PHONETIC ASPECTS OF SPEECH PRODUCED WITHOUT A LARYNX

Lennart Nord, Britta Hammarberg* and Elisabet Lundström

Dept of Speech Communication & Music Acoustics, Royal Institute of Technology, KTH, Box 70014, S-10044 Stockholm, Sweden.
Phone 46 8 7907874, Fax 46 8 7907854
*also at Dept of Logopedics and Phoniatrics, Karolinska Institute, Huddinge Hospital, S-141 86 Huddinge, Sweden

ABSTRACT

The aim of the present report is to compare the different types of alaryngeal voices, esophageal and tracheo-esophageal voices, acoustically and perceptually. A general objective is trying to establish the acoustic cues for naturalness in laryngectomee speech and what constitutes the typical alaryngeal voice quality. Other tasks include intelligibility and acceptability ratings with professional as well as with naive judges. Analysis of selected aspects are reported, such as voice quality features and prosodic features. Differences and similarities between the voice productions are discussed.

1. INTRODUCTION

After a laryngectomy, the patient has to learn to master speech with a new voice source. The sound generator is the upper part of the esophageal entrance, which is set into vibration, either by air that is insufflated into the esophagus from the mouth, or taken from the lungs via a tracheo-esophageal fistula. Acoustic and perceptual aspects of the two kinds of speaking techniques, hereafter called "E-speech" and "TE-speech", were compared. Comparisons were also made using characteristics used for descriptions of normal laryngeal speech ("N-speech") [3]. Previous reports have dealt with acoustic and perceptual aspects, see [4-7,10,11].

2. SPEECH MATERIAL AND SPEAKERS

The speech material contained vowels in carrier phrases, sentences with different prosodic patterns, a short informal conversation and a standard Swedish text of 90 words. So far, 6 TE-speakers, 8 E-speakers and 4 normal laryngeal speakers of the same age group (48 - 80 years) have been analysed. Two of the TE-speakers used Panje voice devices and three low-pressure Blom-Singer devices.

3. PRESSURE AND FLOW MEASUREMENTS

To investigate pressure and flow conditions and also to get an estimate of the voice source shape and spectral content, a flow mask [13] was used in separate readings of /papapa:/, embedded in a carrier phrase. Subjects were asked to produce these words at three loudness levels, subjectively estimated as weak, normal and strong. Inverse filtering and pressure measurements were performed on three E-speakers, three TE-speakers and two normal speakers. Mean values of all /p/ measurements for the three loudness levels and for the three speaker groups were calculated. As can be seen in Figure 1, the normal laryngeal speakers generally produced the words with lower pressure values than what the alaryngeal speakers did, especially when they were asked to produce sounds with low intensity. The alaryngeal speakers could not change their voice levels to the same extent as the laryngeal speakers could, but still managed to vary the loudness level in three steps. Mean values were for the E-speakers 14 cm H_2O, for the TE-speakers 22 cm H_2O and for the normal speakers 7 cm H_2O. This result compares favourably with what is known from investigations of sound pressure levels, e.g. [12], in which TE-speakers were found to speak as loudly as laryngeal speakers. E-speakers usually have weaker voices than the others.

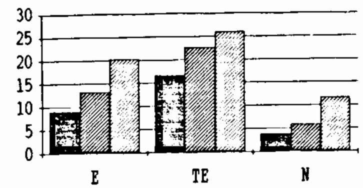

Figure 1. Pressure values in cm H_2O during the production of /p/ for E-, TE- and normal laryngeal speakers, at three loudness levels, weak, normal and strong. (3 subjects in each speaker group; 15 samples/displayed value)

4. INVERSE FILTERING AND VOICE QUALITY

By means of inverse-filtering of the air flow during phonation, the aperiodicity of the wave shapes was analysed and correlated to perceived voice quality.
Flow glottogram curves were obtained for many of the alaryngeal speakers, although they showed a great deal of irregularity. In Figure 2, two examples of automatic inverse filter analysis are shown [2].

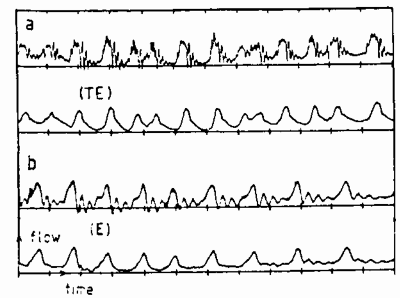

Figure 2. Flow registrations (upper curve) and corresponding inverse filtered flow glottogram (lower curve) of vowel pulses in /papapa:/, uttered by a TE-speaker (a) and an E-speaker (b).

