

Acoustical and Clinical Assessment of Voice Quality with Laryngeal Pathology

Hartono Abdoerrachman

Abstrak

Penelitian mengenai kekacauan gelombang suara telah banyak dilakukan oleh para ahli dan telah dilaporkan sebelum ini, tetapi masalah ini masih banyak menarik minat para klinisi untuk menelaah lebih lanjut. Para klinisi sangat mengharapkan menemukan suatu teknik yang praktis, otomatis dan non-invasif untuk penggunaan medis. Dalam penelitian mengenai kekacauan suara ini, dicakup 672 data suara selama 3 tahun, dari Januari 1990 sampai Desember 1992, yang terdiri dari berbagai kasus dengan penyakit yang menyebabkan perubahan suara. Hasil penelitian ini menunjukkan bahwa beberapa tujuan penggunaan analisis suara ini tercapai, meskipun target utama masih jauh.

Abstract

Acoustical waveform studies have been conducted since a long time, and results have been reported by many experts, but it is still an interesting matter for clinicians to explore. Clinicians are seeking for a practical automatic and non-invasive technique for clinical purposes. Our study on voice perturbation covered a 3 years period data, from January 1990 until December 1992, including 672 voice data consisting of various diseases, which cause vocal changes. The results revealed that some aims of voice analysis were achieved, although the main target is still a little bit far.

Keywords: voice analysis, acoustical assessment, acoustical waveform perturbation, pathological voice analysis.

INTRODUCTION

Besides the eye and the ear, speech is the effective and direct way for verbal communication, and speech disorder will have negative effect to the social life and future career of a person. There are many causes resulting in phonation disturbances, organic or functional disorders, affecting one or more organs involved in voice production, such as the respiratory organ, phonatory and resonance organs, for which it is sometimes difficult to establish the etiological diagnosis. The importance of establishing the exact diagnosis is to understand the problem, so that the proper treatment could be managed. Examination of the larynx by indirect or direct laryngoscopy is one way to visualize the position and movement of the vocal cords. This method is somewhat invasive, traumatic, and not always easy to perform, especially on sensitive and non-cooperative patients.

There has been a growing interest in the objective documentation and assessment of voice quality. This

interest arose because of important advances in techniques and analyses that can be applied to vocal physiology in a scientific manner.¹ Clinicians have recently explored the usefulness of various acoustic factors as well as glottographic reflections of vocal vibration. There is a great need to develop methods for quantification of glottal voice source behaviour in clinical practice for patients with voice disorders. It is well known that vocal fold vibration in modal register involves, in addition to the gross opening and closing movements, a wave-like motion of the looser mucosa relative to the denser underlying tissues. Histological sections indicate that the membranous portion of the vocal fold should be considered as a layered structure.¹

Hirano^{2,3} divided the layers of the vocal cord into 5 components, and from the mechanical point of view, the five histological layers can be reclassified into 3 sections: the cover, consisting of the epithelium and the superficial layer of the lamina propria, the transitional layer, consisting of the intermediate and deep

* Center of Ear Care and Communicative Disorder, Department of Otorhinolaryngology, Faculty of Medicine, University of Indonesia, Dr. Cipto Mangunkusumo National Central Hospital, Jakarta, Indonesia

layer of the lamina propria, and the body, consisting of the muscle.

Phonation is the aerodynamic and acoustic product of an extremely complex vibratory system. Part of the evidence for this complexity is the fact that the cycle-to-cycle consistency of laryngeal vibration in phonation is never completely regular (Boves, 1984, cit.⁴) Baken (1987) stated that the phonatory system is not a perfect machine and every speaker's vibratory cycle is erratic to some extent (cit.⁴) From the acoustical point of view, the characteristic of perturbation in the laryngeal waveform can be detected in both the time domain, called jitter, and in the frequency domain, called shimmer. From the auditory point of view, jitter and shimmer contribute to the rough perceptual effect, which is called harshness. Physiologically a harsh voice can arise either functionally from hyperkinetic adjustments of the phonatory muscles, or pathologically from disorders and diseases on one or both of the vocal cords. Laryngeal waveform perturbation can occur across a wide range in incidence from occasional to habitual, and in degree from slight to severe. Severe perturbations are always signs of either pathological or functional disorder, but slight perturbations are evident in all speaking voices, especially at the border of voicing phase.

