Aspects of laryngectomee speech communication - a progress report

Hammarberg, B. and Lundström, E. and Nord, L.

journal: STL-QPSR
volume: 30
number: 1
year: 1989
pages: 175-178

http://www.speech.kth.se/qpsr
ASPECTS OF LARYNGECTOMEE SPEECH COMMUNICATION
- A PROGRESS REPORT

Britta Hammarberg*, Elisabet Lundström**, and Lennart Nord**
*Karolinska Institute, Dept. of Logopedics and Phoniatrics, Huddinge Hospital, **Dept. of Speech Communication and Music Acoustics, KTH

Abstract

The present progress report focuses on listener evaluation in an ongoing investigation of laryngectomee speech. Also, results from our previous acoustic studies are briefly discussed. Strategies for intelligibility and acceptability ratings are presented.

INTRODUCTION

In a joint project between the Department of Logopedics and Phoniatrics, Huddinge Hospital and the Department of Speech Communication and Music Acoustics, Royal Institute of Technology, we are investigating different speech techniques for the laryngectomee. We are in particular interested in an evaluation of the tracheo-esophageal (TE) fistula speech, and we compare it with esophageal speech, with electrolaryngeal speech (Servox device) and with larynged speech.

The aim of the project is to obtain an objective evaluation of the different speaking methods in terms of communicative efficiency, including measures of acceptability by users and listeners. The acoustic analysis includes a number of measurements. We have to some extent relied on earlier experience of voice description from laryngeal - normal as well as deviant - voices (Hammarberg, 1986).

Prosodic features, such as pitch, intensity, duration and speaking rate are measured. Furthermore, voice quality features, including long time average spectral shape and aspects of spectral energy distributions, are investigated. Other, more detailed voice source characteristics, derived by inverse filtering, are compared with flow registrations from normal, laryngeal voices.

PROGRESS REPORT

Speakers

Until now, eight subjects have been included in the analysis program; three TE speakers, three esophageal speakers, and two normal laryngeal speakers. The six alaryngeal speakers were in the age of 60-73 years (median age 64) and were laryngectomized 2-15 years ago. They were also recorded, using a Servox device. The two normal speakers were 65 and 73 years of age, respectively. Although we plan to record and analyse a number of speakers with varying speaking skills, so far the recorded alaryngeal speakers master the techniques very well. Their speech patterns are therefore considered to be quite consistent.

Acoustic analysis

A recent report of the acoustic analysis results from these eight speakers was given in Nord & Hammarberg (1988). As a general conclusion, the derived TE speech characteristics, such as SPL and F0, were more close to normal laryngeal speech characteristics than to esophageal speech. These findings are in accordance with Robbins, Fisher, Blom, & Singer (1984) and Pindzola & Cain (1988).
Regarding the voice source parameters, one interesting finding was that the intensity level of the fundamental frequency (LO) was weaker for the alaryngeal voices than for the normal laryngeal voices. As is well known, a characteristic of the normal voice source spectrum, is that LO is varying with voice effort. For a strong voice, the formants dominate the spectrum and for a weak voice, the fundamental dominates and constitutes most of the sound intensity. The TE voices and the esophageal voices all had the fundamental weaker than the first formant, irrespective of sound level. See Fig. 1, showing a spectrum of one of the TE voices. The reason for this is probably due to constraints in the pharyngeal-esophageal voice source. The artificial larynx voices (using the Servox device) also had a substantially weaker fundamental than the laryngeal voices. This is due to the acoustics of the sound generating system: a lot of energy would be needed to generate a strong sounding lowest partial through the neck tissue.

In the evaluation of the speaking rate, we wanted to distinguish between parameters reflecting:

(i) production strategies due to constraints and habits and
(ii) parameters reflecting the perceptual impression.

The material to be analysed consisted of a read passage of 89 words ("Ett svårt fall"). A number of parameters were derived. The total number of words per minute (or syllables per second) excluding pauses would seem to reflect articulation rate, irrespective of any problems concerning air supply or phonation. A second measure, based on the entire reading time reflected the impression of fluency. A variety of speaker strategies could be observed. Some subjects made several pauses, not only esophageal speakers with their limited air supply, but also TE speakers and Servox users. This could reflect earlier achieved habits. The number of pauses, longer than 200 msec were also calcu-
lated. A tendency in the speech material was that the TE speakers made few but long pauses, while the esophageal speakers made several short pauses.

EVALUATION OF ALARYNGEAL SPEECH
The aim of the present part of the project is to evaluate the different types of alaryngeal speech from the point of view of the listener, i.e. the acoustic measures will be compared with listener ratings of intelligibility and acceptability. This type of objective data could be of great value for the laryngectomee who must ultimately choose a rehabilitation voicing routine.

Ratings of intelligibility and acceptability by different listener groups
The measure of intelligibility has in many studies been defined as the percentage of test items (sounds, words, or sentences) correctly transcribed or identified by listener groups. One commonly used procedure is to use speech presented in noise.

In a recent study by Cullinan, Brown, & Blalock (1986), experienced (speech pathologists) and inexperienced (hospital volunteers) listeners rated the intelligibility of audio taped alaryngeal speech in 8 esophageal and 5 TE speakers. The speech items were oral reading of a passage and "uninterrupted talk" for two minutes. The TE speakers were rated as more intelligible than were the esophageal speakers. Reliability of both listener groups was good.

This finding was in accordance with Tartly-Mitzell, Andrews, & Bowman (1985), who also found TE speakers to be highly intelligible to naive listeners.

Hubbard & Kushner (1980) compared intelligibility scores obtained for sentences read by good-to-superior esophageal speakers with scores obtained for sentences read by normal speakers via three modes of presentation: visual, auditory and auditory-visual combined. Results indicated that intelligibility scores for the esophageal speakers were significantly higher in the combined than in the auditory condition.

Speech acceptability of good and excellent esophageal speakers and TE speakers was compared by Trudeau (1987). In his study, naive listeners made no difference in their acceptability ratings of esophageal and TE speech. Moreover, data indicated that speaker proficiency and not voice type had a significant effect on the listener judgments.

In a study by Williams & Watson (1985), speaking proficiency of esophageal speakers, TE speakers and electrolarynx speakers was rated by three groups of judges, with varying knowledge about laryngectomes. Among the results, they reported that expertise level of the judges influenced the ratings of alaryngeal speaking mode. The expert listeners preferred TE speech, whereas naive listeners did not rate the TE speakers significantly better than the esophageal speakers on intelligibility and overall communication effectiveness.

Present investigation
The following speech samples will be included in the present investigation:

1) reciting word lists
2) sentences with key words
3) oral reading of a paragraph
4) speaker's description of his/her vacation and hobbies
5) conversational speech samples from an interview
Regarding the findings by Hubbard & Kushner (1980), we have decided to vary the modes of presentation to the listeners so that items 1,2 will be presented auditorily, and items 4,5 auditory-visually combined, i.e. via video tape. Item 3 will be presented auditorily for one listener group, and video-taped for another listener group.

Intelligibility will be rated both under good listening conditions, i.e. in a sound treated room, and in "cocktail party" noise. It is well-known to the laryngectomee that one of the most difficult environment to be heard in is among laryngeal speakers in groups. Items 1, 2, 3 will be rated as regards intelligibility, items 3, 4, 5 for acceptability.

At the symposium, audio and video illustrations of TE speech, esophageal speech and artificial larynx speech will be presented.

Acknowledgement
This work is financed by a research grant from the Swedish Council for Planning and Coordination of Research (Project no A 15-5/49)

References