Electroglottographic and Perceptual Evaluation of Tracheoesophageal Speech

Rehan Kazi, Jeeve Kanagalingam, Ramachandran Venkitaraman, Vyas Prasad, Peter Clarke, Christopher M. Nutting, Peter Rhys-Evans, and Kevin J. Harrington, *London, United Kingdom*

Summary. To optimize tracheoesophageal (TO) speech after total laryngectomy, it is vital to have a robust tool of assessment to help investigate deficiencies, document changes, and facilitate therapy. We sought to evaluate and validate electroglottography (EGG) as an important tool in the multidimensional assessment of TO speech. This study is a cross-sectional study of the largest cohort of TO speakers treated by a single surgeon. A second group of normal laryngeal speakers served as a control group. EGG analysis of both groups using connected speech and sustained vowels was performed. Two trained expert raters undertook perceptual evaluation using two accepted scales. EGG measures were then analyzed for correlation with treatment variables. A separate correlation analysis was performed to identify EGG measures that may be associated with perceptual dimensions. Our data from EGG analysis are similar to data obtained from conventional acoustic signal analysis of TO speakers. Sustained vowel and connected speech parameters were poorer in TO speakers than in normal laryngeal speakers. In perceptual evaluation, only grade (G) of the GRBAS scale and Overall Voice Quality appeared reproducible and reliable. T stage, pharyngeal reconstruction and method of closure, cricopharyngeal myotomy, and postoperative complications appear to be correlated with the EGG measures. Five voice measures—jitter, shimmer, average frequency, normalized noise energy, and irregularity—correlated well with the key dimensions of perceptual assessment. EGG is an important assessment tool of TO speech, and can now be reliably used in a clinical setting.

Key Words: Electroglottography–Tracheoesophageal speech–Laryngectomy–Voice prosthesis.

INTRODUCTION

The restoration of voice after total laryngectomy by creation of a tracheoesophageal fistula and insertion of a one-way valve is considered the gold standard in the voice rehabilitation of laryngectomees.[1] The resultant voice—tracheoesophageal (TO) speech—although superior to other forms of alaryngeal speech is, however, highly variable.[2] This is possible due to the aerodynamic and myoelastic properties of the tracheoesophageal fistula and neoglottis.[3] Aspects of surgical reconstruction of the pharynx, adjuvant treatment modalities for laryngeal cancer, and treatment complications or toxicities can affect these properties and influence the resultant voice.

The *sine qua non* to further understand these associations is a robust method of voice assessment. This may guide surgical techniques during laryngectomy and help produce better TO speech. Objective voice assessment may also be useful in aiding speech therapists coaching TO speakers. At present, there is no single method of assessing TO speech reliably. Much of the analysis of TO speech, thus far, has been the analysis of speech signal of sustained vowels. This method is inadequate as TO speech tends to be aperiodic and low in pitch, resulting in a large proportion of the speech sample being excluded from analysis.[1,4]

It is widely accepted that the analysis of laryngeal speech by analyzing the glottal waveform during connected speech and utterance of a sustained vowel /i/ is robust and clinically useful.[1,4] This is achieved by directly measuring the electrical impedance across the neck of the subject in a method known as electroglottography (EGG).[4,5] This method has never been used before to analyze and validate TO speech and its links to perceptual data. We hypothesized that EGG analysis of TO speech would overcome the problems of low sampling rates and inadequate pitch extraction algorithms that plague speech signal analysis methods. In this study, we aim to analyze EGG as a robust tool for the assessment of TO speech. We have studied a substantially larger cohort of TO speakers using EGG with both connected speech and sustained vowels as compared to our earlier pilot study on the same. To validate this method, we studied a cohort of normal laryngeal speakers and we sought to show that the data we obtained were comparable to those obtained from similar cohorts of TO speakers analyzed by conventional speech signal analysis. Voice parameters were then correlated with treatment variables to establish the "biologically relevant" measures.

Notwithstanding the power of objective voice analysis, the final arbiter of speech quality is its perception to the listener.[6] Correlation of the voice parameters to perceptual indices is important to determine not only the crucial links between objective assessment and auditory perception but also to help optimize the intelligibility.[7,8] To this end, we undertook a perceptual evaluation of our cohort of TO speakers as an important step forward from our earlier pilot study. The data from perceptual evaluation using two trained expert raters were analyzed to look for consistency, and then correlated with the EGG voice parameters. This has enabled us to identify parameters that are meaningful and clinically relevant.