Medical background and aim of the study

The attempt to discover objective acoustic and physiological ways of characterizing laryngeal waveform in terms of perturbation parameters has a history of over 30 years, beginning with the pioneering work of researchers such as Moore and von Leden (1958), Lieberman (1963) and also Michel in 1964. Since then, the topic of waveform perturbation had been spreading and had attracted many experts of different fields, such as laryngology, signal processing, speech science, and speech pathology.

It is important to emphasize that there are many different ways of defining waveform perturbations, and that no wide consensus has yet existed on the definitions of the identity of the most information-rich parameters. When a patient complains of hoarseness or a harsh voice, an important clinical decision must be made whether this is a pathological or a functional symptom. This decision will influence directly the mode of treatment, and so any objective method for supporting treatment decisions, if they can be made automatic, reliable, cost-effective and non-invasive, is of obvious oto-laryngological value.

According to Laver et al.⁵ and Wilson et al.,⁶ there are four major applications of automatic methods of

waveform perturbation analysis that could identify laryngeal pathology:

1. Screening of a given population, to complement existing screening programs in hospitals, such as well person clinic, because the incidence of pathological conditions that need urgent treatment is substantially lower than the incidence of solely functional voice disorders.
2. Priority assessment of patients visiting their general practitioner with complaints of hoarseness or harshness.
3. Diagnostic support, where the acoustic system had demonstrated its ability to provide reliable evidence discriminating pathological versus functional etiology.
4. Follow up tool, a monitor to assess voice changes after the patients received radiotherapy, chemotherapy, surgery, speech therapy or to detect remission or deterioration in progressive disease.

It is probably reasonable to suggest that current acoustic techniques of quantifying perturbatory evidence in laryngeal waveforms could reliably and cost-effectively support the functions of screening, priority assessment, and monitoring. As Laver⁵ stated, it is premature to suggest that such systems can support a discriminative diagnostic function with any satisfactory degree of reliability. Principally there are promising signs that such techniques are capable of providing diagnostic support. A major obstacle is the lack of adequate databases of acoustic and clinical recordings of speakers with relevant laryngeal disorders, and of normal controls.

Techniques

According to Laver et al.,⁵ there are two advantages of acoustic technique: firstly, the recording methods involved are non-invasive, non-frightening to patients, and using fairly low cost technology, and secondly the acoustic perturbation data techniques of analysis are directly executed to automate and hard-copy records that are easy to interpret and can easily be made available for the patient's file.

The disadvantages in using acoustical data-analysing techniques for perturbation analysis are: firstly, acoustic recordings are subjected to contamination by both environmental factors and equipment-derived artifacts. Thus, acquiring the accurate data of the acoustic signals requires a well prepared recording environment as well as equipment, and the hazards derived from those factors can be greatly reduced when appropriate precautions are taken. Secondly, the

process of automating the perturbatory analysis involves relatively high-cost computing equipment.

The ideal acoustic recording equipment for acquiring acoustical data for perturbation analysis should have the features of flat low frequency response, that should be able to record as low as 40-50 Hz or even lower, and should have no signal distortion which is mostly caused by phase distortion.

The equipment should also have high signal-to-noise ratio, because most of time-domain pitch detection algorithms are not very resistant to the effects of contamination by transient noises, and by continuous ambient noises. And at last, the recording equipment must be affordable to the clinic using it.

The acoustical data collecting must be made in a quiet environment to achieve an adequate signal-to-noise ratio, preferably in a sound-treated room which is shielded from electrical signals and mechanical sounds.

MATERIALS AND METHODS

This paper will report the results of the analysis upon 672 voice samples from various diseases, recorded within a 3 years period, from 1990-1992, using DAT tape-recorder, with a constant distance from mic to mouth of 15 cm, following patient's practice. Recordings were made in a sound-treated booth. Four hundred forty two samples consisted of 23 normal voice, 17 tremorous voice, 40 spasmodic dysphonia, 177 vocal-cord polyp, 56 vocal nodule, 97 recurrent nerve paralysis and 32 polypoid of the vocal cord. Voice samples of sustained vowel /e/ were digitized through a 16-bit A/D converter at a sampling rate of 40 kHz and stored on a disk controlled by a computer. A half-second segment was extracted by excluding the initial and final portions from each sample. Saturated recording of sustained vowel, and also too short duration of the /e/ phonation, less than 0.5 second, are excluded.

Using peak picking method, the method recommended by Imaizumi et al.,⁷ local maximum points which could correspond to vocal excitation epochs were detected successively and then cycle by cycle perturbation quotient were calculated and finally the pitch perturbation quotient (PPQ-%) and amplitude perturbation quotient (APQ-%) were determined. The additive noise level (NLvl) were also calculated.

