Abstract: **Objectives**: To determine the usefulness of acoustic spectrography as a diagnostic tool by which to characterize and monitor the esophageal voice. **Material and Methods**: We studied 33 laryngectomized patients (all male), who underwent the following assessments: qualitative acoustic protocol (voice spectrography of the vowel /a/ and a sentence), quantitative acoustic protocol (phonation time, fundamental frequency, maximum intensity of sound pressure, speech rate), and a perceptual evaluation (the psychoacoustic acceptability of the voice). **Results**: There is a significant statistical relationship between the modified Yanagihara spectrogram classification, the psychoacoustic perception of the voice, and quantitative acoustic parameters. **Conclusion**: We consider acoustic spectrography to be a simple, effective, and reliable method for studying the esophageal voice, monitoring the development of oral communication skills and improving the rehabilitation of the laryngectomized population.

**Key words**: Acoustic spectrography. Yanagihara Classification. Esophageal voice. Perceptual analysis.

**INTRODUCTION**

A total laryngectomy is a mutilating operation and although the quality of life of these patients is relatively good,¹² we should aim to improve their social and work-related adaptation by providing training in acquiring a psychoacoustically-acceptable esophageal voice. It should be pointed out that only 30%-40% of laryngectomized patients have an intelligible esophageal voice.³ The work of an ENT specialist should not be limited to periodic check-ups in the surgery purely to diagnose early possible recurrences of the disease. It should go a step further in assessing laryngectomees by developing an objective instrumental method that will allow us to monitor the progress a patient is making with their esophageal voice.

The aim of this study is to assess the degree to which qualitative spectrographic analysis can be used to evaluate the perceptual voice analysis of laryngectomized patients and to characterize voice quality according to the Yanagihara classification⁴ (adapted to the esophageal voice).

Spectrography enables complex sounds to be broken down into a series of sinusoidal component units,⁵ which are multiples of the fundamental frequency. Spectrographic analysis is a useful tool for extracting quantitative parameters...
from voice analysis as well as for classifying voice quality. According to Yanagihara, the acoustic characteristics of dysphonia are determined by the interaction of three factors: noise component in the principal formant of each vowel, high frequency noise component above 3 kHz, and the loss of high frequency harmonic components. Using these parameters, it is possible to produce a four-level classification of the severity of dysphonia based on the narrow-band spectrogram (45 Hz).

Within the spectrographic study of the esophageal voice, papers found in the medical literature indicate a tendency for sound energy to become concentrated in all the frequencies and not in specific areas, unlike in subjects with larynxes. The characteristic monotone of the esophageal voice is due to the appearance of noise components (irregular signal or signal from a continuous spectrum) which are related to turbulent airflow and to the loss of high frequency harmonics, (over 3 kHz), causing the richness of the voice timbre to disappear. By voice timbre we understand the acoustical characteristic that allows us to judge whether two tones presented in the same way and at the same frequency and intensity are different. These characteristics provoke an added difficulty in the perceptual analysis of the esophageal voice, the study of quantitative and qualitative variables (fundamental frequency, speech rate, etc. in the former and narrow-band spectrogram in the latter) being much more precise and objective in the characterization of the esophageal voice.

**MATERIAL AND METHODS**

We studied a sample of thirty-three male, laryngectomized patients of between 47 and 77 years of age (62.16 [8.2]). The average length of time since the total laryngectomy was performed was 58.48 months (ranging from 3 to 240 months). All the patients underwent a total laryngectomy because of laryngeal cancer. A total of thirteen patients, (39%), received complementary radiotherapy treatment following surgery, the remaining twenty patients (61%) did not require such treatment. Patients found to have a respiratory infection, a second primary tumor in the head and neck or lung cancer at the time the tests were to be conducted were not included in the study.

The subjects who participated in the study phonated a sustained /a/ sound before reading out loud the sentence: “papá pinta la pared de color púrpura con la pintura que compró por la tarde” (In English: “Dad paints the wall purple with the paint he bought in the afternoon”). The subjects were seated 30 cm away from a 515SDX-Shure® microphone, which was connected to a personal computer equipped with a sound card, the SoundScope version 1.2 computer program and a 16-bit SoundScope A/D card with a sampling frequency of 22 kHz. The acoustic signal (the spectrogram) of the sustained phonation of the vowel /a/ was recorded and the fundamental frequency was calculated. The calculation of speech rate (the number of syllables/time in seconds), was extracted from the sentence the patients read out, having previously been recorded on the spectrogram. The sound pressure or voice intensity (dB HL) was also calculated using an Isotech® SLM-1352A digital Sound Level Meter with the Soundmeter computer program. The maximum intensity (dB HL) was recorded during the phonation of the consonant-vowel sound /pa/ for 15 seconds.

