THE ALARYNGEAL VOICE SOURCE AS ANALYSED BY VIDEOFLUOROSCOPY, FIBERENDOSCOPY, AND PERCEPTUAL - ACOUSTIC ASSESSMENT

Britta Hammarberg [1,2] and Lennart Nord [2]
[1] Dept of Logopedics and Phoniatrics, Karolinska Inst, Huddinge Univ Hospital, Huddinge, Sweden, and [2] Dept of Speech Communication and Music Acoustics, KTH, Stockholm, Sweden.

ABSTRACT

Vibratory characteristics of the pharyngo-esophageal (PE) segment, which constitutes the laryngectomee voice source, were related to perceived and acoustic voice qualities, with special emphasis on the voiced-voiceless distinction in stop sounds. Four proficient laryngectomee speakers were included, representing both tracheo-esophageal fistula characterised by either of two strategies: (1) an opening gesture of the back wall of the PE-segment in the fiberoptic speech and esophageal speech. Videofluoroscopic and fiberstroboscopic images of the PE-segment were recorded during phonatory tasks. Results indicated that 84% of the voiced - voiceless word pairs were audibly distinguished. The voiced-voiceless contrast was registration prior to the production of a voiceless stop, and a more forceful upward movement of mucus and barium contrast in the voiceless stops, that was not observed as clearly in the voiced cognate sounds, and (2) a slight prephonatory delay in the closing of the back wall towards the front wall in the initial phase of voiceless stops was observed in the videofluoroscopic registrations. The voiced - voiceless contrasts were automatically confirmed by spectrograms.

INTRODUCTION

In the most commonly used laryngectomee speech techniques the voice source is situated in the upper part of the esophageus, the so-called pharyngo-esophageal (PE) segment. This segment is a multi-layered structure of mucosa and muscle, quite similar to the vocal fold structure. The PE-segment is brought into vibration either on air that has been injected into the esophagus from the mouth, so called esophageal (E) speech, or on pulmonary air that is led into the esophageus via a valve placed in a tracheo-esophageal fistula, so called tracheo-esophageal (TE) speech. Although for both methods the PE-segment is used as voice source, the air volume during phonation is much larger in TE-speech, since the lungs are used. Thus, TE-speakers produce louder and more fluent speech [1,2,3]. As regards consonant intelligibility, however, earlier results indicated that the two speaker groups did not differ significantly from each other, with a mean consonant intelligibility score of 83% for the E-speakers and 86% for the TE-speakers [4].

PURPOSE

The purpose of this study was to relate the vibratory characteristics of the PE-segment to perceived and acoustic voice qualities, with special emphasis on the voiced-voiceless distinction in stop sounds. The results aim at enhancing our knowledge of the physiological and structural characteristics of the alaryngeal voice source and its functional constraints.

The present study is part of an ongoing project that also comprises aspects such as communicative efficiency in background noise, ratings by experienced and naive listeners, and aerodynamic/acoustic measurements.

Table 1. The speakers

	Sex	*Speaking technique*	*Age*	*Years since op.*
No 1	man	tracheo-esophageal	67	2
No 2	man	tracheo-esophageal/esophageal	54	3
No 3	man	esophageal	52	12
No 4	woman	tracheo-esophageal	55	2

METHODS

Speakers

Four speakers were recorded, representing both TE-speech and E-speech (see Table 1 for details). One of the male TE-speakers (No. 2) also mastered E-speech. All subjects were regarded as proficient speakers and had maintained their occupations as military officer, foreman, technician and nurse.

Analyses

Videofluoroscopic images of the PE-segment were recorded during phonatory tasks (and swallowing) in frontal and lateral projections with a rate of 25 pictures per second. Prior to the registration the subject swallowed Barium contrast (Mixobar High Density) to cover the walls in the pharynx and the esophagus. *Audio recordings* of the phonatory tasks were simultaneously made on the same video recorder as was used for the videofluoroscopic registrations and on a separate high quality DAT recorder.

Two of the speakers (Nos. 2 and 3) were also recorded by *videofiberstroboscopy* performing the same phonatory tasks. A fiberoptic laryngoscope (3.5 mm Olympus ENF-P) was connected to a stroboscope (Bruel & Kjaer 4914) and to video equipment. The fiberscope was inserted through one of the nostrils and placed with the tip in the pharynx just above the esophageal entrance.