All the magnitude score data are stored in a computer, and coded for the name of each disease: VCP for vocal cord polyp, SPD for spasmodic dysphonia, RNP for recurrent nerve paralysis, VCN for vocal cord nodule, POLPOD for polypoid of the cord, and N for normal voice.

RESULTS AND DISCUSSION

Within three years, from 1990-1992, 672 voice samples from various diseases were collected, and grouped in coded names. The VCP is predominantly found, followed by RNP, VCN, SPD, and POLPOD. From grouped and extracted data, one figure is exposed and histograms are composed to reveal and clarify the results of the investigation.

The present research of acoustical waveform perturbation show the following results:

Figure 1 shows the automatization of the 0.5 second voice sample analysis, direct implementation of the algorithm on the computer, also the waveform and the sonagram, showing the fundamental frequency (F0), the harmonic, non-harmonic components and the fluctuation, all in one. Each taken sample is a 0.5 second segment by excluding the onset and the final portion of the whole sustained vowel /e/.

Figures 2 and 3, show the real frequency (%) of the PPQ and of the APQ, inferring that the peak of the diseased voices are shifted to the right. These evidences assume that the PPQ and APQ of diseased voices have larger amount of scores compared to the normal voice. This situation is more evident and clear in Figures 4 and 5.

Figures 4 and 5 show that the cumulative distribution of PPQ and APQ of the diseased voices are shifted to the right, inferring larger magnitude of scores compared to normal voice, and only about 10% of the POLPOD has smaller score of the APQ. This differences of scores are proven significant by the ANOVA statistical calculation ($P=0.0001$ for the PPQ and 0.0053 for the APQ).

As for the follow up parameter, Figures 6 and 7 show the cumulative frequency of PPQ and APQ of VCP voices before and after surgery. It is evident that the scores before treatment are larger compared to post-treatment, which is also seen in the Percentile Comparison graph. The statistical calculation of pre and post-treatment voice samples is significant for the PPQ ($P=0.0068$), but not significant for APQ ($P=0.113$) and this might be caused by too small data.

We also include the observation on the noise level (NLvl), as we know that hoarse voice is also influenced by this factor. Figure 8 shows the real frequency (%) of NLvl for normal and diseased voices. The graph reveals higher scores in diseased voices compared to normal, as the peak of diseased voices are shifted to the right. This situation is clearer shown in Figure 9, where the cumulative frequency of diseased voices are shifted to the right as well. The difference of scores are significant statistically ($P=0.0001$).

