SPEECH MONITORING OF INFECTIVE LARYNGITIS

F. Plante(+), H. Kessler(*), B. Cheetham(+), J. Earis(*)

(+) Electrical Engineering and Electronics Department
Liverpool University, Liverpool L69 3BX, UK
(*) Dept of Respiratory Physiology, Aintree Chest Centre
Fazarkarley Hospital, Liverpool, L9 7AL, UK

ABSTRACT

Many types of parameters have been proposed for the evaluation of vocal cord abnormalities by speech waveform analysis. However, none of them taken separately allows a reliable assessment of the presence and the degree of abnormality. In this study we proposed to combine three different parameters which take into account different consequences of the hoarseness. To assess the effectiveness of the parameters, a group of subjects during and after acute infective laryngitis are compared with control subjects.

The jitter, the glottal to noise excitation (GNE) and the normalised error prediction (NEP) are the parameters studied. Preliminary results indicate that reliable discrimination between normal and abnormal patients may be possible using these three parameters.

1. INTRODUCTION

A change in the quality of the voice (hoarseness) is a well recognised symptom of diseases involving the larynx. Causes include acute infective laryngitis, chronic non-specific laryngitis, benign and malignant laryngeal tumours, vocal cord paralysis and laryngeal myopathy induced by the long term use of inhaled steroids. Objective analysis of hoarse voice, using acoustic analysis of speech, can provide a non-invasive measure of vocal cord function. This has the potential to allow precise and rapid diagnosis of serious laryngeal disease, to observe disease progress and to monitor response to treatment. The main aim of this study is to develop an objective measure of 'hoarseness' by comparing voiced sounds in previously healthy subjects during and after acute infective laryngitis.

Many parameters have been proposed in order to evaluate the hoarseness. As the hoarse voice has different acoustical characteristics, none of them taken separately has been found to be sufficient to accurately describe the degre of hoarseness. Such assessment may only be achievable by grouping several parameters. These parameters need to be as much as possible non related. In this study, we used three parameters which are based on different analysis methods, and try to characterise the acoustical distortion due to the hoarseness. These parameters are the jitter, the glottal to noise excitation (GNE) and the normalised error prediction (NEP).

The main aim of this study is to describe three different methods of analysing hoarseness of voice and to test them against a small population of normal subjects and patients suffering from acute laryngitis with obvious hoarseness of voice.

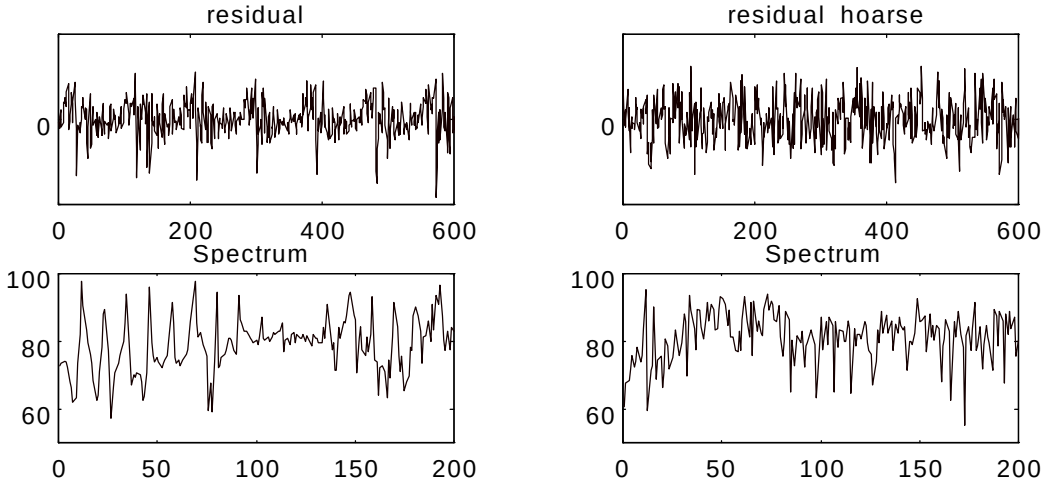

Figure 1 : Residual (top) and its low frequency spectrum (bottom) for a normal voice (left) and hoarse voice (right).

2. THE ANALYSIS METHODS

For hoarse voice, the periodicity and dynamics of the opening and closure of vocal cords vary more from one cycle to another than for normal voice. This distortion is characterised by a variation in the pitch period, and an increased randomness which can be seen as a degradation of the harmonic structure.

