type 1 error.

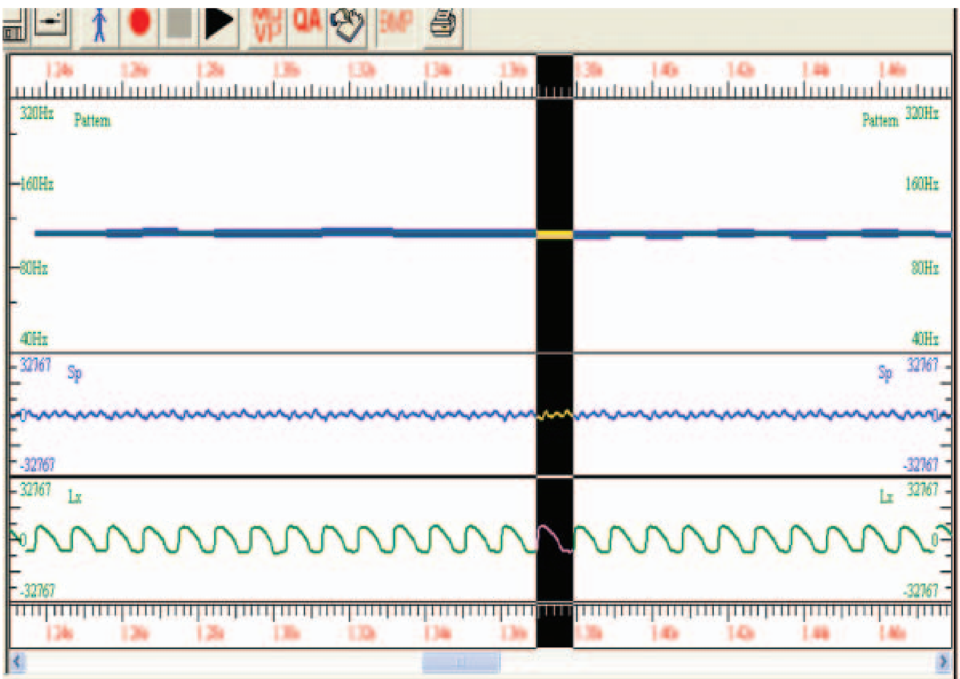

Figure 1. Representative male normal subject: laryngograph (Lx) waveform and highlighted portion for formant analysis. Speech (Sp) (blue), speech signal; Lx (green), laryngograph waveform. (Please see colour online)