This study shows that EGG has a vital place in the objective analysis of TO speech. Specific parameters of EGG analysis correlate well with perceptual evaluation of TO speakers. We

Accepted for publication January 11, 2007.
This work is supported by the Head and Neck Research Fund.
From the Head and Neck Unit, Royal Marsden Hospital and Institute of Cancer Research, London, United Kingdom.
Address correspondence and reprint requests to Rehan Kazi, Head and Neck Unit, Royal Marsden Hospital, 203 Fulham Road, London SW3 6JJ, United Kingdom. E-mail: rehan.kazi@rmh.nhs.uk
Journal of Voice, Vol. 23, No. 2, pp. 247-254
0892-1997/$36.00

doi:10.1016/j.jvoice.2007.01.006

propose that optimal multidimensional voice assessment of TO speech should include an EGG analysis of connected speech and sustained vowels supplemented with perceptual evaluation.

PATIENTS AND METHODS

Patients (TO speakers) and normal controls

TO speakers without evidence of disease recurrence were identified from the speech and language therapy database at our institution. Forty-seven patients agreed to participate in this research project, of whom 36 were males and 11 females. All these patients were operated on by the same surgeon. The mean age at the time of voice assessment was 63.3 years (standard deviation [SD] = 10.5 years), with the median time from completion of treatment to voice assessment being 81 months (range = 3–253 months).

All TO speakers, with the exception of one, were using tracheoesophageal valves of the Blom-Singer variety. Thirteen used indwelling valves and the remaining had ex-dwelling ones. Eight patients had a Blom-Singer duckbill prosthesis, whereas the remaining 39 used standard low-profile Blom-Singer valves. Just over half the TO speakers studied (26/47) had 16 French (Fr) gage valves and the remaining required larger 20 Fr valves. The median valve length was 10 mm (range = 4–14 mm).

Of the 47 TO speakers, 22 had a primary total laryngectomy and 25 had a salvage laryngectomy after radiotherapy with or without chemotherapy. In seven patients, partial pharyngectomy with reconstruction (pectoralis major myocutaneous flap—five, radial free flap—two) was undertaken in addition to total laryngectomy and in three a circumferential pharyngectomy, that is, total pharyngolaryngectomy with jejunal free flap reconstruction was undertaken.

A sample of normal laryngeal speakers was drawn from staff members of the hospital. These normal controls had normal vocal folds—in structure and function—with no history of laryngeal, neurological, or speech disorders. This group comprised 31 subjects of whom 18 were males and 13 females. The mean age of this cohort was 40.9 years with an SD of 13.5 years.

The local research and ethics committee approved the study.

Equipment

Acoustic and EGG analyses were performed using the "Speech Studio" equipment and software (Laryngograph Ltd, UK). All recordings were obtained with the subject in a comfortable seated position and a pair of surface electrodes (gold-plated, 3 cm) attached on either side of the thyroid alae (in normal subjects with an intact larynx) or a few centimeters above the stoma and 3–4 cm apart (in laryngectomees). Sound was received by a small *capsule electret* microphone (Sony Ltd, UK) held in front of the chest at a constant mouth-to-microphone distance of 15 cm and at angle of 45°. The electrodes were supplied with an AC sinusoidal current of high alternating frequency (3 MHz) so as to bypass the overlying skin without the use of additional conductive paste.

The signal was transmitted to the Laryngograph processor that consisted of an amplifier and fundamental frequency ("pitch") extractor linked to a recorder to capture and play back speech and waveforms. The Sp and Lx signals were acquired at 16 kHz, 16 bits signal resolution. Pitch was extracted by use of a specific dedicated hardware circuit and the period was counted by a 12-MHz clock and rounded down to single microsecond to ensure accuracy.

Voice recording protocol

The amplitude of the signal was adjusted to the optimal gain position for each individual subject before the actual recording. After familiarization, recordings were made in the following sequence: (1) sustained vowel /i/ produced at a comfortable pitch and loudness stably for at least 5 seconds duration; (2) maximum phonation time (MPT, seconds): sustained vowel /i/ produced at a comfortable pitch and loudness for as long as they could manage after maximal inspiration; and (3) read the standard text provided ("Arthur the rat") at a comfortable pitch, loudness, and pace.

Measures

All subjects provided synchronous electroglottographic and acoustic recordings of both sustained vowels and connected speech in a single recording session.