From the narrow-band spectrographic analysis of the phonation of the vowel /a/, a qualitative study was conducted on the acoustic signal, grouping each one according to the modified Yanagihara classification. Finally, five members of the ENT department (all with normal hearing thresholds and all used to dealing with laryngectomized patients) participated in the perceptual analysis classification, assessing the acceptability of the voices of all the patients included in the sample.

**Spectrographic Analysis**

A qualitative acoustic analysis of the sustained vowel sound /a/ was conducted using a narrow-band spectrogram. The spectrograms were classified according to the levels proposed by Yanagihara but as we were dealing with laryngectomized patients, we divided groups II and III into subgroups “a” and “b” to be able to better classify the voices. This is because it is very difficult to obtain level I, and IV-quality esophageal voices. This subdivision into groups “a” and “b” refers to the differences in the distribution of the noise component appreciated in the spectrogram. Whether in group II or group III, type “b” presents a high noise component in high frequencies without quite moving up to the next level in the Yanagihara classification.

**Yanagihara Classification**

Level 1: Irregular harmonic components mixed with noise components, particularly in the region of vowel formants.

Level II (“a” or “b”): Moderately hoarse voices, with noise components in the second formants of vowels predominating over the harmonic components, together with the appearance of slight noise components at high frequencies over 3000 Hz (type b: complete absence of harmonics between the first and second formants).

Level III (“a” or “b”): The second formants of the vowels have been completely replaced by noise components, and high frequency noise is intensified, (type “b”: the second formant has completely disappeared).

Level IV: There is a loss of the periodic components of the first formants on account of their being occupied by noise, and high frequency noise becomes even more intensified.

**Table 1. Perceptual Analysis (Voice Acceptability)**

<table>
<thead>
<tr>
<th>Voice Tension</th>
<th>Voice Fluency</th>
<th>Noise Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1. Fluent sentence</td>
<td>1. Low</td>
</tr>
<tr>
<td>Strained</td>
<td>2. Fluent word</td>
<td>2. Medium</td>
</tr>
<tr>
<td>Very strained</td>
<td>3. Little fluency</td>
<td>3. High</td>
</tr>
</tbody>
</table>
Perceptual Analysis
We conducted a perceptual analysis measuring voice acceptability based on the subjective perception of the listener in regard to voice tension, fluency and noise component (Table 1).

Statistical Analysis of the Results
The normality of quantitative variables was studied using the Shapiro-Wilk test. Variables of normal distribution are given as an average ($\bar{x}$) and standard deviation (s). The most common descriptive markers were calculated: averages, standard or mean deviations, percentiles and interquartile increase.

The variables that do not follow a normal distribution are given as means (mx) with Interquartile Amplitude (IQA), which is the interval formed by the 25 and 5 percentiles; that is to say, an interval that covers the middle 50% of the sample.

We used the ANOVA of a factor to compare averages, followed by the Tuckey multiple comparison procedure when there was homogeneity of variance and the Tamhane procedure when there was heterogeneity.

Variables that did not fit a normal distribution pattern were compared using the Kruskal-Wallis test.

Pearson’s test was used to study correlations between parametric variables, and Spearman’s test was used for correlations between non-parametric variables. A bilateral $P$ of less than .05 was defined as significant. The statistical analyses were performed using the SPSS software package for Windows v11.1.

RESULTS
The quantitative results from the acoustic analysis are given in Table 2.

The distribution of the type of spectrogram following the modified Yanagihara classification was as follows: 0% for types I and IV, 35% for type IIa, 42% for type IIb, 17% for type IIIa and 6% for type IIIb. Therefore, the type of voice that predominated in our sample of patients was slightly hoarse, with noise components in the second formants of vowels predominating over harmonic components, together with the appearance of slight noise components in the high frequency regions above 3000 Hz and the total absence of harmonics between the first and second formants (Figure).

In regard to the perceptual analysis, the voice tension assessment was normal in 46.1%, strained in 30.7%, and very strained in 23.1%. Fluency was constant throughout the sentence in 50%, the words were fluent (but not the sentence as a whole) in 23.1%, and fluency was syllabic in 26.9%. The voice noise component was low in 48%, moderate in 28% and high in 24%. A statistically significant association ($P \leq .01$) was proven using the $\chi^2$ test between the most positive and most negative evaluations in each section of the perceptual analysis meaning that each of the three voice characteristics assessed behaves in the same way as the other two.

There was a statistically significant relationship ($P < .05$) between the types of spectrogram and the quantitative acoustic variables studied (fundamental frequency, phonation intensity and speech rate), meaning that the type IIa spectrogram (the most favorable from a psychoacoustic point of view) was associated with high values of maximum phonation intensity (90 [12.5] dBHL), speech rate (3.8 [0.9] syllables/sec, and low fundamental frequency values (92 [11.3] Hz). The differences were statistically significant ($P < .05$) compared to the other types of spectrogram. The ANOVA also revealed differences between the fundamental frequency value and the level of acceptability, (patients with a low fundamental frequency obtained higher classifications of voice acceptability) ($P < .01$). Similarly, the greater the value of maximum intensity, the better is the perception (acceptability) of the voice of laryngectomized subjects ($P < .05$).