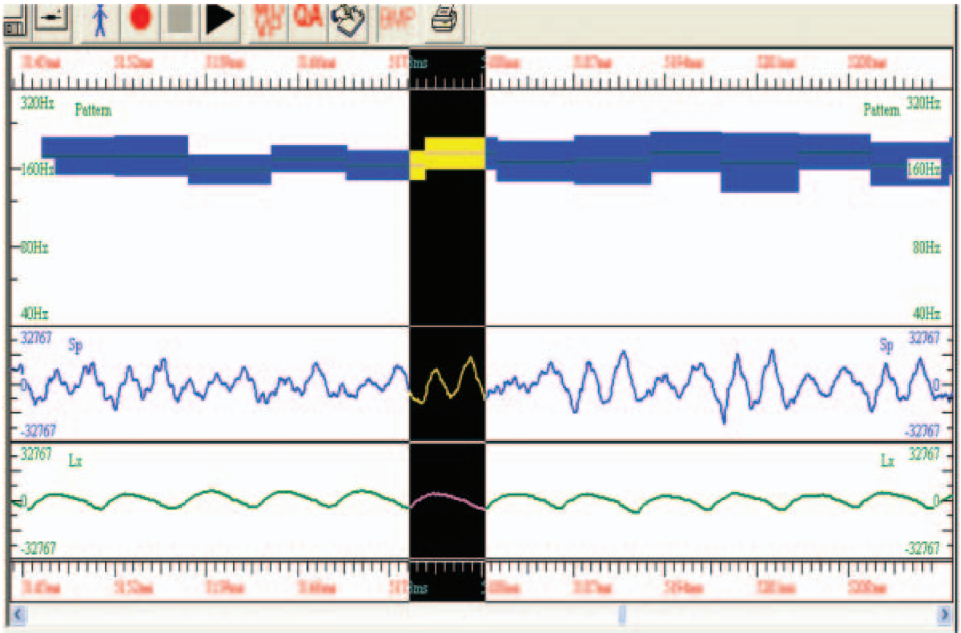

Figure 2. Representative male partial glossectomy patient (T2N0M0, squamous cell carcinoma (SCC) of the anterior two-thirds of the tongue, treated with partial excision and left selective neck dissection): laryngograph (Lx) waveform and highlighted portion for formant analysis. Speech (Sp) (blue), speech signal; Lx (green), laryngograph waveform. (Please see colour online)

Table 1. Comparison of formant frequencies in normal subjects

	Males			Females		
Formants	*n*	Mean	SD	*n*	Mean	SD
F1	18	266.1	37.9	13	368.2	79.1
F2	18	2223.9	158.1	13	2678.7	163.4
F3	18	3226.9	414.6	13	3301.2	244.5

SD, standard deviation.

Table 2. Comparison of formant frequencies in patients who have undergone a partial glossectomy

	Males			Females		
Formants	*n*	Mean	SD	*n*	Mean	SD
F1	19	348.3	76.7	7	357.3	116.7
F2	19	2199.9	340.3	7	1797.1	895.8
F3	19	3172.2	442.1	7	2951.5	285.8

SD, standard deviation.

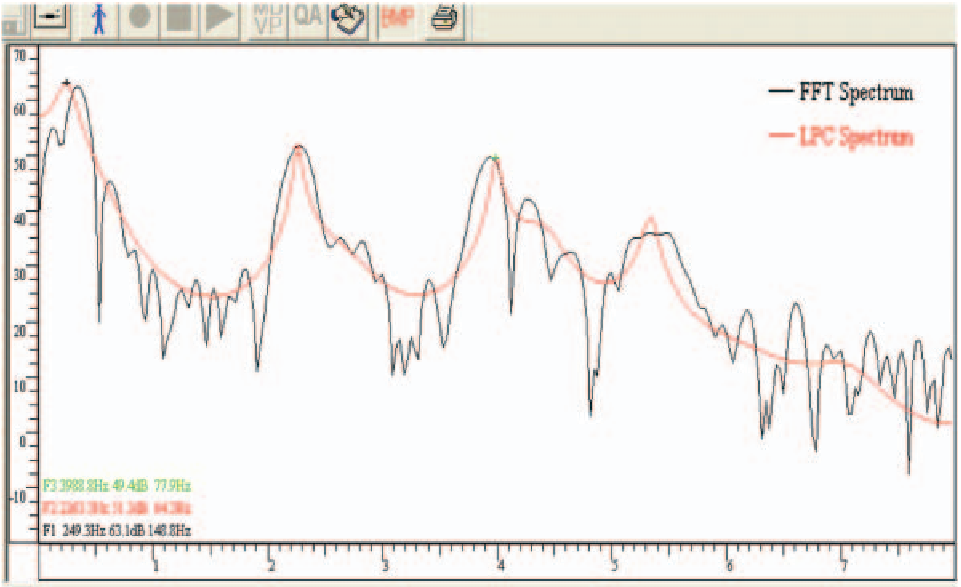

Figure 3. Linear predictive coding (LPC) spectrum of a representative male normal subject. The formant frequencies derived correspond to the peaks in the red waveform in the spectrum. The first three peaks represent the positions of F1, F2 and F3 format frequencies, respectively. *x*-axis, frequency (kHz); *y*-axis, amplitude (dB). The LPC spectrum is shown in red. (Please see colour online)