The resultant electroglottographic waveform (Lx) allows various voice parameters to be accurately determined. However, importantly, the derivation of these parameters is fundamentally different from analysis of the standard acoustic waveform. Voice parameters including the fundamental frequency (Fx—calculated from the Lx waveform on a cycle-by-cycle basis) as opposed to that derived from the Fast Fourier analysis of the acoustic waveform and jitter (short-term frequency perturbation or changes) were obtained from the Lx waveform, and shimmer (short-term amplitude perturbation) and normalized noise energy (NNE)—log ratio of energy contained in the noise to the energy in the signal—were derived from the acoustic waveform. MPT was determined using sustained vowels. NNE is another more select/specific measure of the "Harmonic-to-Noise Ratio" and is useful in determining the noise that may be embedded within pathological voice signals.[8]

Connected speech (standard text) was used to determine the words per minute (WPM) and for other quantitative measures of the larynx such as the larynx frequency distribution (DFx 1st order, Hz) and irregularity (CFx, %) from the frequency cross-plot.[5] These help to quantify the frequency regularity and the speech quality (for the whole sample—and overall intrinsic structure of the voice). DFx 1st order is the larynx frequency distribution for individual vocal fold periods. Here, the frequency (Fx) is plotted horizontally on a logarithmic scale, and vertically on the probability of occurrence of a particular frequency. Because this histogram takes into account all periods (Tx) measured, it is called "first order distribution of excitation fundamental frequencies."

CFx is a scattergram (or cross-plot) that plots successive Fx period by period and looks at the differences in the frequencies of successive periods. It is a reflection of vocal regularity expressed as percentage.[5] The principle is similar to the one applied for jitter measurements in sustained sounds, but now applied to connected speech or sounds. The importance of these voice parameters have been clarified in Tables 1 and 2.

Perceptual evaluation

Two independent trained expert voice raters, blinded to the treatment details of the subjects were asked to rate voice quality using the GRBAS scale and the Overall Voice Quality (OVQ).[9–12] This system consists of five well-defined parameters: G (overall grade of hoarseness), R (roughness), B (breathiness), A (asthenicity), and S (strain). A four-point ordinal scale is used to rate each parameter from 0 to 3 (0—normal, 1—mild, 2—moderate, and 3—severe). The OVQ scale is another perceptual scale that has three ratings: 1—good, 2—reasonable, and 3—poor. This scale is similar to one used by van As in Holland.[2,9] Recordings of the connected speech (standard text—"Arthur the rat," in Appendix 1 and taking approximately 2½ minutes to complete at comfortable pace and loudness) were played from the equipment's two speakers at a standard distance of 1 m in a sound-treated room to each rater as many times as necessary for raters to make a judgment. For the first rater, recordings were played back after an interval of 3 weeks to assess intrarater (test-retest) reliability.

Statistical analysis

All data were analyzed using the *Statistical Package for Social Sciences* (SPSS Inc., Chicago, IL) version 14. The data from the normal subjects and TO speaker groups were compared using the Mann-Whitney *U* test as they did not meet tests for normality. Correlation between sustained vowel and connected speech was assessed using Pearson's correlation coefficient. Perceptual evaluation data from the two expert raters were analyzed for internal consistency by deriving Cronbach's α coefficient. The interrater reliability and test-retest reliability of perceptual evaluation were examined by intraclass correlation coefficient (ICCC) testing. Finally, we studied correlation between EGG voice parameters and treatment variables and then between EGG parameters and perceptual data. The Mann-Whitney or Kruskall-Wallis (analysis of variance by ranks) test was used for these analyses. On account of the multiple correlations across the paper, a Bonferroni's correction was applied.

TABLE 1.
Importance of the Sustained Vowel Parameters

Parameter	Importance
Fundamental frequency	Rate at which voicing source vibrates, correlates with the pitch
Jitter	Frequency perturbation of changes, reflection of fine motor control
Shimmer	Amplitude perturbation of variation, again a reflection of the motor control
NNE	Estimate of the noise that may be embedded within pathological voice signals
MPT (s)	Estimate of the breath support

TABLE 2.
Importance of the Connected Speech Parameters

Parameter	Importance
Larynx frequency distribution (DFx 1st order)	Picks up the frequency regularity
Frequency cross-plot (CFx)	Gives an overview of intrinsic structure and regularity of the voice
WPM	Estimate of the fluency of the speech

RESULTS

Data from both groups, TO speakers and normal volunteers, are presented together.

Objective sustained vowel analysis

Data for normal controls and TO speakers are presented in Table 3. These data are obtained from the use of the sustained vowel /i/ at a comfortable pitch and loudness using the stable mid-portion of the recording for analysis.