As the disorder occurs at the larynx, it is preferable to remove the vocal tract component before computing the parameters. First, linear prediction is used to calculate the parameters of the vocal tract function transfer. Then by inverse filtering, the residual is computed. The residual is an approximation to that of the time derivative of the glottal waveform.

Figure 1 illustrates differences that are observed between the residuals obtained by analysing a normal and a hoarse voice. For the hoarse voice, it is difficult to discern the pitch period, and the harmonic structure disappears. The residual looks more like a random signal than a train of impulse.

2.1 Jitter

In 1963 Lieberman[2] noticed than the differences between consecutive pitch periods is more pronounced for hoarse voice than for normal voice. This variation in the pitch period is refered to as the jitter.
Since this original work, a many techniques have been proposed for computing the jitter. They differ according to the metric used to compute the difference between consecutive pitch periods [5].

In this study we used the definition of jitter proposed by Kasuya [1]. The jitter is defined for a block of speech samples as the average percentage variation between consecutive pitch periods (eq.1). The jitter is expected to be higher for hoarse voice than normal voice.

$$Jitter = \frac{100}{N-1}\sum_{n=1}^{N-1}\left|\frac{p(n)-p(n-1)}{p(n)}\right| \qquad \text{eq.1}$$

where $p(n)$ is the length of the nth pitch period and N is the number of pitch periods detected.

The computation of the jitter necessitates the determination of the instantaneous pitch period. This is achieved by detecting the instants of glottal closure by locating the peaks of the instantaneous envelope of the residual [6].

2.2 Glottal To Noise Excitation (GNE)

Previous approaches to evaluating the degre of randomness use the ratio between the energy at pitch harmonics and the spectral energy between the harmonics [1]. However this approach is highly dependent on the accuracy of a pitch detector (the effect of a small error in the estimated pitch, will become larger for the higher order harmonics).

Michaelis and Strube proposed another approach based on the cross-correlation between the envelopes of three bandpass filtered versions of the residual [4]. For normal speech, the harmonic structure is well defined across a wide frequency band (fig 1) and the spectrum may be well modelled by a sum of sinusoids harmonically related. Then the bandpass filtered signals will have the same periodicity and higher cross-correlation coefficients. On the other hand, for hoarse voice, such harmonic structure is not well defined and the bandpass filtered signals are not highly correlated. The frequency bands used are 0-1kHz, 1-2kHz and 2-3kHz. The maximum cross-correlation coefficient between any paire of bandpass filtered signals is defined to be the glottal to

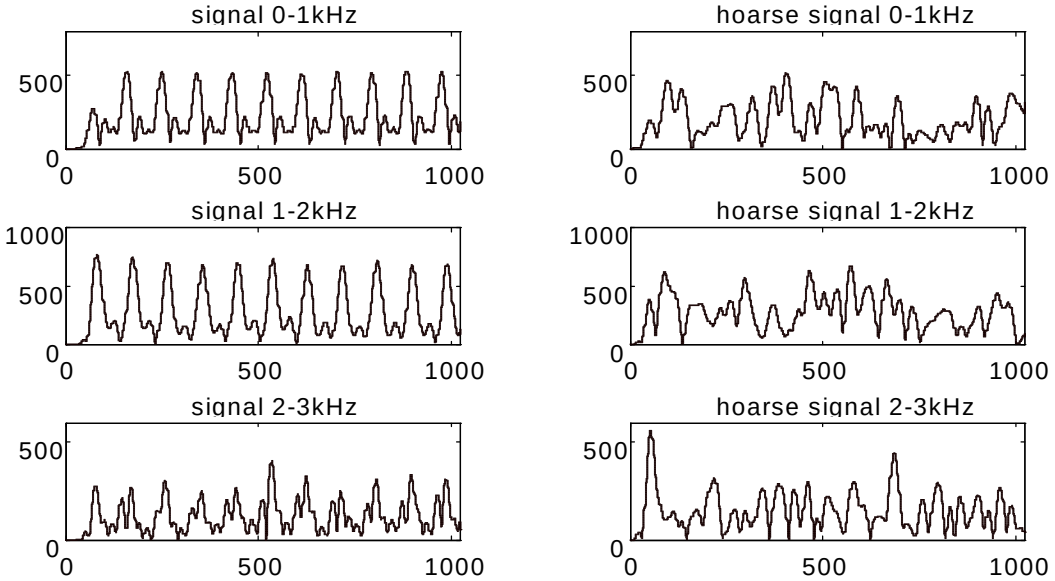

Figure 2: Examples of the three bandpass filtered residual for normal voice (left) and hoarse voice (right).

noise excitation.