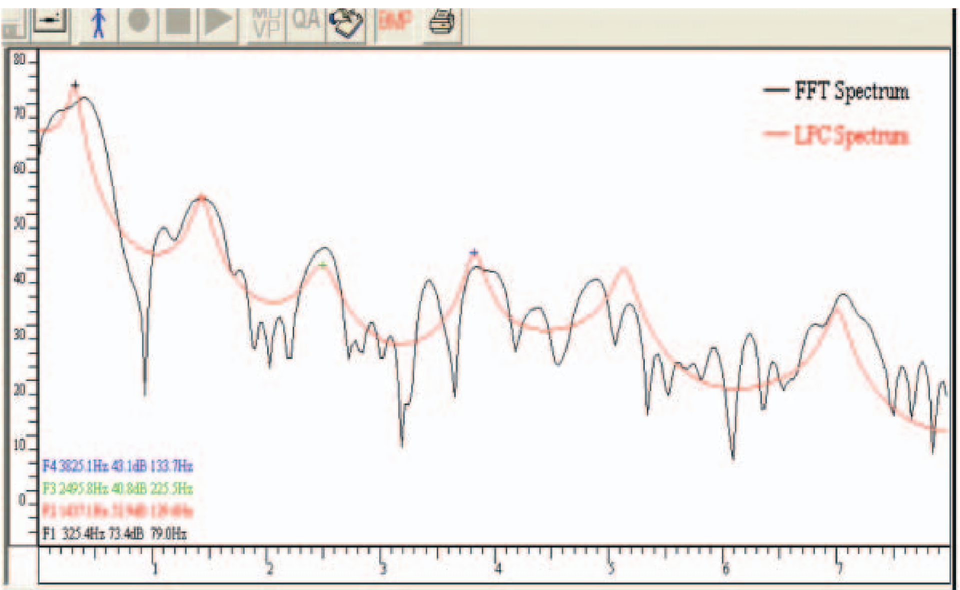

Figure 4. Linear predictive coding (LPC) spectrum of a representative male partial glossectomy patient (T2N0M0, squamous cell carcinoma (SCC) of the anterior two-thirds of the tongue, treated with partial excision and left selective neck dissection). The formant frequencies derived correspond to the peaks in the red waveform in the spectrum. The first three peaks represent the positions of F1, F2 and F3 format frequencies, respectively. *x*-axis, frequency (kHz); *y*-axis, amplitude (dB). The LPC spectrum is shown in red. (Please see colour online)

Table 3. Significant formant correlation work using a Student's *t*-test

Formant	Patient group/variable	*p*-value
F1	Normal males versus females	<0.001
F2	Normal males versus females	<0.001
F2	Normal females versus study females	0.04
F3	Normal females versus study females	0.02
F1	Normal males versus study males	0.01
F1	Complications	0.01

Results

All three formant frequencies (F1, F2 and F3) for the vowel /i/ could be measured using LPC in glossectomy patients and normal subjects. Formant frequencies in both glossectomized and normal subjects for F1, F2 and F3 were distributed normally. The formant frequencies for the two groups are presented in tables 1 and 2, respectively. LPC spectra of a representative male normal subject and a representative male partial glossectomy patient are as shown in figures 3 and 4, respectively.

F1 and F2 formants were significantly different between healthy men and women (Student's *t*-test, $p<0.001$), but there was no significant difference in F3 between these two groups. In contrast to this gender difference in normal subjects, there were no significant differences in formant frequencies between male and female glossectomy patients. Comparison of only women, i.e. controls versus patients, showed significant differences between the two groups for F2 and F3 (Student's *t*-test, $p=0.04$ and 0.02, respectively) but not for F1. This was the opposite in men where F1 was significantly different (Student's *t*-test, $p=0.01$) as opposed to F2 and F3 which were not (table 3).

The formant frequencies according to the subsite of pathology across the study group are presented in figure 5.

Comparison of the effects of age, primary tumour presentation versus recurrence, site of pathology, i.e. oral cavity versus oropharynx, subsite, i.e. the anterior two-thirds of the tongue versus tongue base, radiation, and finally neck