The results show a significantly lower fundamental frequency and poorer acoustic measures for TO speakers as compared to normal controls. The average fundamental frequency for normal male subjects was 127.0 Hz and that for normal female subjects was 231.0 Hz. In comparison, male TO speakers had an average fundamental frequency of 98.2 Hz and female TO speakers had an average frequency of 120.6 Hz. This significant difference in fundamental frequency highlights a key deficiency in TO speech in women, that is, a fundamental frequency that is low and little different from a normal male voice.[13]

Measures of jitter, shimmer, and normalized noise energy were similarly significantly poorer for the TO speakers as compared to the normal subjects reflecting a poorer control of the voice.

Objective connected speech analysis

In this section, data were collected using the entire connected speech material based on the patients' reading of the "Arthur the rat" passage at a comfortable pitch and loudness (Appendix 1).

Results from the normal and TO speakers are presented in Table 4.

These results show statistically significantly poorer values in all quantitative measures and larger variability in TO speakers as compared to normal subjects. Once again, the quantitative measure of connected speech frequency—DFx 1st order and the frequency cross-plot (CFx, irregularity, %) differed considerably between normal female subjects and female TO speakers.

Temporal measures

Data on the MPT using the sustained vowel /i/ expressed in seconds and WPM using the standard reading passage are

TABLE 3.
Sustained Vowel /i/ Analysis in Normal Subjects and TO Speakers

Speech Parameters	Normal Subjects (N = 31)	Males (N = 18)	Females (N = 13)	TO Speakers (N = 47)	Males (N = 36)	Females (N = 11)	*P* Value* (M/F)
Av. fundamental frequency (Hz)	171.3 (91.4 to 313.8)	127	231	103.8 (45.2 to 302.3)	98.2	120.6	(0.001/0.001)
Jitter (%)	0.4 (.06 to 1.39)	0.6	0.3	5.9 (0.2 to 36.3)	5	8.6	0.0001 (<0.001/<0.001)
Shimmer (dB)	0.9 (0.2 to 2.5)	1	0.8	2.1 (0.8 to 6.8)	2.2	2	0.0001 (<0.001/<0.001)
NNE (dB)	−19.1 (−28.2 to −9.4)	−19.4	−18.8	−3.3 (−14.3 to 1.6)	−3.1	−3.5	0.0001 (<0.001/<0.001)
MPT (s)	23.9 (10.8 to 43.8)	25.8	21.2	11.8 (2.2 to 32.4)	12.5	10.2	0.0001 (<0.001/0.006)

Notes: All figures represent medians with range in parentheses.
* *P* value comparing all normal subjects with all TO speakers using Mann-Whitney *U* test with the *P* values of the correlation between normal males and male TL/normal females and female TL in parenthesis.

presented in Tables 3 and 4. Among normal subjects and TO speakers, there is little difference in both temporal measures between males and females. Not surprisingly, TO speakers have significantly lower values for both measures but the magnitude of the difference is stark. MPT is almost halved in TO speakers but the WPM rate is only reduced by 20%.

Perceptual evaluation

Two independent trained expert raters were given voice recordings of the normal subjects and TO speakers and were asked to rate these according to the GRBAS scale and the simpler OVQ scale. The results of this evaluation in median (range) are presented in Table 5.

The raters were clearly able to distinguish normal speakers from TO speakers effectively and appeared to ascribe higher scores (poorer values) to TO speakers as compared to the normal subjects across the scale.

Table 6 shows the results of interrater reliability and test-retest reliability examined using the ICCC testing. When rating TO speakers, there appears to be more reliability between raters when ascribing an overall grade (G) to speech quality on the GRBAS scale and designating the OVQ grade as compared to the other parameters. There appears to be some concordance between our raters when evaluating the breathy, asthenic, and strained nature of TO speech and but less agreement in deciding on roughness.

Data from expert rater 1 (ER1) were used to examine the test-retest reliability. ER1 rated the same voice recordings of our TO speakers 3 weeks apart for this purpose. There appears to be a high degree of test-retest reliability in the G, R, and B dimensions of the GRBAS scale and the OVQ score. Collectively, these analyses suggest that the overall grade (G) of GRBAS, OVQ, and to some extent breathiness (B) of the GRBAS scale are robust perceptual parameters when assessing TO speakers. There was an excellent correlation between G and OVQ (Spearman rank, 0.97, $P < 0.001$).