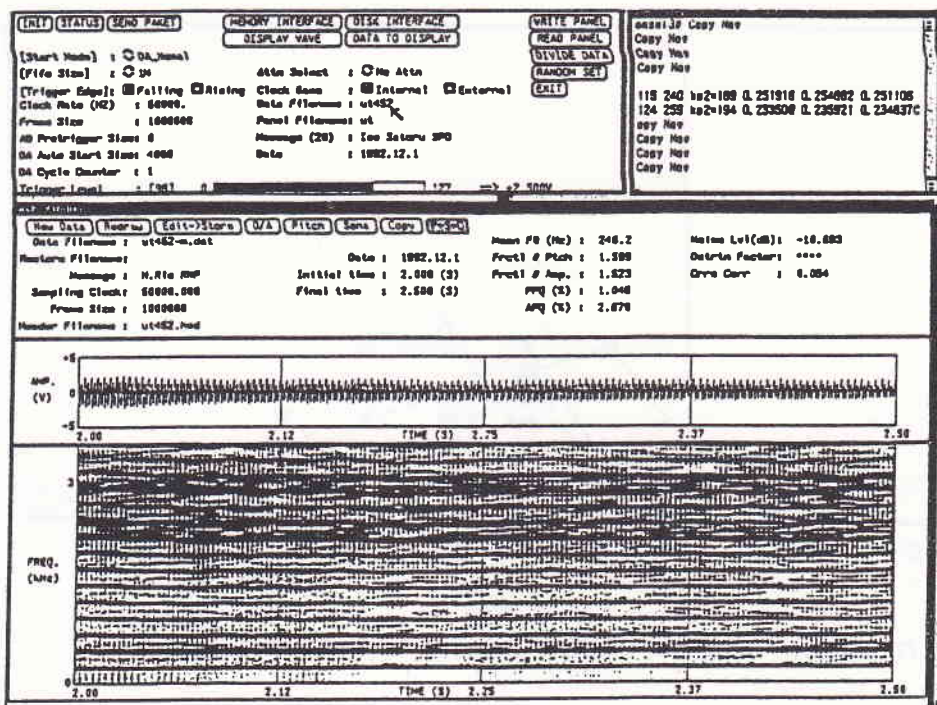


Figure 1. Profile of the automatization for 0.5 second voice sample analysis

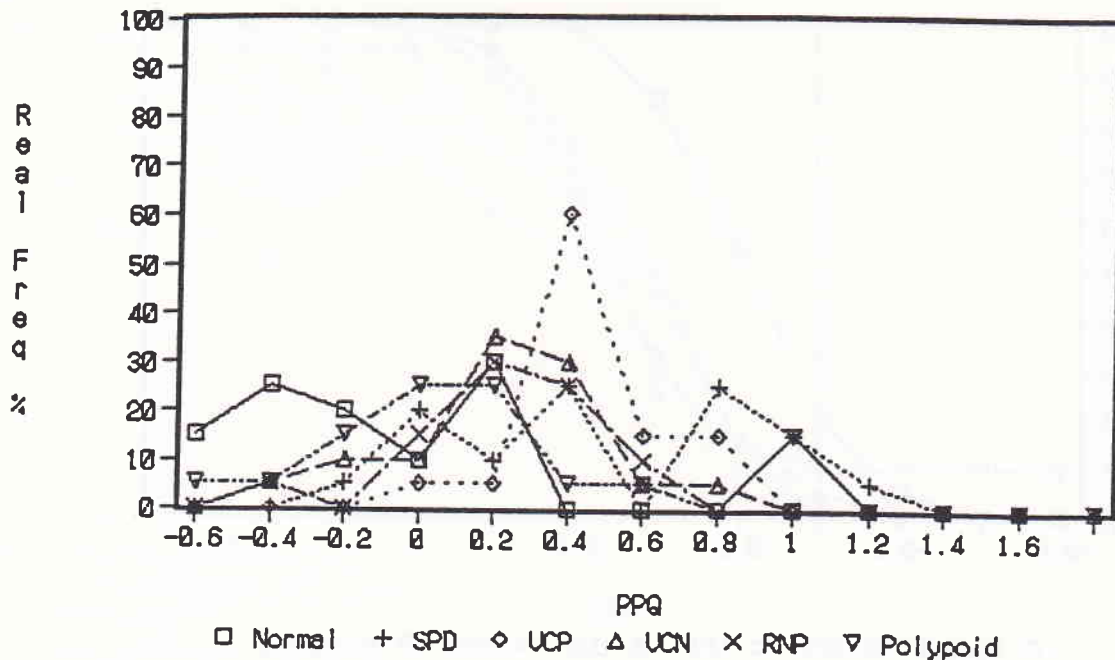


Figure 2. Real Frequency (%) of PPQ for normal and diseased voices

