Proceedings of The Institute of Acoustics

ACOUSTIC ANALYSIS OF VOCAL FOLD PATHOLOGY

John Laver, Steven Hiller, Janet Mackenzie

Centre for Speech Technology Research
Department of Linguistics, University of Edinburgh.

INTRODUCTION

The main aim of this study is to develop a computer-based system of acoustic analysis which is capable of screening voices for the presence of vocal pathologies. The social implications of such a system, which involves a non-invasive and relatively cheap recording technique, are considerable. The early detection of such disorders as laryngeal cancer is highly desirable, since prompt medical treatment has a high success rate.

Acoustic screening can be profitably applied to two different populations. The first of these is an unselected population. For example, routine screening for vocal pathology could be carried out in "well woman" or "well man" clinics, alongside existing screening tests for breast cancer, cardiac function etc. Alternatively, a more limited, pre-selected population could be screened in cases where vocal pathology is already suspected. For instance, patients referred by their general practitioners for laryngeal examinations may face lengthy waiting lists. If these patients were to be recorded at the time of referral, it might be possible to ensure that those cases where acoustic measures suggest the presence of a neoplastic structural abnormality would be seen immediately. Acoustic screening could thus be used to select priority cases, accelerating examination of patients judged to be at serious risk.

A second aim of the project is to investigate the possibility of using acoustic measures to differentiate between various types of vocal pathology. This requires an exploration of the relationships between various classes of structural disturbance of the vocal fold tissue layers and acoustic perturbations of the laryngeal waveform. It is possible to formulate a range of predictions about the acoustic consequences of different types of pathology [1], and these predictions are being tested by collecting detailed information about the status of the individual patient's larynx. This is made possible by collaboration with laryngologists and speech therapists at the Radcliffe Infirmary, Oxford and the Royal Infirmary, Edinburgh.

THE ACOUSTIC SYSTEM

The research results to be discussed derive from computer-based acoustic measurement of individual pitch periods in approximately 40 seconds of tape recorded read text. The measurement system uses an elaborated version of the Gold and Rabiner parallel processing method [2], with phase compensation, low-pass filtering, non-linear smoothing for intonational baseline measurement, and sampling frequency multiplication at waveform peaks by parabolic interpolation for greater resolution of individual pitch periods. This system is described in detail in Hiller, Laver and Mackenzie [3]. There are three types of output data. Firstly, intonational data, such as mean, median and standard deviation of fundamental frequency (FØ), is obtained from the smoothed FØ trend line. Secondly, statistical analysis of excursions of individual pitch period values from the local fundamental frequency trend line provides data for pitch perturbation. Thirdly, an analysis of intensity perturbation is made. The following

measures are used to describe both pitch and intensity perturbation:

1. Mean magnitude of excursion
2. Standard deviation of the magnitude of excursion
3. The rate of excursion (RATEX). This is the percentage of points in the sample where the magnitude of excursion is equal to, or more than 3% of the local trend line value.
4. The directional perturbation factor (DPF). This measure, adapted from Hecker and Kreul [4] is the percentage of changes in algebraic sign in the excursion values. A 3% threshold is also applied to this measure.

SUBJECTS, SPEECH MATERIAL AND ANALYSIS PROCEDURES

Table 1 presents information about sex, age and the percentage of self-reported smokers in the control group and pathological group investigated in this study. It can be seen that the control group speakers are on average younger than the pathological speakers. This reflects a bias in collection of control speakers, who were mostly selected from the university environment. Future work in this project will rectify this bias. None of the control speakers reported any known speech or hearing problems (including cold or sinus ailments) at the time of recording.

Table 1. Subject group information

Group	Sex	N	Mean age (range)	Percentage smokers
Control	M	38	34 (18-63)	19.4%
Control	F	26	26 (18-44)	29.2%
Pathological	M	32	54 (27-82)	30.3%
Pathological	F	31	56 (26-75)	51.7%

Table 2 shows a broad classification of the laryngeal disorders evidenced by the pathological speakers, as determined by laryngological examination.

Table 2. Classification of Laryngeal disorders diagnosed in pathological group

Type of pathology	Number of males	Number of females
Epithelial disorders (e.g. carcinoma, papilloma keratosis)	12	0
Polyps, nodules	5	10
Disorders of the cartilagenous area	4	3
Mild oedema, redness etc	5	14
Palsies	6	4
Total	32	31

A tape recording was made of each speaker as he or she read the first two paragraphs of "The Rainbow Passage" [5]. Forty seconds of each recorded speech sample was digitized at 20 KHz and stored on computer magnetic tape for future processing by the acoustic system. The results of the analysis were stored in a computerized voice acoustic data base for use in screening procedures.