Correlation analyses between speech parameters and treatment variables

We sought to find any correlation between EGG measures and treatment variables. This was analyzed by the Mann-Whitney *U* test when considering dichotomous treatment variables or the Kruskall-Wallis test when analyzing multiple variables. A number of significant correlations (ie, $P < 0.05$) were seen using both sustained vowel and connected speech and they are listed in Table 7.

The strongest correlation is between average fundamental frequency and reconstruction of the pharynx by either pedicled flaps or free tissue transfer. Patients requiring reconstruction appear to have a significantly lower frequency than those who have a primary closure of the pharynx. Frequency is also correlated with the type of pharyngoesophageal (PE) segment

TABLE 4.
Connected Speech Analysis in Normal Subjects and TO Speakers

Speech Parameters	Normal Subjects (M/F)	TO Speakers (M/F)	*P* Value* (M/F)
DFx 1st order (Hz)	149 (91.3–229.2) 120.3/182.3	89.5 (55–153.8) 87/95	(<0.001/0.001)
CFx irregularity %	11.3 (2.8–13.8) 11.5/7.7	62.9 (19.7–89.8) 58.9/71.4	0.0001 (<0.001/<0.001)
WPM	165.84 (54–215) 161.3/172	134.5 (97–214) 136.7/126.2	0.0001 (0.009/0.003)

Notes: All figures represent medians with range in parentheses.
* *P* value comparing all normal subjects with all TO speakers using Mann-Whitney *U* test, figures in parentheses represent the *P* value of the correlation between normal males and male TL/normal females and female TL.

TABLE 5.
Perceptual Evaluation Results

	G	R	B	A	S	OVQ
Normal subjects						
ER1	0 (0–1)	0.5 (0–1)	0 (0–0)	0 (0–0)	0 (0–1)	1 (1–3)
ER2	0 (0–1)	0 (0–1)	0 (0–1)	0 (0–1)	0 (0–1)	1 (1–3)
TO speakers						
ER1	2 (1–3)	2 (1–3)	2 (1–3)	1 (1–3)	1 (0–3)	2 (1–3)
ER2	2 (1–3)	2 (1–3)	2 (0–3)	1 (1–3)	1 (1–3)	2 (1–3)
P value (Mann-Whitney)						
ER1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
ER2	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Notes: All figures are medians with range in parentheses.
Abbreviations: ER1, expert rater 1; ER2, expert rater 2.

(pharyngeal) closure with a T-shaped closure and horizontal closures giving similar and higher frequencies than circumferential ones. Complications and higher T stages also appear to be correlated with a lower frequency.

Reconstruction also appears to correlate with higher NNE. We also found significant correlation of jitter, shimmer, and irregularity with gender, primary site, and primary versus recurrent disease. The correlation between MPT and presence of a surgical myotomy is real and is interesting.

Correlation analyses were also undertaken to identify associations between speech parameters and perceptual evaluation data. The significant associations are presented in Table 8. Given the poor inter- and intrarater reliabilities for R, A, and S dimensions of the GRBAS scale, the useful associations appear to be confirmed to five parameters, that is, jitter, shimmer, average fundamental frequency, NNE, and irregularity (CFx) with overall grade (G), breathiness (B) of the GRBAS and OVQ.

DISCUSSION

TO speech differs significantly from pathological laryngeal voice in that it is more aperiodic and of a lower fundamental frequency.[1] Groups elsewhere have attempted to analyze the voice using a range of methods. One commonly followed method is the use of acoustic signal typing as described by Titze which has been widely applied in TO speakers.[14,15] However, it would be more appropriate if the glottic signal is used to assess the voice. In our unit, we have used EGG of connected speech and sustained vowel /i/ as a tool to assess, document, and investigate TO speech. Using high sampling rates and specific pitch extraction algorithms, EGG allows a more thorough and logical analysis of the important glottic signal.[4,5] It can also be used to provide the speaker with some biofeedback during therapy sessions with the speech therapist.

The cohort of TO speakers that we have studied is to date the largest single cohort of patients from a single institution and treated by a single surgeon. EGG in our cohort has yielded results similar to other studies using speech signal analysis.[9,15–19] This supports our assertion that EGG is a reliable and robust tool in assessing TO speech.[20] Given that normal laryngeal speakers analyzed by the same methodology and by the same assessor produce normative data further lends validity to our technique.