Inspection of the unfiltered speech wave oscillogram revealed unusual excitation traces. In Figure 3, vowel excerpts are shown for one E-speaker and one laryngeal voice. As is clearly evident, there is no well-defined single point of excitation for the alaryngeal voice, compared to what is the case for the normal laryngeal voice.

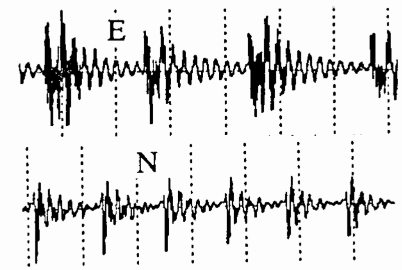

Figure 3. Speech wave oscillograms of vowel samples for an E-speaker and a normal speaker (N).

5. LONG TIME AVERAGE SPECTRA

Long-time-average spectra of these voices have been derived and analysed. A reading of text passage of approximately 45 secs was used as analysis material. The signal was fed into a Hewlett Packard 3562A Dynamic Signal Analyzer and spectral analysis was performed. On the spectral display, it was possible to identify the isolated peak corresponding to the level of the fundamental during the reading. We have not discarded the unvoiced segments from the reading, but still consider the result as representative of the spectral distribution and also the relative measure of level of fundamental in comparison with total spectral energy. In Figure 4, LTAS-spectra for a TE-speaker and a normal laryngeal speaker ("N") are shown.

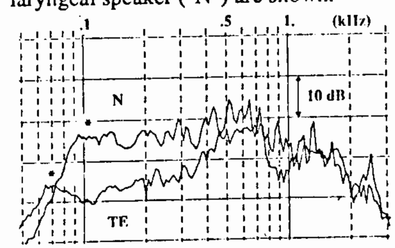

*Figure 4. Long Time Average spectra of a read text passage by a TE-speaker and an N-speaker. The level of the fundamental, L0, is indicated by *.*

The spectral level difference between fundamental and first formant level (L1-L0), seems to be a valid parameter for these alaryngeal voices. So far, preliminary data from seven alaryngeal speakers, suggest that the L1-L0 difference is larger for the alaryngeal voices than for normal voices, i.e. the level of the fundamental is very weak in the alaryngeal voices, see Figure 5. Moreover, it does not vary with loudness to the same extent as in normal laryngeal voices [11].

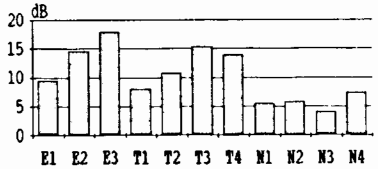

Figure 5. Difference data of total level (L_{tot}) to the level of the fundamental (L0) derived from LTAS-analysis of a text passage, read by three E-speakers, four TE-speakers and four N-speakers.

6. PROSODY

Pitch and duration cues

Prosodic studies of intonation patterns and word emphasis related to overall pitch range and pitch dynamics were made.

In order to evaluate the capability of these speakers to produce acceptable prosodic patterns, a set of sentences with question intonation and emphatic word stress was included in the reading material. In most cases the speakers were able to produce the target sentences. However, they sometimes chose different strategies compared to speakers with laryngeal phonation. Word emphasis was often made by a pausing as well as by a pitch change. In Figure 6 two pitch curves are shown, produced by two alaryngeal speakers, one female E-speaker and one male TE-speaker. As can be seen the pitch patterns are varying in much the same fashion as for normal laryngeal phonation. Note the high pitch produced by the female E-speaker and the very low pitch produced by the male TE-speaker.

Using automatic pitch extraction algorithms, these voices are difficult to analyse, depending on the low voice registers, and irregular vibration patterns. The analysis was made by trying different pitch extraction algorithms, developed by Liljencrants [8]. Visual inspection was also performed on spectrograms and oscillograms. For some of the aperiodic voices it was very difficult to identify any periodic component, although it was still perceivable.

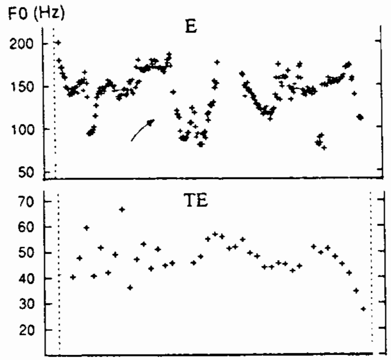

Figure 6. Pitch analysis of a sentence with emphatic word accent (female E-speaker) and a question (male TE-speaker).

