SPEECH INTELLIGIBILITY OF TWO GROUPS OF CORDECTOMIZED PATIENTS AFTER LARYNGOFISSION AND LASER SURGERY

ABSTRACT

Speech intelligibility is inversely related to the noise generated in the vocal folds, in the resonance cavities, and in the environment. In this study the intelligibility of two cordectomized groups of patients, treated with two different surgical techniques, was analysed. One group underwent laryngofissure with conventional surgery; the other underwent surgery by laser. Each group recorded a list of 48 random items, words, and pseudowords. These lists were presented under five different conditions of S/N in an anechoic room with TDH 48 headphones. The stimuli were presented under two conditions of no noise and under three conditions the items were masked with white noise, in which S/N was +3dB, 0dB, -3dB and -6dB. The results show that without masking, intelligibility was very similar for both groups of patients. However, when the speech was masked, intelligibility was better for laryngofissure patients. Moreover, the difference in the recognition scores between the two techniques increased with the degree of masking.

INTRODUCTION.

The source-filter theory of speech production [1] explains speech production as the convolution of two factors: a source of vibration, which is accomplished by the vocal folds; and a resonator, located in the oral tract. So this relation can be expressed thus:

\[ S = F_v \cdot T_r \]  

where \( S \) is speech; \( F_v \) is the source of vibration; and \( T_r \) is the oral tract or the transfer function, radiation of energy from the lips and mouth being negligible for characterization of the speech signal. To a great extent, the mechanism of source generation is independent of the filtering process. The characteristics of the source are not influenced by, nor depend on the acoustic properties of filters.

During prephonation, the vocal folds are adducted, closing the glottis and obstructing the airflow. When the vocal folds abduct they start to vibrate due to the joint action of the pressure exerted by the sub-glottal retained air plus the Bernoulli effect, produced when the airflow passes from the trachea to the open glottis, a smaller opening than the trachea.

Vibration of the vocal folds induces the resonances of the oral tract, an anatomical structure that looks like a 17cm long tube (depending on age and gender). In this tube the tongue modulates areas of different section and length, which acquires different resonance properties, amplifying certain bands of frequencies and attenuating others [2].

In patients with larynx diseases, if a tumour is localized in the vocal cords and it is in the first stages of development, it is possible to operate directly on the cords (cordectomy) preserving the other laryngeal structures. In the case of bilateral cordectomy, the patient can use the ventricular folds to speak or the postoperative rearrangement of the laryngeal structures [3]. The
acoustic signal generated in these structures is then modulated by the vocal tract, which has not been modified by surgery.

Two techniques can be used for these patients: laser CO₂ and conventional surgery or laryngofissure, accessing the tumour from the exterior. (From now on, these two techniques will be referred to as LS and LF, respectively).

The effect of these techniques on speech intelligibility, objectively and subjectively, is analysed in this study. Objectivation of intelligibility was performed by means of quantification of acoustic information that was obtained from the analysis of the signal in the third octave, in accordance with norms [4].

METHOD

Subjects
Four patients who underwent laser surgery and four who underwent laryngofissure, aged between 63 and 70 years were included in our study. None of them received radiotherapy. In all cases, speech evaluation was carried out at approximately twelve months after surgery. At the time of the study, none of them presented other health problems. A normal speaker, without phonation difficulties, was recorded as the control.

Material and procedure
Each patient and the control recorded a list of 48 verbal stimuli digitally, 24 words and 24 non-words. Almost all the stimuli were made up of four-letter sequences (in very few cases, five letters) and the syllabic structure was consonant/vowel/ consonant/vowel.

Objective intelligibility was graded according to the above-mentioned norm [1]

Subjective intelligibility was evaluated by means of a recognition test. The list recorded by each patient was presented to listeners under five different conditions: without noise; masked by white noise with signal-to-noise ratio +3 dB; 0 dB; -3 dB; and -6 dB. Each experimental condition was evaluated by 20 different judges, aged between 20 and 35 years old, without hearing impairment. All the stimuli were equalled in Lₐeq and presented in an acoustic conditioned booth, using TDH 48 headphones set at a constant level of 57 dBA; the level of noise varied under each experimental condition. The inter-stimulus interval was 3 seconds. The subjects were instructed to transcribe what they heard literally, without taking any notice of the meaning of the stimulus.

RESULTS

The judges’ answers were evaluated according to the following criteria: zero was given when the target was recognized without error; 1 when only one phoneme or letter was misidentified; 2 for targets with two or more errors and for stimuli omitted due to masking.

