The Journal of Laryngology and Otology
April 1994, Vol. **108**, pp. 325–328

Acoustic analysis of tracheo-oesophageal versus oesophageal speech

F. DEBRUYNE, M.D., PH.D., P. DELAERE, M.D., PH.D., J. WOUTERS, PH.D., P. UWENTS, M.D.

Abstract

In order to evaluate the vocal quality of tracheo-oesophageal and oesophageal speech, several objective acoustic parameters were measured in the acoustic waveform (fundamental frequency, waveform perturbation) and in the frequency spectrum (harmonic prominence, spectral slope). Twelve patients using tracheo-oesophageal speech (with the Provox® valve) and 12 patients using oesophageal speech for at least two months, participated.

The main results were that tracheo-oesophageal voices more often showed a detectable fundamental frequency, and that this fundamental frequency was fairly stable; there was also a tendency to more clearly defined harmonics in tracheo-oesophageal speech. This suggests a more regular vibratory pattern in the pharyngo-oesophageal segment, due to the more efficient respiratory drive in tracheo-oesophageal speech. So, a better quality of the voice can be expected, in addition to the longer phonation time and higher maximal intensity.

Key words: Speech, alaryngeal; Speech acoustics

Introduction

For the rehabilitation of the voice after total laryngectomy, several methods are possible, but the two most important are oesophageal speech and tracheo-oesophageal speech (TE speech); the latter can be obtained in several ways, e.g. the use of a low-resistance valve such as the Provox® prosthesis.

It is assumed that in both kinds of alaryngeal speech the vocal vibratory source is the same i.e. the pharyngo-oesophageal segment (PE segment). The PE segment consists of the mucosa and the supporting tissues at the transition from the neopharynx to the oesophagus. However, the driving forces are different i.e. pulmonary expiratory air deviated to the pharynx in one kind and ejected air from the oesophagus in the other.

From an acoustical point of view, tracheo-oesophageal speech is known to be better than oesophageal speech, as the intensity and the duration of the phonation are greater and the intelligibility and acceptability are rated higher (Williams and Watson, 1987; Pindzola and Cain, 1988; Ainsworth and Singh, 1992).

In this paper a more detailed acoustic analysis was carried out by studying the waveform and the frequency spectrum of oesophageal speech and tracheo-oesophageal speech. The main question to be answered was whether the acoustic signal reflects differences in the vibratory behaviour of the vocal source, caused by the dissimilar kinds of air supply.

Patients and methods

Two groups of male patients were compared. The first group consisted of 12 well established oesophageal speakers. They were selected at random during a routine follow-up examination. The second group consisted of 12 tracheo-oesophageal speakers who had used Provox® prosthesis for at least two months.

All the patients, although treated previously had been treated for speech rehabilitation by the same speech pathologist.

The patients (seated in a sound-proofed room) were asked to say a vowel 'a' at a comfortable loudness and duration; the sound was recorded using a DAT (digital audio tape) recorder (Sony 55ES), and analysed using the CSL 4300 (computerized speech lab) of Kay Elemetrics. The sampling rate was 40 000 Hz for the waveform study, and 10 000 Hz for the spectrum study.

For the acoustic waveform (i.e. the acoustic signal as it is registered by the microphone) (see Figure 1) the following parameters were measured:

(i) the fundamental frequency (F_0); this is the repetition rate (in Hz) of the waveform; it corresponds with the basic oscillation rate in the vibratory source (the vocal folds or the PE segment), and is related to the perceived pitch of the voice;

(ii) the stability of the acoustic waveform was estimated through the frequency and amplitude perturbation (also called respiratory jitter and shimmer). Jitter is how much (in percentage) every period of the wave differs from the period that immediately precedes it; shimmer is the same as the variation (in dB) in the amplitude of successive oscillations. These measurements (made in the middle of an equally sustained vowel, following Koike's formula

From the Department of Otorhinolaryngology, Head and Neck Surgery, University of Leuven, Belgium.
Accepted for publication: 1 November 1993.