Acoustic analysis included mean and range of fundamental frequency, spectral characteristics, such as the level of fundamental relative to the level of formants, sound pressure level, and segmental observations. Computer-based analysis programs, developed at the KTH department were used [5,6,7].

Perceptual evaluation by the two authors included listening to voice quality and pitch, and rating consonant intelligibility from the recordings.

Speech Material

Speakers were asked to read aloud a standard text and word pairs containing either a voiced or an unvoiced stop consonant initially or finally: /ka:l, ga:l/, /bank, pank/, /dom, tom/, /bʉs:, pʉs:/, /dil:, til:/, /lab:, lap:/, /bu:d, bu:t/, /lok:, log:/, /vi:t, vi:d/, /jø:k, jø:g/.

RESULTS

Voice Source Characteristics

Videofluoroscopic observations showed that in the three male speakers the voice source was situated in the PE-segment, i.e. the bulging semicircular segment in the back wall of the lowest part of the pharynx and the upper part of the esophagus. In the female TE-speaker, however, the voice source seemed to be made up of two structures, a vibrating constriction situated about 2 centimetres below the PE-segment down in the esophagus (a "subsegment") and the PE-segment. The back wall of the PE-segment never seemed to close towards the front wall during phonation, while there was a full closure in the "subsegment." This finding led us to conclude that the lower constriction was the primary voice source.

Fiberendoscopic observations of two of the speakers (Nos. 2 and 3) showed vibrations in the esophageal orifice with vibratory movements from the back wall towards the frontal wall of the PE-segment. The amplitudes of the vibrations were larger in the back wall than in the frontal wall. A clear mucosal wave was also observed, reminiscent of the glottal mucosal wave. Short sequences of regular vibrations of the PE-segment in one of the patients allowed stroboscopic images to be registered. These indicated a pattern of successive closures and openings of the segment. This subject (No. 2), who mastered both TE-speech and E-speech, was also able to open the PE-segment on command, a manœuvre that he used for prephonatory air-intake in his E-speech.

Perceptual evaluation of the voices showed that the two male TE-speakers (Nos. 1 and 2) had rather rough, strong and low-pitched voices, i.e. "classical" laryngectomee voices, and that the E-speaker (No. 3) had a rather high-pitched and somewhat weak voice, whereas the female TE-speaker (No.4) had a hyperfunctional, strained and weak voice.

Data from the *acoustic* analysis confirm these observations, see Table 2.

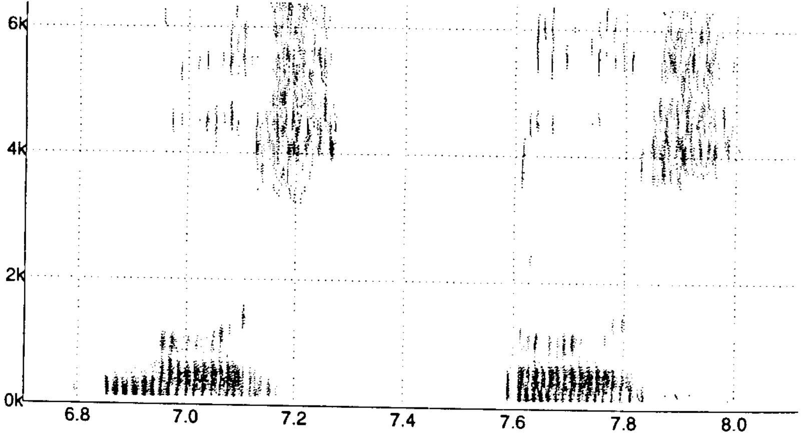

Figure 1. Spectrograms of the voiced-voiceless distinction in the word pair [bœ: pœ:].

Voiced - Voiceless Distinction

As regards the voiced - voiceless contrast, both *videofluoroscopic* and *fiberoptic* registrations seemed to confirm that there was some kind of gesture difference in the PE-segment when the contrast was heard. There seemed to be two strategies for maintaining the distinction. (1) There was an indication of an opening gesture of the back wall in the fiberoptic registration prior to the production of a voiceless stop which was not seen as clear in the voiced cognate sound. The opening gesture consisted of a lifting of the back wall of the PE-segment. Also, there was a more forceful upward movement of mucus and Barium contrast in the voiceless stops. (2) A second strategy was observed in the videofluoroscopic registration: a slight prephonatory delay in the closing of the back wall towards the front wall in the initial phase of voiceless stops.

