Laryngectomee speech in noise - voice effort, speech rate, and intelligibility

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

journal: STL-QPSR
volume: 34
number: 2-3
year: 1993
pages: 057-065

http://www.speech.kth.se/qpsr
LARYNGECTOMIZED SPEECH IN NOISE - VOICE EFFORT, SPEECH RATE, AND INTELLIGIBILITY

Lennart Nord, Britta Hammarberg**, Elisabet Landström

ABSTRACT

Different aspects of alaryngeal speech, both esophageal and tracheo-esophageal speech, are being analysed in a joint project between the Department of Speech Communication and Music Acoustics, KTH, and the Department of Logopedics and Phoniatics, Karolinska Institutet. The purpose of the present part of the project was to evaluate the speech performance of four laryngectomy speakers and one normally speaking subject while they were reading texts aloud with varying amounts of noise in their ears. The noise consisted of a number of voices in a cacophony.

Acoustic speech parameters, such as sound pressure and spectral characteristics, were measured and compared among the subjects. The results showed that the tracheo-esophageal speakers were able to raise their voice level almost as much as normally speaking subjects. The esophageal speakers on the other hand were usually not able to produce as strong voice levels during the text readings. This type of test method with speech in background noise seems promising for assessment of voice effort.

In a second part of the present investigation, intelligibility tests were performed. Normally hearing listeners were asked to adjust the level of noise when exposed to the tape-recorded readings of the laryngectomised speakers, the task being to use a high level of noise while still being able to understand the read passage.

The result of the intelligibility test revealed that the listeners tolerated a higher degree of noise when listening to the normal speaker than to the alaryngeal speakers. Also, the listeners accepted a little more noise when listening to the tracheo-esophageal speakers than to the esophageal speakers.

INTRODUCTION

In an ongoing project, the speech of laryngectomised persons is analysed. Two types of alaryngeal speaking techniques are used by the subjects: (1) esophageal speech, E-speech or (2) tracheo-esophageal fistula speech, TE-speech. In both techniques the voice source is the vibratory upper part of the esophagus, the so called pharyngo-esophageal segment. The two techniques differ as regards air support for phonation. In E-speech, air is injected into the esophagus from the mouth prior to phonation, and in TE-speech pulmonary air is the driving force, and is lead through a fistula in the wall between the trachea and the esophagus.

Within the project different aspects of alaryngeal speech have been in focus, e.g. intelligibility, acceptability and acoustic measurements, see e.g. refs Hammarberg & Nord, 1988; Nord, Hammarberg & Lundström, 1992. As regards intelligibility, results indicated that there was no difference between the two speaker groups (5 male E-speakers and 5 male TE-speakers). A mean consonant intelligibility in CVC words of

* expanded version of a paper presented at EUROSPEECH '93, Berlin, September, 1993
** also Dept of Logopedics and Phoniatics, Karolinska Institute, Huddinge University Hospital
about 85% was found for both groups (Hammarberg, Lundström & Nord, 1990; Hammarberg, Lundström & Nord, 1992). However, other parameters revealed differences between the two groups, i.e. speech rate which was higher for the TE-speakers.

In previous recording sessions, which were made in an anechoic chamber, we asked the subjects to read with different voice efforts in order to assess the effect of loudness on their voices (Hammarberg et al., 1992). However, our results showed that the speakers seldom raised their loudness levels, resulting in more or less equal levels for the two speaker groups. This might be an effect of reading in an anechoic chamber, which is artificially silent.

In the present study, we made a new attempt to make the subjects use stronger voices. The speakers were asked to read aloud while exposed to babble noise at different levels in their ears, the hypothesis being that the amount of noise would influence articulation skill and voice behaviour of the speakers. It was hypothesised that the TE-speakers would be able to raise their loudness level more than would the E-speakers, because of larger air support in TE-speech.

A second aim of the study was to assess the intelligibility of the laryngectomee speech in noise by asking listeners to adjust the laryngectomee speech level against different levels of noise backgrounds, so that the speech could just be followed or understood, a so called "just-follow-conversation" (JFC) task (Hygge, Rönning, Larsby & Arlinger, 1992; Öhngren & Dahlquist, 1988).

MATERIAL AND METHODS

Speakers

The speakers in the present study were four laryngectomised men, 2 esophageal (E) speakers (49 and 67 years of age) and 2 tracheo-esophageal (TE) speakers (53 and 67 years old). All were native Swedish speakers living in the Stockholm area, and recognised as fluent speakers. The TE speakers were using the Blom-Singer low pressure prosthesis and digital occlusion during speech. As a control, a normally speaking man (N1, 46 years of age) was included.