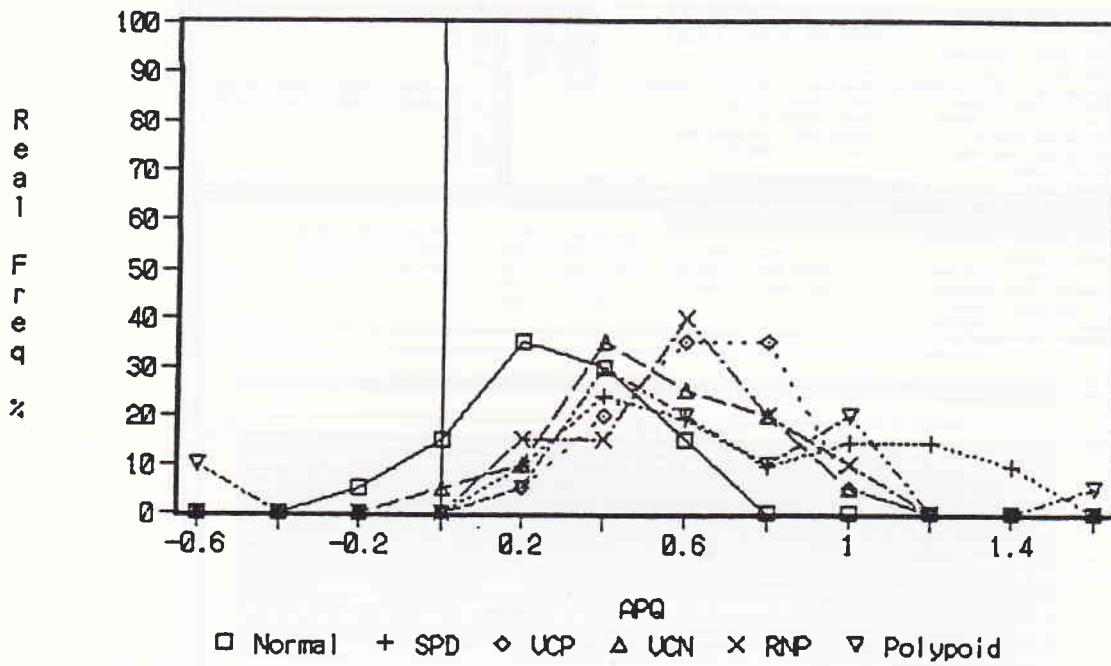


Figure 3. Real Frequency (%) of APQ for normal and diseased voices

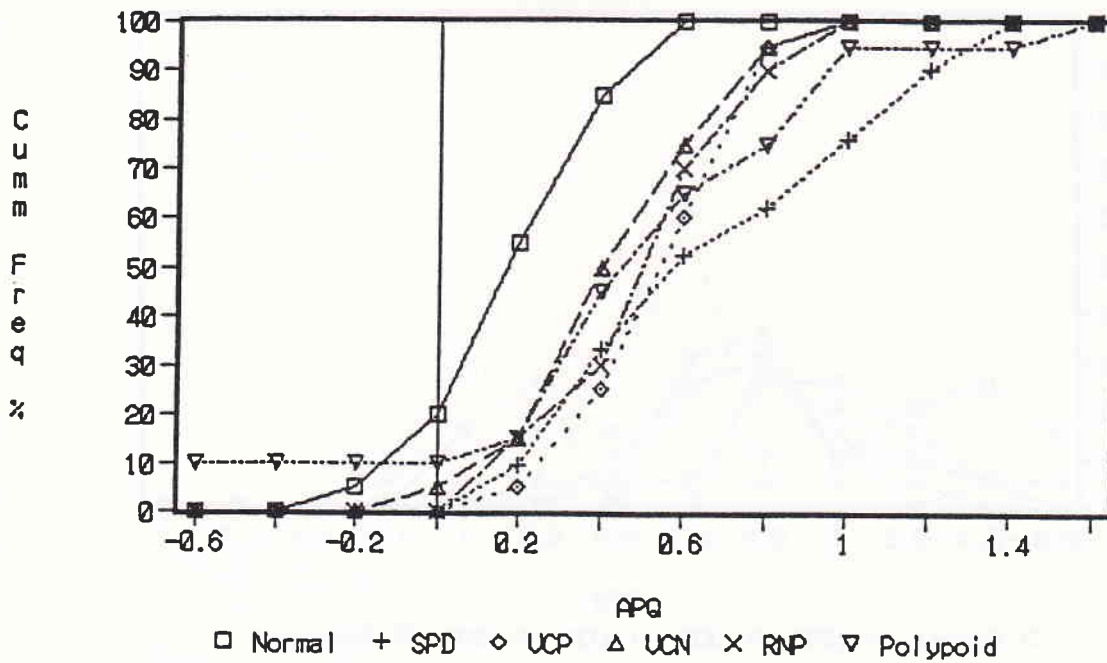


Figure 4. Cumulative Distribution of PPQ for normal and diseased voices