Patients who underwent postoperative radiotherapy obtained higher fundamental frequency values than those who did not require such treatment ($P < .05$).

In regard to the length of time since surgery, there was a (statistically significant) tendency for the fundamental frequency value (Hz) to progressively decrease as the laryngectomized subject became used to using the esophageal

Table 2. Quantitative Acoustic Results

<table>
<thead>
<tr>
<th>Acoustic Variable</th>
<th>Average (Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamental frequency (Hz): /a/</td>
<td>106.23 (13.33)</td>
</tr>
<tr>
<td>Maximum Intensity (dBHL): /a/</td>
<td>81.15 (2.98)</td>
</tr>
<tr>
<td>Speech rate (No. of syllables/total time (seconds))</td>
<td>0.38 (0.19)</td>
</tr>
</tbody>
</table>
voice, going from an average fundamental frequency value of 120 Hz (<1 year since surgery) to a value of 105 Hz (1-3 years after surgery) and to 95 Hz in subjects laryngectomized more than ten years previously (P<.05).

DISCUSSION

This study has tried to assess the utility of spectrographic analysis - a reliable tool and one easy to use in daily clinical practice - in the assessment of the esophageal voice in laryngectomized patients. In fact, our results indicate that there are differences between the spectrogram (of the modified Yanagihara classification) and the acoustic variables often used to quantify the voice. A favorable spectrogram (IIa) is related to greater voice intensity, a faster speech rate and a lower fundamental frequency, facts that are statistically significant as far as the degree of voice acceptability is concerned.

We investigated the degree to which the net value of the fundamental frequency and the speech rate are determining factors when evaluating the acceptability of the esophageal voice. According to our results, a high acceptability in the perceptual analysis is related to low fundamental frequency values. Some authors13,14 talk of a relationship between voice acceptability and fundamental frequency, although others15 report that a high fundamental frequency is suggestive of a more favorable level of voice acceptability and even suggest the need13 to increase the frequency is suggestive of a more favorable level of voice relationship between voice acceptability and fundamental frequency values. Some authors13,14 talk of a relationship between voice acceptability and fundamental frequency, although others15 report that a high fundamental frequency is suggestive of a more favorable level of voice acceptability and even suggest the need13 to increase the frequency.

In fact, in the qualitative acoustic analysis we observed how the “cleanest” spectrograms (with the best Yanagihara classification) are those with low fundamental frequencies, closer to values for the laryngeal voice (100-105 Hz). However, the fundamental frequency value may be a parameter of relative importance in voice acceptability. Some researchers15 do not find differences between the fundamental frequency of the laryngeal voice and the esophageal voice with a voice prosthesis, thus highlighting the poor involvement of this variable in the characterization of the voice. The value of the spectrogram in the analysis of the esophageal voice has been studied by other authors,8,9,16 who have indicated the usefulness of this procedure as a tool in the qualitative analysis of the esophageal voice. Looking carefully at our results, there is a statistically significant correlation between the spectrographic criteria of Yanagihara and speech rate, (a quantitative acoustic measurement related to fluency in words per minute), with greater values for the latter achieved in the cases which also obtained higher scores in the Yanagihara classification. In the spectrographic analysis of the laryngectomized subject, energy is concentrated on all of the frequencies, there is a large noise component, fundamental frequency is an octave lower (than in the laryngeal voice), the frequency of the first formant is higher and the third formant does not always appear. These features make it different to the spectrogram of a laryngeal voice. Furthermore, we have seen a (non-statistically significant) tendency that connects high maximum-voice-intensity values to favorable spectrographic charts, the maximum intensity of an acoustic quantitative variable being directly related to a good esophageal voice, just as some authors state.26

There is, therefore, a statistically significant association between the type of spectrogram (qualitative variable) and the acoustic quantitative variables studied. The spectrographic classification is also related to voice acceptability, thereby reinforcing the use of the spectrogram as a tool in monitoring the esophageal voice in the ENT department. In any case, it is necessary to conduct prospective studies to determine the extent to which the spectrogram is valid in the arduous strategy involved in voice rehabilitation following a total laryngectomy.

CONCLUSIONS

– Despite laryngectomies having a reasonably acceptable quality of life, more than 60% do not achieve a plausible esophageal voice, a fact that considerably slows down their social and/or work-related adaptation. ENT specialists should involve themselves more in improving the health of the laryngectomized patient, the voice in particular, by passing on methodological tools of outstanding reliability to this end.

– The modified Yanagihara spectrographic classification enables the esophageal voice to be objectively classified and could possibly be another useful tool to take into consideration in the voice rehabilitation of these patients.

REFERENCES