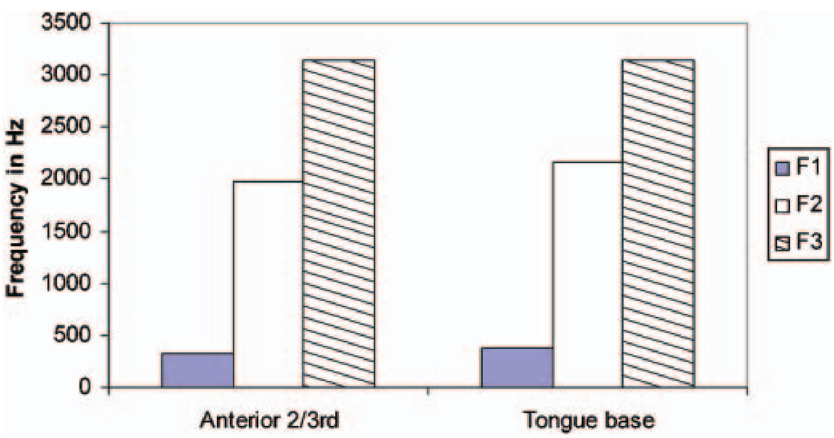

Figure 5. Mean formant frequencies across the subsite of pathology.

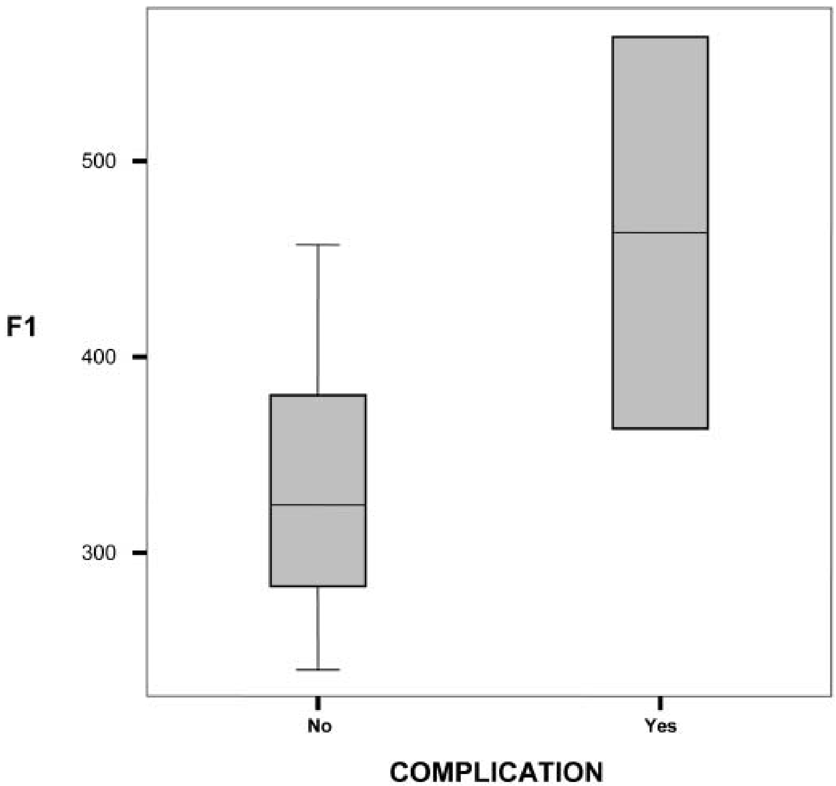

Figure 6. Correlation between the F1 formant value and complications in the glossectomy subjects. Data in the box-plots are the median, 95% confidence interval and interquartile ranges.

dissection on the formant frequencies showed no significant difference (ANOVA, Student's *t*-test, all $p>0.05$). However, it was found that postoperative complications significantly affected the F1 formant frequency (Student's *t*-test, represented as means (SD), no=334 (63.9) and yes=463.4 (140.4), $p=0.01$), as shown in figure 6.

Discussion

Even minor tongue surgery can influence the clarity of speech. Current voice rehabilitation following glossectomy has advanced considerably and many patients have fully intelligible speech after a period of speech and language therapy (Korpijaakko-Huuhka *et al.* 1995, Knuuttila *et al.* 1999, Patel *et al.* 2003). Formants are believed to be responsible for the phonetic characterization of vowel quality and are essential components in the perception of speech (Atal and Hanauer 1971, Baken and Orlikoff 2000). The relationship between vowel space and partial glossectomy has not been addressed adequately in the literature. All of the previous studies in this sphere have been limited by the small patient numbers that significantly impact on their ability to draw clear conclusions. Furthermore, with small studies it is difficult to adjust for variations caused by different surgical techniques. The present study is the largest cross-sectional formant study in the literature using the LPC method. All the patients in this study were treated in a single institution by a single surgeon, which allows uniformity in the analysis. The analysis also includes treatment variables that have allowed one to comment on factors that can affect the formant values. As such the authors believe that this work represents a significant contribution to knowledge in this area.