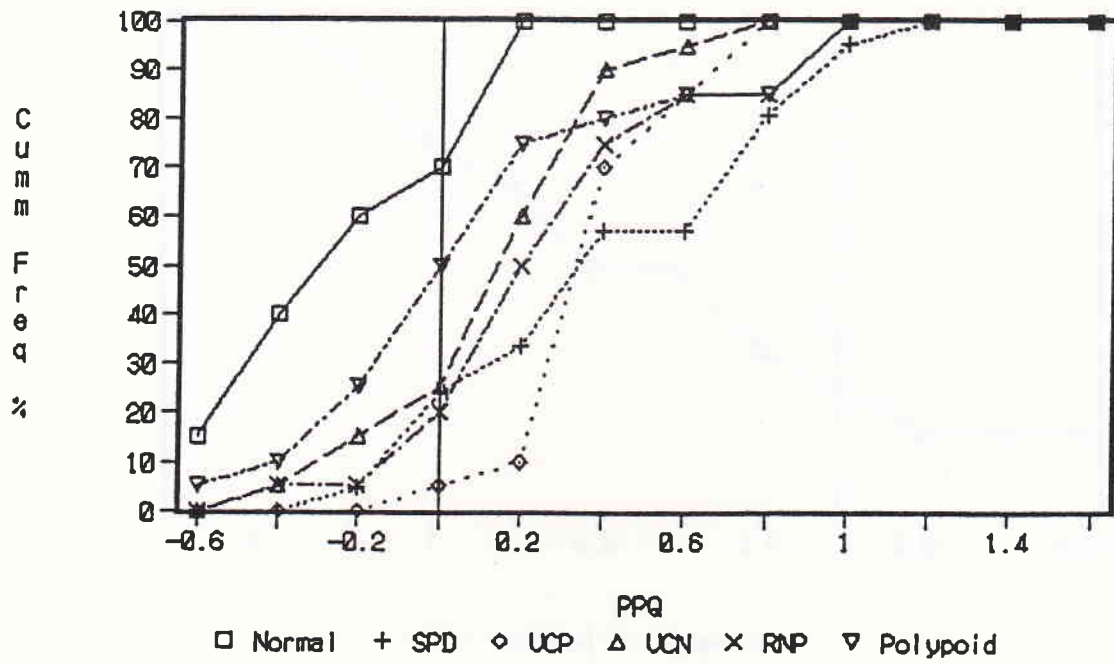


Figure 5. Cumulative Distribution of APQ for normal and diseased voices

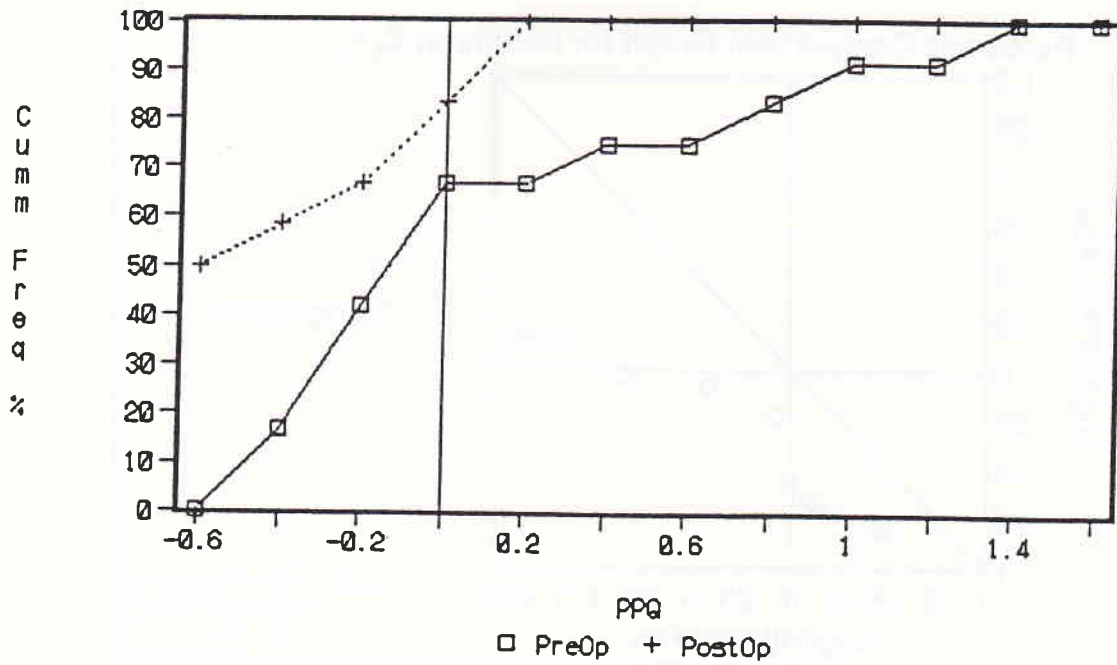


Figure 6. Cumulative Distribution of PPQ for VCP Voices before and after surgery