SCREENING APPROACHES AND RESULTS

There are several possible approaches to acoustic screening for vocal pathology. One simple approach is to focus on single acoustic parameters, and to relate individual values to the means and standard deviations (SD) of control group data. This is easiest if parameter values are normalised by conversion to Z-scores. In other words, they are expressed in terms of units of SD by which they diverge from the control group mean. It seems reasonable to suppose that any individual parameter value which diverges from the control group mean by more than 2SD is indicative of a strong risk of abnormality. The cut-off level for screening can therefore be set at 2SD above or below the control group mean for a given acoustic parameter.

A preliminary evaluation of the screening potential of 10 individual acoustic parameters, using this approach, shows a considerable range of success. This is shown in Table 3. The most effective parameter, in terms of discrimination between control speakers and those with known laryngeal pathology, is shimmer DPF. This picks out 62.1% of male speakers and 53.6% of female speakers with known pathology. The least effective parameter is SD of shimmer excursions, for which only 6.9% of pathological males and 7.1% of pathological females have values which diverge from the control mean by more than 2SD.

Table 3. Percentage of speakers with acoustic values which diverge from the control group mean by 2SD

Acoustic parameter	Male controls	Male pathological	Female controls	Female pathological
Shimmer DPF	... 2.6%	62.1%	3.8%	53.6%
Shimmer SD of excursions	... 5.3%	6.9%	7.7%	7.1%
One or more out of ten parameters .	.. 18.4%	82.8%	30.8%	92.9%
Two or more out of ten parameters .	... 2.6%	69.0%	11.5%	78.6%

An alternative criterion for screening for pathology might be divergence from the control group mean by more than 2SD in at least one of the overall set of 10 acoustic parameters. This gives a rather better detection rate for speakers with laryngeal pathology (82.8% of males and 92.9% of females), but it also identifies an unacceptably high number of the control group as being apparently abnormal (18.4% of males and 30.8% of females). We can call these "false positives". A compromise criterion for abnormality is that at least 2 parameters out of 10 must diverge from the control group means by more than 2SD. This reduces the false positives to 2.6% of males and 11.5% of females, whilst still detecting 69.0% of males and 78.6% of females with known laryngeal pathology.