Electroglottographic analysis of voice has been shown to provide a reliable method of objectively analysing the pathological voice (Fourcin 1981, Childers and Krishnamurthy 1985). This method was used in the present study to analyse speech in partial glossectomy patients and is the focus of further ongoing studies at the authors' institution. For formant analysis the LPC method of spectral analysis was used as it has been shown to be reliable and robust (Atal and Hanauer 1971, Baken and Orlikoff 2000). The vowel /i/ was the present authors' preferred vowel for analysis because it puts the vocal folds in their most stressed configuration. This then gives the greater contrast with normal production. Furthermore, the vowel /i/ specifically has been robustly studied in the literature. In order to be accurate, pitch-synchronous windows for the mid-portion of each sustained /i/ vowel were examined and the formant frequencies derived from the LPC coefficients were averaged to obtain an estimate of F1, F2 and F3 for each speaker. The mid-portion of the vowel was examined as this was the most stable part and thus reduces to some extent the effect of F0 on the measurements. One of the sources of error in LPC analysis can be if the voice signal is markedly aperiodic and noisy (Baken and Orlikoff 2000). However, this was not seen in any of the present speech samples and the present authors were able to measure the first three formants in all of the subjects satisfactorily as originally intended. In fact, most of the patients had reasonably intelligible speech. Reasons for this finding could include the interval post-surgery, the provision of speech therapy and easy recording procedures.

F1 and F2 formant values obtained in the control group of normal speakers agreed, in general, with those obtained in other studies in the normal population (Atal and Hanauer 1971, Baken and Orlikoff 2000). Minor differences could 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 co-articulation effects and analytical method employed (Stevens and House 1955, Georgian *et al.* 1982, Bradlow 1995).

The results showed that the first two formant frequencies, F1 and F2, were no longer significantly different between glossectomized men and women. Interestingly, each mean formant frequency was non-significantly higher in men than in women. This is in contrast to statistically significant higher F1 and F2 scores and also higher mean values for F1, F2 and F3 scores in normal women as compared with normal men. This reversal is difficult to explain and one can hypothesize that surgery has, in effect, altered the anatomy and physiology of the tongue in such a way as to negate the differences in formant frequencies between the genders. Clearly, other factors that may be playing a role in the production of formant frequencies in these patients are the long interval post-surgery and the intervention of speech and language therapists (Heller *et al.* 1991, Korpijaakko-Huuhka *et al.* 1995). Whitehill *et al.* (2006) have detailed very well the relationship between vowel space and intelligibility in glossectomy patients and their findings have a direct bearing on this work. In that work, the speakers with partial glossectomy exhibited significantly lower mean F2 values for the vowel /i/, and restricted F2 ranges, when compared with the control speakers. The significantly smaller vowel space areas for the speakers with glossectomy supported the hypothesis of vowel formant centralization (Whitehill *et al.* 2006).

Mean F1 formant frequencies were raised in partial glossectomy patients as compared with normal controls. This is consistent with the literature where surgery to the tip of the tongue influences the F1 scores (Korpijaakko-Huuhka *et al.* 1995, Wakumoto *et al.* 1996). In addition, mean F2 and F3 scores were reduced as a result of partial surgery. However, these results should be interpreted with caution considering the fact that there are no precise data on the exact position of the tongue during vowel production and some details of surgery are lacking.

The measured differences in formant frequency values were different for male and female glossectomy patients. This is something very interesting and completely unexpected — it appears that men and women process speech after a partial glossectomy differently and their deficits are evident in different formants — F1 for men, F2 and F3 for women.

In the present series, amongst all the treatment variables, a significant correlation was found only for F1 with complications. In other words, significantly increased F1 values were seen in patients with major complications. This perhaps could be the result of altered healing leading to fibrosis and consequent distortion of the remnant vocal tract.