This experimental design makes it possible to analyse intelligibility using the following variables: the kind of speaker (LS, LF, Normal), the kind of stimuli (words, pseudowords) and S/N ratio (without noise, +3dB, 0 dB, -3dB, -6dB). The correlation between objective and subjective intelligibility was also calculated.

As was to be expected, control intelligibility was always significantly better than the intelligibility of both groups of patients (F=2390.6; p<.0001)

<table>
<thead>
<tr>
<th>mean errors</th>
<th>LN</th>
<th>LF</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.54</td>
<td>1.18</td>
<td>1.39</td>
<td></td>
</tr>
</tbody>
</table>
Regarding the S/N ratio, subjective intelligibility for normal speech was practically 100% under the no-noise condition with very few errors and none in target words. In this group, the loss of intelligibility depending on masking agreed with the results obtained for normal listeners in previous experiments [5]. In the LS and LF groups (Table II and Fig. 2) in the absence of noise, the same performance was observed with a loss of about 20% of intelligibility. Nevertheless, the presence of noise impaired the intelligibility of all the patients significantly, particularly the LS group. There was about 50% or more loss of intelligibility with a mask of +3dB: under this condition, normal speech preserved about 90% of the signal making it possible to reconstruct the entire message from the context in spoken language. From higher signal-to-noise ratios the loss of intelligibility grew linearly in both patient groups with very few phonemes or sounds accurately identified.

Figure 1.- Cumulated errors in all groups

Figure 2. Mean errors by type of speaker and S/R
Table II. Mean errors by type of speaker and Signal-to-Noise ratio

<table>
<thead>
<tr>
<th>S/N</th>
<th>LN</th>
<th>LF</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without noise</td>
<td>0.09</td>
<td>0.46</td>
<td>0.48</td>
</tr>
<tr>
<td>3dB</td>
<td>0.26</td>
<td>1.00</td>
<td>1.30</td>
</tr>
<tr>
<td>0dB</td>
<td>0.49</td>
<td>1.16</td>
<td>1.44</td>
</tr>
<tr>
<td>-3dB</td>
<td>0.66</td>
<td>1.53</td>
<td>1.76</td>
</tr>
<tr>
<td>-6dB</td>
<td>1.20</td>
<td>1.77</td>
<td>1.90</td>
</tr>
</tbody>
</table>

Figure 3 presents the errors of words and pseudowords for the two groups of patients. In all the cases, as was to be expected, pseudowords were more affected than words. Clearly this was due to a process of recuperation of the lost segments from the knowledge of the language that the listeners have stored in their memory, in the case of words.

![Figure 3. Mean errors in words and pseudowords in LF and LS.](image)

**Correlations**
The correlation between objective and subjective intelligibility of all speakers and under all conditions was calculated. All the correlations were positive with a 1 per 1000 (p<0.001) of statistical significance, except the correlation under the no-noise condition for the two surgical techniques.

This lack of correlation in the absence of masking noise is because when intelligibility is calculated objectively without any external noise from the source, it always reaches very high values, as this factor takes a value of zero in the calculus formula of the intelligibility index. Only information from analysis in third octave is taken into account to calculate this index. Nevertheless, this acoustic information is not evaluated qualitatively and for this reason the index score reached a maximum. But under the same condition, subjective intelligibility gave many errors in the two groups of patients.

When the index was calculated in the presence of noise, masking affected the two groups of patients in a similar way (Fig. 4. pink and dark blue lines). However, it is interesting to observe the performance of the LF and LS subjective intelligibility. In the four masked conditions the lines follow the same pattern and are parallel.
CONCLUSIONS
1. After both surgical techniques intelligibility was very similar in the absence of noise, with a score of about 20% of errors. This allows good comprehension of language by means of cognitive inferences that reconstruct lost segments of words.

2. These recovering inferences are only possible with words. Pseudowords, which reflect acoustic intelligibility more accurately because they lack meaning, are significantly more affected than words. The impairment of pseudowords increases with masking noise.

3. Masked by noise, even with a good S/N ratio, both LS and LF lost intelligibility at a higher rate than normal; LS presented the worst performance.

In conclusion, this study shows that the laryngofissure technique preserves intelligibility better than the laser technique. An accurate explanation for the difference between LS and LF probably requires a microscopic analysis of the acoustic nature of these two types of voice. Nonetheless, and in this first analysis, it is possible to relate this difference to the fact that laryngofissure presents an external and open access to the anatomical structures, which in turn makes a more detailed reconstruction of the damaged tissue possible.

References

ACKNOWLEDGEMENTS
This study was supported by a research project granted by the Spanish Government, reference SEJ2004-06529/PSIC