The voice quality of the female E-speaker above was quite strained and rough, although she managed well to produce acceptable pitch patterns and had a quite high pitched voice (mode value 148 Hz). The voice breaks easily into a much lower register. Reasons for this are probably to be found in the changes of pressure conditions at the voice source, caused by consonantal constrictions. In Figure 6, such a break occurs at the arrow, /o:/ followed by /l/. The "diplophonic" rough character that this voice exhibits has also been noticed in a male esophageal speaker with a high pitched voice (mode value 121 Hz). It is as if the vibrating PE-segment is very sensitive to the right level of driving pressure; a too weak pressure will not start any oscillatory process and too much force may create deviant vibratory frequencies.

Typically, a very low pitch is common. The problem often is that the aperiodicity creates noise, overlaid on the fundamental. See the second pitch curve, displayed in Figure 6. Although varying in a normal fashion, the pitch does not exceed 60 Hz (mode value 43 Hz).

7. DISCUSSION

As reported in previous studies on pathological voices, there is a correlation between the voice pulse shape and the perceptual impression of voice quality. An irregular and strongly varying voice source pulse often correlates with a harsh voice [1]. One finding in the present study was the unusual excitation patterns of the alaryngeal voices. We still need a better insight into the mechanisms behind these irregular patterns, and a modelling of the structures responsible for these vibrations would be of great value. Work in this area is going on [9].

As a result of the present study so far, two differences between the normal laryngeal and the alaryngeal groups are evident. Firstly, the alaryngeal voices were characterized by a weaker fundamental relative to the total energy level as compared to the normal voices. Secondly, apart from this static aspect of the voice source, a dynamic aspect is observed if speakers are asked to produce sounds with different intensity. Normal, laryngeal voices will have a more pronounced fundamental if they phonate at low intensities. The same does not happen for the alaryngeal speakers.

8. CONCLUSIONS

It was found that the alaryngeal voices, E-voices and TE-voices were characterized by a weak level of the fundamental compared to normal laryngeal voices. Other, more detailed voice source characteristics, such as inverse filtered flow registrations displayed strong irregularities for the alaryngeal voices.

Acknowledgement: This work was financed by research grants from the Swedish Cancer Society and the Swedish Council for Social Sciences.

REFERENCES

[1] Askenfelt A., Hammarberg B. (1986): "Speech waveform perturbation: a comparison of seven measures", *J of Speech & Hearing Research 29*, pp. 50-64.

[2] Gauffin J., Hammarberg B., Imaizumi, S. (1986): "A microcomputer based system for acoustic analysis of voice characteristics", *Proc ICASSP Tokyo Vol 1*, pp. 681-684.

[3] Hammarberg, B. (1986): "Perceptual and acoustic analysis of dysphonia". *Studies in Logopedics and Phoniatrics No 1, Huddinge Univ Hospital*. Doct Dissertation.

[4] Hammarberg, B., Lundström E., Nord L. (1989): "Aspects of laryngectomee speech communication - a progress report." *STL-QPSR 1*, pp. 175-178.

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

[6] Hammarberg, B., Lundström, E., Nord, L. (1990): "Intelligibility and acceptability of laryngectomee speech", *PHONUM, Reports from the Department of Phonetics, Univ of Umeå, 1*, pp. 88-91.

[7] Hammarberg, B., Nord, L. (1989): "Tracheo-Esophageal Speech, Esophageal Speech and Artificial Larynx Speech - Acoustic and Perceptual Aspects", *Proc XXI Congr Int Assoc Logopedics & Phoniatrics, Prag August 6-10*, pp. 426-428.

[8] Liljencrants, J. (1991): "Manual for speech analysis software on the Apollo computer" (unpubl manuscript). Dept Speech Comm & Music Acoustics.

[9] Liljencrants, J. (1991): personal communication

[10] Nord L., Hammarberg B. (1988): "Communicative aspects of laryngectomee speech. Some preliminary phonetic-acoustic results from an interdisciplinary project", *STL-QPSR 4*, pp. 31-49.

[11] Nord L., Hammarberg B. (1989): Analysis of Laryngectomee Speech - A Progress Report. In *'Eurospeech 89', European Conf on Speech Comm & Technology (ed. by J.P. Tubach & J.J. Mariani), Vol II*, pp. 493-496, CPC Consultants Ltd, Edinburgh, UK.

[12] 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 49*, pp. 202-210.

[13] Rothenberg, M. (1973): "A new inverse-filtering technique for deriving the glottal air flow during voicing", *J of Acoust Soc of America 53*, pp. 1632-1645.