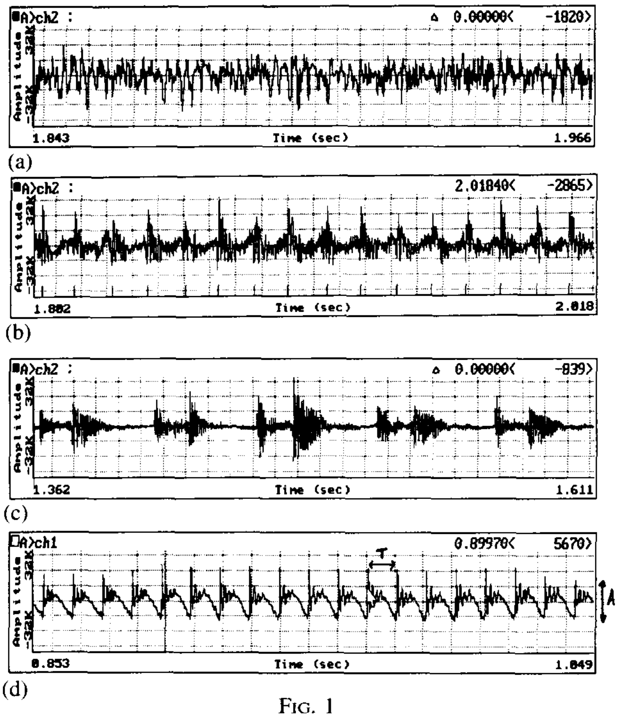

FIG. 1

Examples of acoustic waveforms, in the middle of a vowel 'a'. From top to bottom: (a) nonperiodic waveform, only noise; (b) measurable F_0, in noisy background; (c) paired occurrence of vibrations; (d) for comparison normal laryngeal voice: very regular vibration pattern (T = period; A = amplitude).

(1973)), are not reflecting voluntary changes in pitch or loudness, but the degree of instability of the vibratory source. Also in perfectly normal voicing a certain amount of physiological jitter and shimmer is present. Normal jitter is less than 0.8 per cent, and normal shimmer less than 0.5 dB; in pathological cases, elevated perturbation is related to the roughness of the voice (Hirano *et al.*, 1988);

(iii) the frequency spectrum is the frequency content of the sound. This is obtained by a Fourier analysis of the signal; we used the FFT (Fast Fourier Transform) on 512 points, averaged over 0.25 s in the middle of the vowel. The result is the graphic representation of all frequency components of a sound, in this case from 0 to 5000 Hz; examples are shown in Figure 2. One can see the fundamental, the overtones (or harmonics), the energy distribution and noise level.

The following spectral parameters were measured:

(i) the prominence of the harmonics (or partials). We took the mean difference (in dB) between the peaks and the depths of three successive harmonics in the 500 Hz region of the spectrum. The harmonic prominence, i.e. the distance between the amplitude of the individual harmonics and the interharmonic level (or 'noise level'), is related to the vocal quality (Dejonckere and Lebacq, 1987);

(ii) the slope of the spectrum. This was defined as the difference between the mean spectral intensity below 500 Hz and the mean spectral intensity above 4000 Hz; it can be considered as an estimation of the relative amount of high frequency noise (>4000 Hz) in the spectrum of the voice.

At the end of the procedure, the maximal phonation time and the maximal intensity (30 cm from the mouth) of the vowel 'a' were measured.

Results

Waveform analysis (see Table I)

In half of the oesophageal speakers no fundamental frequency could be determined: the acoustic waveform was completely aperiodic, thus had no regular repetition rate (Figure 1a), or showed only a very weak periodicity. This means that at the level of the sound producing source (i.e. the PE segment) the vibratory behaviour was so variable and inconstant, that no basic rhythm (F_0) was emerging, but only noise was found.