Results obtained from EGG analysis of sustained vowel /i/ clearly show that TO speech has greater variability than normal laryngeal speech. Jitter, shimmer, and NNE were poorer in TO speakers than in normal subjects. This is not surprising as normal speech is characterized by smooth onset, offset, and absence of pitch breaks, whereas the same cannot be said of alaryngeal speech which lacks fine motor control.[2] The results of the quantitative measures from connected speech further illustrate this point and highlight the extent to which TO speech is variable and "pathological" when compared to the normal laryngeal speech. This variability is a characteristic feature of TO speech and could be a result of the larger anatomical and morphologic variation of the neoglottis as compared to the vocal folds. Another possible reason could be the inclusion of patients who have undergone partial or total pharyngeal reconstruction. Female TO speakers in our cohort had an Fx of 120.6 Hz. This is

TABLE 6.
IR and TT Reliabilities of Perceptual Evaluation Using ICCC

Reliability	Rater	G	R	B	A	S	OVQ
Normal subjects (IR)	ER1:ER2	0.8	0.8	1	1	1	1
TO speakers (IR)	ER1:ER2	0.9	0.6	0.7	0.7	0.7	0.9
TO speakers (TT)	ER1	0.9	0.8	0.8	0.5	0.5	0.9

Abbreviations: ER1, expert rater 1; ER2, expert rater 2; IR, interrater; TT, test-retest.

TABLE 7.
Correlation Analyses of Speech Parameters Versus Treatment Variables

Speech Parameter	Treatment Variable	*P* Value*
Ave. frequency		
119.1 (63.6)—T1,2/83.1 (30.2)—T3,4	T stage	0.04
124.5 (64.7)—Horizontal/64.8 (15.1) Circumferential/118.3 (30.2)—T-shaped	PE segment closure	0.006
67.1 (20.4)—Yes/120.3 (62.3)—No	Reconstruction	0.001
68.6 (30.1)—Yes/113.5 (59.0)—No	Complication	0.04
Jitter		
5.9 (1.5)—Oropharynx/13.1 (8.9)—Hypopharynx/5.4 (8.7)—Larynx	Primary site	0.04
5.0 (8.4)—Male/8.6 (7.7)—Female	Gender	0.04
NNE		
−1.4 (2.6)—Yes/−4.9 (4.3)—No	Reconstruction	0.01
DFx 1st order		
[T-shaped/Circumferential] 102.7 (28.4)/66.9 (23.2)	PE segment closure	0.003
[T1,2/T3,4] 95.8 (43.7)/88.5 (34.9)	T stage	0.003
Irregularity (CFx)		
59.3 (20.4)—Male/73.9 (15.6)—Female	Gender	0.04
64.7 (22.8)—Primary/61.5 (17.4)—Recurrence	Primary vs recurrence	0.02
MPT		
13.1 (7.2)—Yes/8.5 (5.6)—No	Myotomy	0.03

Notes: Only significant associations ($P < 0.05$) are presented. Figures are means with SDs in parentheses.
**P* value comparing all normal subjects with all TO speakers using Mann-Whitney *U* and Kruskall-Wallis test.

higher than male TO speakers (98.15 Hz) and may be related to tissue density of the vibrating PE segment. It remains, however, a significantly lower pitch than female laryngeal speakers (231.0 Hz). This is a significant disability and a source of discontent among the female TO speakers.[13]

The poorer temporal measures of WPM and MPT were again significantly poorer in the TO speakers as against the normal subjects. This can be explained by the fact that TO speakers have reduced breath support due to varying amounts of air leakage at stoma occlusion. Also, they have to constantly alternate 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.

TABLE 8.
Correlation Analyses of Speech Parameters and Perceptual Data

Speech Parameter	Perceptual Data (GRBAS and OVQ)	*P* Value*
Jitter	G and OVQ	0.02
Shimmer	G and OVQ	0.02
Irregularity (CFx)	G and OVQ	0.003
Ave. frequency	B	0.04
Shimmer	B	0.05
NNE	B	0.01
Irregularity (CFx)	A	0.02
MPT	S	0.03

Notes: Only significant ($P < 0.05$) associations are presented. Figures are means with SDs in parentheses.
**P* value comparing all normal subjects with all TO speakers using Mann-Whitney *U* and Kruskall-Wallis test.

Methods of objective analysis thus far have relied on the analysis of sustained vowels and not the more practical connected speech.[4,5] Sustained vowels have traditionally been used as they are easy to analyze, are produced in a controlled environment, and have been robustly studied. However, they miss the onset/offsets that are normally present in everyday speech. Moreover, connected speech allows us to examine the intrinsic structure of the voice and is a reflection of the motor control. We therefore advise use of both sustained vowels and connected speech for a true assessment of the voice.