As for the *perceptual evaluation* there were audible contrasts between voiced - voiceless stops in about 84% (75-92%) of the word pairs during the videofluoroscopic and fiberscopic recordings. *Acoustic analysis* of voiced - voiceless contrasts in the present study showed that when the distinctions were mastered, they were realized by the same acoustic cues as in laryngeal speech, i.e. voice bar for the voiced stops and occlusion for the voiceless stops, see Fig. 1. Voice onset times were close to what Hirose et al. [8] have found for Japanese alaryngeal speech.

DISCUSSION

In our earlier studies of consonant intelligibility tests, the distinction voiced - voiceless in stops has proved to be difficult for laryngectomees to master [4]. There is evidence, however that the distinction can be mastered by some laryngectomees. In the present study of four proficient tracheo-esophageal/esophageal speakers, about 84% of voiced - voiceless word pairs were audibly distinguished. The contrasts were acoustically confirmed by spectrograms.

What constitutes the voiced - voiceless distinction in alaryngeal speech? From the preliminary findings of the fiberoptic and videofluoroscopic registrations, we observed an opening gesture in the voiceless stops. In some cases this gesture was followed by a more forceful occlusive phase in the voiceless sounds, realised by a larger amount of upgoing mucus and Barium during the registrations. The latter observation might be interpreted as a *fortis - lenis* contrast, which is one of the phonetic cues of this distinction in Swedish. Also in the spectrographic analyses a minimal *aspiration* phase was seen in unvoiced stops, which corresponds to the prolonged opening gesture observed in some voiceless sounds. This is in agreement with Hirose et al. [8], who observed a transient opening of the PE-segment for the production of voiceless consonants.

The speaking rates were found to be within the normal range, which is in accordance with the general view that these speakers were proficient.

ACKNOWLEDGEMENTS

We are grateful to Nick Edsborg, Dept of Radiology, Huddinge Univ. Hospital, for making the videofluoroscopic registrations, and to Per-Åke Lindestad, Dept of Logopedics and Phoniatrics, Huddinge Univ. Hospital, for the fiberendoscopy.

This investigation was supported in part by grants from the Swedish Council for Social Sciences and from the Karolinska institute.

REFERENCES

[1] Robbins J, Fischer H, Blom E, Singer, M (1984): A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. *J of Speech and Hearing Disorders*, vol. 49, pp. 202-210.

[2] Perry, A. (1988): Surgical voice restoration following laryngectomy: the tracheo-oesophageal fistula technique (Singer-Blom). *British J of Disorders of Co‹‹munication*, vol. 23, pp. 23-30.

[3] Hammarberg, B. & Nord, L. (1988): Communicative aspects of laryngectomee speech. Presentation of a project and some preliminary results, *Phoniatric & Logopedic Progress Report*, Huddinge Univ Hospital, vol. 6, pp. 10-27.

[4] Hammarberg, B., Lundström, E. & Nord, L. (1990): Consonant intelligibility in esophageal and tracehoesophageal speech. A progress report. *Phoniatric & Logopedic Progress Report*, Huddinge Univ Hospital, vol. 7, pp. 49-57.

[5] Liljencrants, J. (1988): Spectrogram Program "SPEG". Custom Computer Program. Dept Speech Comm & Music Acoustics, KTH, Stockholm, Sweden.

[6] Carlson, R. (1988): Waveform Editor "MIX". Custom Computer Program. Dept Speech Comm & Music Acoustics, KTH, Stockholm, Sweden.

[7] Ternström, S. (1992): Soundswell - Signal Workstation Software Manual, Ver.3.0., Sounds-well Music Acoustics HB, Sollentuna.

[8] Hirose H, Sawashima, M, Yoshioka, H (1983):Voicing distinction in esophageal speech - perceptual, fiberoptic and acoustic studies. *Ann Bull Research Inst Log Phon*, vol.17, pp.187-199.

Table 2. Acoustic measures for the four speakers.

	Speaking rate, including pauses (syll/sec)	Speaking rate, excluding pauses (syll/sec)	Pauses, in % of total reading time	Mean F0 (Hz)	SPL, 1m (dB)	L0, 1 m (dB)
No. 1/TE	3.3	5.3	38	75	66	42
No. 2/TE	4.1	6.1	33	78	55	41
No. 3/E	3.0	4.7	37	137	54	44
No. 4/TE	3.0	3.9	24	150	43	28