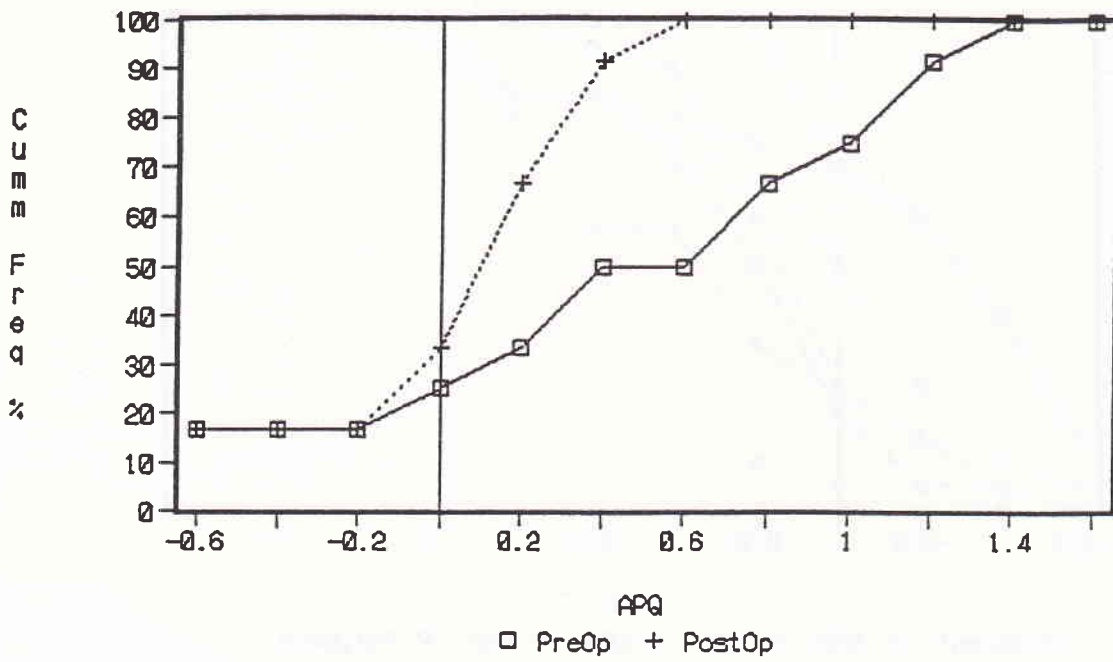


Figure 7. Cumulative Distribution of APQ for VCP Voices before and after surgery

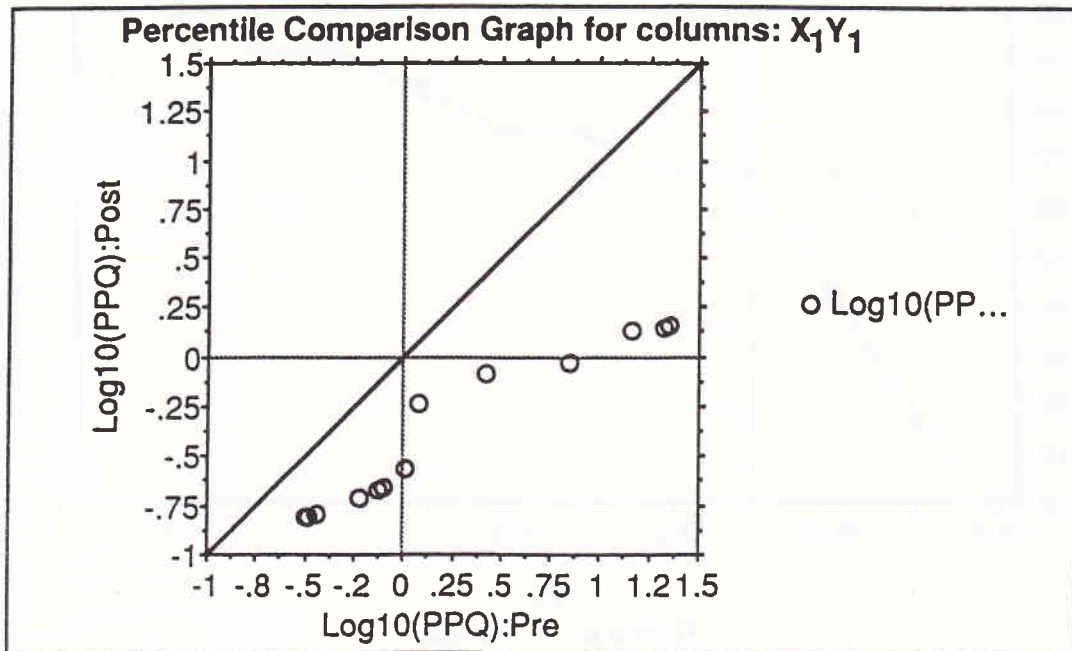


Figure 8. Real Frequency (%) of Noise Level (NLvl) for normal and diseased voices