In the other half of the oesophageal speakers, there was a measurable fundamental frequency, lying between 60 and 80 Hz (in one patient 134 Hz). Here the fundamental frequency is reflecting a basic regularity in the vibrations of the PE segment, in a sense comparable to the periodicity seen in normal vocal fold action.

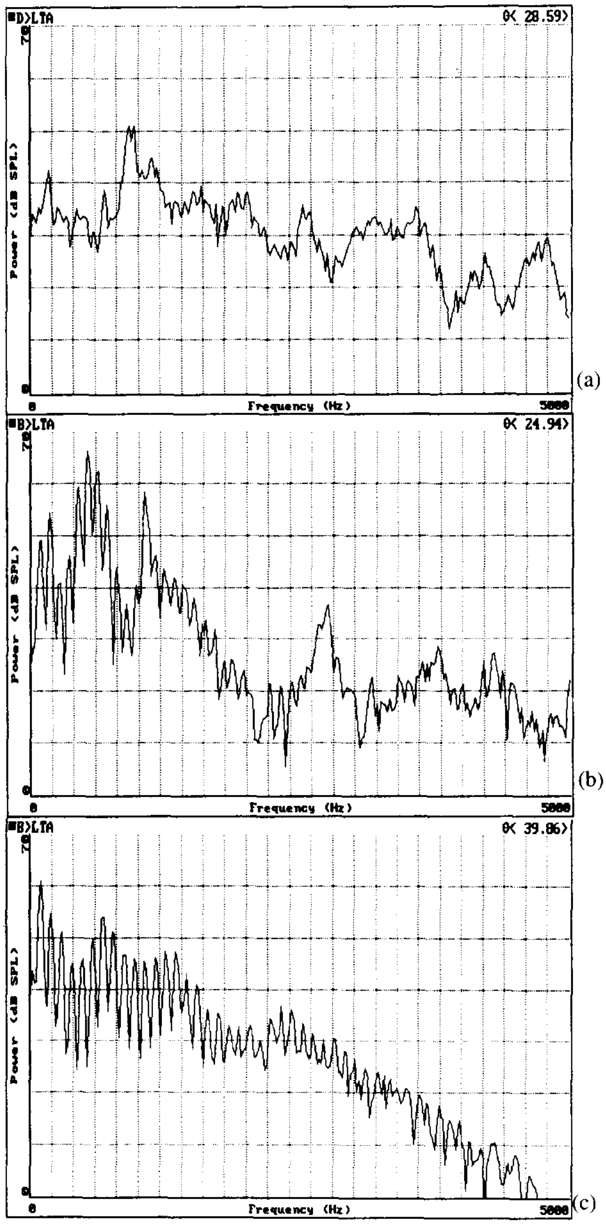

FIG. 2

Examples of frequency spectra (the peaks occurring at regular distances: in the left part of the curves are the harmonics). From top to bottom: (a) spectrum without visible harmonics; (b) spectrum with clear harmonics around 500 Hz; (c) for comparison: spectrum of normal laryngeal voice.

TABLE I
RESULTS OF WAVEFORM ANALYSES

	Oesophageal speech	Tracheo-oesophageal speech
Fundamental frequency (F_0)	In 5/12: F_0 60–80 Hz in 1/12: F_0 134 Hz in 6/12: a periodic or very weak periodicity	In 12/12: F_0 50–110 Hz
Frequency and amplitude perturbation	In 1/12: jitter 2.6% shimmer 1.3 dB In 11/12: F_0 absent or unstable	In 6/12: jitter 0.61–2.6% shimmer 0.4–1.6 dB In 6/12: F_0 unstable

In comparison all tracheo-oesophageal speakers showed a measurable F_0, lying between 50 and 110 Hz (Figure 1b). This seems lower than the pitch of normal male laryngeal voices. In some patients a rather unusual vibration pattern was observed: the oscillations appeared in groups of two, thus in a pairing (Figure 1c); the significance of this observation is not clear at present.