Perhaps a shortcoming of this study is its cross-sectional nature. A prospective longitudinal study is recommended to evaluate specific alterations in the vowel formant structure within individual patients prospectively over time. Such a study would also allow us to analyse the effects of interventional speech and language therapy. It would also be most interesting if future studies are directed towards determining the links between the effect on formant frequencies as a result of a glossectomy with speech intelligibility and perception. In order to clarify the effect of a partial glossectomy on formant frequencies, it is recommend that one performs videofluoroscopy during speech as this may give more information about the length of the vocal tract and the position of the tongue in partial glossectomees as compared with normal subjects. There is also an element of selection and survival bias that is inherently seen in any retrospective cross-sectional design study.

Conclusions

This study found that the formant values in patients of oral and oropharyngeal cancer following a partial glossectomy were altered significantly as compared with the normal control subjects. Only gender and complications and not the site, subsite, radiation, and neck dissection was seen to influence the formant scores. This study shows that robust and reliable data about formant frequencies could be obtained using linear predictive coding in normal subjects and glossectomy patients with a sustained vowel.

References

Atal, B. S. and Hanauer, S. L., 1971, Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America*, **50**, 637–655.

Baken, R. J. and Orlikoff, R. F., 2000, Sound spectrography. In R. J. Baken (ed.), *Clinical Measurement of Speech and Voice* (San Diego, CA: Singular), pp. 258–266.

Bradlow, A. R., 1995, A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, **97**, 1916–1924.

Childers, D. G. and Krishnamurthy, A. K., 1985, A critical review of electroglottography. *CRC Critical Reviews in Biomedical Engineering*, **12**, 131–161.

Fourcin, A. J., 1981, Laryngographic assessment of phonatory function. *ASHA Reports*, **11**, 116–127.

Georgian, D. A., Logemann, J. A. and Fisher, H. B., 1982, Compensatory articulation patterns of a surgically treated oral cancer patient. *Journal of Speech and Hearing Disorders*, **47**, 154–159.

Hamlet, S. L., Patterson, R. L. and Fleming, S. M., 1992, A longitudinal study of vowel production in partial glossectomy patients. *Journal of Phonetics*, **20**, 209–224.

Heller, K. S., Levy, J. and Sciubba, J. J., 1991, Speech patterns following partial glossectomy for small tumours of the tongue. *Head Neck*, **13**, 340–343.

Kazi, R., Prasad, V., Kanagalingam, J., Nutting, C. M., Clarke, P. and Harrington, K. J., 2006, Assessment of the Formant Frequencies in Normal and Laryngectomized Individuals Using Linear Predictive Coding. *J Voice*. 2006, September (Epub ahead of print).

Knuuttila, H., Pukander, J., Maatta, T., Pakarinen, L. and Vilkman, E., 1999, Speech articulation zafter subtotal glossectomy and reconstruction with a myocutaneous flap. *Acta Otolaryngology*, **119**, 621–626.

Korpijaakko-Huuhka, A. M., Lehtihalmes, M. and Soderholm, A., 1995, Speech defects after oral cancer surgery. Functional and acoustical analysis of retrospective data. In K. Elenius and P. Branderud (eds), *Proceedings of the 13th International Congress Phonetic Science*, pp.508–511.

Patel, S. G., Archer, D. J. and Henk, J. M., 2003, Tumours of the oral cavity. In P. Rhys-Evans, P. Q. Montgomery and P. J. Gullane (eds), *Principles and Practice of Head and Neck Oncology* (London: Martin Dunitz), pp. 163–192.

Rentschler, G. J. and Mann, M. B., 1980, The effects of glossectomy on intelligibility of speech and oral perceptual discrimination. *Journal of Oral Surgery*, **38**, 348–354.

Stevens, K. N. and House, A. S., 1955, Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, **27**, 484–493.

Wakumoto, M., Ohno, K., Imai, S., Yamashita, Y., Akizuki, H. and Michi, K. I., 1996, Analysis of articulation after glossectomy. *Journal of Oral Rehabilitation*, **23**, 764–770.

Whitehill, T. L., Ciocca, V., Chan, J. C. and Samman, N., 2006, Acoustic analysis of vowels following glossectomy. *Clinical Linguistics and Phonetics*, **20**, 135–140.