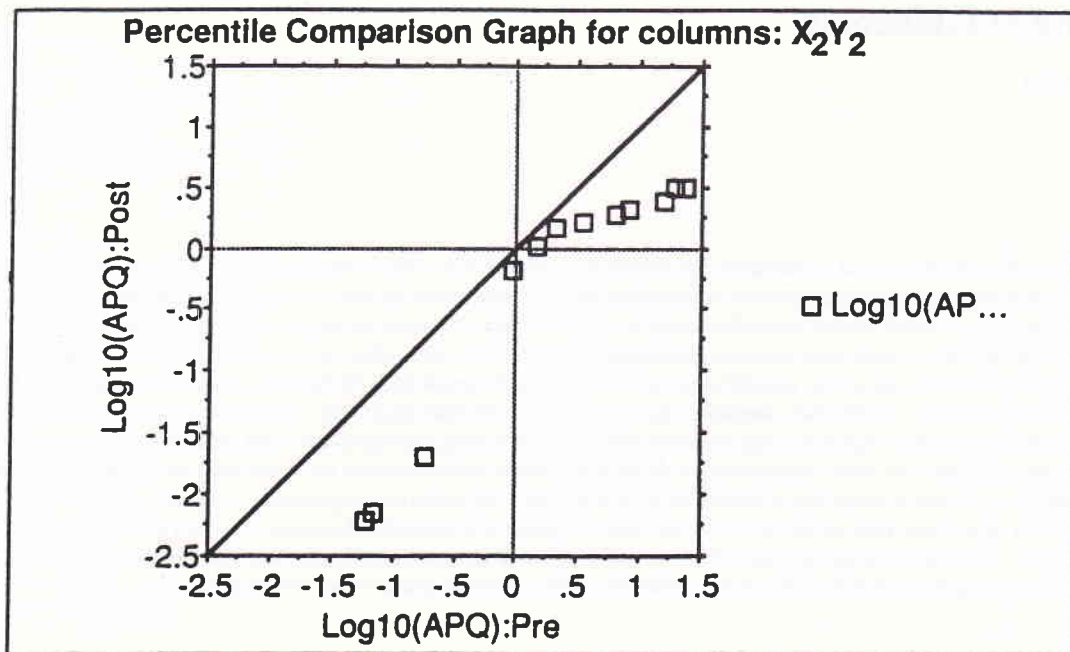


Figure 9. Cumulative Distribution of Noise Level (NLvl) for normal and diseased voices

CONCLUSION

Clinicians are eager and longing for an automatic, practical, easy to operate, inexpensive and non-invasive instrument to detect, if possible to determine diagnosis of the disease. This condition is also noted in the Otolaryngology field, with several diseases that cause voice changes.

After having conducted the study and discussed the results, it seems that we are still far from the final target. Nevertheless, some of the aims as has been mentioned before, were achieved. We see that the significance of the role of the acoustic waveform analysis as supportive tool and adjunctive examination, and also as instrument and measures for follow up and parameter work up has been proven. The significant change in waveform perturbation would also arise the suspicion or early change in the epithelial layer of the vocal cord, that might lead to early detection of a hazardous disease.

These evidences are somewhat beneficial for the clinicians in planning, determining and following up the patient's history of disease. It should be accepted that the application of acoustical analysis method is in

the same level as other medical examinations, such as laboratory findings or X-photo examination. For a more intensive study, it is suggested to include larger amount of samples, divided in different stages of the disease.

REFERENCES

1. Georgiou B. Given and Ear to your Computer. *BYTE The Small Systems journals*. 1978;3:56-91.
2. Hirano M, Hibi S, Yoshida T, Hirade Y, Kasuya H, Kikuchi Y. Acoustic Analysis of Pathological Voice. *Acta Otolaryngol*. 1988;105:432-8.
3. Hirano M. Morphological structure of the vocal fold as a vibrator and its variations. *Folia Phoniat*. 1974;26:89-94.
4. Titze IR. The Human Vocal Cords, A Mathematical model. Part II. *Phonetica*. 1974;29:1-21.
5. Laver J, Hiller S, Beck JM. Acoustic Waveform Perturbation and Voice disorder. *J Voice*. 1992;6:115-26.
6. Wilson FB, Starr CD. Use of the phonation analyzer as a Clinical tool. *J of Speech and Hearing Disorder*. 1985;50:351-56.
7. Imaizumi S, Gauffin J. Acoustical perceptual characteristics of Pathological Voices: rough, creak, fry and diplophonia. *Ann Bull RILP*. 1991;25:109-19.