The stability or instability of the vocal signal was further investigated with the help of the frequency and amplitude perturbation measurements, or vocal jitter and shimmer (Table I). In most of the patients jitter and shimmer were greater than in normal voices; in some cases the cycle-to-cycle perturbation was even above the measurement limit, because the F_0 was either too unstable or simply absent. However there was a difference between oesophageal and tracheo-oesophageal speech, in that in half of the tracheo-oesophageal speakers jitter and shimmer remained below the upper limit of the measurement programme. This was true for only one oesophageal speaker. In other words, we found less instability in tracheo-oesophageal speech.

Spectral analyses (see Table II)

In the tracheo-oesophageal group, harmonics or partials around 500 Hz were recognized in seven out of 12 patients, with a prominence between 3 and 25 dB against the interharmonic level. In the oesophageal group, harmonics were present in only one patient. Examples of spectra without visible harmonics and with clear harmonics, in the 500 Hz region, are shown in Figure 2a and b.

The overall slope of the spectrum (related to the energy level above 4 kHz or the amount of high frequency noise) was not statistically different between the oesophageal and tracheo-oesophageal speakers (*t*-test; $p>0.05$) (see Table II). However the spectra of our patients in general appeared flatter (i.e. contained more high frequent noise) than normal laryngeal voices.

The maximal intensity, and the maximal phonation time were clearly different between the two groups, as seen in Table III (*t*-test: $p<0.001$).

Discussion

The results confirm that, with tracheo-oesophageal speech a greater loudness and a longer phonatory duration can be achieved, as already shown in several reports (Robbins, 1984; Robbins *et al.*, 1984; Sanderson *et al.*, 1993). Because we were interested in a more detailed ('microscopic') analysis, a number of acoustic parameters in the time domain and spectrum domain were examined.

Some of these measurements were similar in oesophageal and tracheo-oesophageal speech, but different from normal or slightly dysphonic laryngeal voices. In comparison with normal voices (Figure 2c), our patients showed on average a flatter spectrum and a relatively higher level of energy below 4000 kHz, corresponding with a greater amount of turbulent noise.

Also common to oesophageal and tracheo-oesophageal speech was the lower pitch, in so far as the fundamental frequency could be determined. The lower pitch of alaryngeal speech, compared with laryngeal voices of men of the same age is not a new observation (Robbins *et al.*, 1984; Baken, 1987; Mendelsohn *et al.*, 1993). This seems to be a characteristic of alaryngeal speech in general. However, we found acoustic differences, as some parameters showed 'better' results in tracheo-oesophageal speech than in oesophageal speech. Firstly, the acoustic waveform contained a detectable fundamental frequency more often. Secondly, the perturbation of the successive oscillations (the 'jitter and shimmer') was lower in tracheo-oesophageal speech. Robbins (1984) reported jitter values of 0.77 per cent in normal voices, 5.1 per cent in tracheo-oesophageal speakers, and 18.2 per cent in oesophageal speakers. Thirdly, the prominence of the harmonics in the spectrum was higher in tracheo-oesophageal speech. Sasserath *et al.* (1992) came to the same conclusion using a different technique.

A more stable fundamental frequency and more clearly defined harmonics suggest a better vocal quality and also a better vibration pattern.

It seems that a more regular and stable vibration mode is obtained in the PE segment, due to the expiratory air flow, which is a more efficient driving force than the short ejection of air out the oesophagus. In this context it is noteworthy that a higher pressure is needed to initiate and sustain vibrations in the PE segment than in the vocal folds. For example, pressures of 10–40 cm H_2O were measured under the PE segment during (tracheo-oesophageal voicing; while in normal or even loud laryngeal phonation subglottic pressures of 4–8 cm H_2O are common. In the same way, resistance values in the PE segment of 155–