Figure 2 shows examples of the three bandpass filtered residual signals in the case of normal voice and hoarse voice. As expected, in the case of hoarse voice, the three signals have little similarity and the cross-correlation coefficients are lower (Table 1).

Correlation	Normal	Hoarse
0-1kHz / 1-2 kHz	0.89	0.34
0-1 kHz / 2-3kHz	0.65	0.27
1-2kHz / 2-3kHz	0.74	0.19

Table 1 : Correlation coefficient find between the filtered signals for the examples of figure 1 and 2.

2.3 Normalised Error Prediction (NEP)

Another way to assess the noise level is to compute the normalised energy of the residual obtained by linear prediction.
The normalised linear prediction error can be expressed as the ratio of the geometric mean of the model spectrum to its arithmetic mean [3]. This is a measure of the spectral spreading of the data. The more the spectrum is spread, the less the normalised error is. For normal voice, the harmonic structure is well defined with high peaks and low vallies, which give a high degree of spreading. In case of hoarse voice, as the harmonics are not well defined, the spreading is smaller, given a larger normalised error.

3. DATA

Recordings were made from 4 control subjects with no evidence of any laryngeal disease and 4 patients with acute laryngitis. Two of the patients were subsequently recorded after their symptoms had disappeared. One of the patients was recorded during the period when his hoarseness was increasing.
The recordings of the 4 patients with acute laryngitis are compared with the 4 controls and the recordings of two of the patients following full recovery of their voices.

Each recording consisted of the vowel 'ah' sustained for 5 seconds. Then two segments of 492 ms were extracted at the middle and the end of the sustained part of the vowel. The Jitter is computed using eq.1 on the length of the segment. For the GNE and NEP, the segments were subdivided into 20 frames of 46ms. For each frame, the cross-correlation coefficient and normalised error was computed. The GNE is the maximal cross-correlation coefficient for the segment. The NEP is the mean of the 20 normalised errors.

4. RESULTS

Figures 3 and 4 represent the position of the subjects of the different groups in the graphs (Jitter-GNE) and (Jitter- NEP) respectively.
Each subject is represented by two points corresponding to the parameters calculated for the two segments extracted in the middle and at the end of the sustained vowel.
Members of the control group are represented by crosses each labelled with the number of the subject. The patients with acute laryngitis are represented by a filled symbol (circle, triangle, square or diamond). After recovery of the voice, each subject is represented by the same symbol, but non-filled. The larger triangles represent the same subject as the smaller, but after an increase of the level of hoarseness.

The patients analysed after recovery of their voices were found to have parameters similar to those of the control group. On the other hand, for one subject whose hoarseness increased, the values of the parameters moved further away from those obtained with the control group.
Each of the three parameters, is correlated with the level of

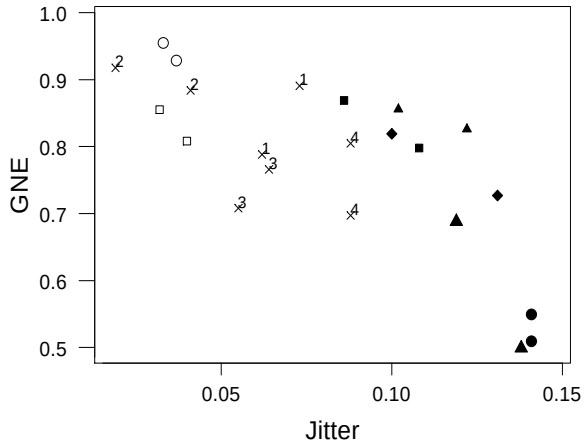

Figure 3: Dispersion of patients in a graph of GNE against Jitter during (filled symbol) and after (non-filled) laryngitis. The crosses represent the control subjects.