The hearing ability of the speakers was tested by tone audiograms. Three of the four laryngectomee speakers had age-related hearing impairment at higher frequencies, which can be considered to be normal. The fourth laryngectomee speaker (E2) had no impairment.

Recordings in noise backgrounds

The subjects were recorded while they read texts aloud with varying amount of noise in their ears. The subjects read excerpts from short stories, under three noise conditions, 0 dB, 70 dB, and 80 dB. Different excerpts were chosen in order to prevent learning effects on the part of the listeners. The noise consisted of ten voices in a cacophony played backwards. Spectrum of the babble noise is shown in Fig. 1. The subjects were instructed to raise their voices in order to be heard. Absolute sound pressure levels at 1 meter distance were calculated.
**Fig. 1:** Mean power spectrum of the babble noise. *(Ordinate: Sound pressure level, 10 dB/div)*

**Instrumental analysis**

In order to measure sound pressure level and spectral characteristics, calibrated long-term average spectra were produced from the tape-recorded readings by a Hewlett Packard 3562A Dynamic Signal Analyzer. We compared the level of the fundamental during the reading with the sound pressure level (SPL), calculated as the mean power spectral level. The fundamental was identified as the peak in the low frequency area of the power spectrum displayed on the screen, Fig 2. However, for the alaryngeal voices identification of the fundamental turned out to be difficult as these speakers have a very weak and aperiodic fundamental. Furthermore, the recording booth introduced some noise in the low frequency region, at about 70 Hz, obscuring part of the spectrum, where a few of the speakers had their fundamental frequency according to earlier measurements. Also, speech tempo and pausing was measured.

**Fig. 2:** Mean power spectrum for one of the laryngectomee speakers. The range of the fundamental frequency (50-90 Hz) is marked by circles. *(Ordinate: Sound pressure level, 10 dB/div)*
Listeners
To evaluate the intelligibility in noise background, five normal-hearing listeners were asked to adjust the level of noise while exposed to tape-recorded readings of the laryngectomised speakers. The listener task was to use as high a level of noise as possible, while still being able to comprehend what was said ("just follow conversation") (Hygge et al., 1992). No separate assessment was made of their actual understanding of the text.

The listener group (2 females, 3 males) was experienced in participating in listening tests, but had no experience of laryngectomee speech. Ages ranged from 25 to 51 years.

![Figure 3. Sound pressure level for the esophageal (E1, E2) and the tracheo-esophageal (TE1, TE2) speakers and one normally speaking man (N1), exposed to various amounts of noise in their ears.](image)

RESULTS

SPL and spectral characteristics
The SPL results are displayed in Figure 3. All dB-values refer to a speaking distance of 1 meter. One of the two tracheo-esophageal speakers (TE1) was able to raise his voice level from 66 to 79 dB with increasing noise in his ears. This is within the range of normal laryngeal speakers (Granström & Nord, 1992). The other TE-speaker (TE2) had a quite strong voice already at normal voice effort and did not change its level (70-72 dB). The two esophageal speakers, on the other hand, were not able to produce as strong voice levels during the text readings (58-63 and 65-68 dB, respectively).

As regards the relation between level of fundamental and SPL, the difference values point in the direction that when a speaker raises his voice level, more energy is attributed to the formant region and less to the fundamental, see Figure 4. This
tendency is expected and found for normal voices as well. One typical characteristic of the alaryngeal voices compared to normal voices is the low level of the fundamental.

![Bar chart showing differences between the sound pressure level (SPL) and the level of the fundamental for the speakers, exposed to various amounts of noise in their ears.](image)

**Figure 4.** Difference between the sound pressure level and the level of the fundamental for the speakers, exposed to various amounts of noise in their ears.

### Speech rate and pausing

As regards speech tempo and pausing, we calculated the number of pauses versus loudness. As the texts were not identical for the speakers, the various text-readings were normalised over a hundred-word paragraph. Our hypothesis was, that when the E-speakers were forced to use a great deal of effort, they spent air rapidly and had to use shorter stretches of speech with many pauses for inhalation.