TABLE II
RESULTS OF SPECTRAL ANALYSES

	Oesophageal speech	Tracheo-oesophageal speech
Harmonic prominence	In 1/12: harmonic prominence 18 dB In 11/12: absence of harmonics	In 7/12: harmonic prominence 3–25 dB In 5/12: absence of harmonics
Spectral slope (ratio >4 kHz >0.5 kHz)	Mean: 14.0 dB (SD 4.39)	Mean: 16.7 dB (SD 7.22)

TABLE III

MAXIMAL INTENSITY AND MAXIMAL PHONATION TIME: GROUP MEANS AND STANDARD DEVIATIONS (SD)

	Oesophageal speech	Tracheo-oesophageal speech
Maximal intensity	Mean: 65 dB (SD 6.75)	Mean: 79.7 dB (SD 3.36)
Maximal phonation time	Mean: 1.54 s (SD 0.62)	Mean: 6.0 s (SD 3.86)

270 cm H_2O/LPS were seen, which is also much higher than the values found at the normal glottis (Weinberg *et al.*, 1982; Wilson and Leeper, 1992).

It can be concluded that the more powerful driving pressure in tracheo-oesophageal speech results in an optimized phonatory process in the PE segment, which is reflected in detectable acoustic advantages. The acoustic advantage is of course only one element in the entire cost/benefit ratio of voice prosthesis in the laryngectomized patient.

References

Ainsworth, W. A., Singh, W. (1992) Perceptual comparison of neoglottal, oesophageal and normal speech. *Folia Phoniatrica* **44:** 297–307.

Baken, R. J. (1987) *Clinical Measurements of Speech and Voice*, College-Hill Press, Boston, pp 154–155.

Dejonckere, Ph., Lebacq, J. (1987) Harmonic emergence in formant zone of a sustained <a> as a parameter for evaluating hoarseness. *Acta Otorhinolaryngologica Belgica* **41:** 988–996.

Hirano, M., Hibi, S., Yoshida, T. (1988) Acoustic analysis of pathological voice. *Acta Otolaryngologica* **105:** 432–438.

Koike, Y. (1973) Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia phonologica* **7:** 17–24.

Mendelsohn, M., Morris, M., Gallagher, R. (1993) A comparative study of speech after total laryngectomy and total laryngopharyngectomy. *Archives of Otolaryngology, Head and Neck Surgery* **119:** 508–510.

Pindzola, R., Cain, B. (1988) Acceptability ratings of tracheo-oesophageal speech. *Laryngoscope* **98:** 394–397.

Robbins, J. (1984) Acoustic differentiation of laryngeal, esophageal, and tracheo-oesophageal speech. *Journal of Speech and Hearing Research* **27:** 577–585.

Robbins, J., Fisher, H., Blom, E., Singer, M. (1984) A comparative acoustic study of normal, esophageal and tracheo-oesophageal speech production. *Journal of Speech and Hearing Disorders* **49:** 202–210.

Sanderson, R. J., Anderson, S. J., Denholm, S., Kerr, A. I. G. (1993) The assessment of alaryngeal speech. *Clinical Otolaryngology* **18:** 181–183.

Sasserath, M., Remacle, M., Ryckaert, M., Trigaux, I. (1992) Vocal rehabilitation after total laryngectomy by phonatory prosthesis. Estimation by high resolution frequency analysis. *Folia Phoniatrica* **44:** 70.

Weinberg, B., Horii, Y., Blom, E., Singer, M. (1982) Airway resistance during esophageal phonation. *Journal of Speech and Hearing Disorders* **47:** 194–199.

Williams, S., Watson, B. (1987) Speaking proficiency variations according to method of alaryngeal voicing. *Laryngoscope* **97:** 737–739.

Wilson, J., Leeper, H. (1992) Changes in laryngeal airway resistance in young adult men and women as a function of vocal sound pressure level and syllable context. *Journal of Voice* **6:** 235–245.

Address for correspondence:
F. Debruyne,
Department of Otorhinolaryngology,
U.Z. St. Rafaël,
Kapucijnenvoer 33,
3000 Leuven,
Belgium.