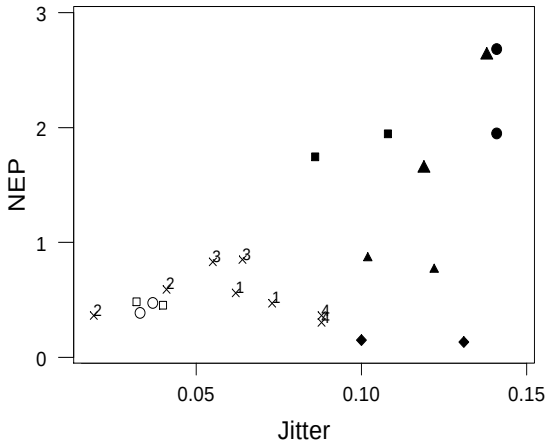

Figure 4: Dispersion of patients in a graph of NEP against Jitter during (filled symbol) and after (non-filled) laryngitis. The crosses represent the control subjects.

hoarseness. This correlation is higher for the jitter and smaller for the GNE. For each parameter it is not possible to fix a threeshold which allows a reliable separation of the normal subjects and hoarse subjects. However in each of the two graphs (Jitter-GNE) and (Jitter-NEP) it is possible to subdivide the graph into two regions corresponding to the normal subjects and hoarse subjects respectively.

It can be seen that the variability intra subject is larger for subjects with hoarse voice than for the subjects with normal voice. The dispersion of the points on the graph for the normal subjects is smaller than for the hoarse group. Nevertheless, the variability in the normal group is quite high.

5. DISCUSSION

The separation between groups of subjects with normal voice and hoarse voice (due to acute laryngitis) is possible using several parameters. However even in this simple case, the discrimination is less pronounced than may be expected from listening to the subjects. This raises several points.

The variability of the parameters between subjects of the same group shows than an absolute comparison is quite difficult. Every subject has a particular phonatory apparatus and physiology which introduce variation in the speech signal. These variations are reflected in the values of the parameters. So prior to assessing the modification due to the hoarseness, it may be better to eliminate these variation by adapting the computation of the parameters to the subject. The monitoring of the speech quality with the same subject will be easiest because we can assume that the phonatory apparatus does not change during the length of the monitoring.

The variability intra-subject is more pronounced for the subjects with hoarse voice. This brings two reflections. The first is that the hoarseness is characterised by the inability to sustain a periodic speech sound. Then it will be more interesting to take into account the variation in the values of the parameters. The hoarseness is thus seen as a "non state" of speech quality. This means that the parameters will not be concentrated in one region of the graph as for normal subjects, but will instead be spread out. This spreading will reflect the level of hoarseness. At some times, the hoarse voice can appear "normal" i.e. have some normal parameters.
If we take the worst parameters for each subject, the discrimination is generally much easier.

In this study the parameters chosen allow us to discriminate two populations. However, they need to be improved for monitoring less severe hoarseness or closely assessing the speech quality during therapy or medication.
As mentioned previously, it is important to further adapt the methods to the subject. For example, to compute the GNE it would be more efficient to design the bands of the filters according to the pitch value (i.e. to have the same number of harmonics in each band).
Also it will be necessary to take into account the variation of the parameters. One notices that the jitter is the more discriminative parameter, because it quantises variation in the pitch period. The GNE and NEP need perhaps to be modified in this way.

The three parameters selected try to take into account different aspects of the hoarse voice. But there are some other acoustical characteristics which need to be investigated with other parameters, like the shimmer (variation in amplitude of the pitch excitations for one pitch cycle to the next).

6. CONCLUSIONS

This study shows that using three parameters (Jitter, GNE and NEP) it is possible to separate subjects with hoarse voice from subjects with normal voice or after recovery. However, the need for some improvements has been identified for the computation of the parameters.

The study of a larger population of control subjects, and subjects with acute laryngitis, and other pathologies (vocal cord paralysis, tumours, etc.) and the monitoring of speech therapy represent the further steps planned for this work.

7. REFERENCES

[1] Kasuya, H., Masubuchi, K., Ebihara, S., Yoshida, H., "Preliminary experiments on voice screening" J. Phonetics, Vol.14, 463-468, 1986.
[2] Lieberman, P., "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges" *J. Acoust. Soc. Am.* Vol.35, 344-353, 1963.
[3] Makhoul, J., "Linear Prediction: A Tutorial Review" *Proceeding IEEE*, 115-134, 1975.
[4] Michaelis , D., Strube H.W., "Empirical study to test the independence of different acoustic voice parameters on a large voice database", *Eurospeech'95*, 1891-1894, Madrid, 1995.
[5] Plante F., "Detection Acoustique des pathologies phonatoires chez l'enfant" PhD Thesis, 1993 (in french).
[6] Plante, F., Meyer, G., Ainsworth, A. "Pitch detection: auditory model versus inverse filtering" Proceeding IOA, Vol.16, 81-88, 1994.