The EGG measures when correlated with treatment variables produced many significant associations. It is clear that some of the associations although significant are difficult to explain. However, some of them are real and deserve further attention. For example, the association of fundamental frequency and treatment variables, such as T stage, PE segment closure, complications, and reconstruction. A higher Fx is seen with patients with a horizontal PE segment closure, with no complications and no reconstruction. Certainly this is something to be aspired to and recreated in TL patients. TO speech is clearly of a lower pitch or fundamental frequency than normal laryngeal speech. Fundamental frequency is determined by the following equation: $Fx = 1/2L\sqrt{(T/P)}$, where L is the length of the vocal fold, T is the mean longitudinal stress, and P is tissue density.[21]

In the case of the neoglottis or PE segment, the myoelastic properties are clearly different. Mean longitudinal stress (*T*) is small and tissue density (*P*) large producing a lower Fx. Our data also show that where there has been pharyngeal reconstruction, Fx is substantially lower. The correlation with T stage is clear, but is probably a factor of the need for reconstruction, that is, large T stage tumors and more likely to require partial or total pharyngectomy in addition to laryngectomy.

Association of increasing jitter and CFx irregularity with female TO speakers is not surprising and expected as jitter increases with increasing frequency.[8] Similarly, the association of an increased MPT with surgical myotomy during laryngectomy is possibly due to its influence on the tonicity of the PE segment. A tonic PE segment would be capable of increased MPT.

The perceptual analysis of TO speakers and correlation with EGG measures were attempted with the intention to provide clinical credibility to our methodology.[6,7] We used experienced raters and a common standard passage as this has been shown in the literature to be the most reliable technique.[10–12] The use of a judgment of an OVQ allows a concise and a simple overall impression of the voice quality.[2] The perceptual evaluation of our cohort of TO speakers by trained raters using the GRBAS and OVQ scales establishes that the overall grade (G) and OVQ are clearly reproducible and reliable measures. These two scales significantly correlated with average fundamental frequency, jitter, shimmer, NNE (from use of sustained vowel), and irregularity (from use of connected speech) thus establishing the links between objective and perceptual data. The authors feel that both OVQ and GRBAS can be used reliably with expert raters to obtain an overall "impression" in TO speakers but this should be supplemented with other methods of assessment.

In summary, this study of a large cohort of TO speakers by EGG using both connected speech and sustained vowel establishes that this method is robust, valid, and reliable.

Correlation of the voice parameters with treatment and perceptual data provides "real-life" significance to these parameters. We propose that these sustained vowel and connected speech measures should be focused on when assessing TO speech. These can be the basis of typing and help therapists categorize TO speakers into groups that will respond favorably to training. It may also help therapists develop targeted strategies directed toward addressing specific deficiencies identified on EGG and to optimize vocal rehabilitation. Perhaps, this may lead to further research into surgical techniques to alter the PE segment and consequently the voice parameters akin to phonosurgery of the vocal cords.

CONCLUSIONS

This study shows that robust, reliable, and valid data could be obtained using EGG in normal volunteers and laryngectomees using both a sustained vowel and connected speech. All electroglottographic and acoustic measures of voice such as fundamental frequency, jitter, shimmer, NNE, and the temporal measures were significantly poorer than normal laryngeal speakers. This was further validated on perceptual evaluation using the GRBAS and OVQ scales by trained expert raters. Although the advent of the prosthetic valve in laryngectomees has greatly improved their speech and communication and is considered the "gold standard," it is still significantly different and highly variable as compared to normal speech.

We propose a multidimensional assessment of TO speech that includes objective and quantifiable EGG sustained vowel/connected speech measures and perceptual evaluation. These measures could be the basis of further typing and help therapists categorize TO speakers into groups for targeted strategies directed toward addressing specific deficiencies and to optimize vocal rehabilitation.