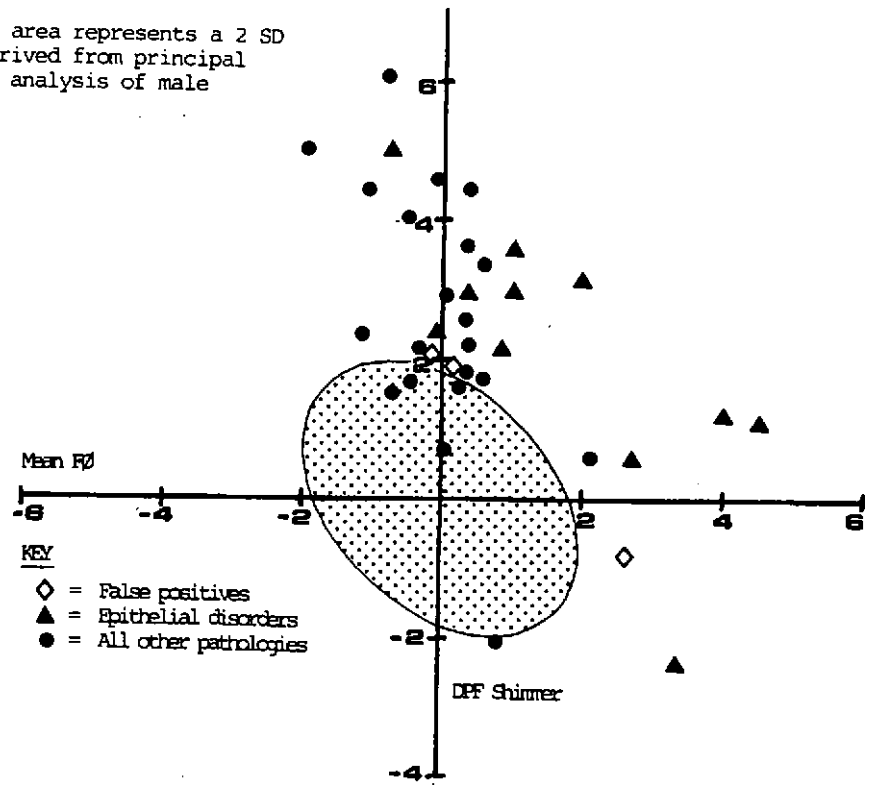
ACOUSTIC ANALYSIS OF VOCAL FOLD PATHOLOGY

Although mean FØ alone does not seem to be a very good screening parameter, it is interesting in that it was the only parameter which was abnormal in one early case of laryngeal cancer. This prompts the hypothesis that in some cases patients may be able to maintain normal levels of perturbation in spite of slight structural abnormality, but only within a rather abnormal pitch range. It therefore seems useful to examine the data in terms of the conjunction between mean FØ and perturbation, to see if this forms a better basis for screening. (A plot of mean FØ against perturbation has the added advantage that it can be related to predictions about the probable consequences for FØ of alterations in stiffness, mass and symmetry of the vocal folds). We can concentrate initially on shimmer DPF, as the perturbation measure which was most effective as a single screening parameter.

On the graph shown in Figure 1, the intersection of the axes at zero corresponds to the control group mean for each parameter. Each unit away from the mean corresponds to one standard deviation. By definition, the control group will cluster around the intersect.

Figure 1: A scattergram of DPF shimmer vs. Mean FØ for male speakers

The shaded area represents a 2 SD ellipse derived from principal components analysis of male controls.



A principal components analysis can be applied to the control group data, and ellipses can be drawn which indicate the covariance between the two parameters. If an ellipse is drawn at the 2SD level, this will, by definition, enclose approximately 95% of the control group. This ellipse can then be used as a screening boundary, so that any data point falling outside the ellipse is interpreted as being potentially indicative of abnormality.

The 2SD ellipse shown on Figure 1 describes the male control group data. Of the 38 control males, only 3(7.9%) fall outside the ellipse, and would thus be picked up as false positives. In contrast, if the male speakers with laryngeal pathology are plotted on this graph, 87.5% fall outside the ellipse. This type of two dimensional analysis thus looks to be a more promising approach to screening than a single parameter approach.

It should be remembered that this 87.5% detection rate applies to a group which includes a wide range of disorders, including some very minor and benign pathologies. Laryngeal cancer and pre-cancerous states, which are obviously the most serious disorders, almost always arise in the covering epithelium, so that we are most concerned with detection of structural tissue alterations of the epithelium. It is encouraging that 100% of this subgroup fall outside the ellipse.

The results for female speakers show a similar pattern, with 15.4% of controls, and 90.3% of pathological speakers falling outside a 2SD ellipse. The slightly higher percentage of false positives may reflect the smaller female control group, which does not yet show a normal distribution of mean FØ values.

It is worth commenting that shimmer DPF does not, however, seem to be the best acoustic measure for differentiation between classes of vocal pathology.

There is some evidence that a plot of jitter Ratex against mean FØ, whilst having less success in initial screening, is a better discriminator for pathology type. For example, there is a clear tendency for epithelial disorders to result in high jitter ratex scores and/or high mean FØ, whereas a large proportion of nodules and polyps, have low jitter ratex scores and low mean FØ compared with the control group.

A more sensitive, but statistically more complex, approach to acoustic discrimination of different vocal pathologies would involve a comparison of overall acoustic profiles. An acoustic profile would include information on all the acoustic measures produced by this system. The average profile shapes for different classes of vocal pathology are recognisably different, and this supports a further investigation of this approach.

CONCLUSION

In conclusion, these results appear to justify attempts to use acoustic analysis in screening for the presence of laryngeal pathology. The task of acoustic discrimination between types of pathological abnormality is more complex, but there are indications that acoustic analysis may provide useful quantitative information in support of the diagnostic process.

ACOUSTIC ANALYSIS OF VOCAL FOLD PATHOLOGY

REFERENCES

[1] J. Mackenzie, J. Laver and S. Hiller, 'Structural pathologies of the vocal folds and phonation'. Work in Progress, Dept.of Linguistics, Edinburgh University, No. 16, 8-116, (1983).

[2] B. Gold and L.R. Rabiner, 'Parallel processing techniques for estimating pitch periods of speech in the time domain', J.A.S.A., Vol. 46, 442-448, (1969).

[3] S. Hiller, J. Laver and J. Mackenzie, 'Automatic analysis of waveform perturbations in connected speech'. Work in Progress, Dept. of Linguistics, Edinburgh University, No. 16, 40-68, (1983).

[4] M. Hecker and E. Kreul, 'Descriptions of the speech of patients with cancer of the vocal folds. Part 1: Measures of fundamental frequency'. J.A.S.A., Vol. 49, 1275-1282, (1971).

[5] G. Fairbanks, Voice and Articulation Drillbook, New York: Harper Brothers, (1960).