The same effect could be true also for the TE-speakers, although not to the same extent. In Table I, the number of pauses during the text reading is tabulated together with the SPL for different noise conditions. As can be seen, the speakers increased the number of pauses with increasing SPL, although individual speaking styles appeared. Thus, speaker TE2 paused quite frequently and also used a strong voice level throughout all his readings, regardless of noise level. The normal speaker used fewer pauses and did not increase the number of pauses with increasing voice effort. In a similar way speaker TE1 also used quite few pauses.
Table I. Number of pauses during the speakers’ text-readings (normalised over a hundred-word paragraph) and the sound pressure level (dB, 1 m), for different noise level exposures.

<table>
<thead>
<tr>
<th>Noise-exposure</th>
<th>E1</th>
<th>E2</th>
<th>TE1</th>
<th>TE2</th>
<th>N1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pauses</td>
<td>dB</td>
<td>pauses</td>
<td>dB</td>
<td>pauses</td>
</tr>
<tr>
<td>0 dB</td>
<td>18</td>
<td>58</td>
<td>18</td>
<td>65</td>
<td>13</td>
</tr>
<tr>
<td>70 dB</td>
<td>29</td>
<td>63</td>
<td>24</td>
<td>66</td>
<td>15</td>
</tr>
<tr>
<td>80 dB</td>
<td>26</td>
<td>63</td>
<td>22</td>
<td>68</td>
<td>16</td>
</tr>
</tbody>
</table>

Intelligibility tests

The test was quite easy to carry out and the listener performances were stable. Figure 5 shows mean values of the relative amount of noise compared to the speaking levels. As can be seen listening to the esophageal speakers allowed the least amount of competing noise, while the tracheo-esophageal speakers allowed more noise. The normal speaker on the other hand was still more intelligible in the sense that listeners allowed even more babble noise on his readings.

![Figure 5. Relative noise level for “just follow conversation” experiment. Five normal-hearing listeners were exposed to the esophageal (E1, E2) and the tracheo-esophageal (TE1, TE2) speakers respectively, and to one normal laryngeal speaker (N1), with various amounts of noise in their ears.](image-url)

DISCUSSION

The results in the first part of the study indicate that the TE-speakers were able to raise their voice levels to almost normal loudness levels, whereas the E-speakers were not
able to produce as strong loudness levels in the different noise backgrounds. These results were expected as TE-speech is produced with a larger air support (from the lungs) as compared to E-speech. Air volume is about 1 litre (out of about 4 litres total lung volume) for TE-phonation and 0.1 litre for E-phonation (Snidecor & Isshiki, 1965).

One of the TE-speakers (TE2) had quite a strong voice already at normal voice effort. He kept his loudness level regardless of the degree of noise background. His voice was about 10 dB stronger than average E-speech level.

To get an estimate of the spectral shape of the two types of alaryngeal speech, spectral characteristics from long-term average spectra were analysed. Regarding level of first formant region versus level of fundamental, findings indicate that there is the same relationship between the two levels in alaryngeal speech as in normal speech (Fig. 4), i.e. (1) an increase in SPL is due to an increase in level of first formant, and (2) the first formant level is always dominating over fundamental level. However, in alaryngeal speech level of fundamental seems always to be rather low, indicating resemblance with pressed phonation in normal voice. This might be explained by anatomical characteristics of the pharyngo-esophageal segment, which morphologically is a half sphere thickening of the walls of the pharyngeal-esophageal tract of about 20-30 mm vertical length during phonation.

A limited number of articles reporting on long-term average spectral analysis of the alaryngeal voices have been published. In Weinberg, Horii & Smith (1980) a comparison was made between esophageal speech and normal laryngeal speech and differences in intensity were found to be about 10 dB.

The type of test method with speech in background noise seems promising for assessment of voice effort. The difference between the two speaking groups seems possible to ascertain: the TE-speakers used a stronger voice and were capable of raising their intensity, while the E-speakers used a lower voice level, with less possibility to raise intensity.

The result of the listening test ("just follow conversation") also revealed a rank order regarding listeners’ ease of following conversation in babble noise: 1) normal speaker, 2) tracheo-esophageal speaker, and 3) esophageal speaker.

ACKNOWLEDGEMENTS
The assistance of Martin Dahlquist concerning the experiments with noise exposure, and Eva Agelfors for audiogram registrations, is gratefully acknowledged. The study was supported by grants from the Swedish Council for Social Research (project No. C90-0172:2) and the Swedish Cancer Society (project No 3000-B91-02XAA).

REFERENCES