REFERENCES

1. Rhys-Evans P, Montgomery PQ, Gullane PJ. *Principles and Practice of Head and Neck Oncology.* UK: Martin Dunitz; 2003.
2. van As CJ, Koopmans-van Beinum FJ, Hilgers FJM, Pols LCW. Perceptual evaluation of tracheoesophageal speech by naive and experienced judges through the use of semantic differential scales. *J Speech Lang Hear Res*. 2003;46:947-959.
3. Moon JB, Weinberg B. Aerodynamic and myoelastic contributions to tracheoesophageal voice production. *J Speech Hear Res*. 1987;30:387-395.
4. Fourcin A, Abberton E. First applications of a new laryngograph. *Med Biol Illus*. 1971;21:172-182.
5. Fourcin AJ. Voice quality and electrolaryngography. In: Kent RD, Ball MJ, eds. *Voice Quality Measurement*. San Diego: Singular Publishing Group; 2000:285-306.
6. Bassich CJ, Ludlow C. The use of perceptual methods by new clinicians for assessing voice quality. *J Speech Hear Disord*. 1986;51:125-133.
7. Carding PN, Steen IN, Webb A, et al. The reliability and sensitivity of change of acoustic measures of voice quality. *Clin Otolangol*. 2004;29:538-544.
8. Baken RJ. *Clinical Measurement of Speech and Voice*. Boston, MA: Little, Brown & Company; 1997.
9. van As CJ, Hilgers FJ, Verdonck-de Leeuw IM, Koopmansvan Beinum F. Acoustical analysis and perceptual evaluation of tracheoesophageal prosthetic voice. *J Voice*. 1998;12:239-248.
10. Dejonckere PH, Wieneke GH. GRBAS-scaling of pathological voices: reliability, clinical relevance, and differentiated correlation with acoustic measurements, especially with spectral measurements. In: Loebell E, ed. *Proceedings of the 22th World Congress IALP, 1992*. Hannover: IALP; 1992.
11. De Bodt MS, Wuyts FL, Van de Heyning PH, Croux C. Test-retest study of the GRBAS scale: the influence of experience and professional background on perceptual rating of voice quality. *J Voice*. 1997;11:74-80.
12. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res*. 1993;36:21-40.
13. Kazi R, Kiverniti E, Prasad V, et al. Multidimensional assessment of female tracheoesphageal speech. *Clin Otolaryngol*. 2006;31:511-517.
14. Titze IR. *Workshop on Acoustic Voice Analysis*. Summary Statement. Iowa City, IA: National Center for Voice and Speech; 1994.
15. Pindzola RH, Cain BH. Acceptability ratings of tracheoesophageal speech. *Laryngoscope*. 1988;98:394-397.
16. Blood GW. Fundamental frequency and intensity measurements in laryngeal and alaryngeal speakers. *J Commun Disord*. 1984;17:319-324.
17. Robbins J. Acoustic differentiation of laryngeal, esophageal and tracheoesophageal speech. *J Speech Hear Res*. 1984;27:577-585.
18. Robbins J, Fisher HB, Blom EC, Singer MI. A comparative acoustic study of normal, esophageal and tracheoesophageal speech production. *J Speech Hear Res*. 1984;49:202-210.
19. Qi Y, Weinberg B. Characteristics of voicing source waveforms produced by esophageal and tracheoesophageal speakers. *J Speech Hear Res*. 1995;38:536-548.
20. Kazi R, Singh A, DeCordova J, et al. Electrographic analysis of valved speech following total laryngectomy. *Logop Phoniat Vocol*. In press.
21. Scherer RC, Guo CG. Laryngeal modelling. Translaryngeal pressure for a model with many glottal shapes. *Proc Acoust Soc Jpn*. 1990;1:1-4.

Appendix 1

Arthur the rat

There was once a young rat named Arthur who would never take the trouble to make up his mind. Whenever his friends asked him if he would like to go out with them he would only answer, "I don't know." He wouldn't say "Yes" and he wouldn't say "No" either. He could never learn to make a choice.

His aunt Helen said to him "No-one will ever care for you if you carry on like this. You have no more mind than a blade of grass." Arthur looked wise but said nothing.

One rainy day the rats heard a great noise in the loft where they lived. The pine rafters were all rotten, and at last one of the joists had given way and fallen to the ground. The walls shook and the rats' hair stood on end with fear and horror. "This won't do," said the old rat who was chief. "I'll send out scouts to search for a new home."

Three hours later the seven scouts came back and said, "We've found a stone house which is just what we wanted. There's room and good food for us all. There's a kindly horse named Nelly, a cow, a calf and a garden with an elm tree." Just then the old rat caught sight of young Arthur. "Are you coming with us?" he asked. "I don't know," Arthur sighed, "The roof may not come down just yet." "Well," said the old rat angrily, "We can't wait all day for you to make up your mind. Right about face! March!" And they went off.

Arthur stood and watched the other rats hurry away. The idea of an immediate decision was too much for him. "I'll go back to my hole for a bit," he said to himself, "just to make up my mind."

That night there was a great crash that shook the earth, and down came the whole roof. Next day some men rode up and looked at the ruins. One of them moved a board, and under it they saw a young rat lying on his side, quite dead, half in and half out of his hole.